NELSON CENGAGE Learning

VCE UNITS 1 & 2

6TH EDITION

COMPUTING

SERIES EDITOR James Lawson
Therese Keane, Mark Kelly,
Colin Potts, Anthony Sullivan

NELSON
CENGAGE Learning

VCE UNITS 1 & 2

6TH EDITION

COMPUTING

SERIES EDITOR James Lawson
Therese Keane, Mark Kelly,
Colin Potts, Anthony Sullivan

NELSON CENGAGE Learning

Computing VCE Units 1 & 2
6th Edition
James Lawson
Therese Keane
Colin Potts
Anthony Sullivan
Mark Kelly

Senior Publishing Editor: Eleanor Gregory
Editor: Nadine Anderson-Conklin
Copyeditors: Gene Anderson-Conklin, Nadine Anderson-Conklin and Valina Rainer
Proofreader: Kate McGregor
Indexer: Russell Brooks
Permissions researcher: Debbie Gallagher
Cover designer: Chris Starr
Text designer: Kim Ferguson
Cover image: Shutterstock.com/melis
Senior Production Controller: Julie McArthur
Typeset by: Q2A Media

Any URLs contained in this publication were checked for currency during the production process. Note, however, that the publisher cannot vouch for the ongoing currency of URLs.

Acknowledgements
VCE Computing Study Design (2016) © The Victorian Curriculum and Assessment Authority (VCAA). Used with permission.
VCE ® is a registered trademark of the VCAA. The VCAA does not endorse or make any warranties regarding this Cengage product. Current and past VCE Study Designs, VCE exams and related content can be accessed directly at www.vcaa.vic.edu.au

© 2016 Cengage Learning Australia Pty Limited

Copyright Notice
This Work is copyright. No part of this Work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means without prior written permission of the Publisher. Except as permitted under the *Copyright Act 1968*, for example any fair dealing for the purposes of private study, research, criticism or review, subject to certain limitations. These limitations include: Restricting the copying to a maximum of one chapter or 10% of this book, whichever is greater; providing an appropriate notice and warning with the copies of the Work disseminated; taking all reasonable steps to limit access to these copies to people authorised to receive these copies; ensuring you hold the appropriate Licences issued by the Copyright Agency Limited ("CAL"), supply a remuneration notice to CAL and pay any required fees. For details of CAL licences and remuneration notices please contact CAL at Level 15, 233 Castlereagh Street, Sydney NSW 2000,
Tel: (02) 9394 7600, Fax: (02) 9394 7601
Email: info@copyright.com.au
Website: www.copyright.com.au

For product information and technology assistance,
in Australia call **1300 790 853**;
in New Zealand call **0800 449 725**

For permission to use material from this text or product, please email **aust.permissions@cengage.com**

National Library of Australia Cataloguing-in-Publication Data
Lawson, James (James Austin), 1971– author.
Computing VCE units 1 & 2 / James Lawson, Therese Keane, Colin Potts, Anthony Sullivan, Mark Kelly.

6th edition.
9780170364744 (paperback)
Includes index.
For secondary school age.

Information technology--Study and teaching (Secondary)
Computer programming--Study and teaching (Secondary)
Victorian Certificate of Education examination.

Keane, Therese, author.
Potts, Colin, author.
Sullivan, Anthony Bernard, author.
Kelly, Mark John, author.

004

Cengage Learning Australia
Level 7, 80 Dorcas Street
South Melbourne, Victoria Australia 3205

Cengage Learning New Zealand
Unit 4B Rosedale Office Park
331 Rosedale Road, Albany, North Shore 0632, NZ

For learning solutions, visit **cengage.com.au**

Printed in Australia by Ligare Pty Limited.
1 2 3 4 5 6 7 20 19 18 17 16

CONTENTS

Preface v
About the authors vi
How to use this book vii
Outcomes viii
Problem-solving methodology xii
Key concepts xiv

UNIT 1 1
Introduction

CHAPTER 1 DATA ANALYSIS 2

Understanding research 3
Data and information 4
Gathering data: Primary data and information 5
Quality of data and information 9
Referencing primary sources 11
Seeking permission 15
Privacy 19
Physical security controls 20
Australian Privacy Principles 22
Ethical dilemmas 24

CHAPTER 2 APPROACHES TO PROBLEM-SOLVING METHODOLOGY: DATA ANALYSIS 30

Approaches to problem solving 31
Purpose of graphic solutions 42
Design principles for graphic solutions 52
Formats and conventions 56
Design tools 66
Types of tests 68
Validation 70
Processing data to create solutions 70
Preparing for Unit 1, Outcome 1 78

CHAPTER 3 NETWORKS 80

Networks 81
Types of networks 82
Network architecture 84
Communications software 96
Internet services 98
Network communications standards 101
Mobile devices connected to networks 104
Communications channel 109
National Broadband Network (NBN) 111
Wireless transmission media 112
Network security 116
Measures to secure networks 118
Network physical designs 121
Legal and ethical responsibilities 122
Preparing for Unit 1, Outcome 2 137

CHAPTER 4 ISSUES IN INFORMATION SYSTEMS 139

Information systems 140
Issues in information systems 140
Expressing opinions 153
Methods and techniques to acquire data and information 155
Data integrity 158
Storing shared files 159
Mobile devices and web design 160

CHAPTER 5 APPROACHES TO PROBLEM SOLVING 167

Creating team solutions 168
Project management techniques 168
Designing websites 171
Information architecture 177
Design principles 181
Specific design considerations 194
Design tools 195
Developing websites 205
Preparing for Unit 1, Outcome 3 214

UNIT 2 217
Introduction

CHAPTER 6 PROGRAMMING 218

Information systems in software 219
Hardware 219
Software 222
The operating system (OS) 222
Programming and scripting languages 222
Software development tools 223
Storage structures 225
Developing software 228
Creating effective user interfaces 235
Fundamental programming concepts 240
Preparing for Unit 2, Outcome 1 261

CHAPTER 7 DATA ANALYSIS AND VISUALISATION 268

Information needs and data visualisations 263
Sources of authentic data 264
Data types and data structures 269
Types and purposes of data visualisation 271
PSM: Analysis 277
Design tools 279
Formats and conventions 281
Software tools and functions 282
File formats 288
Evaluating data visualisations 289
Preparing for Unit 2, Outcome 2 299

CHAPTER 8 DATA MANAGEMENT 300

Applications of database systems 301
Database management systems (DBMS) 302
Database structure 303
Characteristics of data types 307
Data sources and methods of data acquisition 308
Collection tools and user interfaces for data entry 309
Database design tools 312
Development of the DBMS 320
Roles, functions and characteristics of hardware components 329
Communication devices 331
Accidental and deliberate security threats 332
Physical and software controls for protecting security 333
Preparing for Unit 2, Outcome 3 343

Index 345

PREFACE

This sixth edition of *Computing VCE Units 1 & 2* incorporates the changes to the VCAA VCE Computing Study Design that took effect from 2016.

This book looks at how individuals and organisations use, and can be affected by, information systems in their daily lives.

We believe that teachers and students require a text that focuses on the Areas of Study specified in the Study Design, and that presents information in a sequence that allows easy transition from theory into practical assessment tasks. We have therefore written this book so that a class can begin at Chapter 1 and work their way systematically through to the end. Students will encounter material relating to the key knowledge dot points for each Outcome before they reach the special section that describes the outcome. The Study Design outlines key skills that indicate how the knowledge can be applied to produce a solution to an information problem. These Preparing for Outcomes sections occur regularly throughout the book, and flag an appropriate point in the student's development for each outcome to be completed. The authors have covered all key knowledge dot points for the outcomes from the Units 1 & 2 course.

Our approach has been to focus on the key knowledge required for each school-assessed outcome, and to ensure that students are well prepared for these; however, there is considerable duplication in the Study Design relating to the knowledge required for many of the outcomes. We have found that, with an outcomes approach, sometimes we are covering material several times. For example, knowledge of a problem-solving methodology is listed as key knowledge for five different outcomes. In these cases, we have tried to cover the concept generally in the first instance, and specifically apply it to a situation relevant to the related outcome on subsequent encounters.

The authors assume teachers will help students to develop the required key skills within the context of the key knowledge addressed in this book and the resources available to them.

We have incorporated a margin column in the text that provides additional information, clarification of terms and reinforcement of key concepts. The margin column also includes activities related to the topics covered in the text, and a consideration of issues relevant to the use of information systems.

Outcome features are included at several points in the book, indicating the nature of the tasks that students are to undertake to complete the school-assessed Outcomes. The steps required to complete the Outcomes are listed, together with advice and suggestions for approaching the task. The output and support material needed for submission are described. Sample tasks and further advice relating to the outcomes are available at **http://computing1and2.nelsonnet.com.au.**

The chapters are organised to present the optimum amount of information in the most effective manner. The text is presented in concise, clearly identified sections to guide students through the text. Each chapter is organised into the sections described on pages **viii–xi**.

9780170364744

ABOUT THE AUTHORS

James Lawson is Head of Computing at Trinity Grammar School, and has taught VCE Computing since 1995. He has been an exam assessor for Computing, and has on a number of occasions presented at both the Victorian Information Technology Teachers' Association (VITTA, now known as Digital Learning and Teaching Victoria or DLTV) and the History Teachers' Association of Victoria (HTAV).

Dr Therese Keane is a senior lecturer at Swinburne University. Therese has worked in a variety of school settings where she taught IT and was the Director of ICT. She holds a Doctorate in Education focusing on ICT leadership in schools. Therese is a member of the ACS ICT Educators Board and a Committee Member on the DLTV and a former office holder in VITTA and the ICTEV. She has presented numerous seminars and workshops for teachers involved in teaching IT. Therese has written several textbooks in all units of VCE Information Technology since 1996. Therese's current work involves providing professional development to IT teachers, delivering workshops and presentations to secondary students and researching the use of technology and computers in schools for teaching and learning purposes.

Mark Kelly learned to program in FORTRAN in 1975, bought his first computer – a Tandy TRS-80 with 4KB of RAM – in 1978 and has been programming and researching IT ever since. He taught VCE IT for 20 years after 10 years teaching English and Psychology. At McKinnon Secondary College he was Systems Manager and author of Rupert, the college's student reporting database.

Colin Potts is Assistant Headmaster (Academic Programs) at Trinity Grammar School, and has taught computing classes for a number of years. He is also a former president of VITTA.

Anthony Sullivan has been teaching Commerce and Computing for more than 20 years. He has taught in both government and non-government settings in Australia and has also taught courses in international schools and schools in the United Kingdom. Anthony has presented at a range of conferences and events on VCE Information Technology and Computing and was a member of the VCE Information Technology Study Design Review Panel.

HOW TO USE THIS BOOK

Key knowledge

The key knowledge that is covered in each chapter is listed on the first page. The list includes key knowledge specified in the outcome related to the chapter.

For the student

The first page of each chapter includes an overview of the chapter's contents so that students are aware of the material they will encounter.

For the teacher

This section outlines how the chapter fits into the overall study of VCE Computing, and indicates how the material relates to the completion of outcomes.

Chapters

The major learning material in the chapter is presented as text, photographs and illustrations. The text describes in detail the theory associated with the stated outcomes of the Computing VCE Units 1 & 2 course in language that is clear and appropriate for students at this level. The photographs show actual hardware, software and other objects described in the text. Illustrations are used to demonstrate concepts that are more easily explained in this manner.

Throughout the chapter, glossary terms are highlighted in bold, blue text. They are defined at the end of the chapter, in Essential terms.

Margin column

The margin column contains further explanations that support the main text, weblink icons, additional material outside the Study Design, and cross-references to material covered elsewhere in the textbook. Issues relevant to information systems and computing in general that will promote classroom discussion are also included in the form of 'Think about Computing' boxes.

Chapter summary

The chapter summary at the end of each chapter is divided into two main parts to help you review each chapter.

Essential terms are the glossary terms that have been highlighted throughout the chapter.

Important facts are a list of summaries, ideas, processes and statements relevant to the chapter, in the order in which they occur in the chapter.

Test your knowledge

Short-answer questions are provided to help students when reviewing the chapter material. The questions are grouped and identified with a section of the text to allow the teacher to direct appropriate questions based on material covered in class. Teachers will be able to access answers to these questions at **http://computing1and2.nelsonnet.com.au**.

Apply your knowledge

Each chapter concludes with a set of questions requiring students to demonstrate that they can apply the theory from the chapter to more complex questions. Teachers will be able to access suggested responses to these applications at **http://computing1and2.nelsonnet.com.au**.

Preparing for the outcomes

This section appears at points in the course where it is appropriate for students to complete an outcome task. The information provided describes what the students need to do in the outcome, the suggested steps to be followed when completing the task, and the material that needs to be submitted for assessment.

NelsonNet

The NelsonNet student website contains:

- multiple-choice quizzes for each chapter

The NelsonNet teacher website also contains:

- answers for the Test your knowledge and Apply your knowledge questions in the book
- Sample SACs
- chapter tests
- practice exams for each unit.

Please note that complimentary access to NelsonNet and the NelsonNetBook is only available to teachers who use the accompanying student textbook as a core educational resource in their classroom. Contact your sales representative for information about access codes and conditions.

An open-access weblink page is also provided, for all weblinks that appear in the margins throughout the textbook. This is accessible at **http://computing1and2.nelsonnet.com.au**.

OUTCOMES

OUTCOME	KEY KNOWLEDGE	REFERENCE
Unit 1 Area of Study 1 Outcome 1	**Data and graphic solutions** On completion of this unit the student should be able to acquire, secure and interpret data, and design and develop a graphic solution that communicates the findings of an investigation.	Chapter 1 and Chapter 2
Data and information	• types and purposes of qualitative and quantitative data	pp. 3–4
	• sources of, and methods and techniques for, acquiring and referencing primary data and information	pp. 5–9, 12–14
	• factors affecting the quality of data and information such as relevance, accuracy, bias and reliability	pp. 9–11
	• techniques for authorising the collection and use of data and information such as using consent forms	pp. 15–18
	• techniques for protecting the privacy of the providers of data and information such as de-identifying personal data	p. 19
Digital systems	• physical and software controls used to protect the security of stored data such as backing up, usernames and passwords, systems protection software and encryption	pp. 20–2
Interactions and impact	• Australian Privacy Principles relating to the acquisition, management and communication of data and information, including non-identification of individuals (principle 2), information only being held for its primary purpose (principle 6)	pp. 22–4
	• ethical dilemmas arising from data acquisition strategies	pp. 24–5
Approaches to problem solving	• types of graphic solutions suitable for educating, persuading and informing audiences	pp. 43–51
	• design tools for representing the functionality and appearance of graphic solutions such as input–process–output charts (functionality) and annotated diagrams/mock ups (appearance)	pp. 66–8
	• formats and conventions suitable for graphic solutions such as titles, text styles, shapes, lines and arrows, sources of data and legend, colours and contrasts	pp. 56–66
	• software functions and techniques for efficiently and effectively manipulating data to develop graphic solutions, and for validating data	pp. 31–42, 68–73
	• techniques for testing graphic solutions	pp. 68–73
Key skills	• frame an investigation inquiry	pp. 32–42
	• identify, legally and ethically acquire, and reference data and information from primary sources	pp. 78–9
	• devise and implement controls and techniques to minimise risks to the security and privacy of data and information	pp. 20–2
	• interpret selected data, identifying relationships and patterns	pp. 29, 78
	• select and apply appropriate design tools to represent the functionality and appearance of graphic solutions for particular purposes	p. 67
	• use software, and select and apply functions, formats, conventions, data validation and testing techniques to efficiently manipulate data and create graphic solutions	p. 69
Unit 1 Area of Study 2 Outcome 2	**Networks** On completion of this unit the student should be able to design a network with wireless capability that meets an identified need or opportunity, explain its configuration and predict risks and benefits for intended users.	Chapter 3
Digital systems	• applications and capabilities of Local Area Networks (LANs) and Wide Area Networks (WANs)	pp. 82–4
	• functions and characteristics of key hardware and software components of networks required for communicating and storing data and information	pp. 90–100
	• purposes of network protocols	pp. 101–4
	• strengths and limitations of wireless communications technology, measured in terms of data transfer rate, data storage options, cost, security and reliability	pp. 112–16
	• types, capabilities and limitations of mobile devices connected to networks	pp. 104–9
	• security threats to data and information communicated and stored within networks	pp. 116–18, 126
	• technical underpinnings of malware that intentionally threaten the security of networks	pp. 117–18

OUTCOME	KEY KNOWLEDGE	REFERENCE
Interactions and impact	• ways in which people, processes, digital systems and data combine to form networked information systems	pp. 84–90
	• legal requirements and ethical responsibilities of network professionals and users of networks with respect to social protocols and the ownership of data and information	pp. 122–4
	• risks and benefits of using networks in a global environment	pp. 124–6
Key skills	• describe the capabilities of different networks and wireless communications technology	pp. 132–3
	• compare the capabilities of a range of network components to support the communication and storage of data and information	pp. 132–8
	• apply design thinking skills when configuring a network solution with wireless capability, taking into account how data and information are transmitted and secured	pp. 132–8
	• apply systems thinking skills to predict risks and benefits of the implementation of a new or modified network solution with wireless capability for the users	pp. 132–8
Unit 1 Area of Study 3 Outcome 3	**Collaboration and communication** On completion of this unit the student should be able to apply the problem-solving methodology to create a solution using database management software, and explain the personal benefits and risks of interacting with a database.	Chapter 4 and Chapter 5
Interactions and impact	• applications of information systems in a range of settings	pp. 140–52
	• a detailed study in a particular field such as entertainment, agriculture, finance, sport, health, that focuses on: – the nature of a contemporary issue associated with the use of information systems – legal, social, environmental or ethical reasons for a contentious issue – types and capabilities of digital systems associated with the field and issue – key stakeholders such as individuals, organisations and governments, and their responsibilities – positive and negative opinions of each stakeholder about the issue	pp. 140–52
	• ways in which end-users can express opinions on websites about how information systems are used for particular purposes such as writing a review in a text box and a rating system	pp. 153–4
Data and information	• sources of, and methods and techniques for, acquiring and referencing primary data and secondary data and information	pp. 155–8
	• factors affecting the integrity of data, such as correctness, reasonableness and accuracy	pp. 158–9
Digital systems	• advantages and disadvantages of using cloud solutions, and using cloud computing for storing, communicating and disposing of data and information	pp. 159–60
	• impact of growth of mobile devices on website design	pp. 160–1
Approaches to problem solving	• visualising thinking tools and techniques for supporting reasoning and decision making when analysing issues and ethical dilemmas	pp. 142–4
	• key principles of information architecture	pp. 177–80
	• characteristics of effective user interfaces for mobile devices, for example useability, accessibility, tolerance, visibility, legibility, consistency, affordance	pp. 171–6
	• design principles that influence the appearance of websites	pp. 181–94
	• design tools and techniques for representing websites	pp. 195–204
	• formats and conventions suitable for websites	pp. 201–3
	• software functions and techniques for manipulating and validating data, and testing websites	pp. 205–8
	• tools and techniques for coordinating the tasks, people, digital systems resources and time required to create solutions	pp. 168–71
Key skills	• select and apply appropriate methods and techniques to acquire and reference data and information	pp. 155–8
	• use digital systems to document and monitor project plans when creating team solutions	pp. 168–71, 213–15
	• analyse the causes and effects of issues using visualising thinking tools	pp. 142–4
	• synthesise viewpoints to formulate a team's point of view	p. 183
	• evaluate cloud computing as a data storage solution	p. 165

OUTCOME	KEY KNOWLEDGE	REFERENCE
	• select and use digital system components appropriate to a team's needs	pp. 211–15
	• select appropriate design tools and represent the appearance and functionality of solutions, taking into account user interactions	pp. 213–15
	• recommend online techniques for encouraging end-users' support of published viewpoints	pp. 211–15
	• use web authoring software and select and apply functions and techniques to manipulate data and create solutions	pp. 213–15
Unit 2 Area of Study 1 Outcome 1	**Programming** On completion of this unit the student should be able to design working modules in response to solution requirements, and use a programming or scripting language to develop the modules.	Chapter 6
Data and information	• characteristics of data types and methods of representing and storing text, sound and images	pp. 225–8
Digital systems	• functions and capabilities of key hardware and software components of digital systems required for processing, storing and communicating data and information	pp. 219–24
Approaches to problem solving	• functional requirements of solutions	pp. 228–9
	• methods for creating algorithms such as identifying the required output, the input needed to produce the output, and the processing steps necessary to achieve the transformation from a design to a solution	pp. 228–35, 240–4
	• suitable methods of representing solution designs such as data dictionaries, data structure diagrams, object descriptions and pseudocode	pp. 229–35
	• characteristics of effective user interfaces, for example useability, accessibility, structure, visibility, legibility, consistency, tolerance, affordance	pp. 235–9
	• techniques for manipulating data and information	pp. 250–5
	• naming conventions for files and objects	pp. 230–1
	• testing and debugging techniques, including construction of test data.	pp. 244–9
Key skills	• interpret solution requirements	p. 260
	• select and use appropriate methods for expressing solution designs, including user interfaces	p. 260
	• apply techniques for manipulating data and information using a programming or scripting language	p. 260
	• devise meaningful naming conventions for files and objects	pp. 230–1
	• apply testing techniques using appropriate test data	p. 260
Unit 2 Area of Study 2 Outcome 2	**Data analysis and visualisation** On completion of this unit the student should be able to apply the problem-solving methodology and use appropriate software tools to extract relevant data and create a data visualisation that meets a specified user's needs.	Chapter 7
Data and information	• sources of authentic data in large repositories	pp. 264–7
	• factors influencing the integrity of data, for example accuracy, timeliness, authenticity, relevance	pp. 267–9
	• characteristics of data types and data structures relevant to selected software tools	pp. 269–71
Approaches to problem solving	• types and purposes of data visualisations	pp. 271–7
	• problem-solving activities related to analysing needs: functional and non-functional requirements and constraints	pp. 277–9
	• characteristics of file formats and their suitability to be converted to other formats	pp. 288–9
	• design tools for representing data visualisations	pp. 279–81
	• formats and conventions applied to visualisations to improve their effectiveness for intended users	pp. 281–2
	• functions of appropriate software tools to extract targeted data and to manipulate data when developing visualisations	pp. 282–7
	• criteria and techniques for evaluating visualisations	pp. 289–92

OUTCOME	KEY KNOWLEDGE	REFERENCE
Key skills	• analyse needs to define specific requirements	pp. 295–9
	• identify and extract, using software functions, relevant data from appropriate data sources	pp. 295–9
	• prepare data structures relevant to the software tools	pp. 295–9
	• interpret selected data, identifying relationships and patterns	pp. 295–9
	• select and apply appropriate tools to represent the design of selected visualisations	pp. 295–9
	• use appropriate software and select and apply functions, formats and conventions to manipulate the extracted data to create data visualisations	pp. 295–9
	• select appropriate techniques and apply criteria to determine the extent to which data visualisations meet users' needs	pp. 295–9
Unit 2 Area of Study 3 Outcome 3	**Data management** On completion of this unit the student should be able to apply the problem-solving methodology to create a solution using database management software, and explain the personal benefits and risks of interacting with a database.	Chapter 8
Data and information	• data sources and methods of data acquisition	pp. 308–10
	• characteristics of effective data collection tools and user interfaces for the purposes of entering data efficiently	pp. 310–12
	• characteristics of data types	pp. 307–8
Digital systems	• capabilities and limitations of database management software to manipulate data	p. 302
	• roles, functions and characteristics of hardware components used to input, store, communicate and output data and information	pp. 329–31
	• accidental and deliberate security threats to data and information stored within databases	p. 332
	• physical and software controls suitable for protecting the security of stored and transmitted data	pp. 333–5
Approaches to problem solving	• the structure of a database, including fields, records and tables	pp. 303–6
	• design tools for representing input forms to capture data and reports to meet specific needs	pp. 315–16, 318–19
	• design tools for representing the structure of databases	pp. 312–14
	• techniques for manipulating and validating data	pp. 316–18, 319
	• formats and conventions applied to create effective solutions	pp. 311–35
Interactions and impact	• applications of database systems in a range of settings	pp. 311–35
	• personal benefits and risks arising from the use of databases	p. 303
Key skills	• analyse needs or opportunities for database management solutions	pp. 311–35
	• use appropriate techniques to describe data types and database structures	pp. 311–35
	• identify and collect data from appropriate sources, using data collection tools that facilitate efficient data entry	pp. 343–4
	• apply suitable functions to validate and manipulate data efficiently	p. 340
	• construct queries to locate data that matches specific criteria	p. 340
	• apply formats and conventions to create effective forms and reports	pp. 343–4
	• evaluate the value of using a database system in fulfilling a personal need	p. 340

VCE Computing 2016–2019 © The Victorian Curriculum and Assessment Authority (VCAA). Used with permission. The VCAA does not endorse or make any warranties regarding this [Cengage/Nelson] publication. VCE is a registered trademark of theVCAA. Current and past VCE Study Designs, VCE exams and related content can be accessed directly at www.vcaa.vic.edu.au

9780170364744

PROBLEM-SOLVING METHODOLOGY

When an information problem exists, a structured problem-solving methodology is followed to ensure that the most appropriate solution is found and implemented. For the purpose of this course, the problem-solving methodology has four key stages: Analysis, design, development and evaluation. Each of these stages can be further broken down into a common set of activities as shown in Figure 1. Each unit may require you to examine a different set of problem-solving stages. It is critical that students understand the problem-solving methodology, because it underpins the entire VCE Computing course.

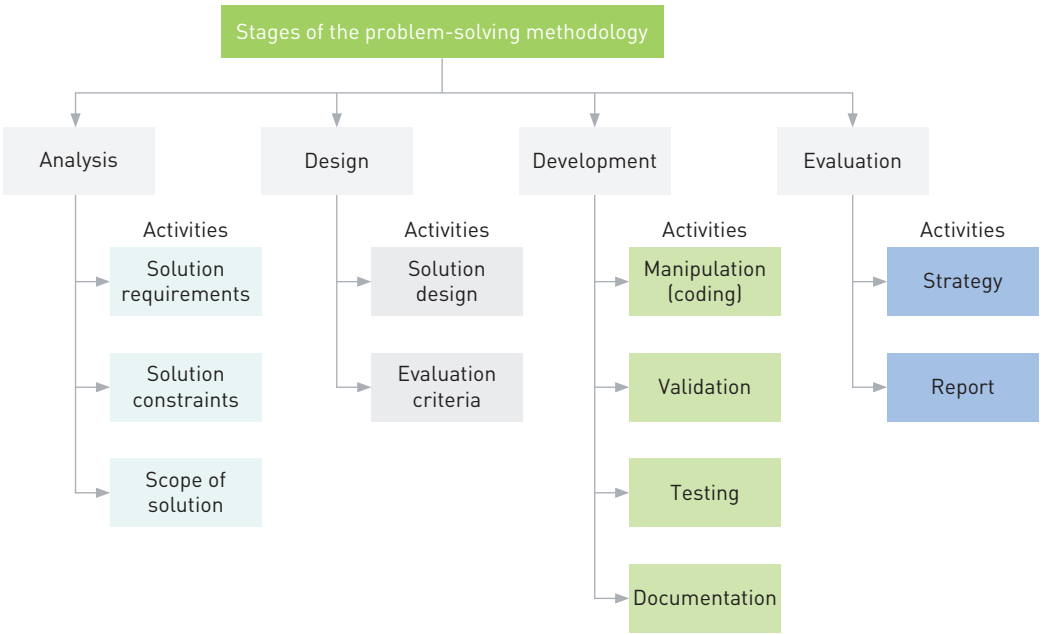


FIGURE 1 The four stages of the problem-solving methodology and their key activities.

Analyse the problem

The purpose of analysis is to establish the root cause of the problem, the specific information needs of the organisation involved, limitations on the problem and exactly what a possible solution would be expected to do (the scope). The three key activities are:

1. identifying solution requirements – features and functionality that the solution needs to include, information it must produce and data needed to produce this information
2. establishing solution constraints – the limitations on solution development that need to be considered. Constraints are classified as economic, technical, social, legal and related to useability
3. defining the scope of the solution – what the solution will and will not be able to do, as well as how the user will benefit.

Design the solution

During the design stage, generate an appropriate design idea. Criteria are also created to evaluate the solution's success once it has been implemented. The two key design activities are:

1. creating the solution design – it must clearly show a developer what the solution should look like and how its data elements should be structured, validated and manipulated. Tools typically used to represent data elements could include data dictionaries, data structure diagrams, input–process–output (IPO) charts, flowcharts, pseudocode (or structured English) and object descriptions. The following tools are also used to show the relationship between various components of the solution: storyboards, site maps, entity–relationship diagrams,

data flow diagrams, structure charts, hierarchy charts and context diagrams. Furthermore, the appearance of the solution needs to be planned so that overall layout, fonts and their colours, for example, can be represented. Layout diagrams and annotated diagrams (or mock-ups) usually fulfil this requirement. A combination of tools from each of these categories will be selected to represent the overall solution design

2 specifying evaluation criteria – during the evaluation stage, the solution is assessed to establish how well it has met its intended objectives. The criteria for evaluation must be created during the design stage so that all personnel involved in the task are aware of the level of performance that ultimately will determine the success or otherwise of the solution. The criteria are based on the solution requirements identified in the analysis stage.

Develop the solution

During this stage, the solution is created by the developers from the designs supplied to them. The 'coding' takes place, while the input data is checked (validation), the solution is tested and any user documentation is created. The four activities involved with development are:

1 manipulating or coding the solution – the designs are used to build the electronic solution. The coding will occur here and internal documentation will be included where necessary
2 checking the accuracy of input data by way of validation – manual and electronic methods are used; for example, proofreading is a manual validation technique. Electronic validation involves using the solution itself to ensure that data is reasonable. Electronic validation, along with any other formulas, always needs to be tested to ensure that it works properly
3 ensuring that a solution works through testing – each formula and function, not to mention validation and even the layout of elements on the screen, need to be tested. Standard testing procedures involve stating what tests will be conducted, identifying test data, stating the expected result, running the tests, stating the actual result and correcting any errors
4 documentation allowing users to interact with (or use) the solution – while it can be printed, in many cases it is now designed to be viewed on screen. User documentation normally outlines procedures for operating the solution, as well as generates output (like reports) and basic troubleshooting.

Evaluate the solution

Sometime after a solution has been in use by the end user or client, it needs to be assessed or evaluated to ensure that it has been successful and actually meets the user's requirements. The two activities involved in evaluating a solution are:

1 working out an evaluation strategy – creating a timeline for when various elements of the evaluation will occur and how and what data will be collected (because it must match the criteria created in the design stage)
2 reporting on the success of the solution – providing feedback to the user about how well the solution meets their requirements. This is based on the findings of the data gathered at the beginning of the evaluation stage when compared with the evaluation criteria created during the design stage.

9780170364744

KEY CONCEPTS

Each VCE Computing subject contains four key concepts whose purpose is to organise course content into themes. These themes are intended to make it easier to teach and make connections between related concepts and to think about information problems. Key knowledge for each Area of Study is categorised into these key concepts, but not all concepts are covered by each Area of Study. The four key concepts are:

1. data and information
2. digital systems
3. approaches to problem solving
4. interactions and impact.

Data and information focuses on the acquisition, structure, representation and interpretation of data and information in order to elicit meaning or make deductions. This step needs to be completed in order to create solutions.

Digital systems focus on how hardware and software operate in a technical sense. This also includes networks, applications, the internet and communication protocols. Information systems have digital systems as one of their parts. The other components of an information system are people, data and processes.

Approaches to problem solving focuses on thinking about problems and ways of creating solutions. Computational, design and systems thinking are the three key problem solving approaches.

Interactions and impact focuses on relationships that exist between different information systems and how these relationships affect the achievement of economic and social goals. Three types of relationships are considered: people interacting with other people when collaborating or communicating with digital systems, how people interact with digital systems and how information systems interact with other information systems. This theme also looks at the impact of these relationships on information needs, privacy and personal safety.

UNIT 1

INTRODUCTION

VCE Unit 1 of *Computing* looks at how individuals and organisations use, and can be affected by, information and networked digital systems in their daily lives.

Throughout the unit, students will apply the design and development stages of the problem-solving methodology. They will acquire and apply the knowledge and skills to work with different data types to create solutions that can be used to persuade, educate, inform and entertain.

This unit also examines the role of networked information systems in the communication of data within a global environment and an exploration of mobile devices.

Several issues relating to the effect of information systems on students themselves are also examined and students are required to work collaboratively to examine these issues.

There are three outcomes to be completed in Unit 1.

AREA OF STUDY 1: DATA AND GRAPHIC SOLUTIONS

Outcome 1

You will collect your own data and information, and design and develop solutions that meet specific purposes using software to create a graphic solution. The information you produce should be in graphic form. You are only required to apply the design and development stages of the problem-solving methodology, but you should undertake your own investigation.

AREA OF STUDY 2: NETWORKS

Outcome 2

You must propose a networked information system with wireless capability for a specific purpose, and explain the security threats that exist within the networked information system. You must also explain the configuration of the network and suggest the risks and benefits of its use for intended users, including potential legal requirements and ethical responsibilities. Throughout this Outcome, you will explore the exchange of data and information within a networked information system. You will also learn the use of mobile devices within networks, and how the security of data and information exchanged within a network can be threatened.

AREA OF STUDY 3: COLLABORATION AND COMMUNICATION

Outcome 3

You will work in teams to design and develop a website that analyses a chosen contemporary issue and supports your team's point of view. In analysing an information systems issue, you will consider the tensions and conflicts between different stakeholders and then, using visualising thinking tools, explore your own opinions. You will manipulate acquired primary and secondary data and, optionally, develop graphical representations that form part of your website. You will also engage in project management and use digital systems to form and monitor plans.

Getty Images/Rafe Swan

CHAPTER 1

DATA ANALYSIS

Key knowledge

After completing this chapter, you will be able to demonstrate knowledge of:

Data and information

- types and purposes of qualitative and quantitative data
- sources of, and methods and techniques for, acquiring and referencing primary data and information
- factors affecting the quality of data and information such as relevance, accuracy, bias and reliability
- techniques for authorising the collection and use of data and information such as using consent forms
- techniques for protecting the privacy of the providers of data and information such as de-identifying personal data

Digital systems

- physical and software controls used to protect the security of stored data such as backing up, usernames and passwords, systems protection software and encryption

Interactions and impact

- Australian Privacy Principles relating to the acquisition, management and communication of data and information, including non-identification of individuals (Principle 2), information only being held for its primary purpose (Principle 6)
- ethical dilemmas arising from data acquisition strategies.

For the student

Students will conduct an investigation into an issue, practice or event and collect primary data, interpret and manipulate this data into a graphical solution to represent their findings.

For the teacher

This chapter is based on Unit 1, Area of Study 1 and, together with Chapter 2, provides the key knowledge required to complete Unit 1, Outcome 1. At the end of Chapters 1 and 2, students should be able to acquire, secure and interpret data and design and develop a graphical solution that communicates their findings of an investigation.

© (2016). The Victorian Curriculum and Assessment Authority (VCAA). Used with permission.

Shutterstock.com/Rawpixel

Understanding research

Some people consume research, and others produce research. Consumers of research spend a lot of time reading other people's research rather than conducting their own. On the other hand, producers of research investigate or explore an area that has relevance to them, interpret their data and then communicate their findings.

Producers of research often start with a theory or a research question, from which they generate a **hypothesis**, test the hypothesis and then interpret the results, as can be seen in Figure 1.1.

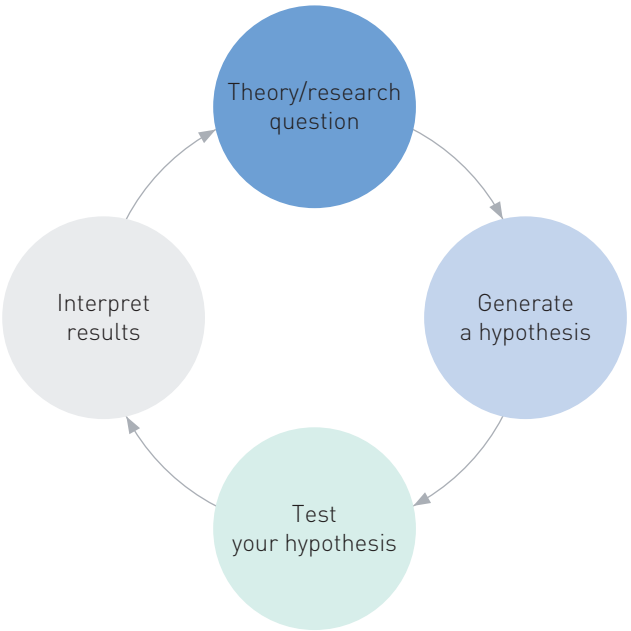


FIGURE 1.1 Producers of research often start with a theory or a research question; then they generate a hypothesis, test the hypothesis and interpret the results

Theories are usually general statements that describe something, provide an explanation of why something happens and can be applied to predict what will happen in the future. Theories are in principle falsifiable or disprovable; that is, they contain information about the sorts of events that, if they were to happen, would show the theory to be false. Some research questions are tied closely with theories. Research questions assist researchers to narrow the focus of the topic of investigation. For example, 'Do science and innovation boost our standard of living and contribute to economic growth?'

A scientific theory summarises a hypothesis that has been supported with repeated testing. A theory is valid as long as there is no evidence to dispute it.

Hypotheses, on the other hand, are based on probabilities about what will happen according to the applied theory. Theories are tested by using data collection tools such as questionnaires and/or interviews, and then the results of the study will either confirm or disprove the hypothesis.

Types of research

An approach to investigate the topic of interest is through quantitative or qualitative research.

Quantitative data is measurable and specific and therefore easier to chart or graph. At a simplistic level, quantitative data gathering is based on verifying theory through the use of statistics and largely numerical data, while qualitative data provides a more in-depth understanding. An example of quantitative data is:

> Of the teachers who teach mathematics to secondary school students (Years 7 to 10), 61 per cent have studied mathematics at university to at least second-year level.
>
> Benchmarking Australian Science, Technology, Engineering and Mathematics, Office of the Chief Scientist, November 2014, p.98

SPSS and MiniTab are statistical software packages used to analyse quantitative data.

When data has been gathered using surveys, focus groups, observation or other methods, quantitative data can be analysed by using software such as Excel, the Statistical Package for the Social Sciences (SPSS) and Minitab. This takes time and often involves hours of data entry, depending on the complexity of the data gathering instrument. For simple data gathering, online surveys such as SurveyMonkey allow users to create surveys and manage the collection and analysis of quantitative data. Most online survey software also permits **qualitative data** to be entered.

Qualitative data is harder to measure than quantitative data. You can gather qualitative data using instruments such as interviews, focus groups, video footage and observation. Generally, qualitative data needs to be recorded accurately and transcribed at a later stage. The analysis of the qualitative data is also quite different from that of quantitative data. With quantitative data, the researcher looks for themes or patterns through the use of numbers, while with qualitative data, the researcher establishes rich descriptions and finds themes through reading the text and classifying these themes. An example of qualitative data gathering is more descriptive:

NVivo is a qualitative data analysis computer software package. It has been designed for qualitative researchers working with very rich text-based and/or multimedia information, where deep levels of analysis of data are required.

> From 2003 to 2012, the mean PISA scores in mathematics declined, while the number of countries performing better than Australia increased. PISA also shows that Australian students' proficiency in mathematics is declining, with the proportion of low performers rising and the proportion of top performers falling.
>
> Benchmarking Australian Science, Technology, Engineering and Mathematics, Office of the Chief Scientist, November 2014, p.100

Advantages and disadvantages of quantitative and qualitative data

Participants are more willing to be part of a quantitative study as it is less demanding of them. Often, quantitative studies use questionnaires, which can capture a large sample size. Having a large sample size provides statistical validity, and helps to accurately reflect the population. Data is then interpreted, relationships identified and findings are then communicated. Conversely, because questionnaires do not have the provision to probe the participants further, the answers provided do not have as much depth and are at times superficial. If too much information were provided, researchers would be overwhelmed by the amount of data collected and would not be in a position to analyse it.

Qualitative research provides for rich, in-depth study of participants. Researchers can ask further questions, especially if something of interest arises. Generally, qualitative studies are small, and provide a narrative description of a sample group. Data gathering tools can include interviews and focus groups. However, because the sample size is small and the sample is not very random, conclusions may not generalise to a larger sample size; therefore, findings may be peculiar to a particular sample.

Data and information

Technically speaking, a datum is a single item of data; however, the term 'data' is commonly used and accepted as both the singular and plural forms of the word.

Information systems focus on both the transformation of **data** into **information** and the management of that information. Critical to using digital systems to solve information problems is an understanding of how data can be input to a computer and then manipulated to create meaningful information. The term 'data' refers to the raw, unorganised facts, figures and symbols fed to a computer during the input process. Data can also mean ideas or concepts before they have been refined. In addition to text and numbers, data also includes sounds, and images (still and moving).

Information is produced when data is manipulated by the computer's processor into a meaningful and useful form, thus becoming information. This can be achieved by organising the data and presenting it in a way that suits the needs of the intended audience. The information produced can be used to inform, entertain or persuade an audience.

Gathering data: Primary data and information

Sources

Primary sources are usually the **stakeholders** in a particular issue – the topic that you are investigating as part of your hypothesis. To question them or survey their opinions can provide different insights and often more in-depth data than information from **secondary sources**. The data will often be more up-to-date and can provide more unusual and important insights into issues, especially at the immediate local level, than secondary sources can, which often present overall conclusions and general summaries. When data is collected, often by non-stakeholders, they frequently use observation and measurement.

Techniques and methods

Collecting data from the stakeholders directly is usually conducted through methods such as surveys using questionnaires, and interviews. While the results of questionnaires are easy to present graphically, interview results often can only be presented as written summaries and conclusions. However, both require analytical discussions to interpret their meaning.

A **questionnaire** is usually a set of questions that ask for a response to be selected from a list of alternatives, such as A, B, C, D; or a range, either 1–5 or very low to very high. Such questionnaires can easily be given to many people, and are easily processed and analysed using computer-based methods since the answers are able to be recorded as numbers.

Another type of questionnaire provides space for short, focused free-form answers similar to those obtained from an interview about particular aspects of an issue; for example, 'Describe the feelings you have when you are playing your favourite computer game' (in an interview, this question would be followed up with 'Why?' or 'What part of the game causes these feelings?'). However, because of their free form, which does not lend itself to being recorded numerically, these answers tend to be more difficult to analyse.

Interviews are usually conducted face to face (technique), sometimes in groups, and can take a substantial amount of time. A major feature of an interview is the opportunity for in-depth follow-up and clarification questions that cannot be done with questionnaires, which are often answered in private. Interviews are very useful for eliciting the feelings, attitudes and opinions of people that are too complex to easily record in a questionnaire.

Other ways to collect data electronically include sensors such as traffic cameras, satellites and online sources, such as websites or data logs. The data collected can also be used for a variety of purposes, including describing, predicting and improving processes within an organisation, or for research.

ISSUE

MH370 search: How new satellite data confirmed Malaysia Airlines plane was lost

[25 March 2014, Nick Miller, Europe correspondent] Sydney Morning Herald

British satellite company Inmarsat analyses seven, hourly pings sent by the missing Malaysian Airlines flight to determine its final resting place.

London: A new satellite tracking technique is what gave Malaysian Prime Minister Najib Razak enough confidence to announce that Malaysia Airlines flight MH370 went down in the remote south of the Indian Ocean.

British firm Inmarsat was behind an earlier analysis that indicated the plane had been flying in one of two big 'corridors', one in the northern hemisphere and one in the southern.

However last week it went back to its data and tried a new mathematical analysis, which concluded on Sunday.

The new analysis allowed them to discard the northern corridor, and focus more precisely on the southern route.

Based on this new information, Mr Najib announced on Monday that MH370's last known position was in the middle of the Indian Ocean, west of Perth.

UK firm Inmarsat was behind an earlier analysis of the path of missing flight MH370.

The nature of the pings indicated that the plane was still moving during that time.

'This is a remote location far from any possible landing sites,' he said. 'It is therefore with deep sadness and regret that I must inform you that … flight MH370 ended in the southern Indian Ocean.'

The aeroplane had Inmarsat's 'Classic Aero' satellite system, which collects information such as location, altitude, body heading and speed, and sends it through Inmarsat's satellites into their network.

This 'ACARS' (aircraft communications addressing and reporting system) was switched off or interrupted early in the flight, meaning no such information was available to track the plane.

However the Classic Aero system still sent hourly 'pings' back to Inmarsat's satellite for at least five hours after the aircraft left Malaysian airspace, the company discovered.

These pings contained no data – they were just a simple 'hello' to keep the link open – however, their timing and frequency contained hidden mathematical clues.

The company looked at the 'Doppler effect' – tiny changes in the frequency of the ping signal, caused by the relative movement of the satellite and the plane (the Doppler effect is the reason why, for example, police sirens are a different pitch or frequency depending on whether they are travelling toward you or away from you).

This analysis allowed Inmarsat to map two huge 'corridors' for the plane's possible location, in big arcs stretching thousands of kilometres north and south of the point where the last radar contact with MH370 was made.

Australian and US experts took this information, added some assumptions about the plane's speed, and narrowed the southern option into an area of ocean that could be realistically searched.

Meanwhile, Inmarsat went back to its satellite data. Its new analysis found that the northern route did not quite correlate with the frequency of the pings from the plane – meaning the plane must have been heading south.

It also suggested that the plane had been travelling at a steady cruising altitude above 30,000 feet.

They compared satellite data from MH370 with that from previous Malaysian Airlines Boeing 777 flights, going back a few weeks, in order to better model the movement of the plane.

'This really was a shot in the dark,' Chris McLaughlin, senior vice president of external affairs at Inmarsat told the BBC. 'It's a credit to the scientific team that they managed to model this.

'Just a single "ping" can be used to say the plane was both powered up and travelling. And then by a process of elimination comparing it to other known flights and established that it went south.'

The UK's Air Accidents Investigation Branch also contributed to the analysis.

THINK ABOUT COMPUTING 1.1

1 List the various data sources used to search for the missing plane.

2 Why is there a need to use a variety of data sources to assist with locating the missing plane?

3 How has the data assisted with estimating the location of the plane?

Data collection methods

Before we can produce information, we first must start with data. Data collection methods such as surveys/questionnaires, interviews or observation provide a means of capturing data.

Surveys and questionnaires

Surveys and questionnaires are common methods used to collect data. They can provide data about what the respondents think is true, or their preferences for consumer goods and political parties. A questionnaire can be a quick way of gathering large amounts of data. Questionnaires and surveys need to be carefully designed, otherwise the participants' responses may not provide suitable data to analyse, rendering them useless. Questions used in a survey must be carefully worded so that the response will provide meaningful and useful data without the need for further clarification.

Focus groups

A **focus group** is the meeting of a small group of individuals who are guided through a discussion by a researcher. The focus group is carefully selected, so it fits a particular demographic and the researcher can obtain the necessary data through a guided discussion that probes the participant's attitudes about the topic. Focus groups often comprise five to 12 people and the discussion is loosely structured to encourage ideas to flow.

Interviews

Interviews are used to elicit the opinions and beliefs of people. They can be used to gather data for research projects. Interviews are usually conducted one-to-one in a quiet, relaxed atmosphere. They should be recorded, with permission of the interviewee, with easily used and unobtrusive audio equipment or video. Writing down the responses during the interview is not helpful to the interviewer or the interviewee. Collating and analysing information can be difficult and time-consuming, and may require the use of someone with expertise.

Open-ended and closed questions

Questions used on a survey and during an interview can be open-ended or closed. **Closed questions** limit the responses available to the respondent (Figure 1.2). They include 'Yes/No' boxes, multiple-choice questions, and scales on which the attitudes and beliefs are measured, such as 'strongly agree', 'agree', 'disagree' or 'strongly disagree'.

Closed (or closed-ended) questions are generally considered to be quantitative in nature. They are called 'closed' because the range of answers the participant can choose is limited. Closed questions are also known as quantitative, as the response options can be converted to numbers. For example:

How often do you feel that you are overworked with homework?
5 I always feel overworked
4 I sometimes feel overworked
3 I occasionally feel overworked
2 I feel overworked once in a while
1 I never feel overworked

Each of these options can have a value placed next to them. However, we do not talk in numbers and we shouldn't create surveys that only have numbers. Questionnaires and surveys should be thought of as a conversation between the person asking the questions and the person answering them.

Open-ended questions do not limit the answers that can be given by the respondent (see Figure 1.3). They should be worded so that the responses received are capable of correct interpretation. For instance, if you asked the question, 'How do you feel about the widespread use of computer games?', the responses would probably be too broad to be usefully categorised and analysed. The wording must therefore limit the scope of the possible responses to specific areas of interest: 'How has the playing of computer games affected the school results for your children?' Open-ended questions also allow for follow-up questions, which are called probing questions, such as 'Why?' or 'Please give an example'. Such questions tend to elicit more detail.

Open-ended questions allow people to answer the question as they want to. They are called 'open-ended' because participants are free to answer in any manner they choose. Unlike closed questions, there are no response options specified. They are qualitative because responses are considered and measured by feel rather than by numbers.

Closed questions are easier to develop, quicker to administer and answer, easier to collate and analyse, and can provide a large and balanced sample; however, they may not be useful for complex issues. In this case, open-ended questions may be needed as they elicit greater detail in the responses, can bring forth unusual ideas and can show links between various aspects of the issues.

Closed questions
1 How long have you shopped at this store?
2 How many times per week do you go shopping?
3 How much do you spend per week?
4 Which of the following sources of information most influences your purchasing habits?
• Advertising pamphlets delivered to the home
• Newspaper advertisements
• Television promotions
• Recommendations from friends
5 Do you use a computer? • Yes • No
6 What time of day do you normally go shopping?

FIGURE 1.2 Closed questions should be designed to elicit short, straightforward answers

Open-ended questions
1 What is your opinion of the games available from this store?
2 How influential do you think the advertising campaign has been?
3 What are some of the errors in data entry that you have observed?
4 Describe the most frustrating experience you have had when using the computer system.
5 What are some of the problems you experience in receiving information on time?
6 What changes would you recommend to improve the billing system?

FIGURE 1.3 Open-ended questions try not to limit the answers the respondent can give

Observation

Observation is a way of finding out about the world around us. Using our senses, such as sight, smell and hearing, we are able to pick up detailed information about our environment. However, as a method of data collection, observation is more than just looking or listening, as we can be selective about what we perceive to be most useful to us. Researchers engaged in observation attempt to learn what life is like for someone in a particular setting, while they remain an outsider. While observing, they make careful notes of what they see, record all accounts including conversations and interactions. Observation generally takes place in community settings such as classrooms or locations believed to have some relevance to the research questions. Observation is unlike other forms of data collection tools, as the researcher approaches participants in their own environment rather than having the participants come to the researcher.

Quality of data and information

One aspect of identifying relevant data from a given data set is ensuring that the data is usable. To be usable, data must be relevant, accurate, free from bias and reliable.

Relevance

To produce usable information, data must be relevant. For example, if a computing department in an organisation is evaluating PC-only software, then surveying people who only use a Mac is irrelevant. The data collected from Mac users would not be relevant to the overall data collection.

Data also needs to be processed while it is current, because decision making should not be based on outdated data.

Accuracy

Data that is entered into a computer must be accurate. Transcription is often a cause of error. Transcription errors occur when the person entering the data misreads the information through, for example, a lapse in concentration, being interrupted or pressing the wrong key. It is easy to make a mistake when entering a large amount of data, particularly numbers with many digits that may not contain spaces or punctuation to signify thousands. Clearly, if the data collected is incorrect, the information produced will be incorrect. If data has been gathered from a primary source, it is a good idea to check it against this. If data has been gathered from a secondary source and is suspect, it is worthwhile verifying the data using other secondary sources.

Freedom from bias

Bias can easily creep into data and make the information processed from it unreliable. Several influences can result in the introduction of bias into data: namely, vested interest, timing, small sample size, bias through sorting and bias through graphic representations.

Vested interest

Bias can enter data if the respondent to a survey or interview has a **vested interest** in the outcome of the research. A common example is celebrities who are paid to promote particular products in commercials or social media. It would be unreasonable to trust their statements that one product is better than others purely based on the fact that they are celebrities; they are only saying what they have been paid to say and may not necessarily be providing an independent judgement that has been derived from research or experience.

Timing

The timing of the data collection may also introduce bias. For example, you plan to survey a sample of the population for their views about Australia becoming a republic. The data you gather may be biased if, just prior to the survey being conducted, a royal tour takes place and there is extensive media coverage about the Royal Family. The timing of the data collection would introduce bias because it coincides with a significant event that could influence the responses.

Note too that bias is not restricted to data gathered from surveys or during interviews. For example, suppose that Qantas needed to decide whether to schedule two new weekly flights to New York. The decision could depend on the demand for existing flights. If the airline collected data from bookings made over a four-week period just before or during a significant event, the data gathered would be biased. Such data should not be relied on for making this decision because the influence of this event on customer demand is irregular and unlikely to occur again.

Small sample size

Choosing a sample size that is too small may also create bias. The sample size and composition must be suitable for the purpose of the data collection and, usually, a larger sample size leads to greater precision, provided the sample composition is suitably representative of the target population. The sample size must be big enough to make any conclusions drawn and information produced credible. For example, if you wanted to determine whether or not the school uniform should be changed, it would be remiss to only survey students in your class. Not only would this sample not be representative of the student body, but it would also not include other stakeholders, such as parents and school administrators. Similarly, if you wanted to gather sales data over a four-day period to predict monthly sales at a fish and chip shop, this time scale would not be sufficient to make a prediction. For instance, by choosing the four Mondays in the month, you may be selecting the quietest trading days in the week. If you pick the four Fridays in the month, you may be picking the busiest trading days. When selecting a sample size, you need to ensure that it is representative of the whole population.

Bias through sorting

The way in which you sort lists can introduce bias, although frequently this is unavoidable. A classroom teacher often consults a class list that is sorted alphabetically; for example, to select students for special tasks. The list is biased towards students whose surnames appear early in the alphabet and thus at the top of the class list. If you need to hire an electrician and consult a paper-based telephone book or an online directory, it is more likely that you will pick an early entry than one from the second page of listings. Bias of this type is difficult to avoid, so it is preferable to educate the user to recognise that the output has built-in bias and to encourage strategies to overcome that bias.

Bias through graphic representations

Bias can occur through your choice of graphic type, scale used and size chosen. Graphic representations should be sized proportionally to avoid overstating or trivialising the importance of one of the **variables** involved. For example, in Figure 1.4, a teacher has created a graph to show the heights of her students. The graph does not really give a clear picture of how the heights vary. The bars look as if they are all of similar height, so it is difficult to see the differences between them. In contrast, the different heights in Figure 1.5 are more distinct. This has been achieved by decreasing the scale of the vertical axis. The variation in the student heights looks much greater, even though the data has not changed. This graph makes it easier to see the small differences.

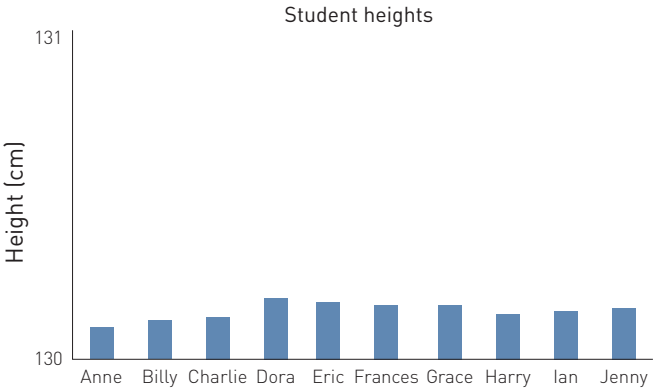


FIGURE 1.4 Column graph with little variation

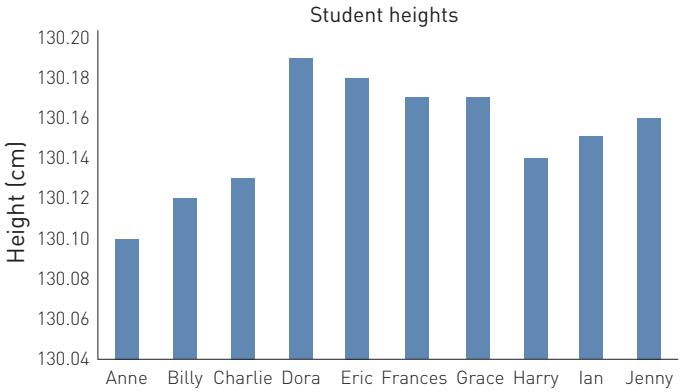


FIGURE 1.5 Column graph with greater variation

Reliability

The internet has made it easier for people to communicate their views and present information in a format that is easily accessible to others. There are many personal websites, homemade videos, wikis, podcasts, vodcasts and a plethora of unchecked information on the internet, and some of the views presented may not be widely accepted or proven to be accurate. Alternatively, there are many sources on the internet that are reliable such as the World Health Organization (WHO), Smithsonian, Australian Bureau of Statistics and universities.

THINK ABOUT COMPUTING 1.2
Find three reliable sources and describe in one paragraph the type of data they produce.

Referencing primary sources

Once primary data has been gathered, details need to be carefully recorded to enable appropriate **referencing**.

Interviews

For an interview, the following details need to be documented so that an interested person can go back to the source for checking, clarification and further information.

- Name of interviewee
- Date of interview
- Place of interview
- Qualification to be an interviewee – that is, whether the interviewee is a stakeholder in the issue and/or an expert about it

- Organisation to which the interviewee belongs (if relevant)
- Contact information for interviewee – phone number, address, email address, online chat handle
- How the interview was conducted; for example, in person, by phone, email or online chat
- Name and contact details of interviewer

Questionnaires

If you want to cite an individual response to a questionnaire or a survey, you need to record these details.

- Name of respondent
- When the questionnaire was completed
- Title of questionnaire
- Organisation to which the questionnaire belongs (if relevant)
- How the questionnaire was conducted – paper/online

Observation

For observation, the following details need to be recorded.

- The name of the person observed
- When the observation was conducted (date/time)
- Where the observation was conducted

Examples of referencing

Citations in a document help readers to find the source of the information and also assist students to avoid plagiarism. There are many ways to cite sources, such as providing footnotes, in-text citations or listing sources at the end of the document through a bibliography or references list.

Footnotes

Footnotes are listed at the bottom of the page on which a citation is made. Some academic disciplines prefer to use footnotes (notes at the foot of the page) to reference their writing. Although this method differs in style from the 'author, date' system, its purpose – to acknowledge the source of ideas, data or quotations without undue interruption to the flow of the writing – is the same. Footnotes are usually sequenced: series of numbers above the text (superscript) are placed in the appropriate part of the text to indicate the cited work and are matched at the bottom of the page after the footnote. A footnote lists the author, title and details of publication, in that order. Here is an example of a footnote:

> For example, in a 2009 article in the *Australian Financial Review*, journalist Jacqueline Maley wrote about the changing pattern of consumers' expenditure on leisure goods and services following the GFEC.[1]
>
> [1] Jacqueline Maley, 'Tough times bring home life's simple pleasures', *Australian Financial Review*, 7 September 2009, p. 3.

FIGURE 1.6 An example of a footnote: a superscript number is inserted in the body of the text; however, the full reference is provided at the bottom of the same page

APA

The American Psychological Association (APA) created a style guide to assist with academic writing such as publications, essays and books. The APA style is widely used and is one of the most common reference styles that students are expected to use. Citations within the text and their corresponding source details in a references list at the end of the work are necessary elements of the APA style. These show the reader where ideas and research have come from. Typically, when referencing using APA, the author's surname and the date of the publication are featured in the text. If quoting directly from the source, then the page number is also included. For example:

In some states, regimes of control on missions and reserves (a core element of the 'protectionist' system that dominated most of the twentieth century) continued into the 1970s or even 1980s. However, from the 1960s the dominant policy approach towards Indigenous Australians began to change, recognising, to some extent, Indigenous peoples' rights to preserve their cultures (Parbury 1999). Key events in this period included amendments to the *Commonwealth Electoral Act* in 1962 to give Indigenous people the vote in federal elections and the 1967 constitutional referendum, which meant Indigenous people would be counted in the national census (Attwood & Markus 2007) and the Commonwealth could legislate for them. By 1973, the new policy approach of Indigenous self-determination was introduced by the Federal Government (Sanders 2002). Along with the introduction of laws for equal wages, land rights and antidiscrimination in the 1960s and 1970s, these changes have sought to ensure fairer treatment of Indigenous Australians and better respect for their human rights.

Attwood, B. & Markus, A. 2007, *The 1967 Referendum: Race, power and the Australian Constitution*, Aboriginal Studies Press, Canberra.

Parbury, N. 1999, 'Aboriginal education: A history' in R. Craven (ed.), *Teaching Aboriginal Studies*, Allen & Unwin, Sydney, pp. 63–86.

FIGURE 1.7 An example of APA-style citation: the author's name and date of publication are inserted in the body of the text; the full reference is provided in the references list at the end of the document

Within the main body of the document, any reference to the publications will need to be cited. As illustrated in Figure 1.7, when using the APA style, the format usually follows author-date-page where the author's last name, the year of the publication, and the page number of the quote are referenced. These are all separated by commas, and are placed within parentheses following the text. The page number is preceded by a lower case 'p' with a period (full stop) after it; for example, (Parbury, 1999, p. 65).

EndNote is a commercial reference management software, used to manage bibliographies and references. It is similar to a database and is used to keep all references in one place. EndNote also integrates with MS Word, so that references can be easily inserted into the text with minimal effort, as shown in Figure 1.8. Alternatively, MS Word has its own built-in referencing capabilities that will allow you to create a bibliography and manage your sources without the need of additional software.

For more detailed information, visit the APA website.

For more information, visit the EndNote website.

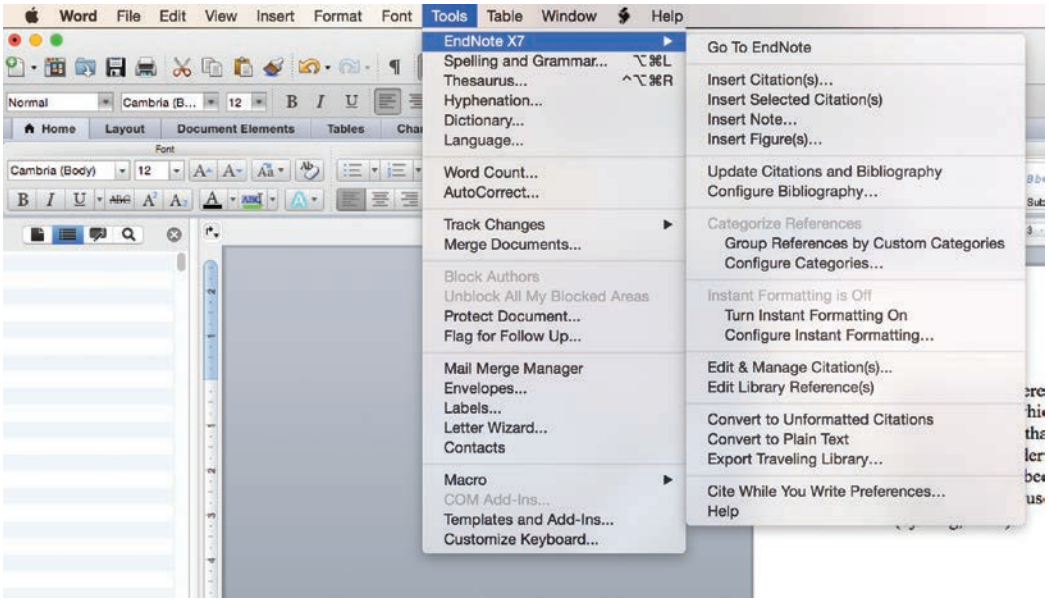
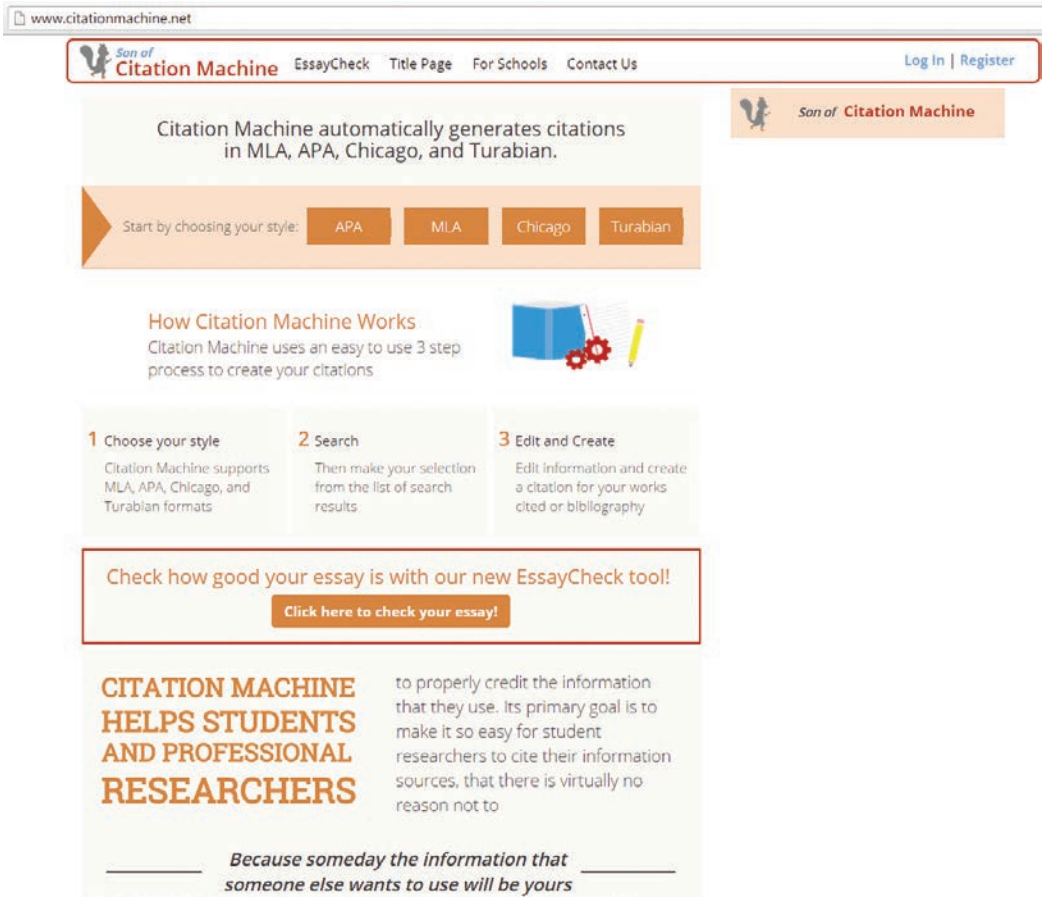


FIGURE 1.8 EndNote integrates easily into Microsoft Word



http://www.citationmachine.net

The free Citation Machine website, shown in Figure 1.9, generates citations in APA style. It can assist students to generate references in the appropriate format.

FIGURE 1.9 The Citation Machine can be used by students to assist with referencing

Seeking permission

Permission must be sought for collecting any data or information that involves people, because of privacy laws such as the Australian Privacy Principles (APPs). For example, to photograph or film individuals or groups, you should obtain permission. The organisation or individual who wants permission needs to let the people photographed or filmed know the purpose of the photographs or film and what it may be used for. Permission is usually provided in written form and is often known as consent. Permission needs to be sought because a photograph or video image in which an individual can be identified is considered to be personal information. Pictures of people can be used in advertisements, or for marketing purposes, and sometimes pictures can be used thoughtlessly and depict people in a false light.

Participants in any research need to be informed about what the research entails. They need to know what they are required to do and how much of their time it will take, and how often they will be required. All details of the research need to be given so that participants can make an informed decision to participate. For example, when researchers want to conduct a questionnaire, they need to specify how much time it will take (for example, 20 minutes) and how many times the questionnaire will need to be completed. Participants need to know whether the questionnaire will be paper-based or electronic, and how they will get access to it. All these details need to be explicitly stated so that each participant has a clear understanding before agreeing to take part. Participation needs to be voluntary. The researchers cannot put pressure on the participants or use coercion, or provide financial or other incentives for them to participate. Participation in research needs to be voluntary and informed.

In universities, research that involves people or animals cannot begin until researchers obtain ethics clearance. Each university has an ethics committee established and they follow the guidelines set out by the National Health and Medical Research Council (NHMRC). Before obtaining ethics clearance, the researcher/s must demonstrate that they have followed correct procedures and processes before collecting data to ensure that all risks have been addressed, and that it is established that participation is informed and voluntary.

Consent forms

One method of obtaining permission for research purposes is to use consent forms. **Informed consent** by all research participants is necessary. This means that before agreeing to participate in research, they are aware of what the research involves, the time required from them and the possible risks that may arise. Participation in research not only has to be informed, but also voluntary; that is, participants are not pressured to be involved and have the capacity to make their own decisions based on their understanding of the research.

You must obtain consent when interviewing or observing participants, creating questionnaires or surveys or collecting any type of data.

Consent forms should have the following information clearly listed.

- The title of the project
- The name of the researcher
- What the project is about and why it is being undertaken
- What is required from the participants in terms of time, effort, resources and costs
- The rights and interests of the participants – that they freely consent to be involved in the research and can withdraw at any time without having to provide a reason
- A statement of whether the participant's identity will be preserved

- Statements of confirmation, such as:
  - 'I have been informed of and understand the purposes of the study.'
  - 'I have been given an opportunity to ask questions.'
  - 'I understand I can withdraw at any time without prejudice.'
  - 'Any information that might potentially identify me will not be used in published material.'
  - 'I agree to participate in the study as outlined to me.'
- Name of participant, signature and date

FIGURE 1.10 A sample consent form

Name of school:

Project title:

Investigator(s):

1 I consent to participate in the project named above. I have been provided a copy of the project consent information statement to which this consent form relates and any questions I have asked have been answered to my satisfaction.

2 In relation to this project, please circle your response to the following:

• I agree to be interviewed by the researcher	**Yes**	**No**
• I agree to allow the interview to be recorded by electronic device	**Yes**	**No**
• I agree to make myself available for further information if required	**Yes**	**No**
• I agree to complete questionnaires asking me about <Insert topic>	**Yes**	**No**

3 I acknowledge that:

a my participation is voluntary and that I am free to withdraw from the project at any time without explanation;

b the project is for the purpose of research and not for profit;

c any identifiable information about me which is gathered in the course of and as the result of my participating in this project will be **(i)** collected and retained for the purpose of this project and **(ii)** accessed and analysed by the researcher(s) for the purpose of conducting this project;

d my anonymity is preserved and I will not be identified in publications or otherwise without my express written consent.

By signing this document I agree to participate in this project.

Participant name: .....

Signature and date: .....

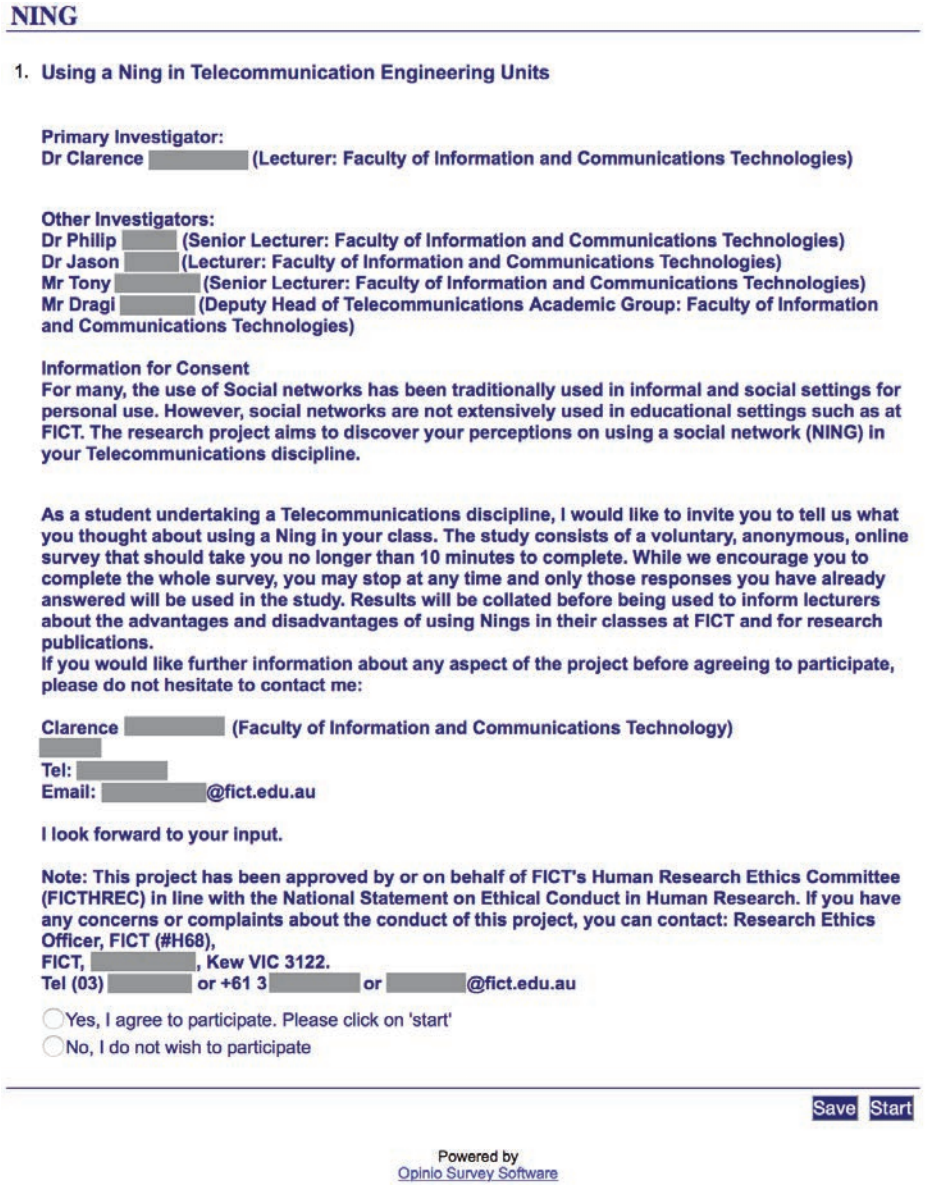
Participation information statements

The most common way to obtain informed consent is to provide a participant information statement written in plain English along with a consent form. This provides participants with information about the research and also provides the scope to answer questions that participants may have. Once they have read and understood the information statement, they are asked to sign the consent form to indicate their agreement to participate. When developing a participation information statement, consider the following.

- Use language your audience can easily understand.
- If technical terms are required, make sure they are clearly explained the first time they are used.
- The statement should be suitable for a 12-year-old reading age.
- Write in a conversational style, as if you were speaking to the participant.
- Language used should be clear, concise, invitational, culturally appropriate and logically set out.

- Use pronouns, such as 'I', 'we' and 'you'. This encourages the use of active voice and will be clearer to the reader. For example: 'You will be asked to participate'.
- Use reader-friendly formatting so that your document is easy to read.
- Ask others to read and edit your document.
- Include a statement on how the data will be stored.
- State where and how the research findings will be published.
- Give further information about the project, such as whom to contact.

Even if the data collection technique is an online survey and the researchers do not meet the participants, respondents must still be informed of the research, and asked to consent online prior to completing the questionnaire. They may be asked to consent by accepting and selecting the 'I agree to participate'. If they choose the 'I do not wish to participate' option, then they do not get access to the questionnaire. See Figure 1.11 for an example of an integrated online information statement and consent form.



NING

1. **Using a Ning in Telecommunication Engineering Units**

Primary Investigator:
Dr Clarence [redacted] (Lecturer: Faculty of Information and Communications Technologies)

Other Investigators:
Dr Philip [redacted] (Senior Lecturer: Faculty of Information and Communications Technologies)
Dr Jason [redacted] (Lecturer: Faculty of Information and Communications Technologies)
Mr Tony [redacted] (Senior Lecturer: Faculty of Information and Communications Technologies)
Mr Dragi [redacted] (Deputy Head of Telecommunications Academic Group: Faculty of Information and Communications Technologies)

Information for Consent
For many, the use of Social networks has been traditionally used in informal and social settings for personal use. However, social networks are not extensively used in educational settings such as at FICT. The research project aims to discover your perceptions on using a social network (NING) in your Telecommunications discipline.

As a student undertaking a Telecommunications discipline, I would like to invite you to tell us what you thought about using a Ning in your class. The study consists of a voluntary, anonymous, online survey that should take you no longer than 10 minutes to complete. While we encourage you to complete the whole survey, you may stop at any time and only those responses you have already answered will be used in the study. Results will be collated before being used to inform lecturers about the advantages and disadvantages of using Nings in their classes at FICT and for research publications.
If you would like further information about any aspect of the project before agreeing to participate, please do not hesitate to contact me:

Clarence [redacted] (Faculty of Information and Communications Technology)
[redacted]
Tel: [redacted]
Email: [redacted]@fict.edu.au

I look forward to your input.

Note: This project has been approved by or on behalf of FICT's Human Research Ethics Committee (FICTHREC) in line with the National Statement on Ethical Conduct in Human Research. If you have any concerns or complaints about the conduct of this project, you can contact: Research Ethics Officer, FICT (#H68),
FICT, [redacted], Kew VIC 3122.
Tel (03) [redacted] or +61 3 [redacted] or [redacted]@fict.edu.au

Yes, I agree to participate. Please click on 'start'
No, I do not wish to participate

Save Start

Powered by
Opinio Survey Software

FIGURE 1.11 Example of an integrated online information statement and consent form. Note that personal information has been redacted for privacy reasons.

The consent information statement should be clearly labelled as such (or quite similar) and contain key pieces of information as follows:

Project title
Give the project a full title

Investigators and other project personnel
List clearly all researchers.

Introduction to project and invitation to participate
Both introduce the project and invite participation.

What this project is about and why it is being undertaken
Sufficiently explain what the project is about, its aims, why it is being conducted.

Project and researcher interests
Researcher and project interests should be sufficiently disclosed as applicable; for example, project is partly, mainly or wholly to satisfy the requirements for a student's VCE subject.

What participation will involve – time, effort, resources, costs
Give sufficient clear detail as to what is being asked of participants – voluntary consent to their time, effort, supply of information/body tissue/records/personal effects, etc. Avoid language that can be read as orders or directives (i.e., not 'You will do this or that'; but better as 'We will ask you', etc.) and presumptuous language (e.g., 'Dear Participant').

Participant rights and interests – risks and benefits/contingencies/back-up support
Outline realistically any potential risks (minimal or otherwise) and what preventative, minimisation or redress arrangements are in place. If some research questions or issues can be considered particularly sensitive, give sample questions or topics as an indication of the information that will be discussed or requested. Describe any benefits pertaining to individual participation or more generally. Avoid grandiose claims.

Participant rights and interests – free consent/withdrawal from participation
Participation should be voluntary, free from any coercion or perceived coercion. Detail on this matter should be clear – that an individual is free to participate or not and the circumstances. If, for example, the participants are students, patients or employees, it will help to clarify that their decision to whether or not to participate will have no bearing on their results, treatment or employment (in some cases this may need further explanation, such as details about recruitment).

A statement about the participant's right to withdraw participation, data or material contributed, ordinarily without question or explanation, needs to be included.

Don't forget to outline how valid consent is to be obtained (by signed consent form, completion and return of an anonymous questionnaire, any witnessing procedure if applicable, etc.). Often it helps to highlight or bold this part.

Participant rights and interests – privacy and confidentiality
Give clear information about secure arrangements for data access, collection, use, retention and/or disposal. This needs to comply with mandatory Australian Privacy Principles.

If signed consent forms are required, state whether they will be stored separately from any data collected and who will have access to them.

Remember, people are increasingly concerned about data access and data matching. Clear information will help allay any concerns.

Research output
Outline intended or anticipated publication or reporting of research findings. If need be, reiterate or refer to privacy arrangements for confidentiality/anonymity. Offer to make available any report or article or summary, where appropriate, and indicate how this will occur.

Further information about the project – whom to contact
You will need to nominate at least one person to contact regarding further information about the research activity or participation in the project.

FIGURE 1.12 Information that should be included in a participation information statement

Privacy

When undertaking research, it is very important to uphold the privacy of the participants. Privacy is a fine balance between the interests of researchers and participants. Privacy laws attempt to stop inappropriate intrusion into the lives of individuals. However, the collection of data is often not the problem, but how the data is used or misused by people entrusted with it. To maintain privacy, and to de-identify the data, personal identifiers such as names and birthdates that are associated with individuals need to be removed so that information cannot be traced or identified.

TABLE 1.1 Data that has been collected but not de-identified

Given name	Surname	Height (cm)	Weight (kg)	Sex	Birthday	Postcode	Age	Active
Charlotte	Maine	145	48	F	22/10/2003	3103	11	No
Claudette	Shine	153	40	F	10/10/2003	3105	11	Yes
Abdul	Mensur	160	37	M	19/10/2002	3122	12	Yes
Eric	Anton	142	38	M	12/12/2003	3040	11	Yes
Jaspreet	Singh	148	41	M	31/12/2002	3041	12	No
Sienna	Megane	152	42	F	2/12/2003	3103	11	Yes
Elijah	Wu	144	43	M	6/01/2003	3122	11	No
Josephine	Wu	149	47	F	6/01/2003	3122	11	Yes
Hamish	Green	154	37	M	2/02/2003	3044	11	No
Jacinda	Black	143	46	F	4/04/2003	3111	11	No

TABLE 1.2 Data that has been de-identified: personal information such as birthdates, first names and surnames has been removed so that data cannot be identified or traced

Participant	Height (cm)	Weight (kg)	Sex	Postcode	Age	Active
ID1001	145	48	F	3103	11	No
ID1002	153	40	F	3105	11	Yes
ID1003	160	37	M	3122	12	Yes
ID1004	142	38	M	3040	11	Yes
ID1005	148	41	M	3041	12	No
ID1006	152	42	F	3103	11	Yes
ID1007	144	43	M	3122	11	No
ID1008	149	47	F	3122	11	Yes
ID1009	154	37	M	3044	11	No
ID1010	143	46	F	3111	11	No

Care must be taken to de-identify data. The example we have provided of de-identifying data only removes the names and dates of birth of survey participants. Depending on the remaining data and the information generated, this may be insufficient to protect the privacy of individual research participants from being identified (even accidentally) by at least some users of the information. This is particularly likely if the users know who the members of the sample group are, or if they are familiar with the wider population that is being sampled. Users of the information may be able to identify a survey participant even after his or her name and date of birth have been removed. This could be caused by a characteristic known to be rare in the sample group or wider population, because this helps to narrow possibilities. Combinations of rare characteristics would help to narrow them even further, so that users of the information could make an intelligent guess as to a person's identity.

Physical security controls

Researchers must ensure that data and materials generated and collected as part of their research, regardless of the format, are stored securely in a durable and accessible form. Data must be stored in a way that meets all legal and confidentiality requirements. Stored data can be protected with both physical- and software-based controls such as backing up data or shredding confidential documents.

Data needs to be securely retained and then securely disposed of or destroyed when no longer needed. How long the data or material need to be retained depends on the type of research or data. The minimum standard is set out in Section 2 of the *Australian Code for the Responsible Conduct of Research 2007*. The minimum timeframe for data to be retained is five years. For some specific types of research, other timeframes may be required. For data such as identifiable health research data, the minimum period may need to be seven years. For identifiable health research data involving children, the data may need to be kept until the individuals involved turn 25, or five years after any research outcome based on the data is published or made available for release, whichever is the longer.

Encryption

Encryption is the process of translating data into a secret code that can only be read by authorised users. To read an encrypted file, you must have access to a secret key that you use to **decrypt** the data. **Unencrypted data** is also known as 'plaintext'. Data that has been collected for research purposes needs to be secured so that only authorised people have access. Encryption is one way to ensure that the data is secured.

Backups

Backups form an essential step in data management. Regular backups protect against a number of risks including human error, computer crashes and software faults. Critical data files or data that are used regularly should be backed up frequently.

It is not unusual for someone to accidentally delete an important file or edit a document and later realise that some important information was removed. One strategy to minimise loss of data by an authorised user is to maintain a backup system. Important files inadvertently lost can be retrieved from the backup media.

A **full backup** copies all of the files from a device to a storage medium. It can take considerable time and is usually performed once over a time period (such as a week, fortnight or month). A **differential backup** copies only those files that have been changed since the last full backup. Restoration of data would involve restoring files from the full backup and then from the differential backup. An **incremental backup** is similar to a differential backup, the difference being that it uses more than two backup media, while a differential backup uses only two media. An incremental backup only copies files that have been changed since the last incremental backup. It is the most complicated strategy from which to restore files since it requires restoration from a full backup and then from a series of incremental backups. It is a good practice to clearly label all backup media so that you know when the backup was made and what is on it.

Location of backup files

Once you have created backups, where do you put them? Ideally, your backups should be stored in a location that is safe from theft and damage caused by extremes of temperature or disasters. Most small businesses have a fireproof and waterproof safe in which valuable company

documents are stored. This might also be used to store backups. It is preferable, however, to store backups at a remote location, perhaps even in the cloud. This means that if there is a large natural disaster, such as a huge flood or an earthquake, the backups will be safe.

One last point to remember is to ensure that backups actually work when you want to restore the data. It is important to test the effectiveness of your backup files by running a disaster recovery simulation. If files cannot be restored from the backup or the system refuses to recognise them, it is better to discover this before a real emergency.

Cloud-computing companies provide offsite storage, processing and computer resources to individuals and organisations. These companies are typically third party and they store data to a remote database in real-time. The internet provides the connection between this database and the user's computer. One of the advantages of cloud storage include the ability to access data from any location that has internet access, eliminating the need to carry a USB, or a hard-drive to retrieve and store data. The ability to share files with other people and collaborate simultaneously, such as by using Google Docs, is also an advantage. Finally, if something were to happen to the computer, such as a fire or natural disaster, and the data on it were to be destroyed, having the data saved offsite in the cloud would prevent the data from being lost.

Google Docs allows users to upload documents, spreadsheets and presentations to Google's servers (Figure 1.13). Users can edit files using a Google application and work on them at the same time as others, so they can read or make edits simultaneously.

The word 'cloud' is used because the internet was originally, and still is, represented by a cloud in network diagrams. An example of cloud computing is Google's Gmail, which uses cloud-computing processing powers and storage facilities.

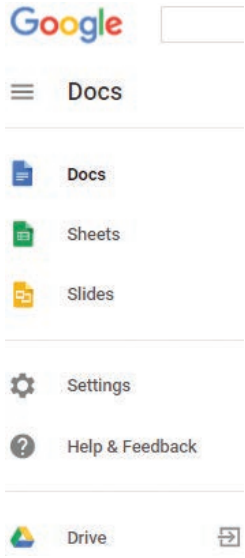


FIGURE 1.13 Google Docs

Usernames and passwords

A username usually is identifiable as belonging to a particular person and can be easily remembered. Usernames are uniquely assigned to users. Passwords are set by the user and should be known only to that user. To maintain high levels of security, user passwords should:

- be at least eight digits long
- include non-alphabetical characters
- not be easily guessed (for example, a favourite pet's name is not suitable)
- be changed every month.

Some network policies force passwords to be changed on a regular basis, and do not allow passwords to be repeated.

In terms of securing individual files, a password can be placed by an individual to prevent unauthorised people accessing the data. Password protection can be placed on files to prevent them from being viewed by unauthorised people, as shown in Figure 1.14.

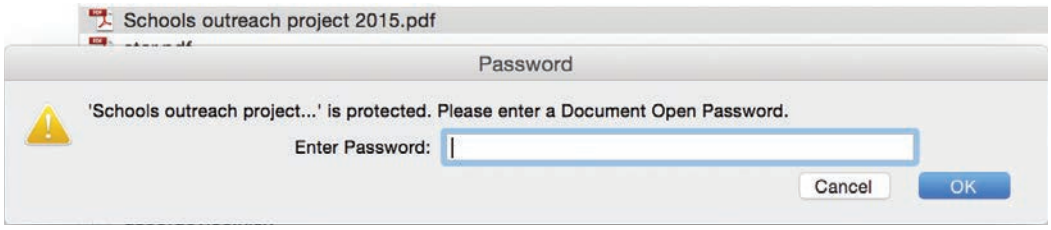


FIGURE 1.14 Password protection has been placed on the file, preventing access to it unless a password is entered

Firewall

A **firewall** is a server and software combination that filters the information coming through an internet connection into an organisation's internal network. Any packet of data that is flagged by the filters as unwanted is not allowed through.

The filters used by a firewall include examining the IP address of computers that request information from an internal server, blocking all access to certain domain names, banning certain protocols (e.g. file transfer protocol, mail protocol or Telnet protocol) from accessing particular servers, and searching for certain words and phrases included in packets of information, and blocking transfers that contain them.

A firewall can also be used to restrict employees' access to sensitive information. For example, a firewall can be used to stop some personnel from accessing the payroll database.

Most firewalls use two separate network interface controllers (NICs); one is connected to the internal network and the other to the outside world. Material can only move from one card to the other through the CPU of the server computer that is acting as the firewall. While the data or information is being checked for authenticity, it is also examined for viruses and other malicious codes. Everything that comes in from outside is examined for danger.

Firewalls are explained in more detail in Chapter 3.

Malware protection

Malware is malicious software that includes spyware, adware, Trojan horses, worms and viruses. Spyware and adware use cookies to track the internet sites that a user might visit. Trojan horses can leave your computer open to others to read your personal information by creating backdoor access to your system. Viruses and worms can hijack your system to send multiple emails to others or perform other acts of mischief. Both can use up essential system resources, which may result in the computer freezing.

Network administrators usually require workstations to run virus protection software. The antivirus software is often updated automatically via the network. A firewall is also useful to block malware from sending personal information over the internet. Anti-adware programs should also be run on workstations.

Australian Privacy Principles

Originally, the *Privacy Act 1988* only dealt with the handling of data by Federal government agencies. Many people criticised this limitation because it seemed that private organisations were not required to apply even the most basic of safeguards on data they collected. Even worse, there were no regulations preventing non-government organisations from collecting data by any method and using it for any purpose without the consent of the people whose private details were concerned. In particular, the rapid growth of electronic transactions, especially over the internet, led many people to demand some sort of legal protection from those who might gather data about internet browsing habits. The government was keen to encourage the development of electronic commerce, while protecting the confidentiality of consumers and increasing public confidence in electronic transactions. These amendments have now been incorporated into the *Privacy Act 1988* (Cth) and are the most significant changes to privacy laws since the inception of the legislation.

There have been several additional powers included within this Act since 1988, but its main purpose has remained unchanged. The *Privacy Act 1988* (Cth) was amended by the Privacy Amendment (Enhancing Privacy Protection) Bill in 2012. This came into effect in 2014. As part of this Act, the Australian Privacy Principles replaced the National Privacy Principles and the Information Privacy Principles so that Australia now has one set of privacy principles. The Australian Privacy Principles (APPs) generally apply to Australian Government agencies. They do not apply to Local Councils or State or Territory Governments. Some States have their own privacy laws. For example Victoria has The *Privacy and Data Protection Act 2014* (PDPA).

The changes to the *Privacy Act* include 13 new privacy policy principles known as the Australian Privacy Principles (APPs). The APPs were devised to set out the standards, rights and obligations for collecting, handling, holding, accessing, using, disclosing and correcting personal information. The APPs oversee the handling of personal information by:

- Australian and Norfolk Island Government agencies
- all private health service providers
- businesses that have an annual turnover of $3 million or those that trade personal information.

Of particular interest are Principle 2, *Anonymity and pseudonymity*, and Principle 6, *Use or disclosure of personal information*.

TABLE 1.3 The Australian Privacy Principles

APP 1 Open and transparent management of personal information Ensures that APP entities manage personal information in an open and transparent way. This includes having a clearly expressed and up-to-date APP privacy policy.
APP 2 Anonymity and pseudonymity Requires APP entities to give individuals the option of not identifying themselves, or of using a pseudonym. Limited exceptions apply.
APP 3 Collection of solicited personal information Outlines when an APP entity can collect personal information that is solicited. It applies higher standards to the collection of 'sensitive' information.
APP 4 Dealing with unsolicited personal information Outlines how APP entities must deal with unsolicited personal information.
APP 5 Notification of the collection of personal information Outlines when and in what circumstances an APP entity that collects personal information must notify an individual of certain matters.
APP 6 Use or disclosure of personal information Outlines the circumstances in which an APP entity may use or disclose personal information that it holds.
APP 7 Direct marketing An organisation may only use or disclose personal information for direct marketing purposes if certain conditions are met.
APP 8 Cross-border disclosure of personal information Outlines the steps an APP entity must take to protect personal information before it is disclosed overseas.
APP 9 Adoption, use or disclosure of government related identifiers Outlines the limited circumstances when an organisation may adopt a government related identifier of an individual as its own identifier, or use or disclose a government related identifier of an individual.
APP 10 Quality of personal information An APP entity must take reasonable steps to ensure the personal information it collects is accurate, up to date and complete. An entity must also take reasonable steps to ensure the personal information it uses or discloses is accurate, up to date, complete and relevant, having regard to the purpose of the use or disclosure.
APP 11 Security of personal information An APP entity must take reasonable steps to protect personal information it holds from misuse, interference and loss, and from unauthorised access, modification or disclosure. An entity has obligations to destroy or de-identify personal information in certain circumstances.
APP 12 Access to personal information Outlines an APP entity's obligations when an individual requests to be given access to personal information held about them by the entity. This includes a requirement to provide access unless a specific exception applies.
APP 13 Correction of personal information Outlines an APP entity's obligations in relation to correcting the personal information it holds about individuals.

Office of the Australian Information Commissioner website – www.oaic.gov.au, Creative Commons Attribution 3.0 Licence

APP 2 Anonymity and pseudonymity

Australian Privacy Principle 2 provides individuals dealing with organisations the option of using a different name or a **pseudonym** in relation to a particular matter. This measure is in place so that individuals cannot be identified. In addition, individuals can also remain anonymous. For example, when an individual calls an organisation, often a message states that the call will be recorded for training purposes. If the individual objects, the call is not recorded. At a later date, when staff from the organisation receive training, the names of the individuals whose voices have been recorded must be changed in order to protect their identity when these real examples are used.

APP 6 Use and disclosure of personal information

Australian Privacy Principle 6 states that the information that is being held is in line with the primary purpose it was intended for. Information cannot be used for a secondary purpose unless the holders of the information have received consent from the individuals concerned. For example, if a sporting organisation collected information about their players for the purpose of organising competitions and making the details available to the coaches and captains, then that would be the primary purpose of the data. However, if the sporting organisation wanted to provide the details of their members to a shop that specialised in sporting merchandise, then it could not do so, as this would be using the information for a purpose for which it was not intended.

Ethical dilemmas

Everyone wants the benefits of digital systems; however, intended and unintended negative effects can impinge upon people's rights. As a result, those who design, control and use digital systems have a responsibility to consider the real and potential negative effects and to eliminate or lessen them as much as possible. Sometimes even this may not be enough to justify the proposed collection or creation of data. It is important to take into account legal objections and ethical considerations when creating or acquiring data. The purpose for collection needs to be clear. This also needs to be articulated in the participant information statements and the consent forms provided to the people from whom information will be sought.

Ethics refers to behaving in ways that are based on our morals and accepted standards. These standards may be common in a particular society or specific to a single organisation. They apply to questionable activities over and above any legal requirements. Ethics often provide us with a set of guidelines of appropriate behaviour. If we choose to ignore these guidelines, we may not be committing a crime, but we may be sacked by an employer or shunned by society. For example, the impact of violent video games on children has long been debated. Some people have voiced their concerns that video game writers should not include animated violence in their games because it has a negative impact on children. A system of classification exists for games, similar to television and film classifications, but there is no legal restriction preventing these games from being created. These examples demonstrate how ethics hinge on society's values and standards. In this example, there are two competing principles. On the one hand, some proponents would argue that protecting children from possibly harmful video games is the right thing to do. On the other hand, others would argue that it is more important to maintain freedom of expression. Often, then, questions of ethics become debates over which

of two principles is more important. Such conflict may be said to constitute a **dilemma**. This is especially the case when the consequences of action are open to debate or interpretation.

The standards or guidelines that determine whether an action is good or bad are known as ethics. Ethics are the moral guidelines that govern, among other things, the use of data collection. Often ethical principles/guidelines have an accompanying law, but the ethical principle is usually broader and the law applies only to certain circumstances or applications of the principle.

For example, it is ethical to obtain permission to publish photos of people on websites or in promotional material. Sometimes people may object to their images being used for these purposes. The purpose for taking the photo and how it is intended to be used need to be made clear. Ethically it is wrong to use a photo for a different purpose from that for which it was originally collected. Similarly, when using data-collection tools such as surveys, interviews and questionnaires, it is important to reassure participants that the data provided, within the limits of the law, will remain anonymous and that their individual comments will not be able to be identified by others. It is not simply that it is important to put participants' minds at rest regarding their concerns about protecting privacy; it is also important to ensure that their privacy is in fact protected (and also to ensure that non-participants in the larger group of which the sample is supposed to be representative, are not put at unacceptable risk of suffering as a result of mistaken identification).

This may be of particular concern when deciding what must be removed to de-identify data sufficiently to protect all its potential users. It is also important when reporting personal information anonymously or using pseudonyms in a newspaper report, for example.

Ethical frameworks assist people to work through ethical dilemmas. A six-step framework as described below can be used for handling ethical dilemmas to solve a legal, ethical or social tension. This framework will provide support and guidance in making a decision when presented with an ethical dilemma.

1 Identify the problem: What decision has to be made and what facts are required?
2 Identify the stakeholders: Who are they? What interests do they have? What power do they have? Who is vulnerable? How are the vulnerable to be protected?
3 Identify possible alternatives: What options are available? What are the likely consequences?
4 Identify ethical standards: Are there any applicable laws? Are there any morals or standards that could be applied? Is there a precedent?
5 Evaluate options: Identify strengths and weaknesses. Identify the option that causes least harm. Can the decision be reversed?
6 Make a decision: Select the preferred option. Justify the option.

CHAPTER SUMMARY

ESSENTIAL TERMS

backup copying files from an information system to some type of storage device to guard against possible data loss

bias a prejudicial or unreasoned judgement

closed (or closed-ended) questions queries that are restricted in the range of options provided so that only specific answers are elicited

data unprocessed, unorganised and distinct facts or ideas; in addition to text and numbers, data also includes sounds, images and video

decrypt to translate encrypted data back into ordinary text that can be read by anyone

differential backup used in conjunction with a full backup, only files that have been altered since the last full backup are copied; restoration requires the full backup to be restored first, followed by files from the differential backup

dilemma when people must choose between two (or more) equally desirable (or undesirable) options; for example, between allowing the sale of violent video games to preserve freedom of expression and banning their sale in order to protect children from possible harm

encryption the process of translating data into a secret code that can only be read by authorised users; to read an encrypted file, you must have access to a secret key that you use to decrypt the data; encrypted data is also known as 'ciphertext'

ethics guidelines based on our morals and accepted standards; these standards may be common in a particular society or specific to a single organisation, and apply to questionable activities over and above any legal requirements – ethics often provide us with a set of guidelines of appropriate behaviour

firewall hardware and software that restrict access to data and information on a network

focus group the meeting of a small group of individuals who are guided through a discussion by a researcher; focus group members are carefully selected, so that each one fits a particular demographic and the researcher can obtain the necessary data through a guided discussion that probes the participant's attitudes about the topic

footnotes a type of reference that is listed at the bottom of the page on which the citation is made

full backup copying all chosen files to a backup device; it can be slow to perform, but is the easiest and quickest way to restore data form

hypothesis a statement based on probabilities proposing what will happen according to the applied theory; some hypotheses are tested using data collection tools such as questionnaires and/or interviews, and the results of the study will either support or disprove the hypothesis

incremental backup similar to a differential backup in that it works in conjunction with a full backup, but only backs up files that have been altered since the last incremental backup; it is the most complicated strategy from which to restore files

information processed, organised and value-added data, which can be paper-based (hard copy) or digital (soft copy)

informed consent a necessity for all participants before agreeing to take part in research, participants must be informed of what the research involves, the time commitment expected and the possible risks that may arise

malware programs designed to infiltrate and cause harm to a computer or network without the owner's knowledge or consent, including viruses, worms, Trojans, adware, spyware, logic bombs and keyloggers

observation a way of finding out about the world around us; using our five senses, we pick up detailed information about our environment

open-ended questions queries that allow people to answer the question in the manner they wish; they are called open ended because the person responding is free to answer in any manner he or she chooses

participant information statement a document that provides participants with information about the research in an unbiased way, and also provides the scope to answer questions that the participant may have

primary sources original, uninterpreted data and resources; that is, information that has not been processed, analysed or interpreted in any way, such as interviews, speeches, emails, debates and meetings; primary data usually come from stakeholders

pseudonym a fictitious name that is given to a person, or that is chosen by a person, to hide or protect their identity

qualitative data collected data that is harder to measure because it is based on subjective data collection techniques such as interviews, focus groups, video footage and observation

quantitative data collected data that is measurable and specific; quantitative data gathering is based on verifying theory through the use of statistics and largely numerical data

questionnaire or **survey** usually a set of questions that ask for a response to be selected from a list of alternatives, such as A, B, C, D; or from a range, 1–5 or very low to very high; questionnaires can easily be given to many people, and are easily processed and analysed using computer-based methods because the answers can be recorded as numbers

referencing citations in a document that assist readers to know where to locate the source of an original idea or quote in a piece of work, and assist students to avoid plagiarism

secondary sources information that has been processed, interpreted or analysed in some way by others, such as textbooks, websites, magazines, newspapers and TV programs

stakeholder an individual or group who either has an interest in or is affected by the decisions and actions of an organisation

survey see **questionnaire**

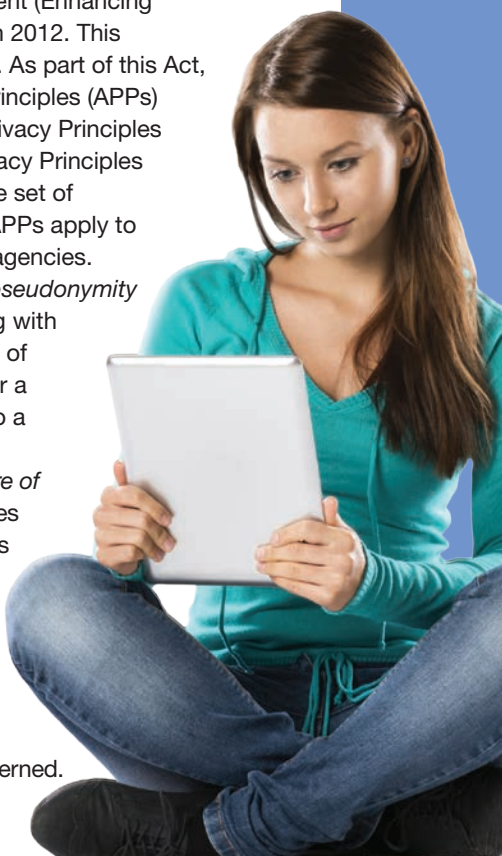
unencrypted data also known as 'plaintext'

variable in programming, a key word, phrase or symbol that represents a value that may change

vested interest arises when an individual, group or organisation has a strong personal interest because there is an advantage to be gained

IMPORTANT FACTS

1 Data must be relevant to produce usable information. Data needs to be processed while it is current because decision-making should not be based on outdated data.
2 Data that is entered into a computer must be accurate. Transcription is often a cause of error. Transcription errors occur when the person entering the data misreads the information through, for example, a lapse in concentration, being interrupted or pressing the wrong key.
3 Interviews are usually done one-to-one, but can sometimes be done in groups, and can take a substantial amount of time. A major feature of an interview is the opportunity for in-depth follow-up and clarification questions that cannot be done with questionnaires, which are often answered in private. Interviews are very useful for eliciting the feelings, attitudes and opinions of people that are too complex to easily record in a questionnaire.
4 Bias can infiltrate data if the respondent to a survey or interview has a vested interest in the outcome of the research, if the timing of the data gathering is inappropriate, or if the sample size chosen is too small.
5 Timing of events needs consideration when collecting data as it can cause skewed results, which can lead to inaccurate or misleading conclusions.
6 Sample size must relate to the purpose of the data collection and, generally, a larger sample size leads to greater precision.
7 There is a plethora of unchecked information on the internet; however, some of the views presented may not be widely accepted or proven. Sources cited should be reliable.
8 The American Psychological Association (APA) created a style guide to assist with academic writing such as essays, books and other publications. The APA style is widely used and is one of the most common styles that students are expected to reference.
9 Privacy is a fine balance between the interest of researchers and the participants. Privacy laws attempt to stop inappropriate intrusion into the lives of individuals. However, the collection of data is often not the problem, but how the data is used or misused by people entrusted with it.
10 Researchers must ensure that data and materials generated and collected as part of their research, regardless of the format, are stored securely in a durable and accessible form. Stored data can be protected with both physical and software-based controls such as backing up and shredding confidential documents.
11 Cloud-computing companies provide offsite storage, processing and computer resources to individuals and organisations. These companies are typically third party and they store data to a remote database in real-time.
12 To verify users' rights to access a network, security features are required. A system of establishing usernames (or user IDs) and passwords allows for the identification and authentication of each user.
13 The *Privacy Act 1988* (Cth) was amended by the Privacy Amendment (Enhancing Privacy Protection) Bill in 2012. This came into effect in 2014. As part of this Act, the Australian Privacy Principles (APPs) replaced the National Privacy Principles and the Information Privacy Principles so Australia now has one set of privacy principles. The APPs apply to Australian Government agencies.
14 APP 2, *Anonymity and pseudonymity* offers individuals dealing with organisations the option of using a different name or a pseudonym in relation to a particular matter.
15 APP 6, *Use and disclosure of personal information* states that the information that is being held is in line with the primary purpose it was intended for. Information cannot be used for a secondary purpose without consent from the individuals concerned.



CHAPTER SUMMARY

TEST YOUR KNOWLEDGE

Review quiz

DATA AND INFORMATION

1 What can occur if information is produced from incorrect or incomplete data?
2 Why is it important to ensure that data is accurate?
3 What are the properties of usable data?
4 What is a common cause of inaccurate data?
5 How can the accuracy of a primary source be determined?
6 With an example, explain the importance of timeliness in ensuring the quality and usability of data.
7 What influences the introduction of bias into data?
8 What is the difference between quantitative and qualitative data? Provide two examples.
9 What strategies could be used when gathering quantitative data?
10 Provide an example of referencing based on APA.
11 Why is it important to obtain permission when collecting data?
12 What is the purpose of consent forms?

DIGITAL SYSTEMS

13 Why is encryption important in data security?
14 How do usernames and passwords protect data?
15 What makes a strong password? Provide an example of a very strong password.
16 Describe a strategy on backing up data?
17 What is the difference between a full backup and an incremental backup?
18 Why is it important to secure data when conducting research?

INTERACTIONS AND IMPACT

19 How do the Australian Privacy Principles affect the individual?
20 Under which legislation does the Australian Privacy Principles fall?
21 Why is it important to de-identify personal data?
22 What is an ethical dilemma in the context of data collection for research purposes?

APPLY YOUR KNOWLEDGE

STREET TRAFFIC

1 Arthur believes that there is too much traffic in his street, particularly on weekdays – in the morning and evening. Arthur lives in a small residential street, but cars use this street as a shortcut to avoid traffic lights and the main roads. Arthur wants to do some research to support his theory about the traffic on his street, and present a case to the local council.

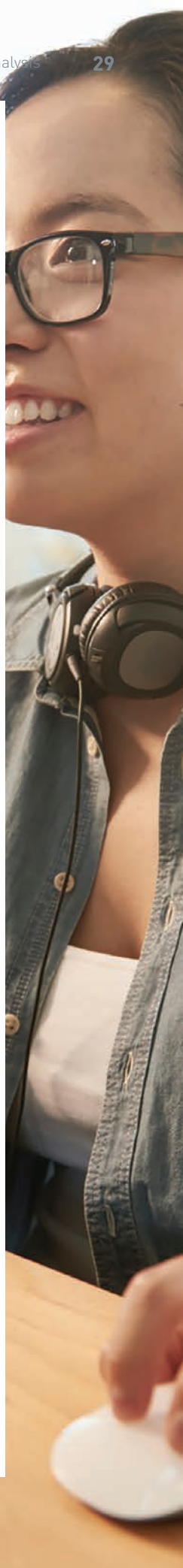
- **a** Clearly state the topic Arthur will investigate.
- **b** What type of data will Arthur need to assist in his investigation?
- **c** Identify an appropriate data gathering technique Arthur could use.
- **d** Justify the selected data gathering technique.
- **e** How will Arthur keep the data safe?
- **f** Does Arthur need to get permission to conduct his research and if so, from whom?
- **g** What tools will Arthur use to interpret the results?
- **h** What types of relationships and patterns is Arthur looking for?
- **i** How will Arthur present the data to the local council?

INTERNET USAGE

2 Go to the UN Data website and find the Information and Communications Technology database. Select to view the data on the percentage of individuals using the internet.

UN Data

- **a** Filter the data to show only the percentage of individuals using the internet.
- **b** Copy and paste this data into a spreadsheet program.
- **c** Create a column chart, a scatter diagram and a bubble chart that depict these statistics.
- **d** Discuss which graphic representation best conveys the data and why.



CHAPTER 2

APPROACHES TO PROBLEM-SOLVING METHODOLOGY: DATA ANALYSIS

Key knowledge

After completing this chapter, you will be able to demonstrate knowledge of:

Approaches to problem solving

- types of graphic solutions suitable for educating, persuading and informing audiences
- design tools for representing the functionality and appearance of graphic solutions such as input-process-output charts (functionality) and annotated diagrams/mock ups (appearance)
- formats and conventions suitable for graphic solutions such as titles, text styles, shapes, lines and arrows, sources of data and legend, colours and contrasts
- software functions and techniques for efficiently and effectively manipulating data to develop graphic solutions, and for validating data
- techniques for testing graphic solutions.

For the student

Following on from Chapter 1 where you conducted an investigation into an issue, in Chapter 2 you will create a digital solution that graphically presents the findings of the investigation. Graphical data solutions are necessary for data analysis and interpretation. Graphic solutions can draw attention to trends and patterns in numerical data and information that might not otherwise be obvious.

For the teacher

This chapter is based on Unit 1, Area of Study 1 and, together with Chapter 1, provides the key knowledge required to complete Unit 1, Outcome 1. At the end of Chapters 1 and 2, students should be able to acquire, secure and interpret data and design and develop a graphic solution that communicates their findings of an investigation. Students are able to use any software tool to create their graphic solution.

Approaches to problem solving

When an information problem exists, a structured **problem-solving methodology (PSM)** needs to be followed to ensure that the most appropriate solution is found and implemented.

Without a structured approach and methodology, problem-solving can become a hit or miss affair. It is therefore important to adopt an agreed structure to solving problems. The obvious advantage of consistency is that when new problems arise, individuals or groups know which approach to follow.

The four stages of problem solving

1 Analysis: Understand all aspects of the problem and state what is required of the solution.
2 Design: Decide what has to be done to achieve the solution.
3 Development: Build the solution with either off-the-shelf or custom-made hardware and software.
4 Evaluation: Determine whether the solution has solved the original problem, and check the requirements of the user to see if they have been met.

Within these four stages of the problem-solving methodology, various activities take place. In the previous chapter, we explored how to acquire, secure and interpret data through investigating an issue. This chapter continues the investigation of an issue and focuses on how to design and develop a graphic solution that communicates the findings of the investigation.

The VCAA has not prescribed specific software to create the graphic solution, and this choice is up to schools. In this chapter, both spreadsheet and web-authoring software will be covered. They will both be used to create an **infographic**.

Functions of a spreadsheet

Spreadsheets can be used for many purposes that involve calculations (see Figure 2.1). For example, spreadsheets can be used for account keeping or stock control, to present budget information as part of a project, or to store and manipulate data from surveys or results of an investigation.

Spreadsheet functions include:

The ability to perform complex mathematical functions. Mathematical functions include the ability to:

- perform basic arithmetic operations (+, –, /, *)
- perform statistical or other mathematical functions (average, minimum, maximum, median, standard deviation).

The ability to perform complex logical functions. Logical functions include the ability to:

- create decision statements such as `IF()` or `SUMIF()`
- use a `LOOKUP()` table to extract data needed for a calculation from another worksheet
- use `AND()`, `OR()` or `NOT()` to create complex formulas.

The ability to produce different types of graphs and charts, including the ability to:

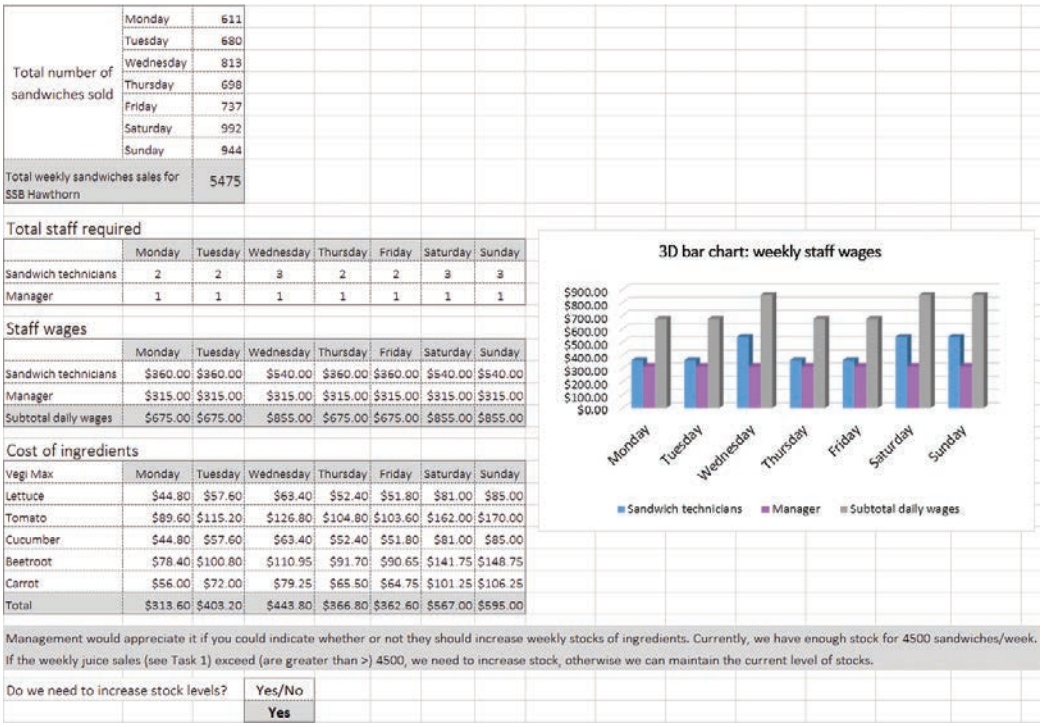
- graph a series of data using a range of graph types, including bar and pie
- format a plot area and gridlines.

The ability to format data to meet the graphics needs of the user. Formatting functions include the ability to:

- insert labels (such as headings and subheadings)
- insert headers and footers (filename, page and date)
- insert notes and comments (to explain a function or provide help)
- insert borders and shading to add more meaning to the layout.

The rows and columns collectively are called a worksheet. A number of worksheets can be included in a spreadsheet file.

Spreadsheets will help you to manipulate data to develop graphs and charts. Spreadsheets can also accommodate small to massive data sets, include a large variety of charting tools and types, include chart style galleries that allow users to format each component of their graphic solution, allow users to enhance graphic representations, provide tools that enable users to highlight data trends and present appealing, persuasive graphical summaries.



Total number of sandwiches sold	Monday	611
	Tuesday	680
	Wednesday	813
	Thursday	698
	Friday	737
	Saturday	992
	Sunday	944
Total weekly sandwiches sales for SSB Hawthorn		5475

Total staff required

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Sandwich technicians	2	2	3	2	2	3	3
Manager	1	1	1	1	1	1	1

Staff wages

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Sandwich technicians	$360.00	$360.00	$540.00	$360.00	$360.00	$540.00	$540.00
Manager	$315.00	$315.00	$315.00	$315.00	$315.00	$315.00	$315.00
Subtotal daily wages	$675.00	$675.00	$855.00	$675.00	$675.00	$855.00	$855.00

Cost of ingredients

Vegi Max	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Lettuce	$44.80	$57.60	$63.40	$52.40	$51.80	$81.00	$85.00
Tomato	$89.60	$115.20	$126.80	$104.80	$103.60	$162.00	$170.00
Cucumber	$44.80	$57.60	$63.40	$52.40	$51.80	$81.00	$85.00
Beetroot	$78.40	$100.80	$110.95	$91.70	$90.65	$141.75	$148.75
Carrot	$56.00	$72.00	$79.25	$65.50	$64.75	$101.25	$106.25
Total	$313.60	$403.20	$443.80	$366.80	$362.60	$567.00	$595.00

Management would appreciate it if you could indicate whether or not they should increase weekly stocks of ingredients. Currently, we have enough stock for 4500 sandwiches/week. If the weekly juice sales (see Task 1) exceed (are greater than >) 4500, we need to increase stock, otherwise we can maintain the current level of stocks.

Do we need to increase stock levels?	Yes/No
	Yes

FIGURE 2.1 Modern spreadsheet packages allow calculations of formulas, experimenting with 'what–if?' scenarios, as well as the arrangement and graphical representation of statistical data.

To make manipulating the data more efficient and to make the resulting information more effective, spreadsheets may involve features such as the ability to:

- sort rows of data based on the values in a specified column
- program **macros** to automate calculations or formatting
- use hyperlinks to link worksheets together for ease of navigation and use
- use relative and absolute cell references to ensure accuracy of formulas
- name a range of cells to make formulas more meaningful when testing
- protect cells to ensure that the integrity of the formula is not affected
- check the range to ensure that accurate data is entered into the solution.

Analysing the problem

Refer to the PSM on pages xii–xiii.

In any problem-solving task, you should follow the PSM and analyse the problem. Once you have defined the problem, the remainder of the analysis stage can commence. Within the problem-solving methodology, the analysis stage is often considered to be the most crucial. Getting the analysis stage correct and having a clear picture of what is required will provide you with the appropriate building blocks for the latter stages.

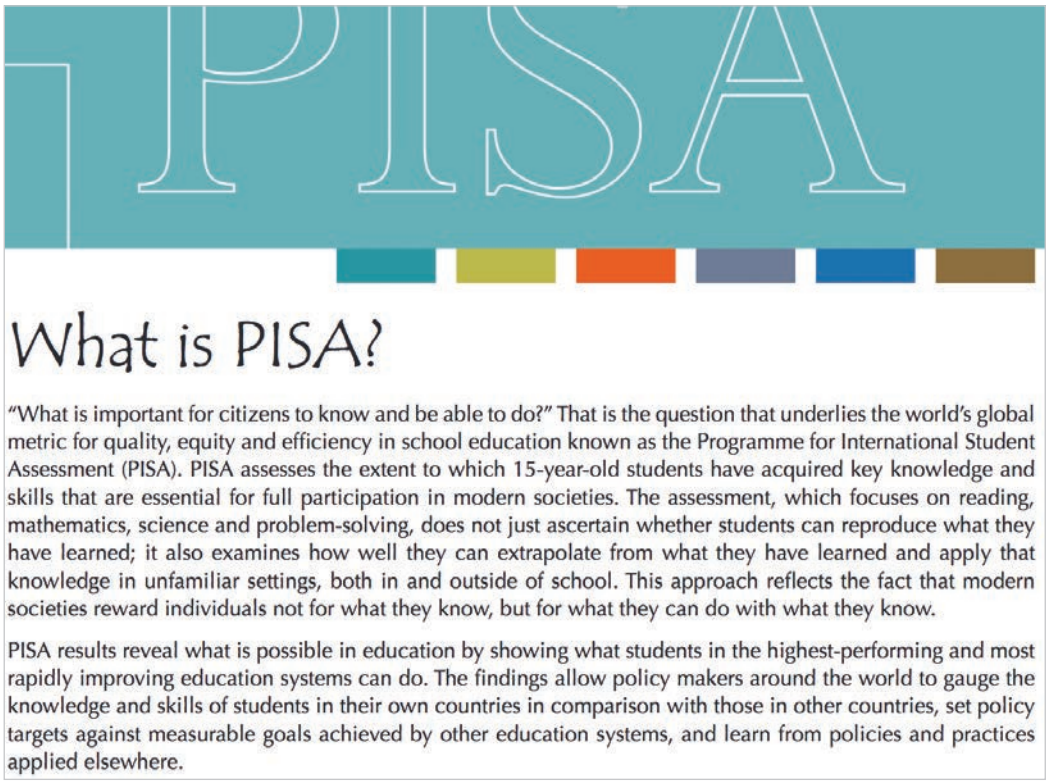
CASE STUDY THE NUMBER OF STUDENTS UNDERTAKING STEM SUBJECTS IS INSUFFICIENT FOR AUSTRALIA'S FUTURE NEEDS

There is a worldwide shortage of Science, Technology, Engineering and Mathematics (STEM)graduates. The situation includes Australia, where the Australian Industry Group report (2013) suggests skills learned through each STEM discipline are critical for national productivity and global competitiveness, but warns that 'Australia's participation in STEM

skills at secondary school and university are unacceptably low' (p. 1). There is clearly a shortage of students studying STEM subjects at secondary school and continuing them at university. When looking at Australia's PISA and TIMSS results (Table 2.2), there is evidence to show that Australia is not doing as well in this space. To test out the hypothesis that not many students study STEM subjects, we will undertake an investigation of the Year 11 and 12 students at RubyMede College and their STEM choices and rationales (continued in Part 2, on page 35).

Two key international assessments are the Programme for International Student Assessment (PISA), shown in Figure 2.2, and the Trends in International Mathematics and Science Study (TIMSS). PISA, coordinated by the Organization for Economic Cooperation and Development (OECD) assesses reading, scientific and mathematical literacy, focusing on knowledge and skills gained through education but applicable to everyday choices and the solving of real world problems. PISA tests 15-year-olds from 74 different countries, while TIMSS tests students in Grades 4–8 and is the longest running and most extensive international test of science and mathematics learning.

Australian Industry Group (Producer). (2013). *Lifting our Science, Technology, Engineering and Maths (STEM) Skills*. Retrieved from http://www.aigroup.com.au/portal/binary/com.epicentric.contentmanagement.servlet.ContentDeliveryServlet/LIVE_CONTENT/Publications/Reports/2013/Ai_Group_Skills_Survey_2012-STEM_FINAL_PRINTED.pdf



What is PISA?

"What is important for citizens to know and be able to do?" That is the question that underlies the world's global metric for quality, equity and efficiency in school education known as the Programme for International Student Assessment (PISA). PISA assesses the extent to which 15-year-old students have acquired key knowledge and skills that are essential for full participation in modern societies. The assessment, which focuses on reading, mathematics, science and problem-solving, does not just ascertain whether students can reproduce what they have learned; it also examines how well they can extrapolate from what they have learned and apply that knowledge in unfamiliar settings, both in and outside of school. This approach reflects the fact that modern societies reward individuals not for what they know, but for what they can do with what they know.

PISA results reveal what is possible in education by showing what students in the highest-performing and most rapidly improving education systems can do. The findings allow policy makers around the world to gauge the knowledge and skills of students in their own countries in comparison with those in other countries, set policy targets against measurable goals achieved by other education systems, and learn from policies and practices applied elsewhere.

PISA 2012 Results in Focus: What 15-year-olds know and what they can do with what they know, OECD 2014, p. 3

FIGURE 2.2 What is PISA?

Designing solutions

Once the analysis stage is complete and a general idea of the solution is in mind, it is time for the design stage. Designing a solution often occurs without using a computer at all – the solution may just be drafted on paper. The main advantages of designing a solution on paper are that you can sketch your ideas quickly and easily, and that you can make changes easily without affecting the development of a solution.

ISSUE

What students know and can do: student performance in mathematics and science

- Shanghai-China has the highest scores in mathematics, with a mean score of 613 points – 119 points, or the equivalent of nearly three years of schooling, above the OECD average. Singapore, Hong Kong-China, Chinese Taipei, Korea, Macao-China, Japan, Liechtenstein, Switzerland and the Netherlands, in descending order of their scores, round out the top 10 performers in mathematics.
- On average across OECD countries, 13% of students are top performers in mathematics (Level 5 or 6). They can develop and work with models for complex situations, and work strategically using broad, well-developed thinking and reasoning skills. The partner economy Shanghai-China has the largest proportion of students performing at Level 5 or 6 (55%), followed by Singapore (40%), Chinese Taipei (37%) and Hong Kong-China (34%). At the same time, 23% of students in OECD countries, and 32% of students in all participating countries and economies, did not reach the baseline Level 2 in the PISA mathematics assessment. At that level, students can extract relevant information from a single source and can use basic algorithms, formulae, procedures or conventions to solve problems involving whole numbers.
- Between 2003 and 2012, Italy, Poland and Portugal increased their shares of top performers and simultaneously reduced their shares of low performers in mathematics.
- Boys perform better than girls in mathematics in only 37 out of the 65 countries and economies that participated in PISA 2012, and girls outperform boys in five countries.
- Shanghai-China, Hong Kong-China, Singapore, Japan and Finland are the top five performers in science in PISA 2012.
- Between 2006 and 2012, Italy, Poland and Qatar, and between 2009 and 2012, Estonia, Israel and Singapore increased their shares of top performers and simultaneously reduced their shares of low performers in science.
- Across OECD countries, 8% of students are top performers in science (Level 5 or 6). These students can identify, explain and apply scientific knowledge and knowledge about science in a variety of complex life situations.

PISA 2012 Results in Focus: What 15-year-olds know and what they can do with what they know, OECD 2014, p. 4

Snapshot of performance in mathematics, reading and science

Countries/economies with a mean performance/share of top performers above the OECD average
Countries/economies with a share of low achievers below the OECD average

Countries/economies with a mean performance/share of low achievers/share of top performers not statistically significantly different from the OECD average

Countries/economies with a mean performance/share of top performers below the OECD average
Countries/economies with a share of low achievers above the OECD average

TABLE 2.1 A sample from the snapshot of performance in mathematics and science

	Mathematics				Science	
	Mean score in PISA 2012	Share of low achievers in mathematics (Below Level 2)	Share of top performers in mathematics (Level 5 or 6)	Annualised change in score points	Mean score in PISA 2012	Annualised change in score points
OECD average	494	23.0	12.6	–0.3	501	0.5
Shanghai-China	613	3.8	55.4	4.2	580	1.8
Singapore	573	8.3	40.0	3.8	551	3.3
Hong Kong-China	561	8.5	33.7	1.3	555	2.1
Chinese Taipei	560	12.8	37.2	1.7	523	–1.5
Korea	554	9.1	30.9	1.1	538	2.6
Macao-China	538	10.8	24.3	1.0	521	1.6
Japan	536	11.1	23.7	0.4	547	2.6
Liechtenstein	535	14.1	24.8	0.3	525	0.4
Switzerland	531	12.4	21.4	0.6	515	0.6
Netherlands	523	14.8	19.3	–1.6	522	–0.5
Estonia	521	10.5	14.6	0.9	541	1.5
Finland	519	12.3	15.3	–2.8	545	–3.0
Canada	518	13.8	16.4	–1.4	525	–1.5
Poland	518	14.4	16.7	2.6	526	4.6
Belgium	515	19.0	19.5	–1.6	505	–0.9
Germany	514	17.7	17.5	1.4	524	1.4
Viet Nam	511	14.2	13.3	m	528	m
Austria	506	18.7	14.3	0.0	506	–0.8
Australia	504	19.7	14.8	–2.2	521	–0.9
Ireland	501	16.9	10.7	–0.6	522	2.3
Slovenia	501	20.1	13.7	–0.6	514	–0.8

PISA 2012 Results in Focus: What 15-year-olds know and what they can do with what they know, OECD 2014, p. 5

TABLE 2.2 Singapore, Korea, and Hong Kong SAR were the top-performing countries in TIMSS 2011 at the fourth grade. For multiple comparisons of average mathematics achievement, read across the row for a country to compare performance with the countries listed along the top of the chart. The symbols indicate whether the average achievement of the country in the row is significantly lower than that of the comparison country, significantly higher than that of the comparison country, or if there is no statistically significant difference between the average achievement of the two countries.

Country	Average scale score	Singapore	Korea, Rep. of	Hong Kong SAR	Chinese Taipei	Japan	Northern Ireland	Belgium (Flemish)	Finland	England	Russian Federation	United States	Netherlands	Denmark	Lithuania	Portugal	Germany	Ireland	Serbia	Australia
Singapore	606 (3.2)				▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Korea, Rep. of	605 (1.9)				▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Hong Kong SAR	602 (3.4)				▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Chinese Taipei	591 (2.0)	▼	▼	▼		▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Japan	585 (1.7)	▼	▼	▼	▼		▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Northern Ireland	562 (2.9)	▼	▼	▼	▼	▼		▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲	▲
Belgium (Flemish)	549 (1.9)	▼	▼	▼	▼	▼	▲					▲	▲	▲	▲	▲	▲	▲	▲	▲
Finland	545 (2.3)	▼	▼	▼	▼	▼	▲							▲	▲	▲	▲	▲	▲	▲
England	542 (3.5)	▼	▼	▼	▼	▼	▲								▲	▲	▲	▲	▲	▲
Russian Federation	542 (3.7)	▼	▼	▼	▼	▼	▼										▲	▲	▲	▲
United States	541 (1.8)	▼	▼	▼	▼	▼	▼	▼							▲	▲	▲	▲	▲	▲
Netherlands	540 (1.7)	▼	▼	▼	▼	▼	▼	▼							▲	▲	▲	▲	▲	▲
Denmark	537 (2.6)	▼	▼	▼	▼	▼	▼	▼	▼								▲	▲	▲	▲
Lithuania	534 (2.4)	▼	▼	▼	▼	▼	▼	▼	▼	▼		▼	▼						▲	▲
Portugal	532 (3.4)	▼	▼	▼	▼	▼	▼	▼	▼	▼		▼	▼						▲	▲
Germany	528 (2.2)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼					▲	▲
Ireland	527 (2.6)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼					▲	▲
Serbia	516 (3.0)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼		
Australia	516 (2.9)	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼	▼		

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2011, Grade 4 ©2003 IEA, The Netherlands

CASE STUDY

THE NUMBER OF STUDENTS UNDERTAKING STEM SUBJECTS IS INSUFFICIENT FOR AUSTRALIA'S FUTURE NEEDS

Hypothesis

Are there insufficient students undertaking STEM subjects to meet Australia's future needs?

Identify the type of data that will be collected

The data will be collected from a variety of sources including:

- VCE students studying STEM subjects at RubyMede College – numerical data
- PISA and TIMSS results – numerical data
- government and industry reports – qualitative data.

Identify the data collection tool that will be used

- Questionnaires will be used to elicit primary data from students at RubyMede College.
- The internet will also be used to support any findings and supplement the investigation.

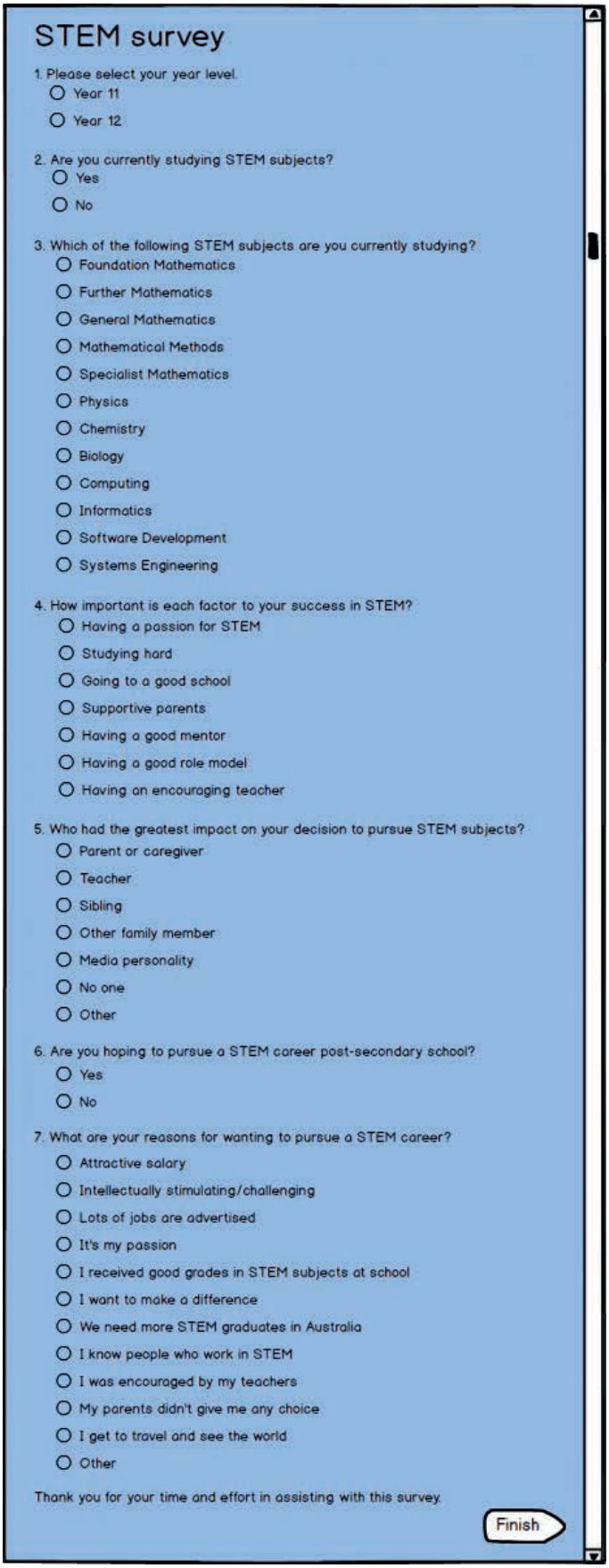


FIGURE 2.3 A questionnaire for students about STEM subjects

How data will be gathered

The questionnaire used for this investigation can be seen in Figure 2.3. It has seven questions and some space to add free text if participants want to clarify or add some more information. The questionnaire will be completed electronically over the internet and will take no longer than 10 minutes to complete.

Create participation information statement and consent forms

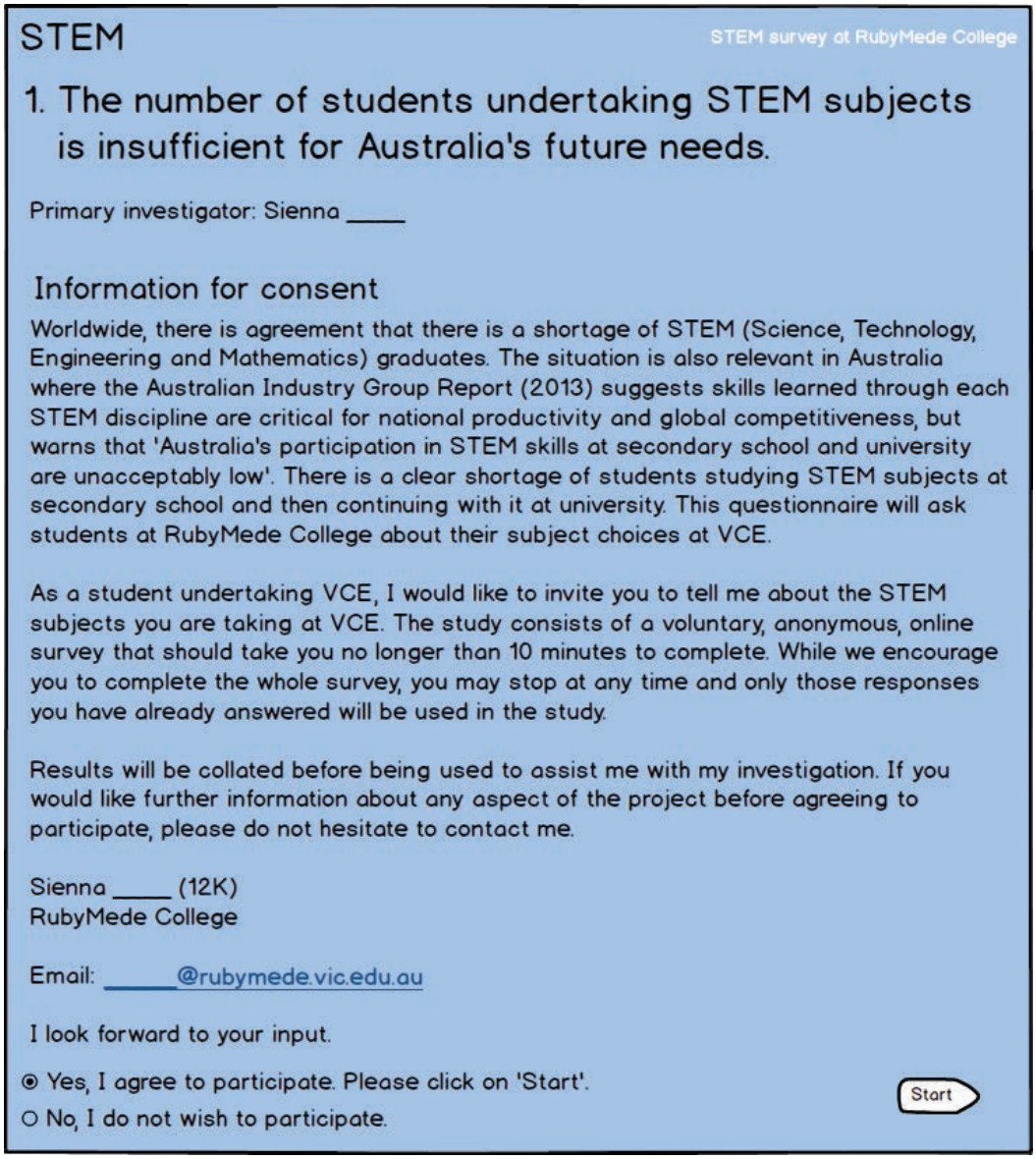
As the participation information statement and consent forms are combined, they can be incorporated into the start of the questionnaire. Participants of the questionnaire will need to provide consent to take part (Figure 2.4). This can be done by agreeing to take part in the questionnaire and then pressing 'start'. If the participant chooses not to take part, they just need to press 'no.' The survey is anonymous and will not collect information that can identify the participants.

Identify risks to privacy and security and have strategies in place to minimise them

The researcher will only have access to the data she will be collecting for her study from the participants who have agreed to be part of the research. The researcher will not have access to any other data collected by RubyMede for its own purposes. Identifiable data will be coded and de-identified and only collated data will be reported. The research will be carried out online and all data collected will be stored on the computer's hard drive and backed up to the school's network drive on a password-protected computer. The data will also be stored in a password-protected file to ensure that the data collected cannot identified.

Collecting data

The researcher sent the URL of her questionnaire to all VCE students in her school. There are more than 210 students in Years 11 and 12. Many of the students did not want to participate, but the researcher secured 43 completed questionnaires from students at her school. The participation rate represents approximately 20 per cent of students from the population. Even though 44 students participated, one of the respondents did not complete the questionnaire. It is acceptable for questionnaires to be incomplete because participation is voluntary. See the report information in Figure 2.5.


STEM

STEM survey at RubyMede College

1. The number of students undertaking STEM subjects is insufficient for Australia's future needs.

Primary investigator: Sienna ____

Information for consent

Worldwide, there is agreement that there is a shortage of STEM (Science, Technology, Engineering and Mathematics) graduates. The situation is also relevant in Australia where the Australian Industry Group Report (2013) suggests skills learned through each STEM discipline are critical for national productivity and global competitiveness, but warns that 'Australia's participation in STEM skills at secondary school and university are unacceptably low'. There is a clear shortage of students studying STEM subjects at secondary school and then continuing with it at university. This questionnaire will ask students at RubyMede College about their subject choices at VCE.

As a student undertaking VCE, I would like to invite you to tell me about the STEM subjects you are taking at VCE. The study consists of a voluntary, anonymous, online survey that should take you no longer than 10 minutes to complete. While we encourage you to complete the whole survey, you may stop at any time and only those responses you have already answered will be used in the study.

Results will be collated before being used to assist me with my investigation. If you would like further information about any aspect of the project before agreeing to participate, please do not hesitate to contact me.

Sienna ____ (12K)
RubyMede College

Email: ____@rubymede.vic.edu.au

I look forward to your input.

◉ Yes, I agree to participate. Please click on 'Start'.
○ No, I do not wish to participate.

Start

FIGURE 2.4 Respondents must consent to participate in this questionnaire.

Report info

Report date:	Wednesday, March 4, 2015 2:12:07 PM EST
Start date:	Monday, March 2, 2015 8:27:00 PM EST
Stop date:	Wednesday, March 4, 2015 12:27:00 PM EST
Stored responses:	44
Number of completed responses:	43

FIGURE 2.5 The report generated from the online survey shows number of respondents who completed the questionnaire. Even though there were 44 responses, there were only 43 completed responses.

Data capture

Part of the data that was captured from the questionnaire is shown in Figure 2.6. At this stage, it is important to just present the data from the questionnaire.

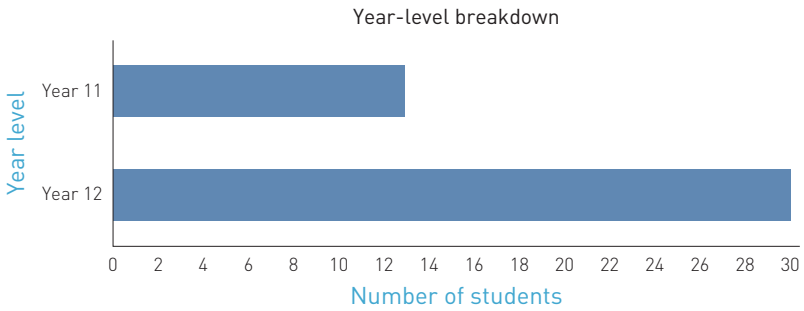


FIGURE 2.6 Student year-level breakdown of those who participated in the questionnaire

TABLE 2.3 Frequency table

Choices	Absolute frequency	Relative frequency (%)	Adjusted relative frequency (%)
Year 11	13	29.55	30.23
Year 12	30	68.18	69.77
Sum	43	97.73	100
Not answered	1	2.27	–
Total answered: 43			

The **absolute frequency** is the total number of times a variable is observed in a given range. For example: Year 11 = 13 Year 12 = 30.
The **relative frequency** is the number of times a variable is observed, in relation to the total number of values for the variable. For example: Year 11 = 13 ÷ 44 = 0.2955. Expressed as a percentage, this is 29.55 %. In each calculation, you would move the decimal place two places to the right. For example: Year 12 = 30 ÷ 44 = 0.6818 = 68.18 %. The same rule applies for the tables that follow.
The **adjusted relative frequency** removes the 'not answered' variable, so that: Year 11 = 13 ÷ 43 = 30.23 Year 12 = 30 ÷ 43 = 69.77.

Approximately 70 per cent of the respondents of the questionnaire were Year 12 students and 30 per cent were Year 11. Of those who participated, all were studying STEM subjects.

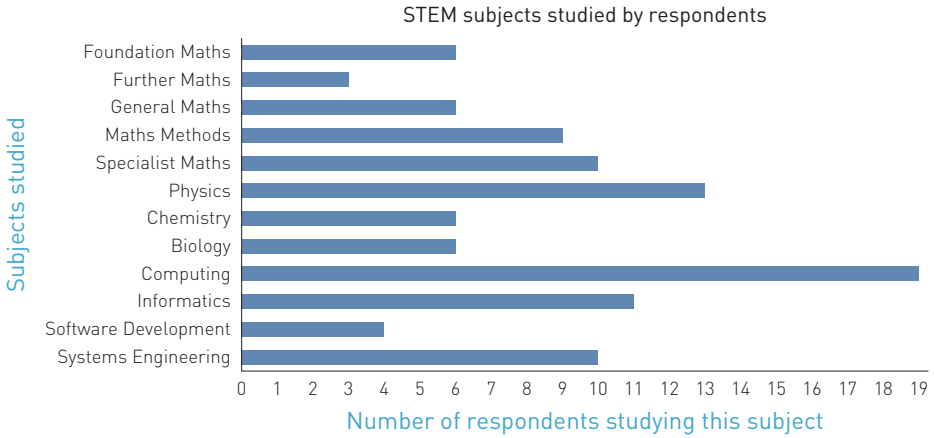


FIGURE 2.7 STEM subjects students studied

TABLE 2.4 Frequency table

Choices	Absolute frequency	Relative frequency by choice (%)	Relative frequency (%)	Adjusted relative frequency (%)
Foundation Maths	6	5.83	13.64	13.95
Further Maths	3	2.91	6.82	6.98
General Maths	6	5.83	13.64	13.95
Maths Methods	9	8.74	20.45	20.93
Specialist Maths	10	9.71	22.73	23.26
Physics	13	12.62	29.55	30.23
Chemistry	6	5.83	13.64	13.95
Biology	6	5.83	13.64	13.95
Computing	19	18.45	43.18	44.19
Informatics	11	10.68	25	25.58
Software Development	4	3.88	9.09	9.3
Systems Engineering	10	9.71	22.73	23.26
Sum:	103	100	–	–
Not answered:	1	–	2.27	–
Total answered: 43				

Absolute frequency in Table 2.4 shows the number of times a subject was selected in the survey. For example, Computing was selected 10 times, as shown in Figure 2.7.
Relative frequency by choice (%) in Table 2.4 is the total number of times a subject was selected divided by the sum in the absolute frequency column. For example: Computing = 19 ÷ 103 = 18.45 (rounded up).
Relative frequency % in Table 2.4 is the total number of times a subject was selected divided by the total number of participants. For example: Computing = 19 ÷ 44 = 43.18.
The **adjusted relative frequency (%)** in Table 2.4 is the total number of times a subject was selected divided by the total number of participants, with the 'not answered' participant removed. For example: Computing = 19 ÷ 43 = 44.19 (rounded up).

It appears that Computing is very popular among students, with 44 per cent of this cohort studying it. Physics was the next most popular, with 30 per cent studying it, with Informatics next on 26 per cent. The least popular subjects, according to the questionnaire, were Further Maths and Software Development.

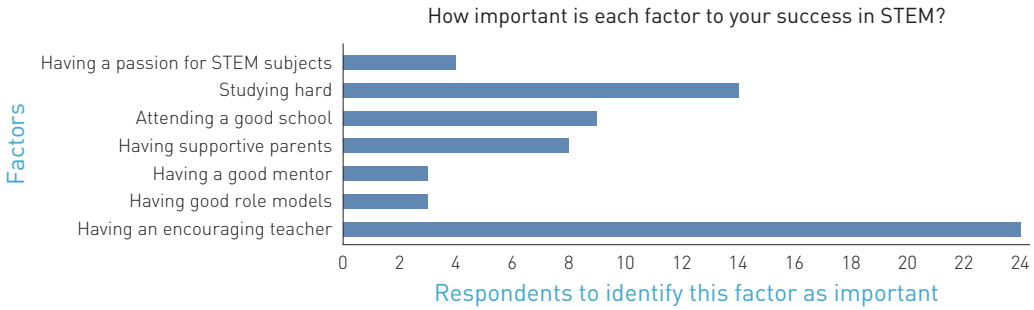


FIGURE 2.8 Identifying factors to experiencing success in STEM

For advice on interpreting frequency tables, see the margin notes on pages 38 and 39.

TABLE 2.5 Frequency table

Choices	Absolute frequency	Relative frequency by choice (%)	Relative frequency (%)	Adjusted relative frequency (%)
Having a passion for STEM subjects	4	6.15	9.09	9.3
Studying hard	14	21.54	31.82	32.56
Attending a good school	9	13.85	20.45	20.93
Having supportive parents	8	12.31	18.18	18.6
Having a good mentor	3	4.62	6.82	6.98
Having good role models	3	4.62	6.85	6.98
Having an encouraging teacher	24	36.92	54.55	55.81
Sum:	65	100	–	–
Not answered:	1	–	2.27	–
Total answered: 43				

The top three factors identified as important to success in STEM were having encouraging teachers, studying hard and going to a good school. From the data, respondents believed that having an encouraging teacher was important.

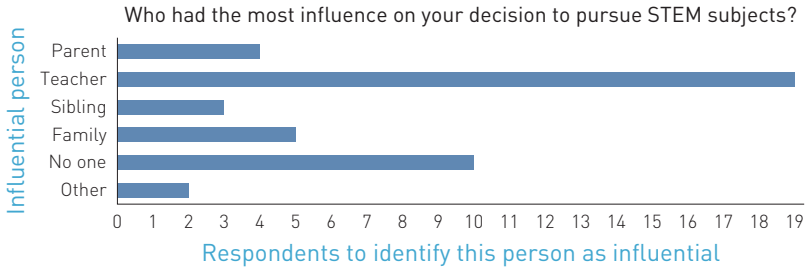


FIGURE 2.9 Influence on decision to pursue STEM subjects.

TABLE 2.6 Frequency table

Choices	Absolute frequency	Relative frequency (%)	Adjusted relative frequency (%)
Parent	4	9.09	9.3
Teacher	19	43.18	44.19
Sibling	3	6.82	6.98
Family	5	11.36	11.63
No one	10	22.73	23.26
Other	2	4.55	4.65
Sum:	43	97.73	100
Not answered:	1	2.27	–
Total answered: 43			
Text input:			
My neighbour			
Sheldon from *The Big Bang Theory*			

Respondents indicated that their teacher had the most influence on their decision to study STEM. The next largest category was 'no-one.' Family and parents rated around the same.

FIGURE 2.10 Reasons for pursuing a STEM career

TABLE 2.7 Frequency table

Choices	Absolute frequency	Relative frequency by choice (%)	Relative frequency (%)	Adjusted relative frequency (%)
Attractive salary	1	1.96	2.27	2.33
It is intellectually challenging/stimulating	2	3.92	4.55	4.65
High level of job vacancies	4	7.84	9.09	9.3
I am passionate about it	3	5.88	6.82	6.98
Good grades in STEM subjects at school	6	11.76	13.64	13.95
I want to make a difference	4	7.84	9.09	9.3
We need more STEM graduates in Australia	2	3.92	4.55	4.65
I know people who work in STEM	5	9.8	11.36	11.63
My teachers encouraged me	5	9.8	11.36	11.63
My parents didn't give me any choice	2	3.92	4.55	4.65
I will be able to travel and see the world	11	21.57	25	25.58
Other	6	11.76	13.64	13.95
Sum:	51	100	–	–
Not answered:	1	–	2.27	–
Total answered: 43				

Respondents were asked why they wanted to pursue a STEM-based career after secondary school and many indicated that they were able to see the world and travel through their chosen career. The next most popular response was because they received good grades in STEM subjects at school.

Data analysis

Once data has been collected, it needs to be analysed. The researcher needs to identify any patterns, give meaning to the data and draw conclusions.

As noted previously, only 20 per cent of the possible VCE students from the College participated in this questionnaire, which represents a very small percentage of students. The data obtained from the questionnaire alone does not provide a clear picture of STEM subject enrolments at the school. However, when the respondents are compared with actual enrolments, it reveals that 53 per cent of STEM students responded to this questionnaire, indicating that fewer than 40 per cent of students undertake STEM subjects at RubyMede College.

TABLE 2.8 The following table represents the questionnaire data compared with the *actual* enrolments in STEM subjects at RubyMede College.

Subjects	RubyMede subject enrolments	Questionnaire participants
Foundation Maths	14	6
Further Maths	20	3
General Maths	17	6
Maths Methods	19	9
Specialist Maths	13	10
Physics	15	13
Chemistry	13	6
Biology	19	6
Computing	21	19
Informatics	22	11
Software Development	8	4
Systems Engineering	12	10
Total	193	103

Purpose of graphic solutions

See David McCandless's 'Information is Beautiful' TEDTalk.

Graphic solutions are pictorial diagrams that show interdependencies between **variables**. They are one of the most commonly used methods of representing data and information, and can make reading data and information more interesting, less time-consuming and more understandable.

They allow data to be compared at a glance. Graphic solutions are particularly useful for summarising a series of numbers and their interrelationships, and they can assist in identify trends and patterns in your data. The following section discusses the types of graphic solutions that may be appropriate for your Outcome.

If you need greater **clarity** from a set of information or to identify patterns or relationships with data sets, you may require a graphic solution. A graphic solution visually represents analysed data. It is an elegant way to represent complex information or large amounts of data.

For example, if we want to:

- compare data, we might use a bar chart
- show a distribution of data, we might use a histogram

- show a relationship between two data sets, we might use a scatter diagram
- show a composition of data that changes over time, we might use an animation or simulation.

Often, before we start this process, we have a rough idea of what the solution might look like based on the desire to inform, persuade or educate. We inform when we provide information or tell facts, such as reporting on a current issue. We persuade when we influence a decision or action, or change a person's opinion. We educate when we teach and provide knowledge and skills through a learning process. For example, data obtained about the number of elderly pedestrians injured crossing a major intersection can inform the local government about the black spot. That same data can be visually represented in another way to persuade the engineering department of the local government to make some adjustments to the intersection to prevent further injuries. The data can also be used to educate the elderly about the need to be vigilant when crossing the intersection. The appropriate graphic solution will need to be selected based on how the data will be used to persuade, inform or educate.

Types of graphic solutions

Charts

Charts, also called graphs, represent detailed data in visual form. The most common forms of charts are column graphs, bar graphs, pie graphs and line graphs. Charts generally use a scale and/or a series of data sets that are indicated along two axes. However, in the case of pie graphs, they use the divisions of the whole to communicate the quantities of data. Charts may use two-dimensional or three-dimensional representation for visual impact.

You should choose the correct type of chart for your data.

Columns and bar graphs arrange data vertically (columns, Figure 2.11) or horizontally (rows, Figure 2.10 on page 41). They are handy for presenting data changes over a period of time or for showing comparisons across different times. They enable visual comparisons easily so that differences can be recognised quickly. The STEM case study in this chapter uses many (horizontal) bar graphs.

Other types of charts you may find useful for your infographic include histograms, bubble charts, doughnut graphs and area charts.

Microsoft Word and Excel offer a large range of charts that you can easily add to any document via the Insert → Chart menu.

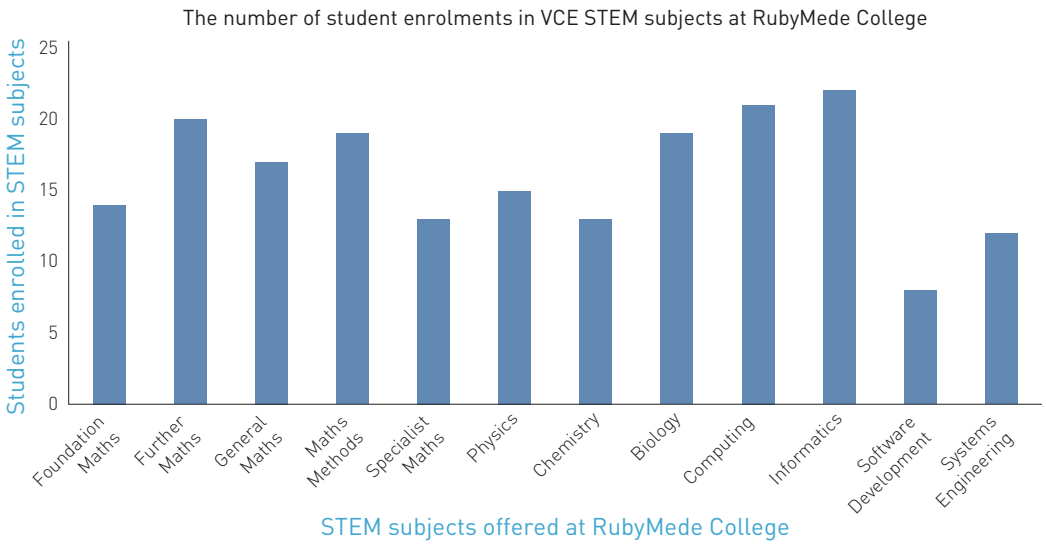


FIGURE 2.11 An example of a column chart that depicts the number of student enrolments in VCE STEM subjects at RubyMede College

Choose a line graph when you need to display continuous data over time, set against a common time. It is useful for showing trends in data at equal intervals. If more than one line is shown, use colours to distinguish each line. The 'Pesticide Planet' infographic, (Figure 2.22, page 55) includes several line graphs, among other types of graphs.

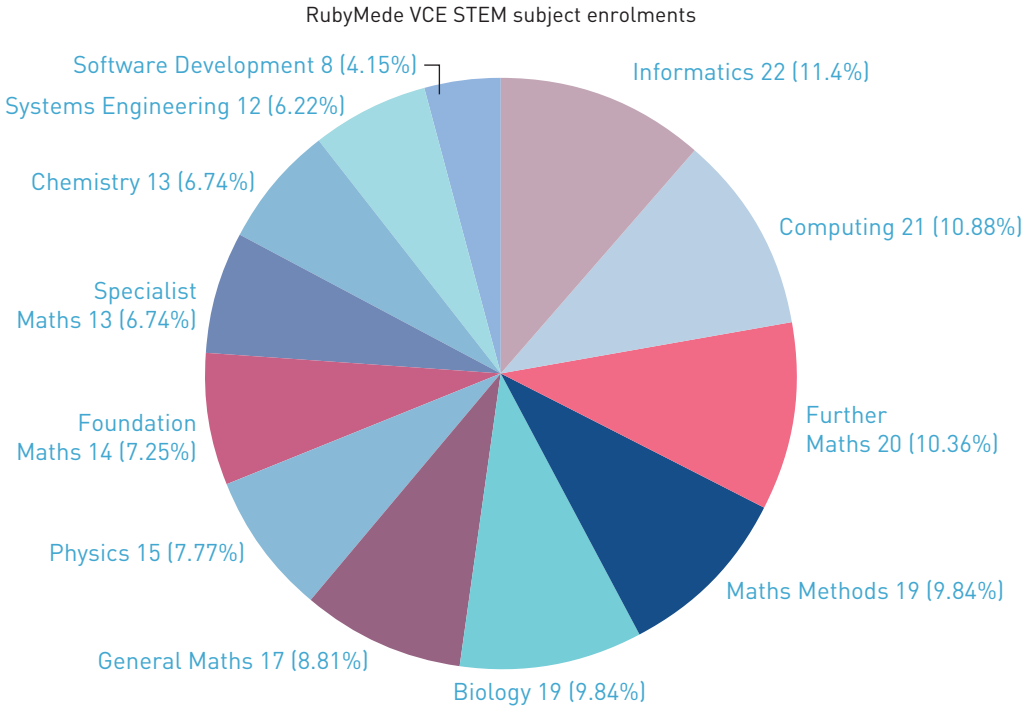


FIGURE 2.12 This pie chart depicts the number of student enrolled in VCE STEM subjects at RubyMede College. Each segment of the pie chart represents the subject and the number of students studying each subject.

A pie graph is divided into 'slices' proportional to the percentage of the whole pie (Figure 2.12). Consider choosing a pie graph if you need to depict approximate proportional relationships (relative amounts) or compare part of a whole at a given point in time. The full circle represents 100 per cent. The angle of each 'slice' is found by multiplying its percentage value by 360°.

You should follow some simple rules with charts.

- They must have titles.
- The *x*-axis and the *y*-axis must be labelled.
- Use a key if more than one set of data is provided on the same graph.
- Include author identification and/or the source of the data, the date and a filename (if appropriate).
- Include the unit of measurement on the relevant axis.
- Label each segment of a pie chart.
- Arrange segments of pie chart (starting at 12 o'clock position) from largest to smallest.
- Consider including absolute figures as well as percentages.
- Choose colours that match the information being discussed.
- Use graphs to show trends or relationships between values on each axis.
- If possible, limit the number of items represented in a chart to five or six.

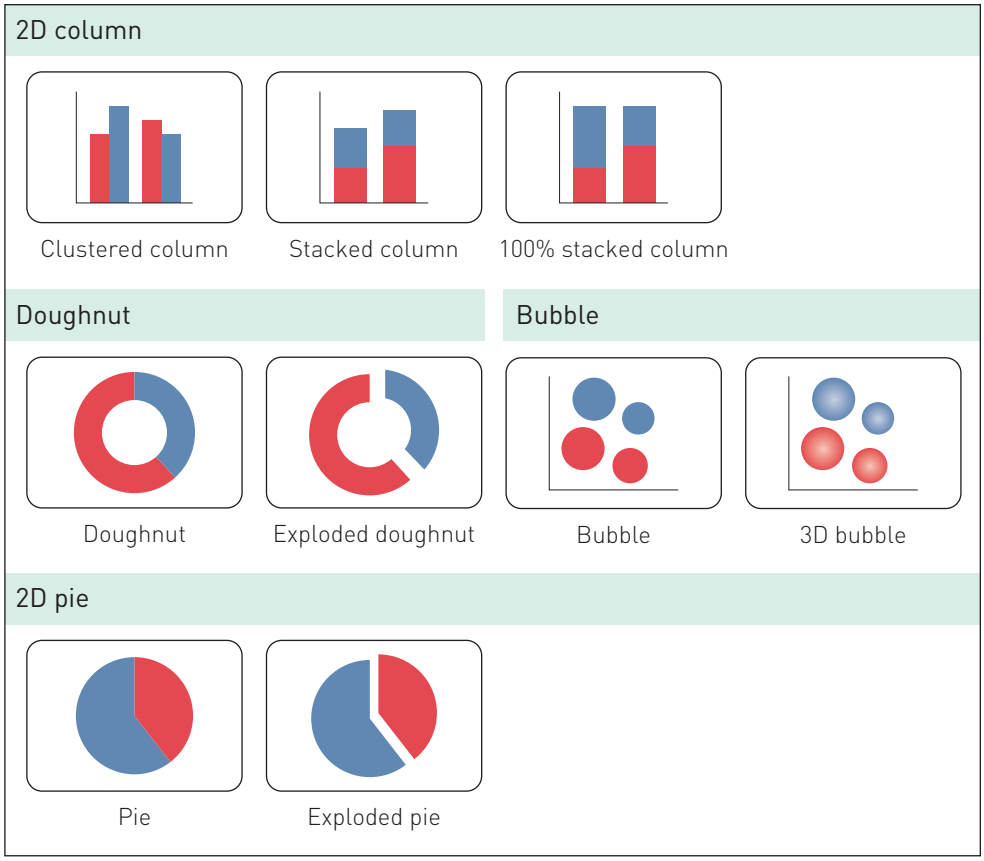


FIGURE 2.13 Examples of charts used for graphic representation of data.

Flowcharts

Flowcharts graphically represent the steps in a process or system. Boxes of data identify the steps to help the user understand what happens next. Flowcharts are used a great deal in information systems because they help explain very complex, technical processes. Many types of flowcharts exist, and some are more visual than others, but the fundamental step-by-step idea remains the same.

Flowcharts should generally be read from the top down or from the left to the right. They also often use lines and arrows to visually direct the user to the next step. In some cases, decision trees with yes/no options instruct the user where to go next in the flowchart.

Website site maps are also a type of flowchart.

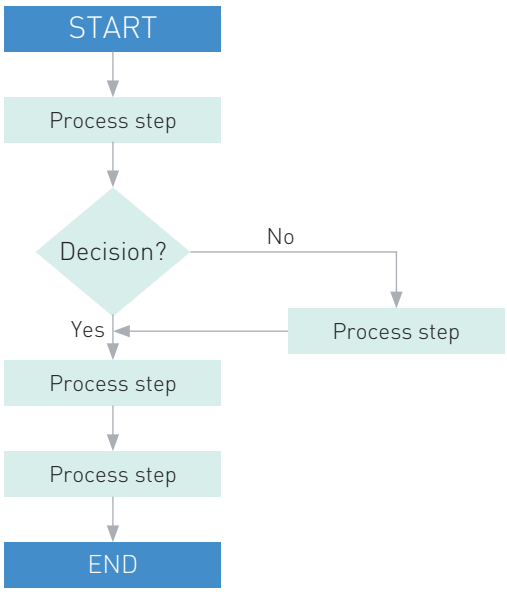


FIGURE 2.14 An example of a flowchart with a decision tree

Lucidchart – free flowchart software

You will find it very useful to become familiar and proficient with flowcharts throughout your study of *Computing VCE*. They are a valuable tool for helping you visualise your solutions and work them out step-by-step.

Diagrams

Diagrams present data or information in a visually clear, accessible way using formats and conventions such as colours, symbols, points, lines, shapes and explanatory drawings. They are usually two-dimensional geometric, but they can be three-dimensional. A diagram is usually a symbolic visual representation of what may be very complex data or information. Common

Blueprints are technical drawings or design plans.

Schematics are diagrams that represent parts of a system using graphic symbols instead of realistic images.

Orthographic drawings are drawings from a series of views and are designed to show every part of an object clearly. They are thought of as multiview drawings.

Axonometric drawings use lines that remain parallel and do not converge at any given point. They are sometimes called paraline drawings.

Isometric drawing is a technique where all lines remain parallel and length and width are drawn at 30°.

types of diagrams include blueprints, schematics and annotated diagrams. An infographic is an increasingly popular form of a diagram. A single diagram, especially an infographic, may include multiple diagrams within it.

Images

Images may include illustrations and photos. Illustrations may be two-dimensional or three-dimensional drawings, drawn by hand using techniques such as orthographic, axonometric, isometric or perspective drawing or using software such as Adobe Illustrator. Illustrations may also be art, cartoons or caricatures, with varying levels of detail. What makes them different from other types of graphic representations in this context is that they (usually) are not representing data.

Hierarchies

Hierarchies are a type of chart that usually depict the relationships and **hierarchy** between roles in an organisation, institution, project team, site map or process chart. While it is most common to use hierarchies to represent relationships between people, they do lend themselves to other uses. Family trees are another type of hierarchy. Figure 2.20 on page 52, which depicts the design principles, is also a hierarchy.

Animations

An animation combines a series of images (either two-dimensional or three-dimensional) to create a sequence of moving images. Animations may include sound and some even offer a level of interactivity. You could use an animation to show the workings of part of a solution or

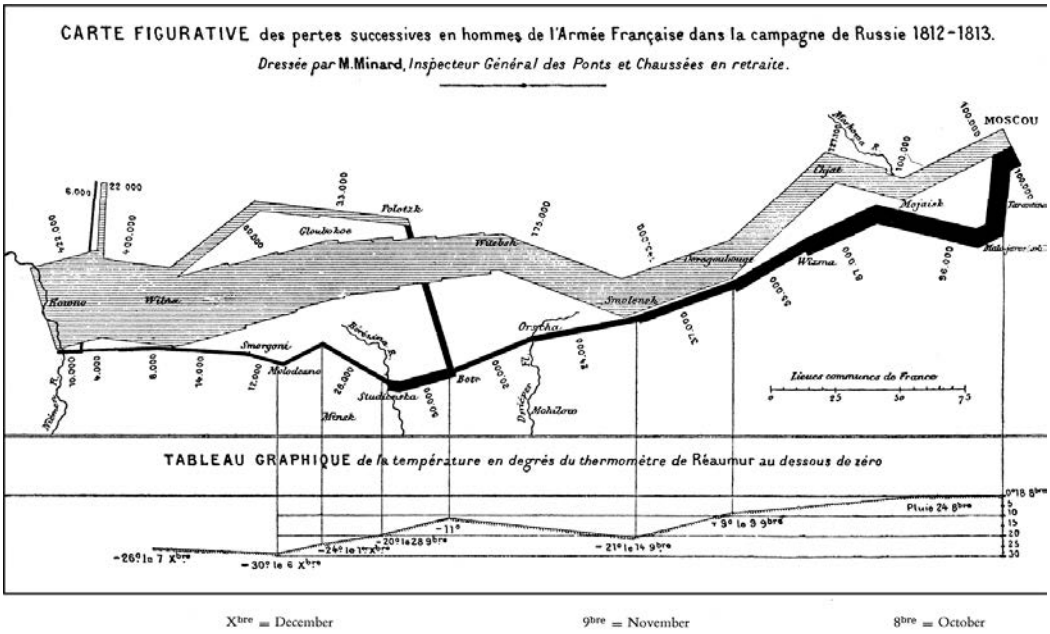


FIGURE 2.15 Widely regarded as the best graph ever drawn, Charles Minard's visual representation of Napoleon's march to Moscow gives the user information on several levels at once: geographical, climatic and numerical. The top line represents the troops moving towards Moscow in 1812–1813. The figures on the side of the line show the dramatic fall in troop numbers, and the corresponding thinning of the line shows the fall visually. It can be seen that Napoleon lost large numbers of troops each time he crossed a river. The line at the bottom of the graph shows the temperature. As the temperature dropped, still more troops were lost. The black line shows the troops retreating. One can only speculate about how different history might have been if Napoleon had had Minard in his midst during the troop campaign.

NelsonNet additional resource: Figure 2.15 Graph, p. 46

perhaps to instruct on the functionality of your design. Animation may be used when printing is not suitable to communicate your message.

Maps

A map represents an area of land or sea as a diagram, showing features such as roads, cities and other markers, such as peaks (mountains), deserts, and borders between states and countries. Another type of map you are likely to be familiar with is a transport map, which shows each individual route of a train or tram line as a single line, to provide a clear guide to destinations and a visual scale of distance between stations.

Signage and other wayfinding devices applied to maps create effective directional design. A map alone is just a map. Adding symbols and other conventions, such as arrows, lines and text, conveys a great deal more information.

Timelines

Timelines show events, images or data in chronological order. A timeline may include both images and text. They are often used to graphically depict historical development over a period of time. You may find them in history and science texts, as well as newspapers, magazines and online. Timelines are often used in infographics to show comparisons or changes over time.

The 'Farewell To A Genius' infographic, Figure 2.33 on page 66, shows an example of a timeline in an infographic that includes both text and images. This infographic may appear to place textboxes haphazardly, but in order to fit the amount of detail in such a small space, each box is carefully placed, making it more like a jigsaw puzzle.

Infographics

An infographic is a graphic solution that shows data and information using different elements that include:

- graphs
- pictures
- diagrams
- narrative
- timelines.

Infographics often communicate multiple findings for a topic. Infographics allow us to tell a more complete story of the data and are more engaging than most traditional ways of communicating data and information.

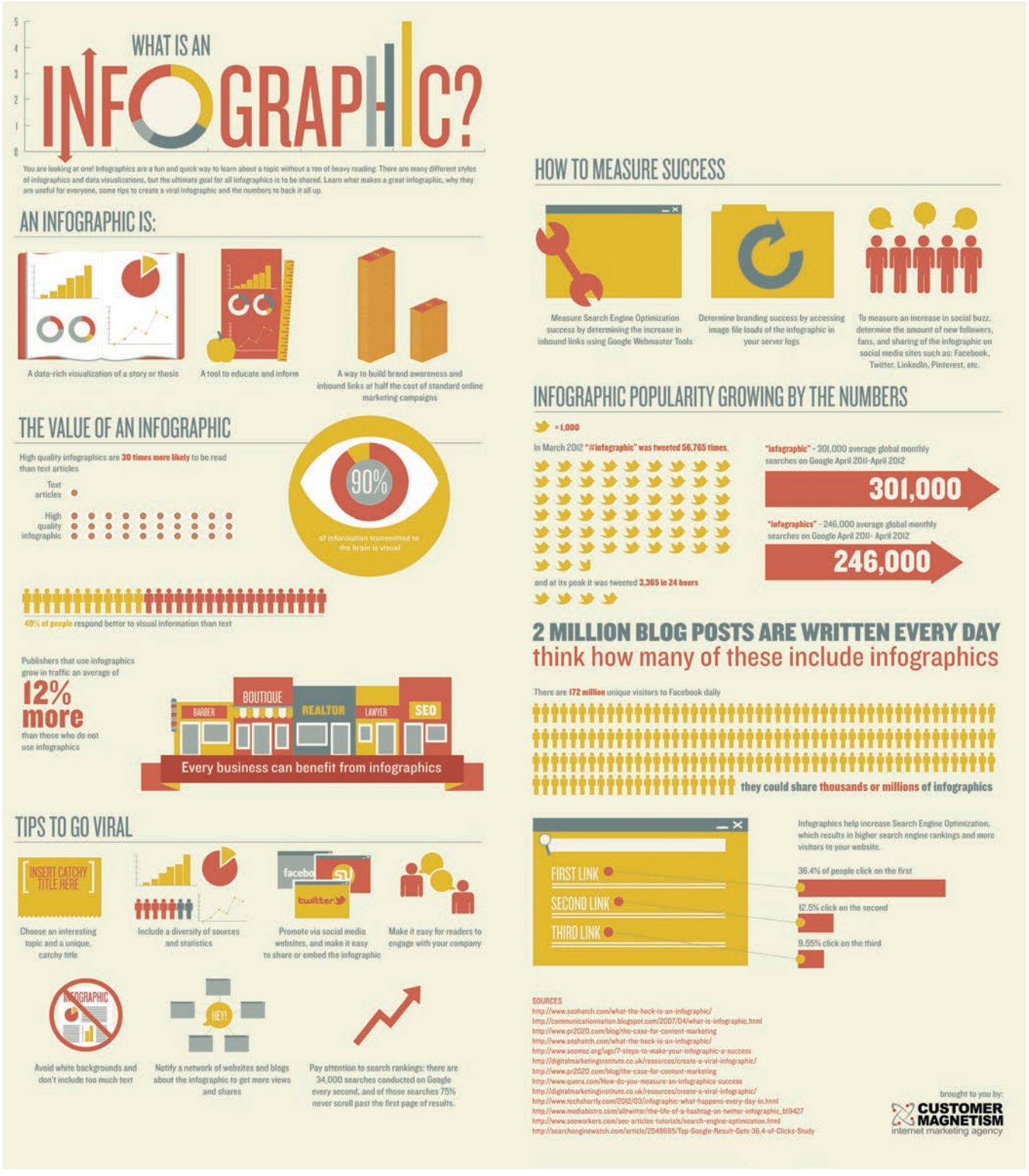
It is important to be creative when devising your infographic. An infographic is more than building a chart in a spreadsheet. Think of it as a creative poster that uses illustrations, icons, graphics and different fonts and colours to make a point. Pages 48–68 show examples of infographics that you may find useful as you prepare for your Outcome.

The steps to creating an infographic are as follows.

1 Conduct an investigation into an issue, practice or event.
2 Interpret the primary data.
3 Familiarise yourself with different ways to visualise data.
4 Mock up a design of the infographic.
5 Create the infographic.
6 Include references/citations/sources in the infographic.
7 Apply formats and conventions to the infographic.
8 Test and validate the infographic. You can follow the same process to create most graphic solutions.

Rather than writing the outcomes of research into written documents, the main points of research can be presented in a visual format. This can be more engaging for the reader.

The final product of Unit 1, Outcome 1 will likely be an infographic. Your infographic should be visually engaging, contain a subject matter that is appealing to the target audience, and be supported by other engaging content.



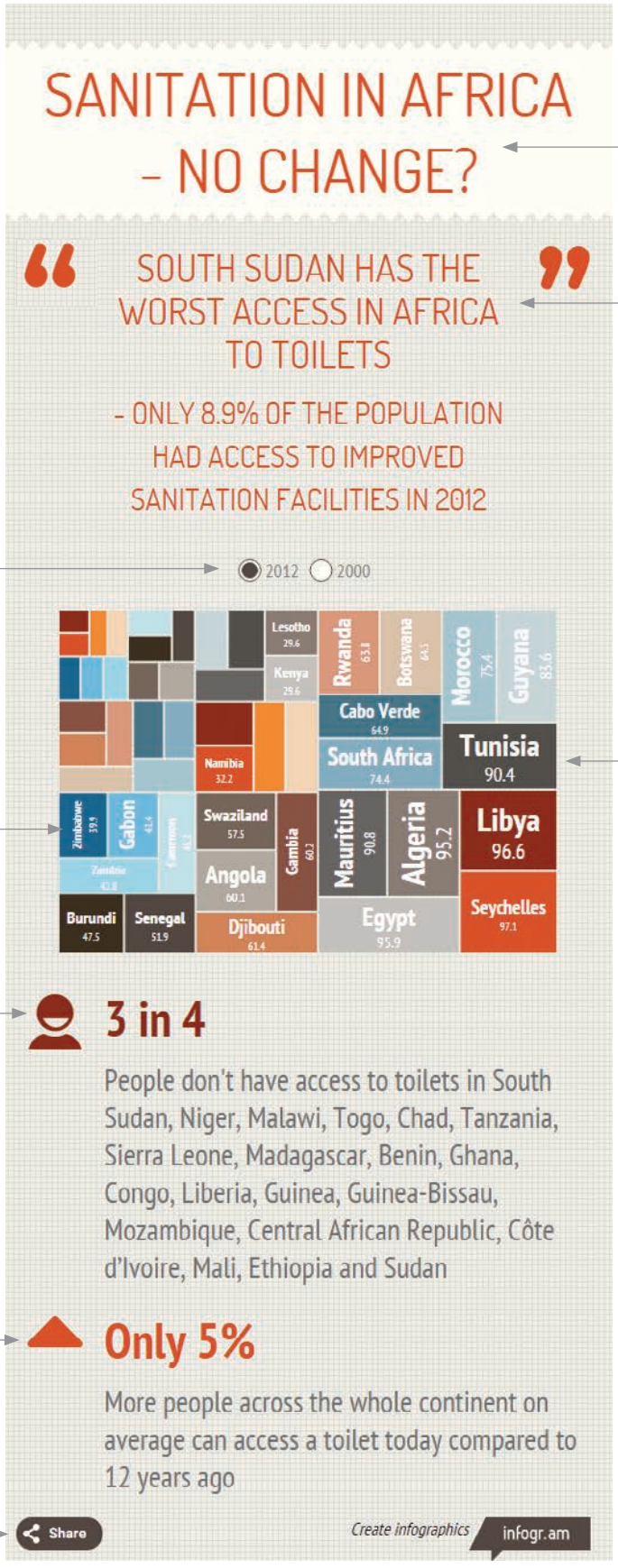
Customer Magnetism

FIGURE 2.16 What is an infographic?

THINK ABOUT COMPUTING 2.1

Summarise what you think an infographic is, making reference to Figure 2.16.

NelsonNet additional resource: Figure 2.16 Infographic, p. 48



The topic appears at the top

Shows the most important findings at the top in the largest fonts, with large quotation marks to emphasise that it is a quote

Provides two different time periods

Shows the data charted in an unusual, interesting way that draws the user's eye; countries that are largest have improved access to sanitation; the smaller the country in the chart, the worse their access is

Uses a complementary colour palette of oranges and burnt reds with subtle hints of contrasting blues

Key findings are highlighted. Notice that these are the only two instances where text is bold.

Social media links are included

NelsonNet additional resource: Figure 2.17 Infographic, p. 49

Copyright Guardian News & Media Ltd 2015

FIGURE 2.17 A simple infographic on sanitation in Africa

NelsonNet additional resource: Figure 2.18 Infographic, p. 50

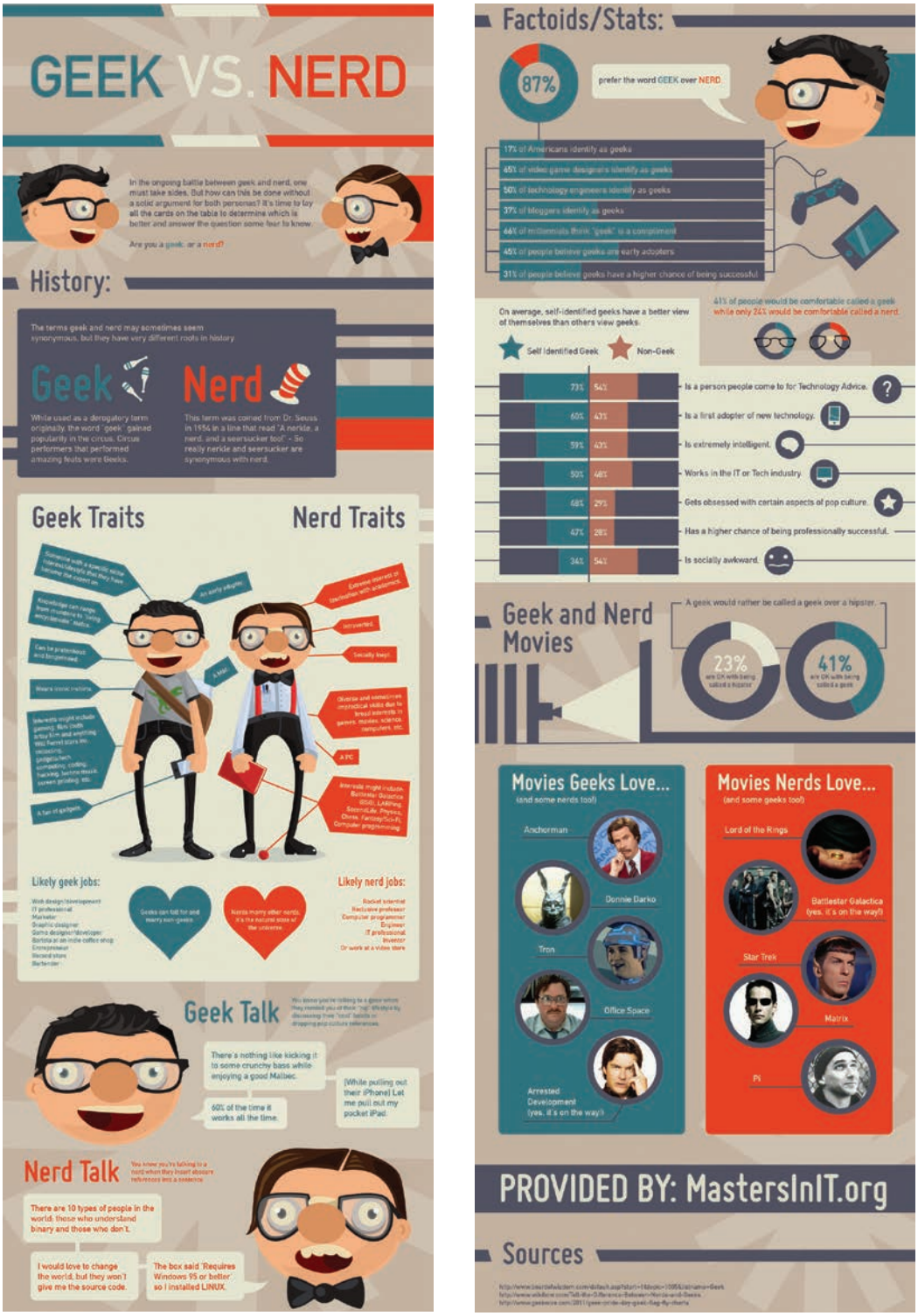


FIGURE 2.18 Another example of an infographic

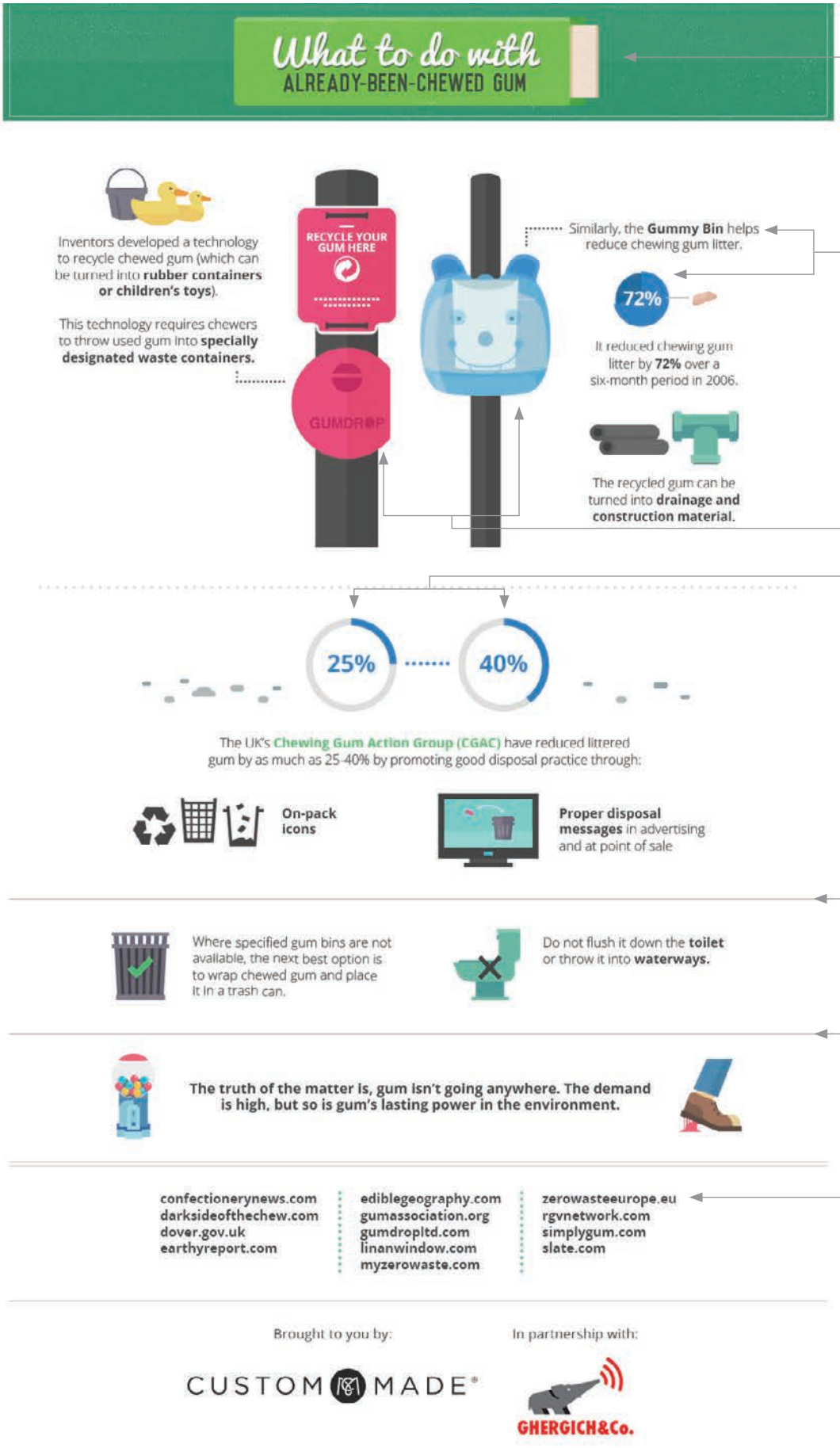
Christian Behrens, The Data Visualisation Catalogue

THINK ABOUT COMPUTING 2.2

1 Annotate Figure 2.18 to identify:

a the specific different types of charts and graphs used

b the general types of graphic solutions used.



Section titles (Titles)

Different text styles (Text styles)

Vibrant, contrasting colours (Colours and contrasts)

Shapes (Shapes)

Ad

NelsonNet additional resource: Figure 2.19 Infographic, p. 51

Faint lines act as section dividers (Lines and arrows)

Sources of data are cited (Sources of data and legend)

FIGURE 2.19 This infographic depicting what to do with chewing gum that has already been chewed has been labelled to indicate how it uses formats and conventions suitable for graphic solutions. See if you can add more labels.

CustomMade.com

Design principles for graphic solutions

Design principles are guidelines to help you enhance the appearance and functionality of solutions. Graphic solutions that are displayed onscreen need to be easily understood and accessed with minimal time and effort.

To communicate effectively, graphic solutions need to be clear and functional. You need to ensure that facts are obvious and your message is unmistakable. Your solution must be carefully designed, taking the design principles into account.

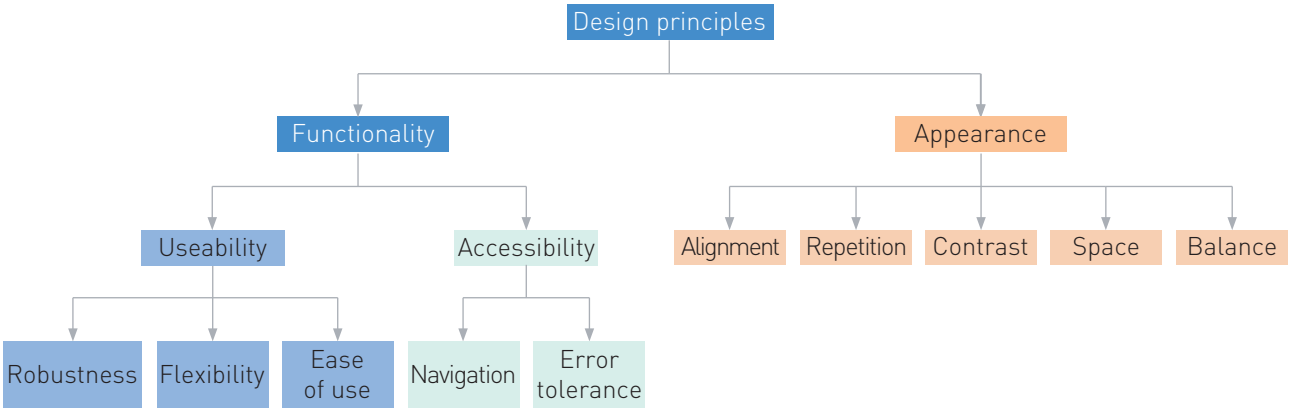


FIGURE 2.20 Design principles are accepted characteristics that contribute to the functionality and appearance of solutions. For VCE Computing, the principles that relate to functionality are **useability**, **accessibility** and **appearance**.

The principles shown in Figure 2.20 are discussed in the following section, in terms of how they may apply to your solution.

Functionality

Useability

All graphic solutions that represent data need to be easy to use, flexible and robust. The user should be able to acquire the required information easily and efficiently. When designing your infographic for useability, consider the following.

- Can the user understand the graphic form and the information it conveys?
- Can the user quickly identify the purpose of the graphic form and interpret its critical data?
- Is the graphic form accurate and a true reflection of the data that has been analysed?
- Have you saved the final infographic in a lossless format to maximise legibility, but in a size that will be quick to load?
- Have you chosen sensible dimensions? For example, you may want to think carefully before designing an infographic that is very narrow, that is very long and requires a great deal of scrolling, or that is wider than most website body panes (around 800px).
- Have you chosen easily readable font sizes?
- Have you tested the infographic in multiple browsers, and also printed it out in colour and greyscale to check that it works?

Accessibility

A user should be able to access meaning and relevant information from a graphic solution efficiently and easily. Users should not have to puzzle over the meaning or placement of text and symbols used.

To help users along, use simple and relevant labels and commonly understood symbols. For example, use $ to represent money.

When using your graphic solution, users should be able to navigate their way through text and images in a standard way – from top to bottom, or left to right. Use arrows if you literally want their eyes to move a certain direction.

Do not make anything so small or hard to read that it goes unnoticed. Important facts should be treated with appropriate formats and conventions that highlight why they are crucial.

Essentially, make it easy for users to glean meaning immediately. When you are creating an infographic, it can never be *too* obvious.

Appearance

Alignment

Alignment refers to arranging text, images and objects vertically or horizontally in either straight lines or correct relative positions. When using word-processing software, you may have already used text alignment tools, which allow you to justify (align) your text to the left, right or centre of the page. Horizontal alignment can be either left, right, centre or full justified. Full justification refers to the text being aligned on both the left and right margins with spacing distributed evenly across the line to achieve this. Vertical alignment can be top, middle or bottom.

Using alignment cleverly gives your text, images and objects a sense of order and organisation that helps to communicate your message clearly. For example, putting something in the centre of the page vertically and horizontally suggests immediately that it is of central importance. Putting something in a small font size at the bottom right would seem to be of little importance.

Smart use of alignment also helps to imply relationships between different elements of your solution. You will find alignment tools in all graphics software packages.

When planning your solution, manage the elements of your infographic by drawing a visual 'grid' (Figure 2.21). This will help you to work out where you can place items in your infographic, how much space you have, and what kind of alignment works best for each element.

Repetition

Repetition refers to the use of the same or similar visual elements repeatedly within a graphic solution. It is used to unify elements of a layout, and is achieved by repeating patterns, textures, fonts, colours and page elements. For example, a set of bullet points creates a sense of repetition that connects those points. On a graphic solution, repetition usually means that each section or module uses similar headings, colours, font styles or other visual cues so that it is easier for the user to understand the relationships between elements within the solution.

Contrast

Contrast refers to the visual difference in colour or tone between objects (both text and images) in a graphic solution. Greater contrast will make objects appear to stand out more from one another. If there is not enough contrast between two objects, they may appear to blend into each other, making it difficult for the user to see each of them clearly. Contrast between the background of your graphic solution and text should make the information clearly visible and legible.

The use of white space can enhance contrast around objects within your graphic solution.

Space

Space refers to the areas around and between objects – text and images. If your graphic solution is cluttered it may be difficult to follow. You may want to include lots of detailed information in

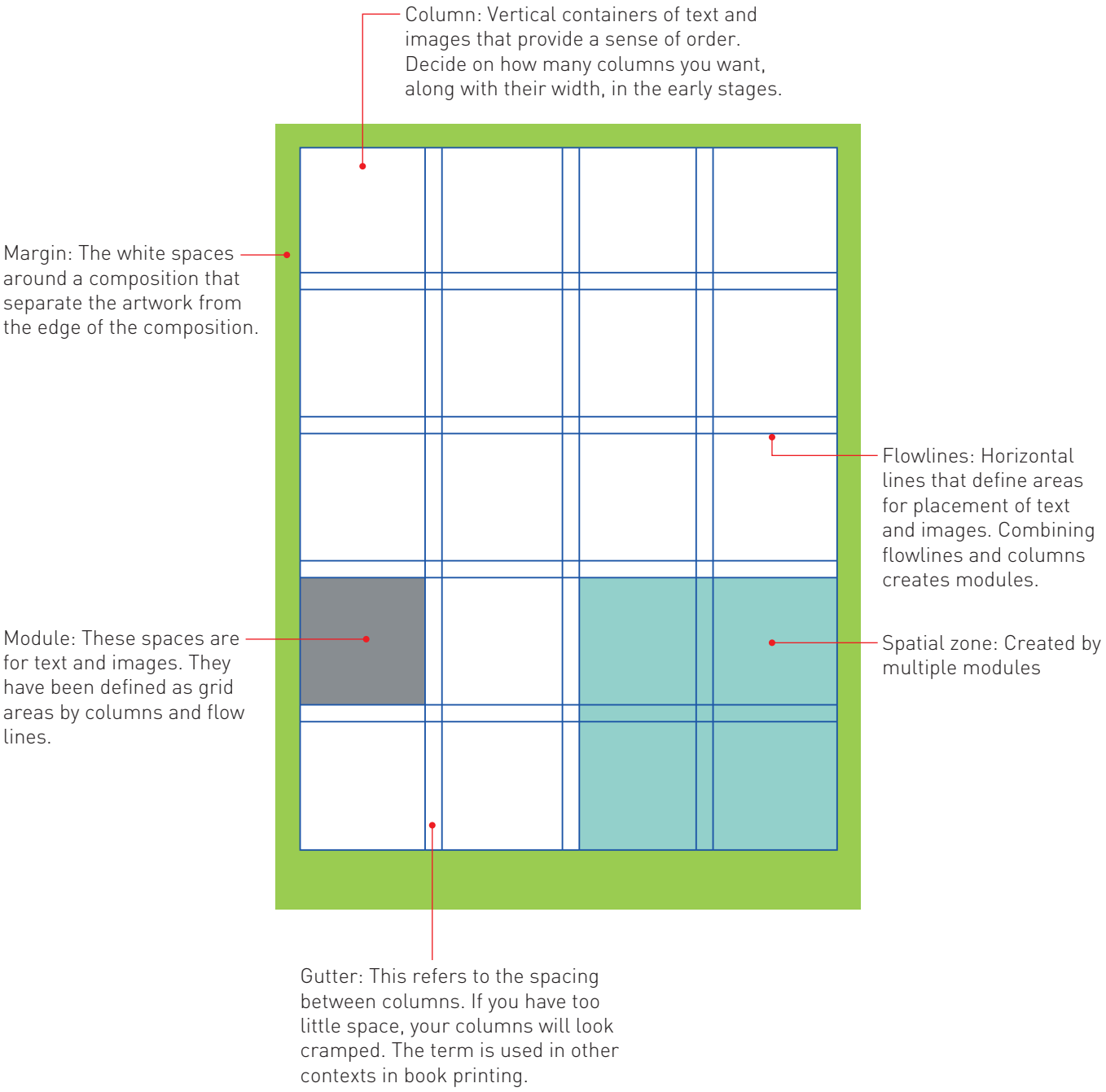


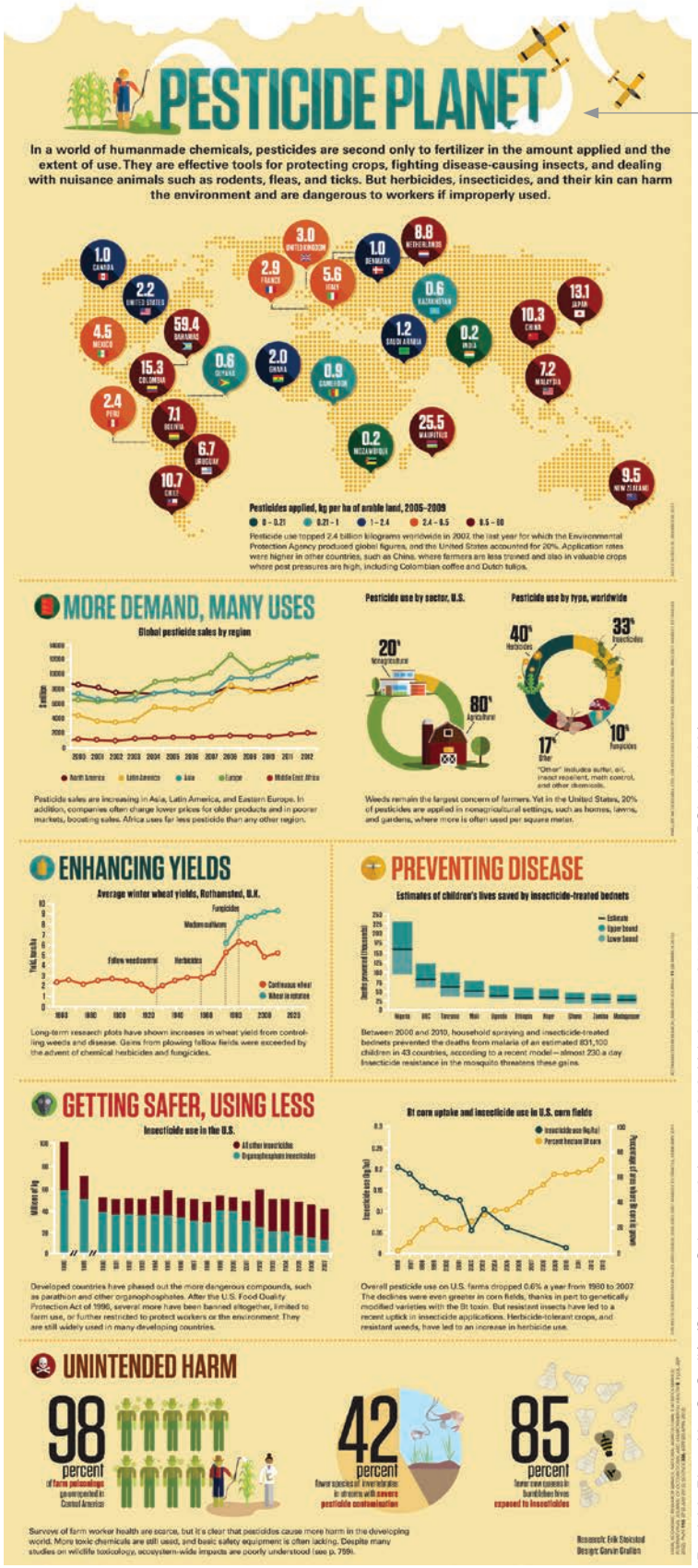
FIGURE 2.21 The visual grid will not just help you with alignment – it also helps with balance, space and accessibility. A downloadable version of this grid is available on NelsonNet.

NelsonNet additional resource: Visual grid template

your graphic solution, but it is important to put as much space between objects as is necessary for them to be distinguished as separate and navigated through in the correct order. That means that if you arrange your space carefully and correctly, it will control user behaviour.

In theory it may sound very trying, but you are mostly going to be using areas of space that are only between a few millimetres and a few centimetres. A visual grid will be of use.

A large area of **white space** may be used to balance a section that contains an equally large area of text, because it will be of equal visual 'weight'. Working with space means also working with balance.



'Pesticide Planet' by G. Grullón/Science, Science 16 August 2013: Vol. 341 no. 6147 pp. 730-731 DOI: 10.1126/science.341.6147.730. Reprinted with permission from AAAS

Succinctness
This infographic tries to convey data and information about a specific topic and problem. It visually displays the research concisely without too much extra information.

Orientation
Orientation refers to the direction and aspect of elements within a graphic solution. When designing your graphic solution, you must consider all elements with regard to the visual hierarchy of various objects. It is important to keep in mind the aspect of each object in relation to other objects and to the graphic form overall. When deciding on the orientation of objects, you should consider the direction of data labels (particularly if there are too many to fit horizontally), the alignment of data labels and text, and whether the title will appear above or below the graphic form.

NelsonNet additional resource: Figure 2.20 Infographic, p. 55

FIGURE 2.22 'Pesticide Planet' is the title for an infographic created by http://www.sciencemag.org/.

THINK ABOUT COMPUTING 2.3
Annotate the Sanitation in Africa infographic, Figure 2.17 on page 49, to show how it applies the appearance design principles.

Balance

A balanced solution will appear harmonious, and this creates visual appeal. Whether you understand or recognise it, you already like to see balanced designs. Do you ever straighten something that is crooked? Do you try to divide things into even pieces because somehow it just seems appropriate?

Unbalanced designs can lack the appropriate emphasis, can look untidy and they may end up discouraging users from viewing them.

In symmetrical balance, an object mirrors the elements on opposite sides of the visual axis, from one side to the other – you see the same amount of white space on the left side as on the right, for example. A person's left hand exactly mirrors their right hand.

Being balanced is not always about being symmetrical – using asymmetrical balance is also an option! This is characterised by an arrangement of elements that is not mirror or equal in appearance, but still uses the central visual axis.

Whether you experiment with asymmetrical balance or stick to symmetrical balance, both left-right balance and top-bottom balance are equally important, so use balance and alignment together when placing objects in your graphic solution.

Formats and conventions

In addition to the basic design principles you should follow when creating your graphic solution, there are a number of useful formats and **conventions** suitable for graphic solutions, such as titles, text styles, shapes, lines and arrows, sources of data and legend, and colours and contrasts.

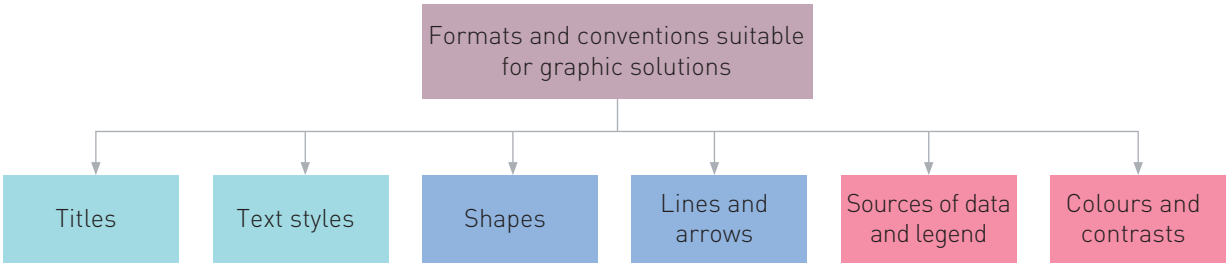
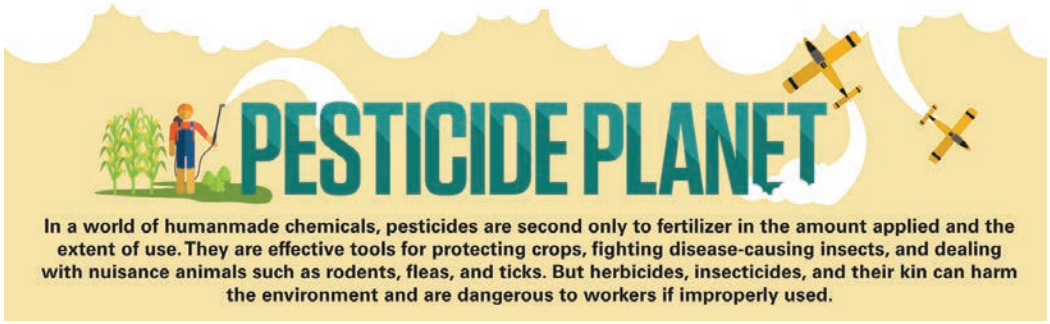


FIGURE 2.23 Formats and conventions

Titles

In the simplest of terms, adding a title to a document makes it a dominant element. Titles are generally styled as headings, with type that is bold and larger than the body text or subheadings. Titles make an impression. Titles should be concise, to-the-point and easy to say. Your title should be in larger text than the rest of your solution – perhaps at least 20pt if the body text is 10pt.



'Pesticide Planet' by G. Grullón/ Science, Science 16 August 2013: Vol. 341 no. 6147 pp. 730-731 DOI: 10.1126/science.341.6147.730. Reprinted with permission from AAAS

FIGURE 2.24 'Pesticide Planet' has a large, visually interesting title.

Text styles

When we discuss text styles, we are essentially talking about fonts. A font is a typeface (such as Times New Roman, Arial or Calibri) plus its attributes (20pt, bold, red). You may already know a few standard, familiar typefaces, such as:

- Times New Roman, a serif typeface. Serif typefaces have tiny marks or 'tails' on the end of the horizontal and vertical strokes of each letter. Serifs are used in books for body text and especially for long passages of text.
- Arial, a sans serif typeface. Sans serifs do not have the serifs on the strokes of each letter. They work best for short paragraphs, large headings and online, but not for long passages of printed text.
- `Courier New, a slab serif typeface, is often used in programming. Slab serifs are best used when the focus is on function and not appearance. Slab serif fonts ensure your characters are legible and unmistakable.`

When you are choosing text styles for your graphic solution, keep things simple. Use a few well-chosen typefaces, perhaps three at most, and use bold, italic, colour and point size to set out heading levels and distinguish between different types of text.

There are other typeface styles, such as handwriting, script and decorative. You will know decorative fonts such as **Impact** because it is the typeface predominantly used in memes online. *Comic Sans*, a casual script or handwriting typeface, is also well known.

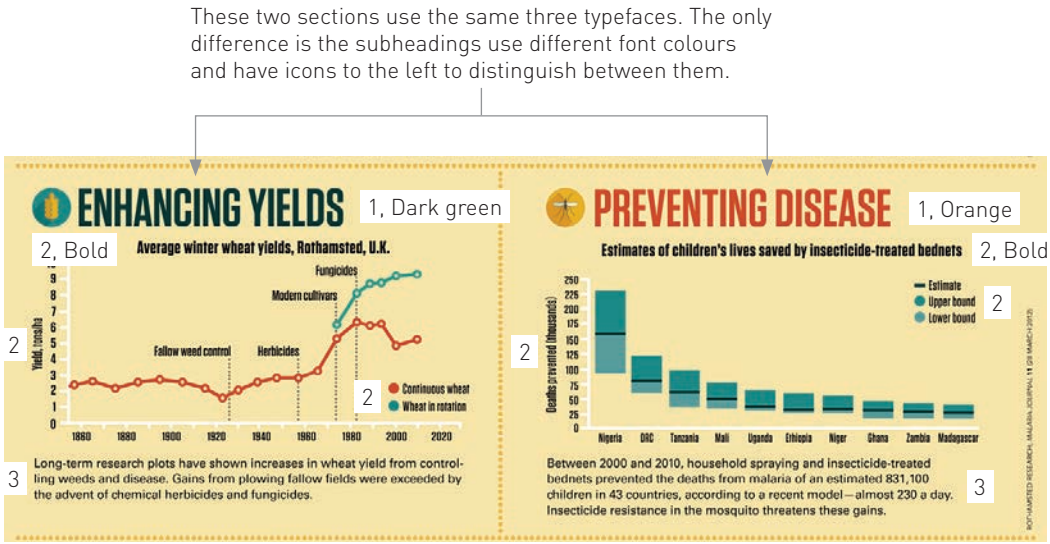


FIGURE 2.25 Subtle changes make a difference in Pesticide Planet

Text styles will apply some contrast while promoting a streamlined, professional appearance. However, using many different typefaces in one space can be untidy and overwhelming.

Remember: less is still more – bigger is not always better. Really think about what needs to be emphasised the most and what needs to be highlighted. Not everything needs to be bold, italic and 40pt.

Shapes

Using shapes in your solution can help to create patterns, contrast, hierarchies and backgrounds. You can use shapes as containers for sections of text in your graphic solution, and as dividers.

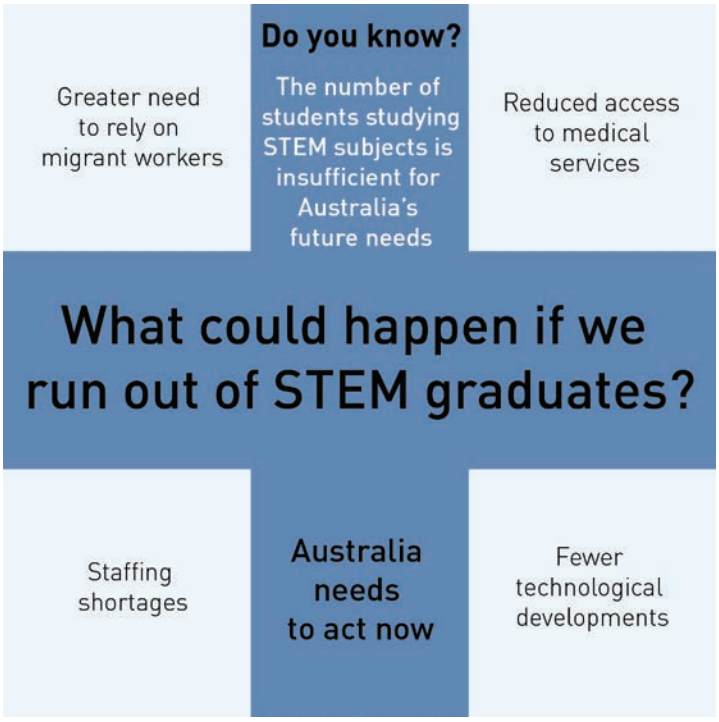


FIGURE 2.26 The cross shape here acts as a divider and container

You do not have to stick to standard, two-dimensional geometric shapes such as squares, circles, rectangles and triangles. Other types of shapes, such as irregular, abstract and freeform shapes can evoke reactions in the user. You can also use shapes to develop logos, symbols and icons.

FIGURE 2.27 These familiar shapes may not be what you immediately think of when shapes come to mind, but they may still evoke strong reactions that could make them powerful in your graphic solution.

THINK ABOUT COMPUTING 2.4

1 Make a list of what you immediately associate with each of the shapes in Figure 2.27.

2 Compare your list with another of your classmates. How much do they overlap?

Lines and arrows

A line is a versatile visual element that uses only length and width. Lines can be:

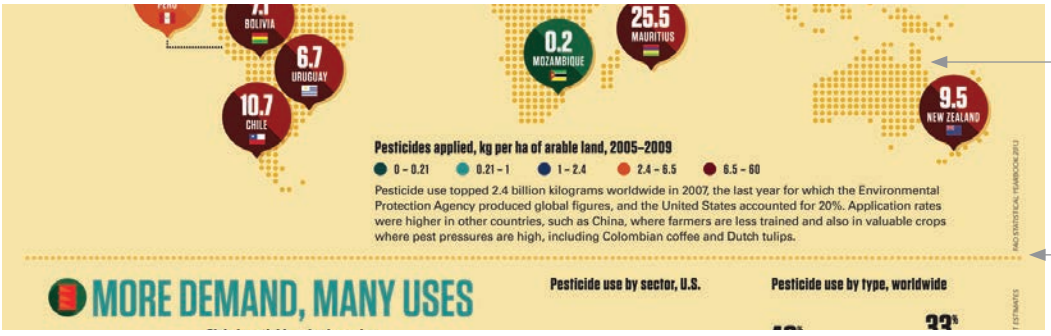
Solid	
Dashed	
Dotted	
Broken	
Double	
Thick	
Thin	
Curved	
Freeform	

- Bold or thick lines work well for emphasis and for representing a structure within a space. The thicker the line, the more it will draw the eye to the space, but the more crowded and boxed in it will look, so use a thick line carefully.
- Light and fine lines can suggest technical details but also retain a sense of minimalism. Line thicknesses you might use range from:

0.25pt
0.5pt
0.75pt
1pt
1.25pt
1.5pt
1.75pt
2pt

You can use lines in your graphic solution as borders or containers for sections of text or images.

The 'Pesticide Planet' infographic uses dotted lines as dividers, but also as a form of repetition from the maps at the top. By using lightly coloured lines that are similar to the background colour, and dots rather than solid lines, the infographic allows the user to read the infographic in the correct order and tell sections apart, but without a sense of crowding that solid or darker coloured lines could have created.



Yellow dots form shapes

Yellow dots form lines

'Pesticide Planet' by G. Grullón/ Science, Science 16 August 2013: Vol. 341 no. 6147 pp. 730-731 DOI: 10.1126/science.341.6147.730. Reprinted with permission from AAAS

You can also use arrows as pointers in your graphic solutions. There are a variety of arrows and arrowheads to choose from:

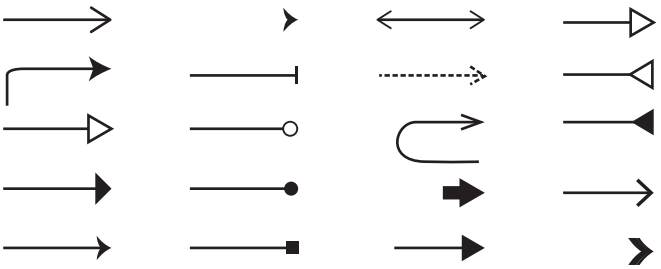


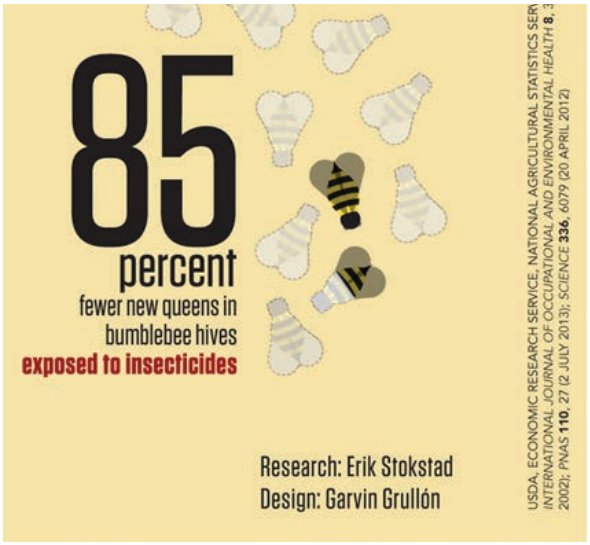
FIGURE 2.28 Popular arrows and arrowheads

Some arrows are sophisticated and elegant, while others are basic. The colour of your lines plus the colour of your arrow and choice of arrowhead can make a difference in the overall appearance of your diagrams.

Sources of data and legend

If you are using data in your graphic solution you need to identify the source in an appropriate way. If you are designing an infographic for your solution you could include a list of all of your sources in the footer of the infographic, similar to 'What is an infographic?' on page 48.

Alternatively, you could cite your source when it is used, similar to 'Pesticide Planet'. Note that 'Pesticide Planet' runs the source vertically up the side of the infographic:



Citing information resources is a must when working with infographics because the data behind them is research-based. Citing sources also provides those who view the infographic the opportunity to further research the topic.

Make sure you cite all of your sources correctly. Include the name of the source, the author, the URL (if applicable), the page number, the date, the publisher and any other relevant information.

You should also make use of legends in your graphic solution when needed to identify the facts shown in charts or graphs clearly. In general, a legend or key explains the symbols used in a chart, diagram, map or table. In terms of your graphic solution, a legend will mostly be used as a patterned marker with blocks of colour that represent different groups of data in a chart.

'Pesticide Planet' includes multiple colour-coded legends – one for each 'module' that has a chart. The legends in 'Pesticide Planet' make it easy to understand what the data stands for and thus what each chart means, which is why legends are so useful. If you do not include a legend for a chart with a complex idea, the user may become confused about what is being shown on each axis and interpret your chart incorrectly. You can design them to take up very little space, as shown below.

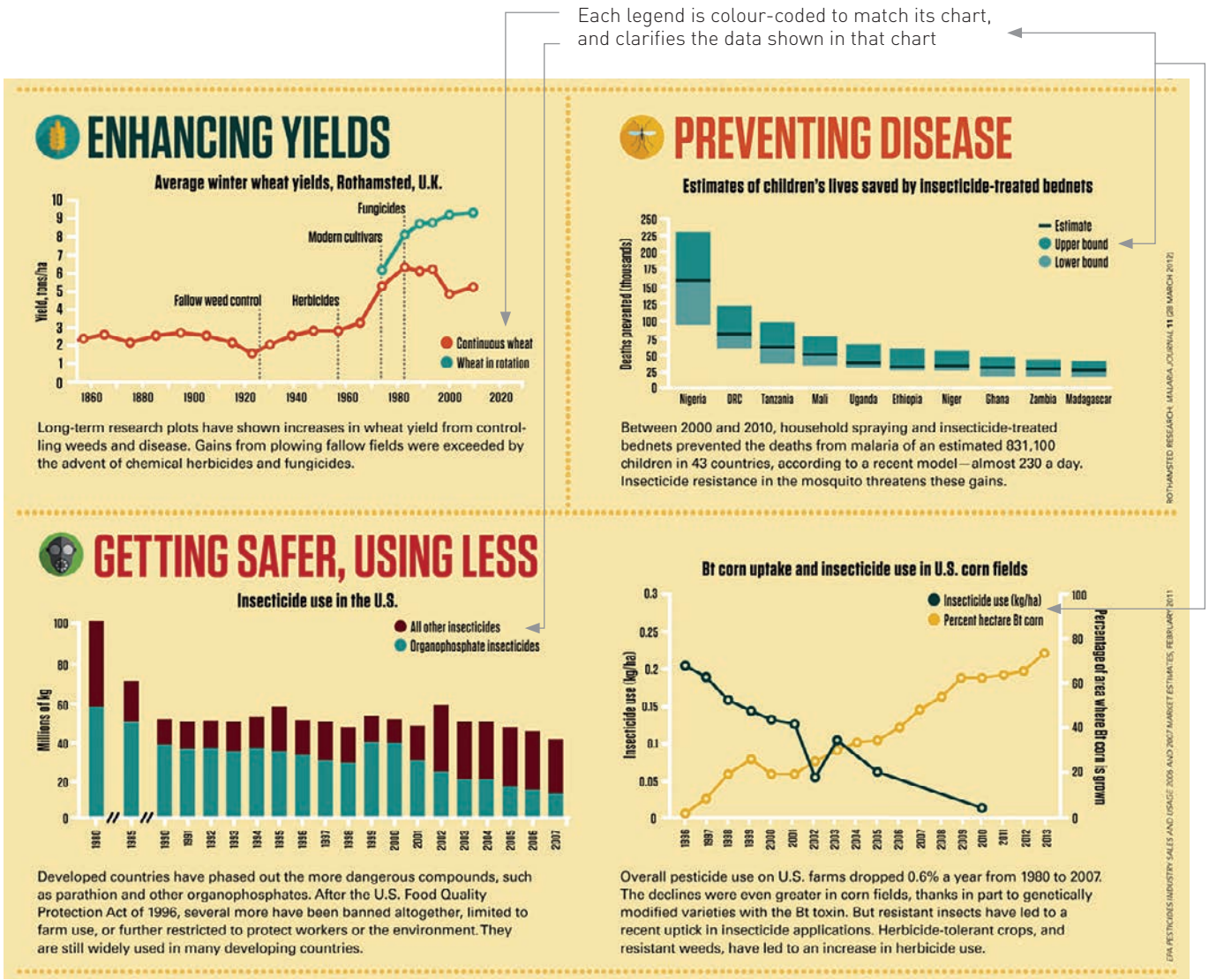


FIGURE 2.29 Legends in 'Pesticide Planet'

'Pesticide Planet' by G. Grullón/Science, Science 16 August 2013: Vol. 341 no. 6147 pp. 730-731 DOI: 10.1126/science.341.6147.730. Reprinted with permission from AAAS

Colours and contrasts

Colour should be used so that it makes the information clear, readable and attractive. The colours should emphasise important features, and a colour scheme should be used to ensure consistency. The following conventions for onscreen colour can be useful in determining colour schemes.

The most easily readable colours for text are black writing on a white background. Avoid using red and green together because people who are colourblind have difficulty distinguishing between them. Blue and brown together can also be difficult to read.

Light shades are best used for backgrounds. Avoid using yellow or other light colours for text on a white background.

Avoid using bright, neon or vivid colours, except where you wish to highlight an object or piece of information.

Limit the number of different colours used in your graphic solution.

As discussed on page 53, contrast refers to the visual difference in colour or tone between objects in a graphic solution. Greater contrast will make objects appear to stand out more from one another. If there is not enough contrast between two objects, they may appear to blend into each other, making it difficult for the user to see each of them clearly. Contrast between the background of the graphic representation and text should make the information clearly visible and legible. The use of white space can enhance the contrast around objects within the graphic representation.

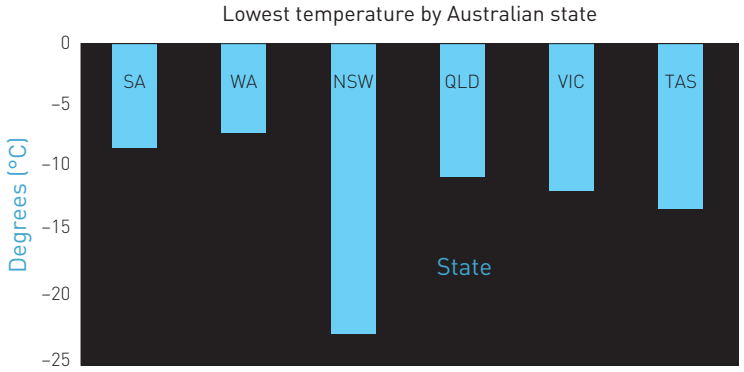


FIGURE 2.29 This graph shows use of clear contrast.

FORMATS AND CONVENTIONS

1 Identify the infographic in this chapter that you think has the most effective **title** and justify your choice.

2 **a** Identify the infographic in this chapter that you think uses **text styles** most effectively and justify your choice using examples.

b Which, if any, of the infographics in this chapter could be improved by more effective use of text styles? Suggest how improvements could be achieved making reference to point sizes, typefaces, bold/italic, and so on.

3 **a** Identify the infographic in this chapter that you think uses **shapes** most effectively. Justify your choice using examples.

b Which, if any, of the infographics in this chapter could improve upon their use of shapes? Justify your choice and suggest how to achieve this by hand drawing or mocking up the infographic with the new shapes that could be used.

4 Other than Pesticide Planet, identify the infographic in this chapter that you think uses lines and arrows most effectively. Justify your choice and outline the types of **lines and arrows** used, with reference to thicknesses and arrowheads.

5 Describe how the **sources of data and legend** shown in the different infographics in this chapter differ from the citation methods described in Chapter 1. Explain the correct way to cite your sources in your own work.

6 Identify the infographic in this chapter that you think uses **colours and contrasts** least effectively. Justify your choice and suggest ways to improve the colour palette.

THINK ABOUT COMPUTING 2.5

Annotate 'What Happens in an Internet Minute' by identifying formats and conventions used. Identify one type of each of the following formats and conventions: titles, text styles, shapes, lines and arrows, and colours and contrasts.

Visualisation

Infographics should not be text-laden. Lengthy text can take longer to digest, while there is some immediacy with images. By using a variety of tools such as shapes, charts, icons and diagrams, you can can assist readers with visualising the data. Given that the purpose of an infographic is to provide information in a visual format, using visual cues will assist the reader with interpreting the infographic.

NelsonNet additional resource: Figure 2.31 Infographic, p. 63

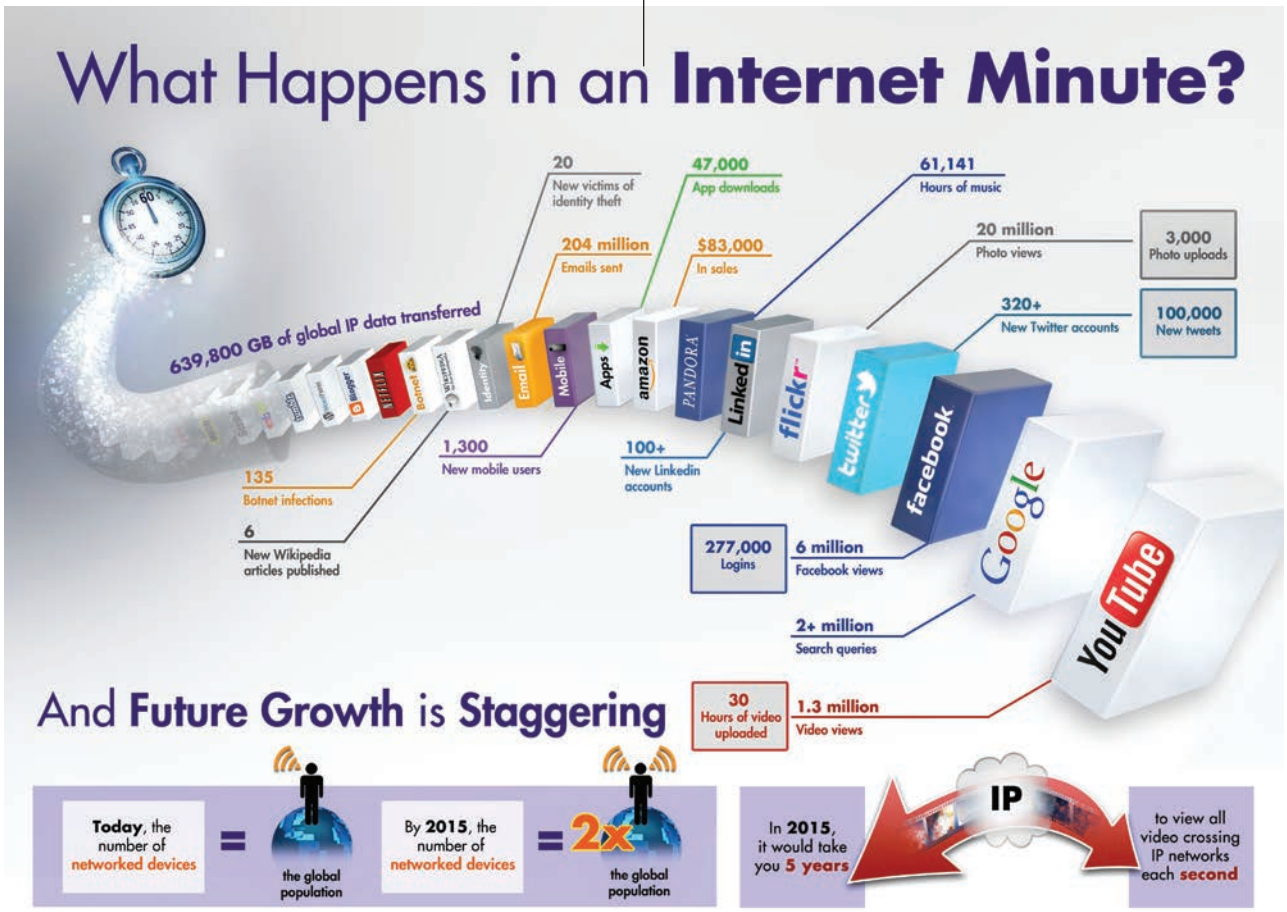


FIGURE 2.31 A creatively drawn infographic using colour, shapes, icons, lines and arrows, graphics and illustrations, with minimal text.

THINK ABOUT COMPUTING 2.6
Annotate two of the psychology of colour infographics: one for design principles (discussed in Figure 2.20), and one for formats and conventions (discussed in Figure 2.23).

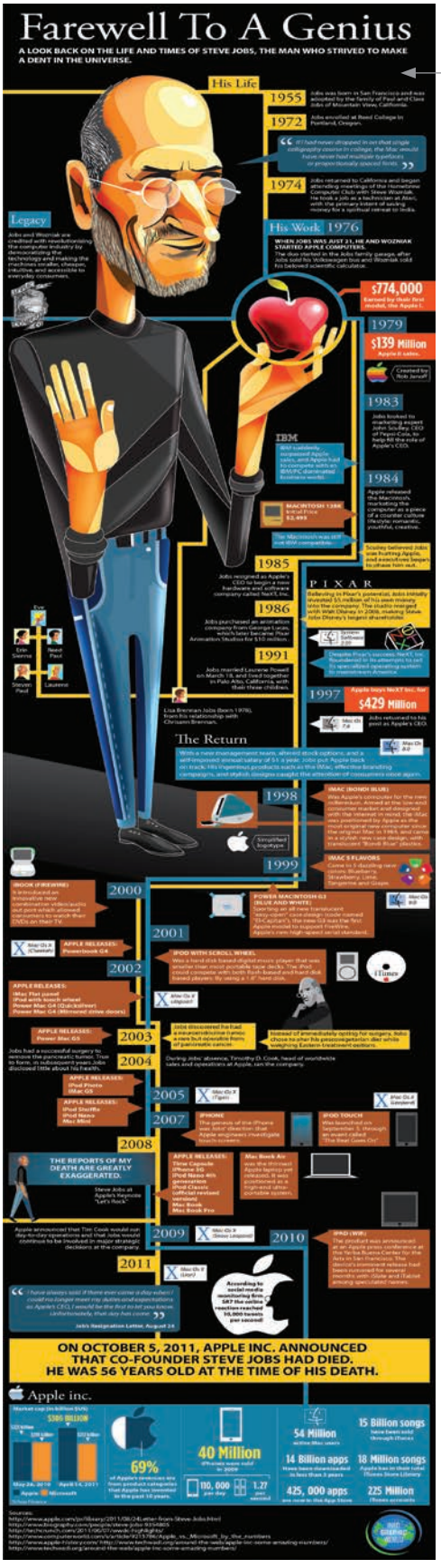
NelsonNet additional resource: Figure 2.32 Infographic, pp. 64–5

WebpageFX

FIGURE 2.32 The psychology of colour

WebpageFX

FIGURE 2.32 *Continues*



Info Graphic World

FIGURE 2.33 Sources are listed at the bottom of the infographic, making them unreadable in printed form.

Proportion
A hierarchy is an arrangement of items in which each is ranked above, below or at the same level as another to show their importance. Proportion refers to the visual hierarchy of objects within a graphic representation. The proportions (or relative sizing) of objects affect the visual hierarchy of the graphic solution. This infographic does not function unless it is viewed entirely on-screen and can be zoomed in to 100% because of its unusual proportions.

Design tools

Design tools are used to represent the functionality and appearance of graphic solutions; normally, the solution is first sketched on paper. Several design tools can be used to represent the design of graphic solutions; however, **annotated diagrams** are the most apt for planning a solution and presenting it.

Software such as spreadsheets can assist with the manipulation of data to develop graphs and charts. Spreadsheets include functions that enable users to work quickly and with less effort, thereby minimising cost.

Spreadsheet functions and design capabilities are discussed on page 31.

A wide range of design tools can be used to design the appearance of a graphic solution. They include:

- **input–process–output (IPO) charts**
- annotated diagrams or mock ups.

IPO chart

An IPO chart (Table 2.9) is used during the design stage to clearly identify the solution's input and output, and the processing steps required to transform the data into information. By completing an IPO chart, the developer gets a sense of how much formula development work might occur during the manipulation stage of the project. An IPO chart can be used to show how data is processed into meaningful information. An IPO chart, also called a 'defining diagram', identifies what data is required for the solution (input), what information the solution needs to produce (output), and the processing manipulation activities required to transform the data into information, or the function of the solution. In the case of creating a graphical solution, the input requires an understanding of what data is needed and where it is coming from, the process focuses on the functionality of the solution and the output refers to the graphical solution that is being created.

NelsonNet additional resource: Figure 2.33 Infographic, p. 66

TABLE 2.9 An example of an IPO chart for investigation into whether the number of students undertaking STEM subjects is sufficient for Australia's future needs.

Input (data)	Processing (calculations)	Output (information)
Selection of year level of participant (i.e. Year 11 or Year 12)	Tally the results in each category	Aggregate total of each category
Selection of STEM subject studied (ie Foundation Maths, Further Maths etc)	Tally the results in each category	Aggregate total of each category
Identification of factors to experiencing success STEM is selected	Tally the results in each category	Aggregate total of each category
Identification of influence on decision to pursue STEM subjects	Tally the results in each category	Aggregate total of each category
Selection of reasons for pursuing a STEM career	Tally the results in each category	Aggregate total of each category

Annotated diagrams and mock ups

Annotated diagrams provide a visual depiction of how graphic solutions should look (Figure 2.34).

The designs indicate features such as differences in font size, colour and positioning of objects. The placement of objects must be planned so that a balanced, visually appealing and clear effect is attained.

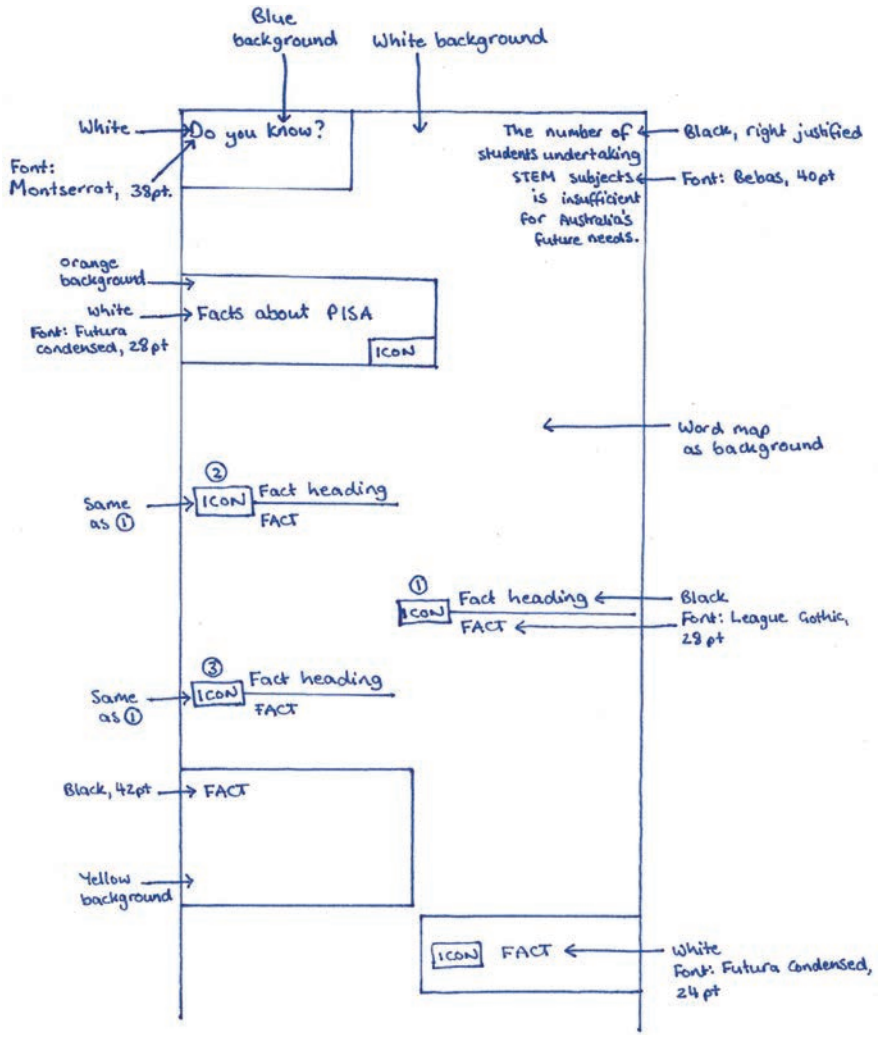


FIGURE 2.34 An example of an annotated diagram for the STEM infographic

THINK ABOUT COMPUTING 2.7

1 Hand draw one of the infographics from this chapter as an annotated diagram similar to Figure 2.34. (You will not be able to identify specific typefaces when you do this.)

2 Hand draw Figure 2.21 from page 54. Try to use it to identify modules, flowlines, columns and gutters in the infographic you have hand drawn.

3 Label the columns, gutters, margins and flowlines on the hand drawn infographic.

The annotated diagram should clearly indicate:

- the shape of graphic solutions
- titles and layout of the diagram
- formats and conventions that are used in graphic solutions.

Before using an annotated diagram, you should consider the data set that will be used as the basis of the graphic solution. The amount and type of data, and purpose for developing a graphic solution of the data will help you decide on the type of graphic solution to create.

Types of tests

It is important to check that what you are trying to produce meets the specified need and fulfils the specified purpose. The **effectiveness** of a solution is determined by its quality. A solution is effective if it is complete, readable, attractive, clear, accurate, accessible, timely, relevant and usable; and if it communicates the message. Each of these attributes needs to be tested once the solution has been built.

Validation versus testing

Validation is involved with input, while testing is concerned with the solution itself, or output. **Testing** can, however, also involve ensuring that **electronic validation** works correctly. In the design stage you need to plan the testing that you will conduct after the development stage. Therefore, after the designs for the graphical solution have been created, we make the test plan or test table. Testing and validation serve different purposes. Testing is performed to ensure that processes, functions and formulas in a solution work as expected. Validation checks that input data is appropriate and reasonable.

Completeness testing

A solution is considered incomplete if the intended user has to find additional information. A graphic solution that does not provide sufficient information will not fulfil its purpose. To check the completeness of a graphic representation, the developer should consider the following.

- Does the graphic representation provide the information required?
- Is the graphic representation clearly and accurately labelled?
- Will the intended users easily find what they need?

Reliability testing

The digital revolution has opened up limitless ways in which information can be communicated, and access to this information is becoming easier and cheaper. But is all of the information reliable? For many different reasons, the data and information presented to us may be untrustworthy. To determine the reliability of data, the reliability of the source should be checked. One way in which the information given in a solution can be verified is to confirm the information from more than one source. Unreliable graphic solutions can cause confusion and information about a particular topic.

Presentation testing

During the design phase of the problem-solving methodology a decision must be made on the appropriate format of the solution; for example, how a solution should be communicated. It is

important to consider its overall look and format. The solution must be attractive and clear. One of the most important considerations should be the intended audience and intended purpose.

Some of the formats and conventions suitable for a graphic solution that can be tested are:

- appropriate use of white space and fonts
- the use of consistent font styles and sizes
- readability – that the information presented is easy to read and can be understood by the intended audience
- consistent and appropriate use of colour and contrast
- careful choice of background colours (generally they should be white or grey)
- consistent use of a colour scheme (perhaps matching the corporate colours or image of the client)
- easy-to-read charts and labels
- adherence to the formats and conventions discussed earlier in the chapter.

No matter what format is chosen, the presentation of the solution can be a difficult attribute to test. What looks good to one person may appear ugly to someone else.

Functionality testing

The functionality of a solution relates to its ability to perform tasks and functions. If a solution is deemed functional, it will provide accurate output that is reliable.

When testing the function of a graphic solution, it is prudent to test that the representation accurately depicts the data as it was meant to. A useful way to organise this testing is to develop a test plan.

A test plan is a method for recording the tests to be executed and the results of the tests. Usually, it includes the type of test, what test data will be used, what results are expected and the end results (Table 2.10).

When constructing your test table, take your time. Complete it while developing your solution, using any mistakes you make as a basis for good testing.

TABLE 2.10 A test table

Feature tested	Test/sample data	Expected result and why	Actual result

Relevance testing

Relevance describes how applicable a solution is to its intended purpose and its intended audience. A solution that lacks relevance can be misleading and uninteresting. A question to consider in checking the relevance of a solution is: 'Does the infographic match the user's search for information?'

Communication of message

All of the required information presented in a solution must be clear and obvious. For instance, a graphic representation that is meant to show the unemployment rates in Australia over the last ten years must convey the trend and the figures plainly, without the user getting lost among other insignificant or irrelevant information.

Questions to ask

To test the quality of onscreen information that is represented in graphic form specifically, ask the following questions.

- Does the graphic representation depict the information required and fulfil its intended purpose?
- Is the overall look and tone of the graphic representation appropriate to its intended audience?
- Is the graphic representation accurate? That is, has the data source been validated and verified?
- Is the type of graphic representation the most appropriate type for the data?
- Is there anything that is misleading, confusing or unclear?
- Are the axes correct?
- Are there any unnecessary elements or information in the graphic representation?
- Is the numerical scale of the value axis identified (for example, thousands or millions)?
- If the chart uses two value axes, can the audience easily identify the appropriate axis for each series?
- If more than one chart is being developed for the same solution, are the charts consistently formatted and presented?
- Is all of the text readable?
- Can any of the information presented be further summarised?
- Are font styles, sizes and colours consistent?
- Is all text spelled correctly?

Validation

To make the solution accurate and reasonable, all data that is used should be validated. This means that the original data should be manually checked for illegal data types, for reasonableness, for correct spelling, to ensure that data fall within a correct 'range' or that any codes that are used are consistent (follow similar style) and reasonable (similar codes relate to similar products). Although **validation** is actually used when entering data during the development stage, it needs to be planned for in the design stage.

It is easy for data entry errors to occur during the input phase, particularly if a large amount of data is involved. Data should therefore always be validated before any processing occurs.

There are two types of **validation**: manual and **electronic validation**.

Manual validation occurs when the data entered are checked for accuracy by a person rather than by a machine. Proofreading is one manual validation technique. When you proofread data, you look for transcription errors. The data entered should be compared with the source document to ensure that they match. Any differences observed must be followed up.

Data may be validated by computer if a validation function is contained in the software or built into the solution to a problem. Electronic validation techniques, sometimes called machine-validation techniques, ensure the accuracy of data and are built into software, such as spell checkers. Types of electronic validation include range checking, existence checking and data type checking.

Manipulating data to create solutions

The development stage of a problem-solving methodology involves using the appropriate software to create the solution. This is the stage at which the data (in this case, numbers, text and audio data from interviews) becomes information. Manipulation occurs when the data

is transformed into information; that is, it is the process of making sense of the data. Before computers existed, manipulation took place by hand and was subject to human error and interpretation. With the aid of computers, however, data manipulation now requires far less effort. The nature of the manipulation usually depends on the software being used to create the infographic.

CASE STUDY THE NUMBER OF STUDENTS UNDERTAKING STEM SUBJECTS IS INSUFFICIENT FOR AUSTRALIA'S FUTURE NEEDS

Questionnaires are often left open for a finite amount of time to ensure that as many participants have the opportunity to respond. Once the questionnaire has closed, data can then be manipulated to assist with the development of the solution. Software packages such as Opinio (online survey) can assist with filtering data to create totals of categories found in the survey. For example, you may want to find out the number of Year 12 students studying Informatics. To do this, you need to filter the remaining data to ensure that you receive the data that you are after. In this case study, there were 10 students who satisfied this criterion. The software was able to make a report on those 10 students, exploring the other STEM subjects they were studying as well as Informatics (see Figure 2.35). Various reports need to be made to assist with data interpretation so that the research question is being answered thoroughly. Various data sources such as the PISA and TIMSS results are also referenced to provide supplementary data to help answer the research question. Once all the data has been gathered, the mock-ups or annotated diagram of the infographic needs to be drawn (Figure 2.34).

As there is no prescribed software in Outcome 1, students can use a variety of software to create their infographic such as a spreadsheet, a word cloud generator, illustrating software, or an online infographic software generator.

Software that will be used to assist with the development of the infographic needs to be selected such as Excel and Piktochart. These packages have been selected to assist with the development of the infographic (Figure 2.36). The infographic needs to be put together in a creative and logical way by starting with one section at a time. The first section looks at the PISA results (Figure 2.37) and provides some background information about the results and Australia's placement in mathematics. The next section focuses on the questionnaire results and is creatively displayed in a Venn diagram (Figure 2.36). The other sections of the infographic needs to be developed and then tested to ensure that the infographic is complete and answers the question.

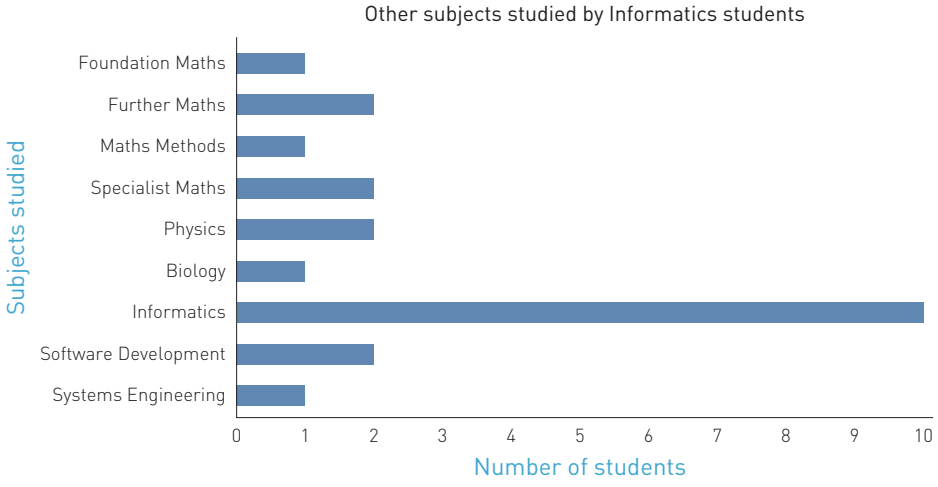


FIGURE 2.35 Data has been filtered to only include the Year 12 students who are studying Informatics; this report wanted to see what other subjects those students were studying.

TABLE 2.11 Frequency table

Choices	Absolute frequency	Relative frequency by choice (%)	Relative frequency (%)	Adjusted relative frequency (%)
Foundation Maths	1	4.55	10	10
Further Maths	2	9.09	20	20
Maths Methods	1	4.55	10	10
Specialist Maths	2	9.09	20	20
Physics	2	9.09	20	20
Biology	1	4.55	10	10
Informatics	10	45.45	100	100
Software Development	2	9.09	20	20
Systems Engineering	1	4.55	10	10
Sum:	22	100	–	–
Not answered:	0	–	0%	–
Total answered: 10				

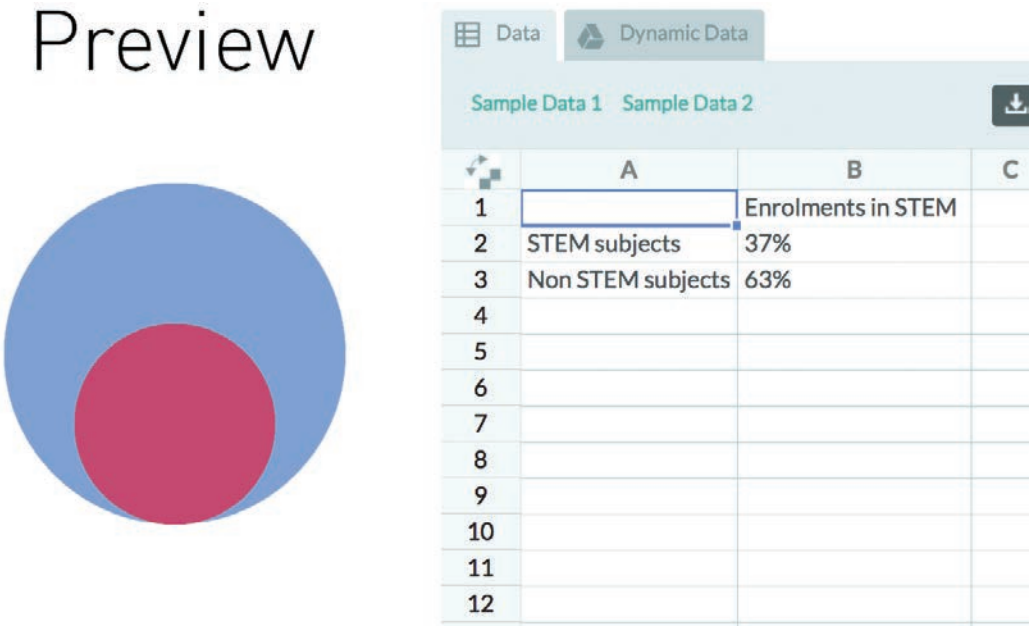
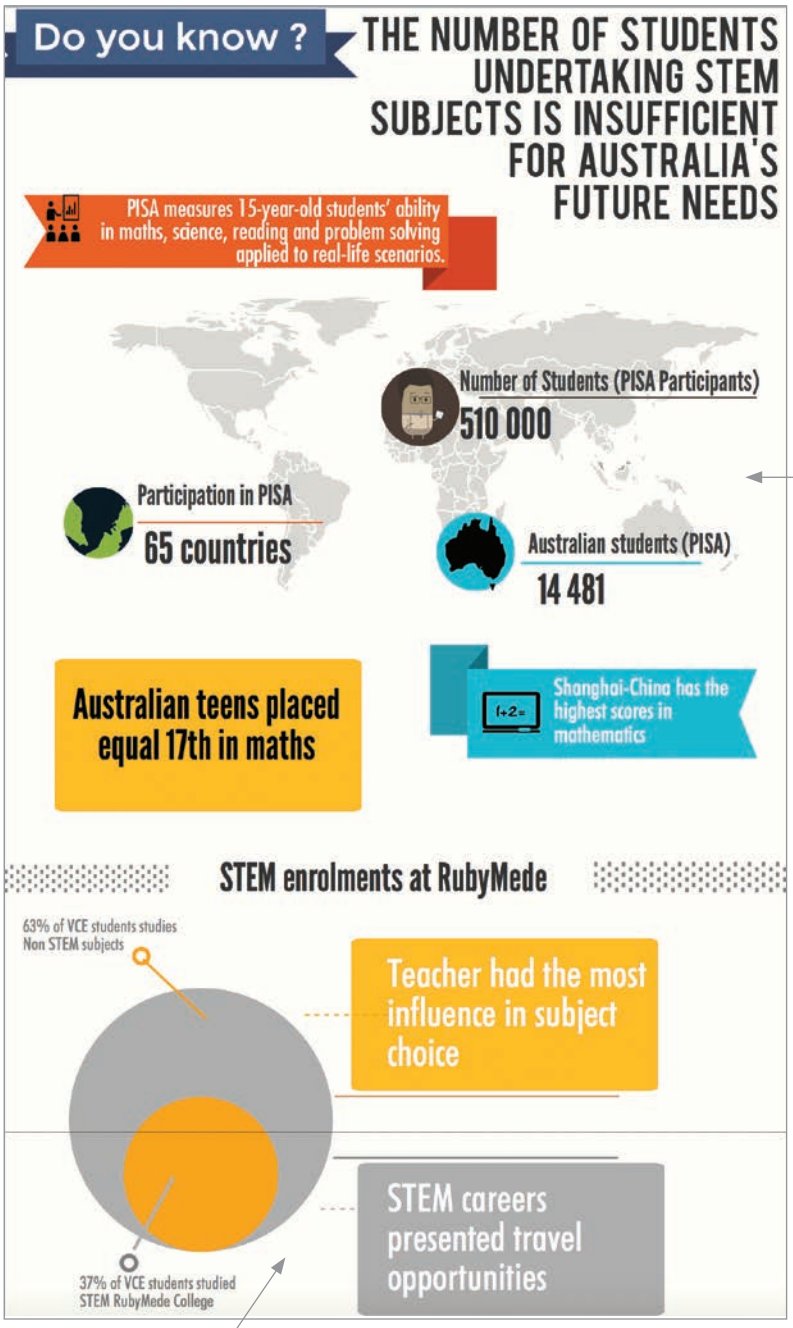


FIGURE 2.36 The creation of a simple Venn diagram using Piktochart



The first section of this infographic looks at the PISA results and provides some background information about the results and Australia's placement in mathematics.

STEM enrolments at RubyMede data derived from the questionnaire and data from the school.

TIMSS and PIRLS International Study Center (2011) TIMSS 2011 International Results in Mathematics. In *IEA's Trends in International Mathematics and Science Study*, Boston College.
OECD (2012), What 15-year-olds know and what they can do with what they know. *In PISA 2012 Results in Focus*.

FIGURE 2.37 A section of the completed STEM infographic used to inform the public that there are too few students undertaking STEM subjects

CHAPTER SUMMARY

ESSENTIAL TERMS

alignment refers to text: can be left, right, centre or full. Full alignment, which is also known as full justification, refers to text being aligned on the left margin of a column and spaced appropriately so that the last letter in a word on the same line is aligned with the right margin.

annotated diagrams a visual depiction of how a graphic representations should look; designs indicate features such as differences in font size, colour and positioning of objects; the placement of objects must be planned so that a balanced, visually appealing and clear effect is attained

clarity how clearly the information in the graphic representation is presented and how well a user can differentiate between objects in the graphic form

contrast the visual difference in colour or tone between elements on a screen; there should be sufficient contrast between background and text or other page elements to make the information plainly readable

conventions standards that have been developed to determine the presentation of documents and other output produced using information systems; rarely static, they change over time as changes in technology and business occur; a convention is first decided on and adopted

effectiveness the measure of how well a solution works and the extent to which it fulfils its purpose; measures of an effective solution include completeness, readability, attractiveness, clarity, accuracy, accessibility, timeliness, communication of message, relevance and reliability

electronic validation techniques, sometimes called machine-validation techniques, that ensure the accuracy of data are built into software such as spell checkers

graphic representation a pictorial diagram that shows the interdependencies between variables; common types derived from numerical data are column charts, line charts, pie graphs, bar charts, area charts, scatter diagrams and bubble charts

graphic solutions pictorial diagrams that show interdependencies between variables.

hierarchy an arrangement of items in which each is ranked above, below or at the same level as another to show their importance

infographic a graphic representation of data and information using different elements

IPO chart (input–process–output) a defining diagram that shows how data is processed into meaningful information; it identifies what data is required for the solution (input), what information the solution needs to produce (output), and the processing steps required to transform the data into information, or the function of the solution

macro an automated series of tasks. In spreadsheet programs such as Microsoft Excel, you can create macros by using the macro recorder. Macros are used to automate series of tasks that are performed frequently. Macros can be run directly from a key combination or from a button added to the user interface

manual validation entails the data entered being checked for accuracy by a person rather than by a machine

orientation the direction and aspect of elements within a graphic representation

problem-solving methodology a structured approach to creating a solution; the method applicable to our course comprises the following steps: analysis, design, develop and evaluate

proportion the visual hierarchy (relative sizes) of objects within a graphic representation

spreadsheet functions include performing mathematical and logical functions, producing various sorts of graphs and charts, inserting labels, headers footers, notes and comments

testing a step in which the solution created is checked for accuracy

useability the ease of use of the graphic solution; all spreadsheet solutions, including graphic representations of numerical data, need to be user-friendly; the user should be able to access the required information with ease and efficiency

validation a step in which the data that has been entered is checked for accuracy

variable in programming, a key word, phrase or symbol that represents a value that may change

white space a section of a graphic representation that is empty of any colour or object that is used to create a clean, uncluttered look and is not considered wasted space by designers

IMPORTANT FACTS

1. Spreadsheet functions enable the user to format data, calculate simple and complex mathematical calculations, and produce different charts and graphs.
2. We inform when we provide information or tell facts, such as reporting on a current issue. We persuade when we influence a decision or action, or change a person's opinion. We educate when we teach and provide knowledge and skills through a learning process.
3. Colour should be used so that it makes the information clear, readable and attractive. The colours should emphasise important features, and a colour scheme should be used to ensure consistency.
4. Clarity is important for onscreen products. All elements on the screen should able to be seen clearly in order to convey the message effectively to the user.
5. Consistency of navigation links, colour schemes and other repeatable features allows users to navigate an onscreen product comfortably with minimal confusion.
6. Colours selected should make an onscreen product readable and attractive.
7. There are a set of formats and conventions to use when producing a graphic solution, including appropriate labelling to ensure that the information communicated is clear and concise.
8. A user should be able to efficiently and easily attain meaning and relevant information from an onscreen graphic representation.
9. When testing a solution, the following areas can be tested for appropriateness: functionality, presentation, usability, accessibility and appropriateness of communication.
10. It is important to test what you are trying to produce meets the specified need and fulfils the specified purpose.

TEST YOUR KNOWLEDGE

Review quiz

PURPOSE OF GRAPHIC SOLUTIONS

1 What is a graphic representation?
2 Find three different types of graphic representations suitable for educating, persuading and informing audiences.
3 Describe the context of each of these three graphic representations and how they could best be used.

DESIGN PRINCIPLES FOR GRAPHIC SOLUTIONS

4 Select one of the infographics from Question 2 and comment on the data that is displayed in terms of the clarity of the message and the evidence provided to support the case.
5 What is the difference between a format and a convention?
6 List five conventions that are used in graphic representations.
7 Describe the types of data you would include in an infographic.
8 How would you test an infographic?
9 List five different software packages that could be used to create an infographic.

DESIGN TOOLS

10 What is a data set?
11 What does the acronym IPO stand for? Explain what each component means.
12 What is usually depicted in an annotated diagram?

TYPES OF TESTS

13 Why is testing undertaken?
14 What should be considered when undertaking completeness testing?
15 What is one result of an unreliable graphic solution?
16 What do you need to ensure by undertaking relevance testing?

VALIDATION

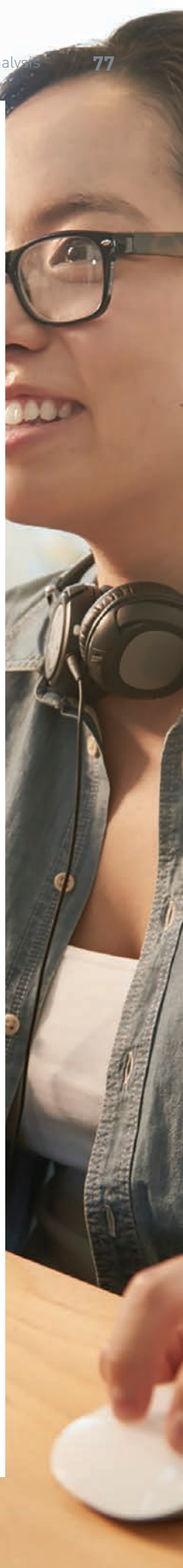
17 What is validation? Identify two ways to perform validation.

APPLY YOUR KNOWLEDGE

Using the 'Infographic: The global impact of wireless internet connectivity' weblink, answer the following questions.

Infographic: The global impact of wireless internet connectivity

1. Focusing on the infographic on the page, rather than the introductory article at the beginning, describe the data displayed in the infographic.
2. List 10 of the subsections shown in the infographic.
3. For each of the subsections listed in Question 2, describe how the data is represented and summarise the findings.
4. What is the 'INTERNET OF THINGS (IOT)'?
5. Provide two additional examples of the IOT, other than the ones listed in the infographic.
6. Using the infographic as a source, explain whether you believe that wireless is more than a luxury, and why, in 500 words.



PREPARING FOR UNIT 1 OUTCOME 1

Acquire, secure and interpret data, and design and develop a graphic solution that communicates the findings of an investigation

In Unit 1, Outcome 1, you are required to conduct an investigation into an issue, practice or event through the systematic collection, interpretation and manipulation of primary data and then create a graphic solution, such as an infographic, that represents your findings.

OUTCOME MILESTONES

1. Identify a topic to investigate.
2. Identify the types of data that will be gathered.
3. Decide which data-gathering methods and techniques to use.
4. Identify, legally and ethically acquire, and reference data and information from primary sources.
5. Identify the techniques to minimise risks to the security and privacy of data and information.
6. Interpret selected data, identifying relationships and patterns.
7. Select and apply appropriate design tools to represent the functionality and appearance of graphic solutions for particular purposes.

STEPS TO FOLLOW

1. Choose a topic to investigate. Topics could be as diverse as social networking, shopping habits or public transport usage.
2. Decide on an aspect of the topic to investigate in detail. For example, the social networking habits of different age groups, shopping preferences of a particular sex and age group, or monitoring public transport usage of a specific group of people.
3. Identify the type of data you will collect.
4. Based on the type of data that will be collected, identify the data-collection method you will use in the research (for example, observation, questionnaires/surveys, interviews or focus groups).
5. Create content for data gathering.
6. Create the participation information statement.
7. Create consent forms.
8. Identify risks to the privacy of participants and security of data and have strategies in place to minimise them.
9. Once permission has been granted from participants, collect the data.
10. Once data has been collected, it needs to be analysed. Identify any patterns, look for meaning in the data and then draw your conclusions.

Getty Images/Rafe Swan

11 From the interpreted data, you will need to create a graphic solution for the purpose of informing, educating or persuading an audience.

12 Use software, and select and apply functions, formats, conventions, data validation and testing techniques to manipulate data and create graphic solutions efficiently. Note: no restrictions are placed on the software tools used to create your solution.

CHAPTER 3

NETWORKS

Key knowledge

After completing this chapter, you will be able to demonstrate knowledge of:

Digital systems

- applications and capabilities of local area networks (LANs) and wide area networks (WANs)
- functions and characteristics of key hardware and software components of networks required for communicating and storing data and information
- purposes of network protocols
- strengths and limitations of wireless communications technology
- types, capabilities and limitations of mobile devices connected to networks
- security threats to data and information communicated and stored within networks
- technical underpinnings of malware that intentionally threaten the security of networks

Implications and impact

- how people, processes, digital systems and data combine to form networked information systems
- legal requirements and ethical responsibilities of network professionals and users of networks with respect to social protocols and the ownership of data and information
- risks and benefits of using networks in a global environment.

For the student

This chapter provides an overview of communications, with an emphasis on the communication of data and information locally and within a global environment. The chapter explains some of the terminology, equipment, procedures and applications that are required to connect and maintain computers so that files, programs and resources can be shared. It also discusses the advantages and disadvantages of using wireless networks, and the role of portable devices. The ways that the security of data and information can be compromised by accidental or deliberate acts are also explored.

In the Outcome at the end of this chapter, you will propose a wireless networked information system that meets a particular need, explain its configuration and predict outcomes for intended users.

For the teacher

This chapter focuses on networks used in a global environment. The capabilities of local area networks (LANs) and wide area networks (WANs) are discussed. An overview of hardware, operating systems and wireless network protocols is provided. A range of wireless transmission media is considered, and several factors that influence the design of a networked information system are identified. The roles and responsibilities of network professionals in terms of their legal requirements and ethical responsibilities are considered. At the completion of this chapter, students will have examined the roles and functions of the components of wireless networks and will be able to recommend a networked information system for a specific use. They will have also considered security threats to data and information communicated via networks. The contents of this chapter will help students to demonstrate key knowledge required to complete Unit 1, Outcome 2.

Students must use a graphics software tool to depict the components of a network.

Networks

A **network** is a collection of computers and devices connected by communications channels that facilitates communications among users and allows users to share resources with one another. Examples of resources are data, information, hardware and software. Networks can be internal to an organisation or cover the whole world by connecting to the internet.

Networks exist for sharing information, such as spreadsheet files, database records, email – indeed anything that helps someone get their job done. The ability to share resources, such as servers, printers and software, also makes a network valuable.

As shown in Figure 3.1, for successful communications, a network needs:

- a **sending device**, such as a notebook computer, which initiates an instruction to transmit data, instructions or information
- a communications device, such as a wireless adaptor inside a notebook computer, to forward packets of data, instructions or information from a sending device via signals carried by a communications channel
- a communications channel or transmission media, such as a cable or radio waves, through which the digital signals travel
- a communications device, such as a wireless router, which receives the signals from the communications channel and forwards the packets to the receiving device
- a **receiving device**, such as a printer, which accepts the data, instructions or information.

Notebook computers, tablets, smartphones and other sending devices usually have a built-in communications device.

The primary function of a communications device, such as a broadband router, is to transmit data, instructions and information between a sending and a receiving device. Data, instructions and information travel along a communications channel in digital form.

A **digital signal** consists of individual electrical pulses that represent the bits grouped together into bytes. Early networks used analog signals that consist of a continuous electrical wave. Computers process data as digital signals, so a modem was used to convert between analog and digital signals.

To check the download speed of your home internet connection, visit the speed test weblink and run the speed test.

Sending devices can usually also serve as receiving devices. Examples include notebook computers, personal computers and mobile phones.

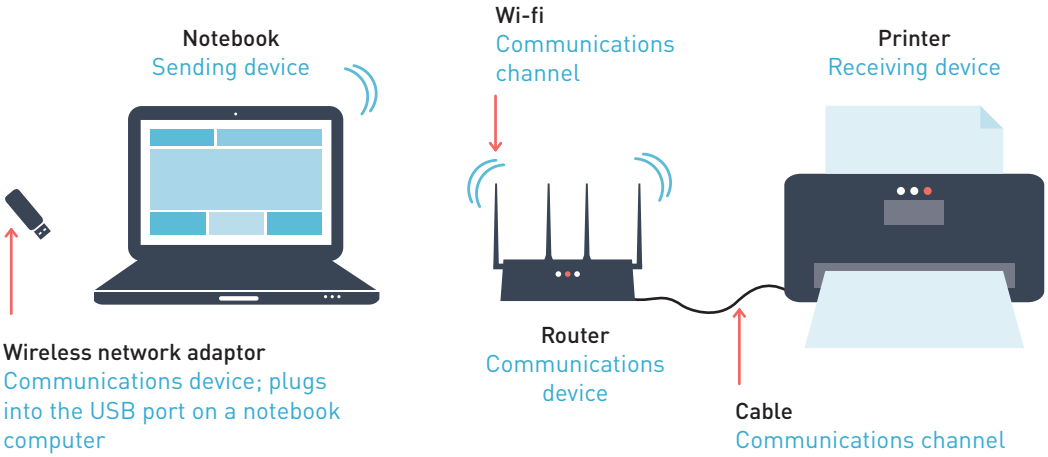


FIGURE 3.1 The notebook sends an instruction to the wireless adapter (communications device), which sends a signal over radio waves (communications channel). The router (communications device) receives the signal and sends an instruction to print via cable (communications channel) to the printer (receiving device).

Types of networks

Local area networks and wide area networks are types of networks that are in common use. The type of network refers to the area over which the network provides connectivity.

Local area networks

A **local area network** (LAN) connects computers and devices in a limited geographical area, such as a home, school, office building (Figure 3.2) or closely positioned group of buildings. Each computer or device on the network is a **node**. In many networks, the nodes are connected to the LAN via cables. Many new networks use wireless transmission media. A wireless LAN (WLAN) uses no physical wires; instead it uses wireless media, such as radio waves. Computers and devices that access a WLAN must have built-in wireless capability. Usually a WLAN communicates with a wired LAN for access to its resources, such as software, hardware and the internet.

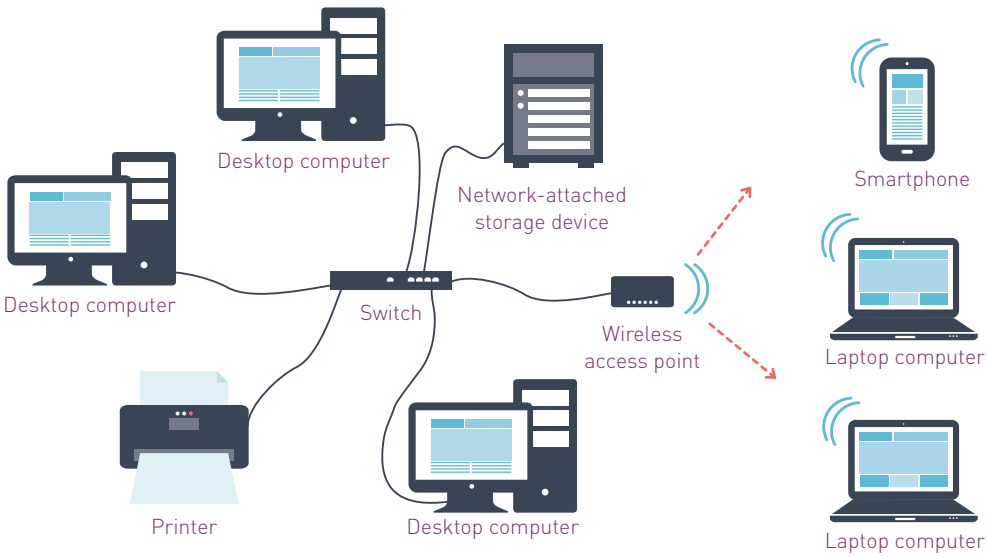


FIGURE 3.2 A local area network operating within a confined geographical area.

Network World
Reviews of network operating systems and general information on networks

The logical design of the components of the network, including the number and types of servers, workstations and network resources, is known as the **network architecture**. It includes the communications devices and the types of physical and wireless transmission media used to connect components.

Intranets

Organisations typically use an **intranet** to publish their event calendars, policies, procedure manuals and technical support files, and to allow access to documents required for group work. An intranet will often include a connection to the internet, allowing employees access to information from the Web. Intranet pages will often include links to internet sites, with information relevant to the organisation.

An intranet uses a **web server**, supports multimedia webpages coded in **HTML** and is accessible via a **web browser**, such as Google Chrome, Mozilla Firefox, Microsoft Edge and Internet Explorer.

An intranet provides the following efficiencies.

- It facilitates communication by allowing employees to work in groups.
- Users can access information faster since data does not need to pass through a router, and loading graphics and images becomes much quicker than on the internet.
- An intranet reduces paper waste because companies are able to move their documents and processes onto the intranet. This can greatly reduce the need for centralised printing

and distribution. If a hard-copy version of a document is needed, on-demand printing at the end-user level is all that is required.

- An intranet improves ease of use. Point-and-click technology allows easy access to hyperlinked company documents.

The effectiveness benefits of an intranet include the following.

- An intranet allows restricted access to company information.
- An intranet provides dynamic information. Company documents stored on an intranet can be updated more easily and quickly than hard-copy versions. The documents can be kept up to date, providing more accurate and timely information to decision makers.
- An intranet allows connection across different network platforms. Users of complex networks that employ a number of different operating systems – such as Microsoft Windows, Macintosh and UNIX – are able to communicate easily within an intranet using their browser software. The HTML code used by the webpages is universal across all platforms.
- An intranet makes data more accessible. Information stored on an intranet can be accessed from anywhere in the world via the internet, if the user is authorised to connect to the internal network. An employee travelling overseas can use the internet to access files that otherwise would have been locked away in a filing cabinet back in the office.
- An intranet supports the ability to communicate using audio and video files rather than just traditional text and graphic images.

Home networks

If you have multiple computers in your home or home office, you can connect all of them together with a home network (see Figure 3.3). The advantages of a home network include the following.

- Desktop computers, notebooks, tablets and smartphones can all communicate with each other.
- All the computers can be connected to the internet at the same time.
- All computers can share peripherals, such as a scanner, printer or a network-attached storage device.

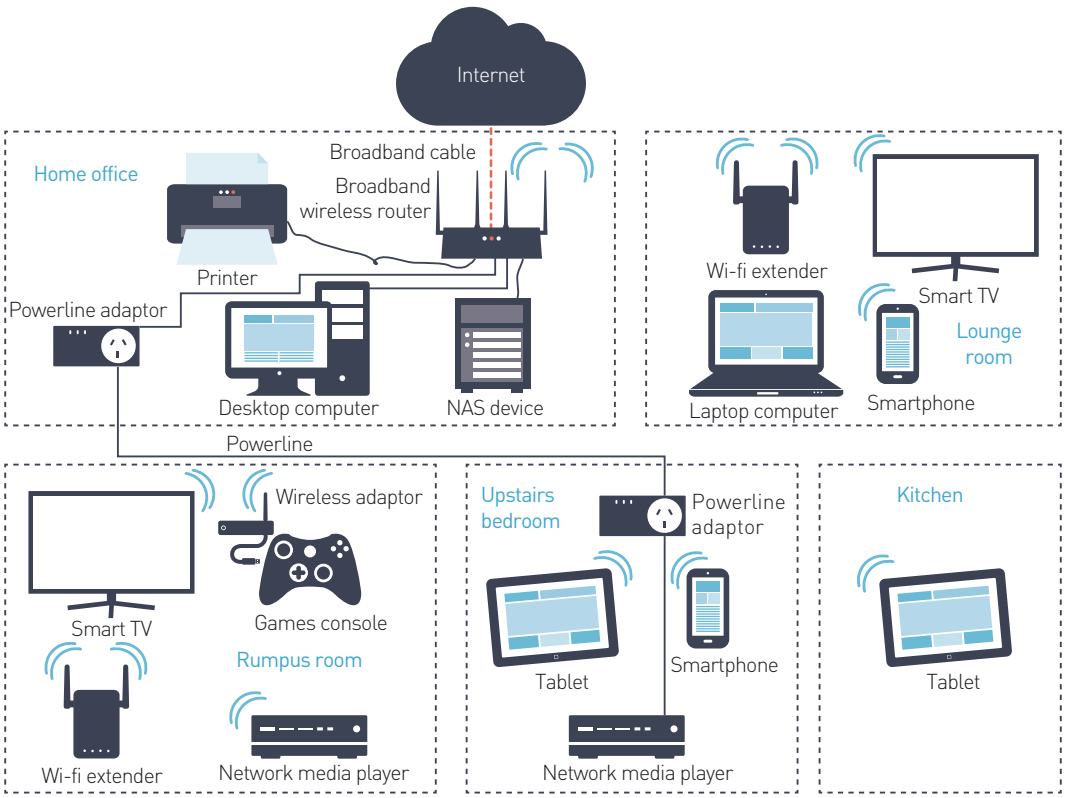


FIGURE 3.3 A home network.

- Each networked computer is able to play multiplayer games with players on other computers in the house.
- Smart TVs can connect to the internet.

Wide area networks

A **wide area network (WAN)** covers a large geographical area – for example, a city, a country or the world – using a communications channel that combines many types of media, such as telephone lines, cables and radio waves. A WAN can be one large network, or it can consist of two or more LANs connected together. The internet is the world's largest WAN. Mobile phones could be considered to be operating in their own WAN.

Network architecture

Each type of network can be further categorised by its architecture. Network architecture refers to the layout or design of the network. In this section we will consider client–server, peer-to-peer and virtual private network (VPN) layouts.

Client–server networks

THINK ABOUT COMPUTING 3.1
Suggest reasons why a peer-to-peer network would not be used in organisations with more than 10 computers.

In a **client–server network**, a server, sometimes called the host computer, controls access to the hardware and software on the network and provides a centralised storage area for programs, data and information. Besides storage capacity, a server allows for file sharing, website hosting, email management and access to shared printers. The other computers on the network, called clients, rely on the server for these resources (Figure 3.4). For example, a server in a school's administration might store a database of student details. Every client on the network can access this database on the server.

The costs associated with a server-based network are significantly higher than those for a peer-to-peer network. Not only is the start-up equipment more expensive, but a client–server network also requires ongoing technical support to maintain the sophisticated hardware and software. On the positive side, however, there are clear economies of scale, as the cost of adding clients that share the server's resources becomes relatively less.

The major difference between the server computer and the client computers is that the server has more storage space and power. Some servers, called **dedicated servers**, perform

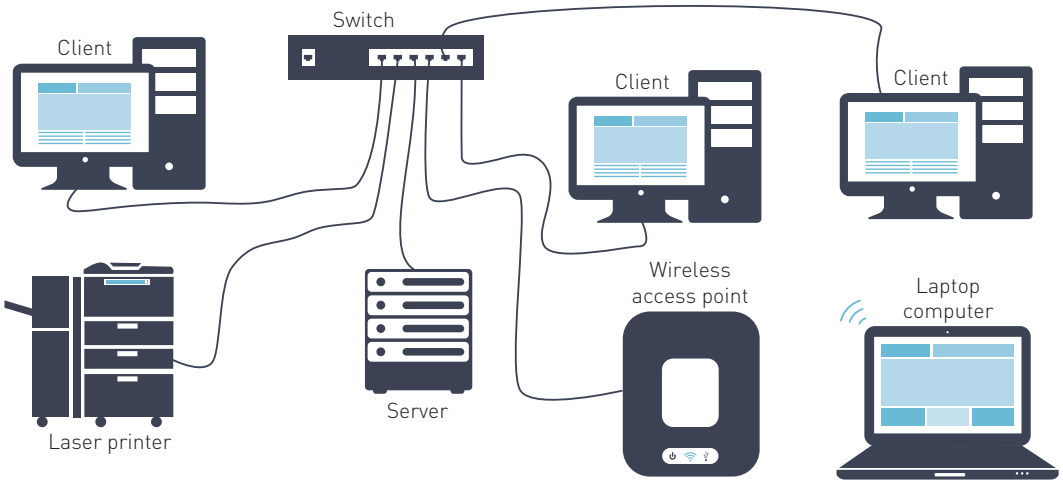


FIGURE 3.4 On a client–server LAN, one or more computers acts as a server and the other computers on the network are called clients.

a specific task. For instance, a file server stores and manages files. Each user on the LAN can share files or programs stored on the file server. A print server manages printers and print jobs. Print jobs received from users on the LAN are queued on the print server in order of their arrival and fed to the various network printers one document at a time. A database server stores and provides access to a database. A network server manages network traffic. A number of servers are often configured in a type of rack, which makes it easier to manage the cables and power supplies. The server rack is often located in a specialised room where the temperature is kept cool to offset the amount of heat generated by the power-hungry servers.

A client–server network provides better security than other configurations, because user access can be managed and logged.

In the past, network administrators used to dedicate each server to a particular task. One application per server made it easier to track down any problems that arose. This approach, however, does not take advantage of the processing power of modern servers. Also, a larger storage room is required as the number of servers is increased. Many servers in networks are now virtual rather than physical. Specially designed software is used to convert one physical server into multiple virtual machines.

Although it can connect a smaller number of computers, it is typically most efficient for a client–server LAN to connect 10 or more computers. Most client–server LANs have a network administrator because of their larger size. The **network administrator** is the operations person in charge of the network.

Peer-to-peer networks

A **peer-to-peer network (P2P)** is a simple, inexpensive network that typically connects fewer than 10 computers. Each computer on a peer-to-peer network can share the hardware (such as a printer), data or information located on any other computer in the network (Figure 3.5).

Each node in a peer-to-peer network shares hardware and data with all other computers on the network.

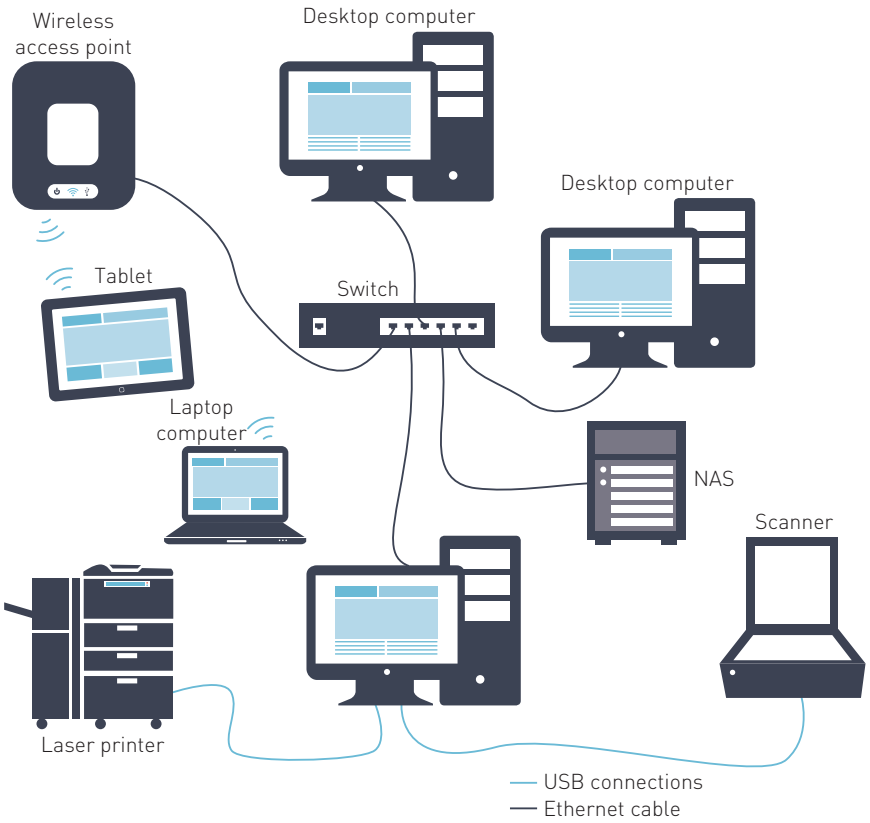


FIGURE 3.5 Each computer on a peer-to-peer network can access data from other users and share resources such as printers. A peer-to-peer network differs from a client–server network in that files can be directly transmitted between nodes rather than from a server.

Each computer can store files on its own storage device or on another computer. Each computer in the network contains both a client operating system (like the 'Home' versions of Windows) with basic networking capability, and application software. All computers on the network share any peripheral device attached to any computer. For example, one computer may have a laser printer and a scanner, whereas another may have an ink-jet printer. They may also share a network-attached storage device (NAS) with a movie and music repository. Peer-to-peer networks are popular in homes where a login server is neither necessary nor practical.

Network-attached storage devices (NAS) are discussed later in this chapter.

Internet peer-to-peer

Another use of peer-to-peer (called **P2P**) involves an **internet peer-to-peer network**, which enables users with the same **networking software** to connect to one another's hard disk drives and exchange files directly (Figure 3.6).

Early P2P programs, such as Kazaa and Limewire, stirred up much controversy about the copyright infringement of music because they allowed users to copy MP3 music files easily from one computer to another. To help reduce copyright infringement, music-sharing services, like iMesh, typically are fee-based, and music files are often encrypted as they travel across the internet.

Many businesses see an advantage to using P2P – that is, companies and employees can exchange files using P2P, freeing the company from maintaining a network for this purpose.

Allowing internet peer-to-peer file sharing exposes your computer to possible security violations. For files to be shared within a P2P network, a specific TCP port must be opened through the firewall on the computer. Once the port is opened, the computer is no longer protected from malicious software, such as viruses and Trojans, capable of causing computers to malfunction or seize. Recognised sites, such as BitTorrent, often have built-in security measures, but their dependability is open to question, and using them can be risky at best.

THINK ABOUT COMPUTING 3.2
Kazaa was a popular Internet P2P application for finding, downloading, playing and sharing files with millions of other users. The Kazaa client could be downloaded free of charge, but it came bundled with adware. After numerous legal proceedings against Kazaa by the recording industry, resulting in damages payments in excess of $100 million, it now operates as a music subscription service. Users pay around $25 per month for unlimited downloads of songs.

The P2P acronym has been adapted by some to mean people-to-people. The development of social media software that allows individuals on the internet to meet each other and share ideas is an example of people-to-people technology.

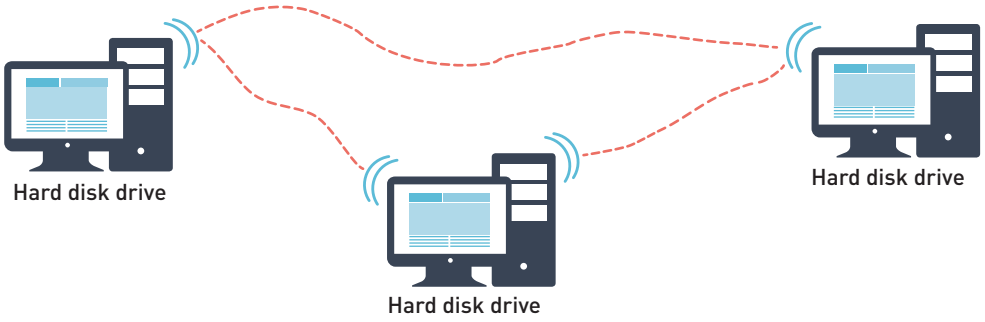


FIGURE 3.6 P2P describes an internet peer-to-peer network that enables users with the same networking software to connect to one another's hard disks and exchange files directly.

Virtual private networks (VPN)

Businesses with offices and branches across Australia or around the world need a fast, secure and reliable way to share data and information across networks. A **virtual private network (VPN)** allows businesses to use a public WAN, (the internet) to create a private network that links remote sites and users to the business's head office. The VPN uses encryption (we discuss encryption later in this chapter) to ensure files and messages are secure.

A VPN provides secure connections over the internet that allow businesses to extend their private network, which in turn allows communication and information sharing with remote branches and travelling salespeople.

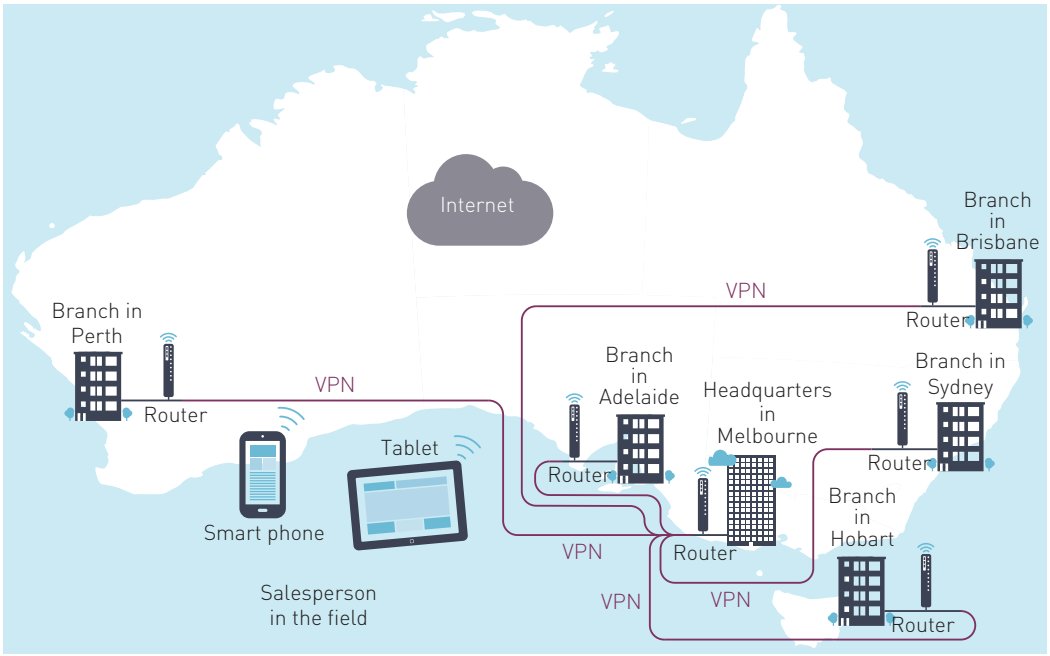


FIGURE 3.7 An example of a wide area network in Australia. The business is using a virtual private network (VPN) to communicate over the internet with branches around the country.

Prior to businesses using VPNs over the internet, the common way to connect computers between remote offices was to use a leased telephone line. Leased lines provided organisations with a means to expand their private network beyond its immediate geographic area. These connections formed a wide area network for the organisation. The telephone lines were leased from telecommunications companies such as Telstra. While the leased lines proved to be reliable and secure, they were expensive to operate, particularly the further the WAN had to reach.

P2P DOWNLOADS: TORRENTS

Internet P2P networks provide a popular means of downloading music, TV, video and software files. Downloading a file from a single site like iTunes typically involves a stream of sequential fragments of the file being sent to the requesting computer. Since the fragments are sequential, the user can play the file as it is being downloaded. If multiple people want to download the same file then that single source can become overworked and possibly crash. The more users on the network, the slower downloads become.

An alternative to single source file downloads is to use a torrent **protocol**. With torrents, there are multiple users who download a particular file at the same time, or who have previously downloaded the whole file. These users act as a source for each other. Together the downloaders become a network of multiple sources all providing pieces of the file to each other.

Here is how it works (Figure 3.8).

A user, for example Jack, installs a torrent client application on his computer. Jack searches the web to find a torrent provider. A **torrent** is a small file that holds information

Data sent or received over the internet is sent to the IP address of the remote computer and a specific Transmission Control Protocol (TCP) port on that computer. We discuss TCP/IP further on page 102.
A firewall is hardware and/or software that restricts access to data and information on a network. We discuss firewalls on page 120.

about, say, a video file. It includes metadata (for example, the name and size) about the file to be shared and identifies the tracker. The tracker is a specific computer that coordinates the file distribution by making links to **peers** who have pieces of the required file. On the torrent provider's website, Jack locates the file he is after and downloads the torrent to his computer. The torrent is read by the torrent application which uses the specified tracker to locate other computers that are downloading the requested video file (peers) or already have the whole file (**seeds**). The tracker makes links between Jack's computer and the identified peers and seeds. With the connections in place, Jack's computer starts to receive the pieces of file he has requested. Since the file is being sent in non-sequential pieces from multiple sources, the video cannot be watched until all pieces are received and the video assembled.

Jack's computer also can start to send pieces of the file to other peers who are downloading the same video even while his computer is still receiving pieces. The torrent application therefore turns downloaders into sources. The more downloaders means the more sources and hence the faster the downloads. Using torrents to download files means that home computers, rather than servers, can act as sources, and a lower bandwidth connection to the internet is not the problem it would be with single source downloading.

If Jack does not allow his computer to pass pieces of the file to peers (that is, he only downloads) he is referred to as a **leecher**.

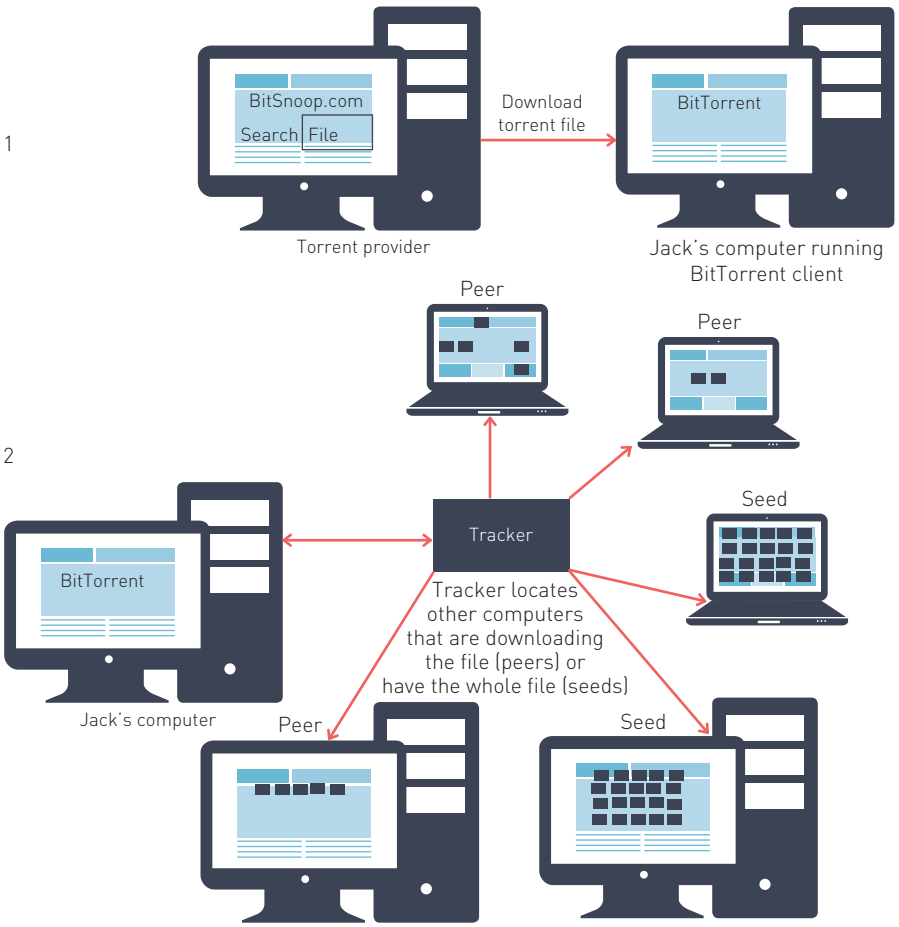
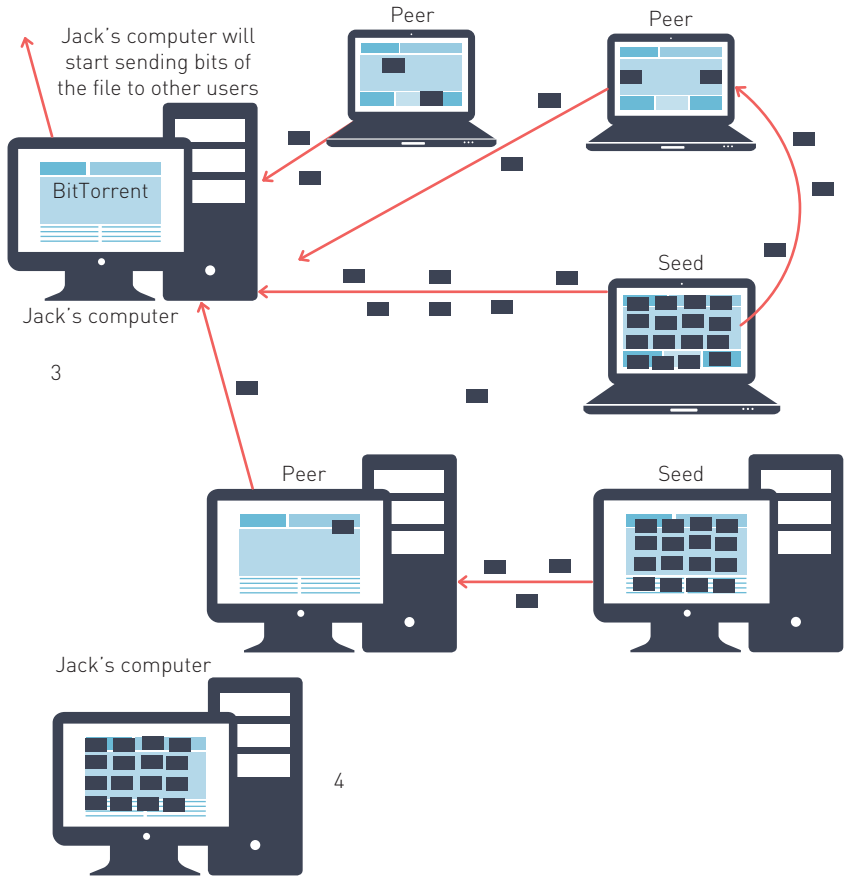


FIGURE 3.8

1 Jack's computer downloads a torrent from a provider and loads it into BitTorrent on his computer.

2 BitTorrent uses the torrent to locate a tracker that makes links with peers and seeds who have part or all of the requested file.



3 The peers and seeds transfer pieces of the requested file to Jack's computer. Meanwhile, Jack can also act as a peer by allowing some of the pieces he has downloaded to be sent to someone else requesting the same file. The more downloaders of the file, the more sources and the faster the downloads become.

4 The requested file has been downloaded to Jack's computer.

ISSUE Is it legal to download music and video files?

The *Copyright Act 1968* allows users to download music and video files from the internet via peer-to-peer transfers only if they have the permission of the copyright holder. Downloading a TV show, a film or a music file is in breach of the law unless the owner has approved the transfer, usually for the payment of a fee. If you pay to download a file, or even If you are not required to pay a fee, you often have to agree to certain conditions before you can download it. Any such conditions override the default provisions in the *Copyright Act* that allow private copying of the file. Sites like The Pirate Bay allow torrents to be downloaded, which can then be used with an application like BitTorrent or uTorrent to collect pieces of the original file from a number of peer sites and put them together. If the user plays the assembled file without permission they have broken the law.

THINK ABOUT COMPUTING 3.3

1 For each file-sharing Internet site that has been forced to shut down because of copyright violations, many more have been launched in their place. How can the rights of artists, film makers and TV producers be properly protected?

2 Should governments be doing more? Or is it the responsibility of society to do the right thing?

3 Look for recent examples of attempts to regulate or control torrent sites. How successful have these measures been?

The Pirate Bay website operates from Sweden, where the copyright laws are possibly less strongly enforced than other countries like the United States. Nevertheless, in 2009 the founders of the site were found guilty of making copyright content available to others. The site was shut down for a short period in 2010, but has since reappeared. In 2012 the High Court of the United Kingdom ruled that British-based internet service providers must block The Pirate Bay. The British Phonographic Industry stated that sites like The Pirate Bay destroy jobs in the UK and undermine investment in new artists.

Critics of site blocking argue that such measures are ineffective as they can be circumvented using proxy servers and other techniques.

Pirate Bay must be blocked, High Court tells ISPs, by Matt Warman, Consumer Technology Editor, The Telegraph, 30 April 2012

Communications devices

Communications devices enable computer users to communicate and exchange items such as data, instructions and information with another computer. For example, a broadband router is a communications device that enables computers to communicate via telephone or cable. At the receiving end, the communications device receives the signals from the communications channel.

Today, thousands of networked digital systems exist, ranging from small wireless networks operated by home users to global networks operated by numerous telecommunications firms. To interconnect these networks, various communications devices exist. Some of the more common types of communications devices are switches, basic routers, broadband routers, network interface cards and wireless access points.

Switches

A **switch** is a device that provides a connecting point for cables in a LAN.

Network traffic in a LAN typically follows specific paths that connect members of a work group, such as the accounts department of a business. A switch stores the address of every device down each cable connected to it. When a packet enters the switch, it uses simple logic to detect the destination of the packet and sends the packet only down the cable that contains the addressed device. The result is that the packet reaches its destination quickly and without colliding with packets being sent to other nodes.

As packets are sent straight to the destination device through a switch, the two devices act as if they are directly connected. On a 100 **Mbps** switch, data can be sent simultaneously to all nodes at 100 Mbps uninterrupted.

Encryption of data packets enhances the security of wireless transmissions. Setting up complex passwords for file sharing, routers and access points, as well as using encryption keys that are not obvious, greatly improves security.

Routers

A **router** is a device used to connect multiple networks – several LANs or a LAN and a WAN, such as the internet. A router will connect LANs and WANs irrespective of the hardware and network communications protocol used in each segment. On the internet or a large corporate network, for example, routers receive TCP/IP packets, look inside each packet to identify the source and target IP addresses, and then forward these packets as needed to ensure the data reaches its final destination. Figure 3.9 looks at how a router can be used to communicate between two LANs operating within one organisation.

While a switch uses a node's MAC (Media Access Control) address to identify which path to send the packet, a router uses the IP, IPX or Appletalk address. The algorithm that switches use

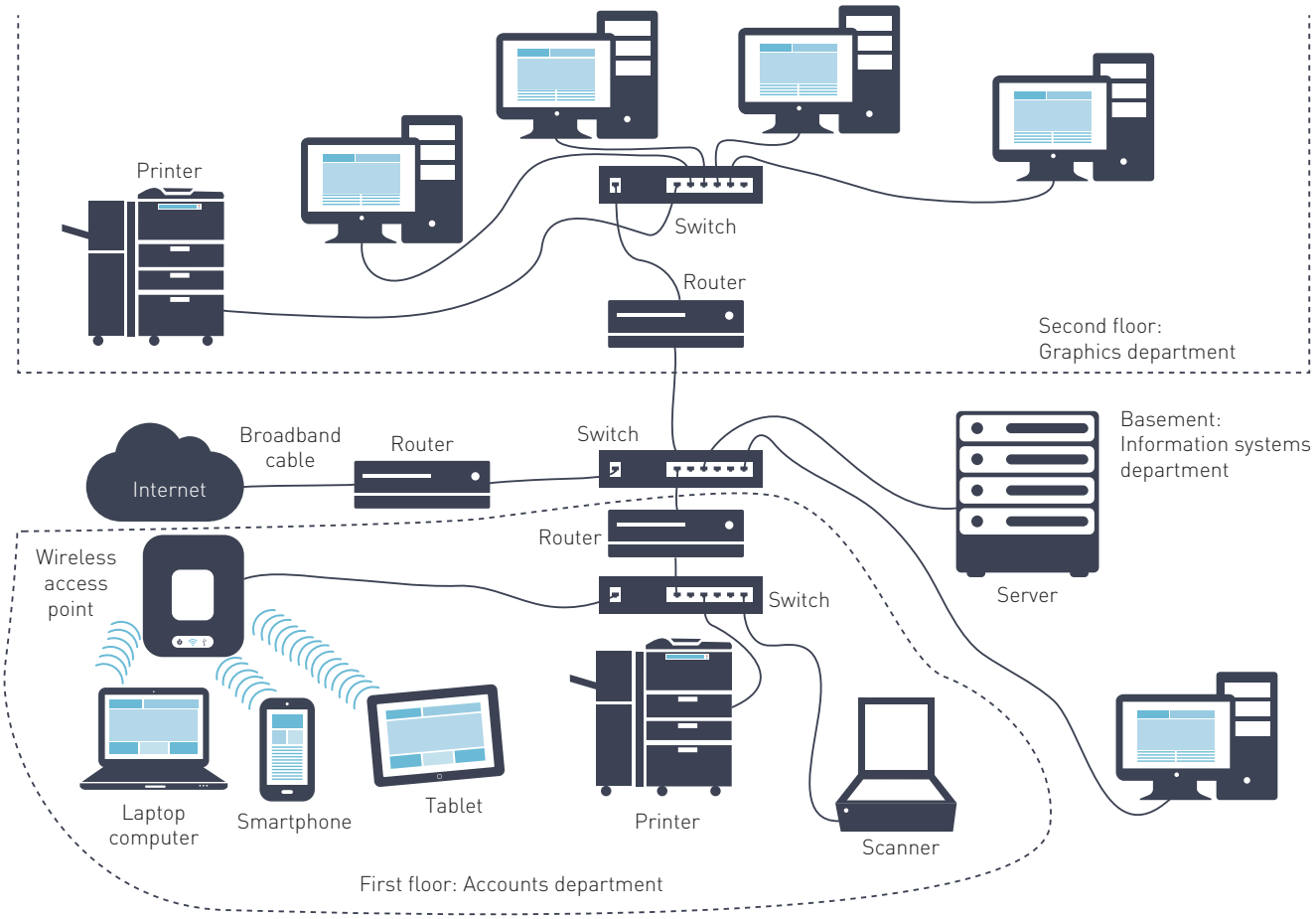


FIGURE 3.9 How a router can connect two LANs and provide internet access to both networks.

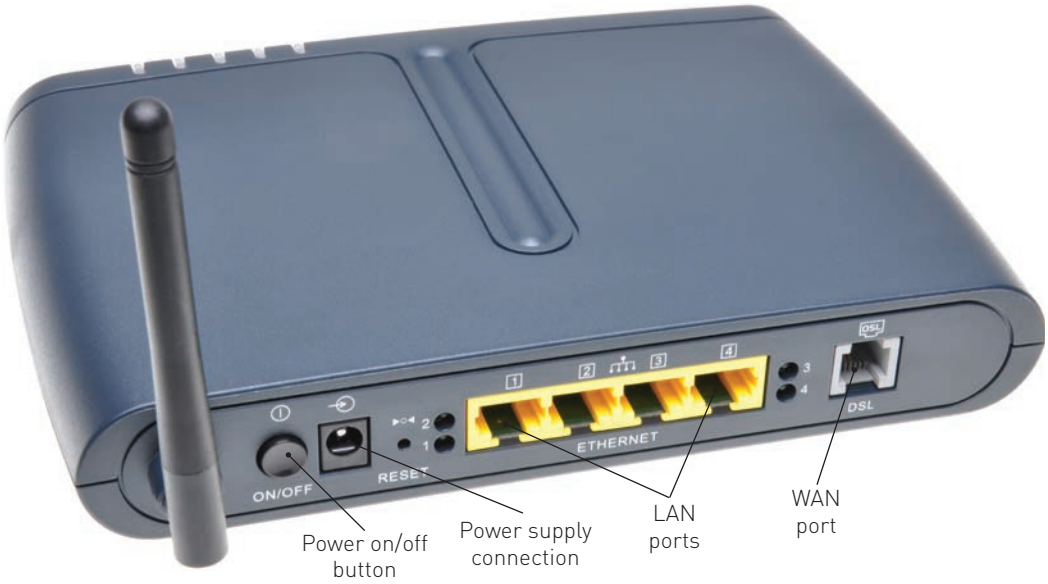
to decide how to forward packets is different from the one used by routers. Most carriers, such as Vodafone, offer a 4G USB modem stick. The small size makes it ideal to use with a mobile device, such as a laptop computer, to connect to the internet using the 4G phone network.

In Figure 3.9, a file sent from one member of a graphics department to another member of the same group stays on the graphics LAN and does not affect network traffic on the accounts LAN. The router detects if a file needs to move from one LAN to another and allows its passage. For example, the graphics department may need to send large animation files between members. The size of these files means that allowing all members of the two work groups to have unlimited communication on the one cable through a single LAN can consume the available resources of the cable. The network will become sluggish and frustrating for users. To overcome this, each work group is given its own LAN and a switch with a router used to link the work group LANs. In this way, large files that are needed only by the graphics department stay on the one LAN, while files that need to be shared between departments can move from one LAN to the other.

Broadband routers

Routers for home networks (often called **broadband routers**) also can join multiple networks. These routers are designed specifically to join the home (LAN) to the internet (WAN) to share internet connections.

Wireless broadband routers combine the functions of a basic router (connecting the LAN to the internet), a switch (for devices, such as a desktop computer, connected by cable), a firewall (security measure) and a wireless access point (to allow wireless connectivity). Figure 3.10 shows a broadband router with an antenna for wireless connections, LAN ports for wired connections and a WAN port for connection to the internet.



Shutterstock.com/Flegere

FIGURE 3.10 Wireless broadband router

The type of broadband router used to act as a bridge between a LAN and the internet depends on the type of connection provided by the Internet Service Provider (ISP). A cable connection uses the high bandwidth available through the same broadband connection that delivers information to a television via a provider such as Foxtel. An **asymmetric digitial subscriber line (ADSL)** connection provides internet access using copper wiring in telephone lines. The cost of an ADSL broadband router is around $100. A cable broadband router is usually sourced from an ISP, since it has to be registered on the ISP's network before it will work.

NBN devices

The National Broadband Network (NBN) is a national network of communication infrastructure that uses lightning-fast fibre-optic, fixed wireless and satellite technology. The original NBN plans were to reach 93 per cent of Australian homes with download speeds up to 100 megabits per second. Revised plans by the federal government, whereby connections are made to a local node rather than directly to a home or business, will result in a speed of 25 megabits per second. Current ADSL 2+ connections are 24 megabits per second depending on line quality and capability of components in local telephone exchanges.

Homes connected to the National Broadband Network (NBN) will require a router that is capable of supporting the faster download speeds that are available.

The NBN utility box (also known as the Premises Connection Device, or PCD) is installed to the outside of a house during the street roll-out of NBN cable. Fibre-optic cable from the roll-out is connected to the utility box (see Figure 3.11). The home-owner's ISP then arranges for the NBN connection box (also known as a Network Termination Device, or NTD) to be installed in the home (see Figure 3.12). A fibre-optic cable is used to connect the PCD to the NTD.

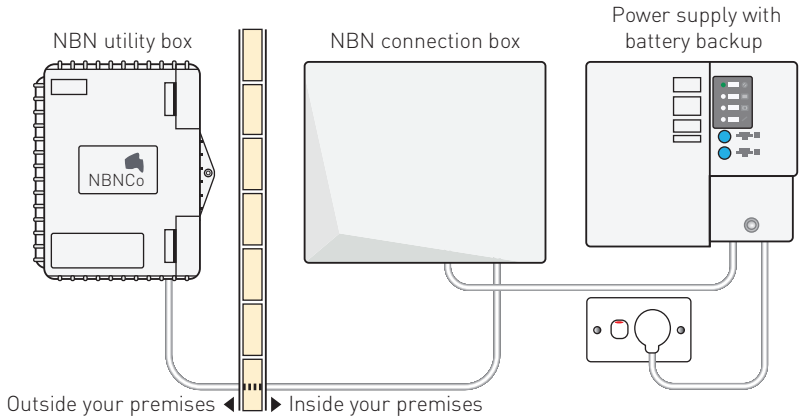


FIGURE 3.11 The NBN utility box

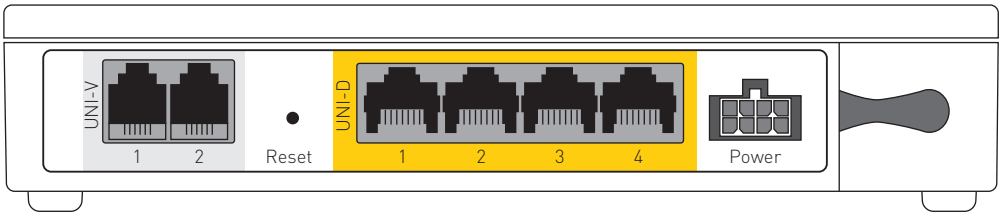


FIGURE 3.12 NBN connection box

On the left-hand side of the NBN connection box are two voice ports. When copper-based telephone connections are phased out, one of these ports will be used for phone connection over the fibre-optic NBN network. The other voice port can be used for other services such as a personal alarm service (for the elderly or disabled). One of the four data ports is used to join the connection box to the home network's router. The other ports allow a home-owner to connect to other service providers.

Network interface cards and wireless adaptors

A **network interface card (NIC)** is a card inserted into an expansion slot of a personal computer, or it may be built in to the computer's motherboard. A **wireless adaptor** performs the function of a NIC for notebook and other mobile computers (Figure 3.13).

Wireless adaptors and NICs work with a particular network technology, such as Ethernet. An Ethernet card is the most common type of NIC for hard-wired networks. Depending on the type of wiring used, the transfer rate on an Ethernet network can be up to 1000 Mbps.

Shutterstock.com/Jouke van Keulen

iStockphoto/connect11

FIGURE 3.13 a A network interface card slots into the motherboard on a desktop computer. **b** A wireless adaptor can plug into the USB port on a desktop computer or notebook that does not have wireless capability.

A NIC performs three functions: it increases the strength of the signal from the computer, packages the data for transmission and controls access to and from the network cable or the wireless channel. An external USB wireless adaptor is available for a desktop computer or portable device that does not have built-in wireless connectivity.

Mobile devices, such as tablets and mobile phones, can connect to a LAN using a built-in **wi-fi** adaptor.

Wireless access point

A **wireless access point** is a communications device used on wireless local area networks. It acts as a central transmitter and receiver of wireless radio signals. A wireless access point is often connected to a wired network backbone for faster transmission of data back to the network server.

Wireless access points are mostly used in business networks where larger buildings and spaces need wireless coverage. Home networks are small enough that a wireless router can provide sufficient coverage without the need for an access point.

If an area is too large to be covered by a single wireless access point, multiple access points can be used. There can be a momentary loss of connectivity when a user moves from the vicinity of one access point to another. Figure 3.14 shows a wireless network with overlapping access points. Overlapping access points provide a seamless area for users to move around in, using a feature known as 'roaming'. When a user moves from one area to another, the wireless networking hardware automatically jumps to the access point that gives the strongest signal.

To guard against hackers stealing valuable files from your wireless network, consider using a protocol other than TCP/IP, such as Ethernet, for file sharing. Access points are usually installed on a LAN behind the firewall. If a hacker successfully connects to your access point, they will have open access to your LAN. Since the hacker will be using TCP/IP to connect to your LAN, you can deny them file-sharing access by using a protocol other than TCP/IP for that service. The hacker may still be able to use your Internet connection, but they will not be able to access your files.

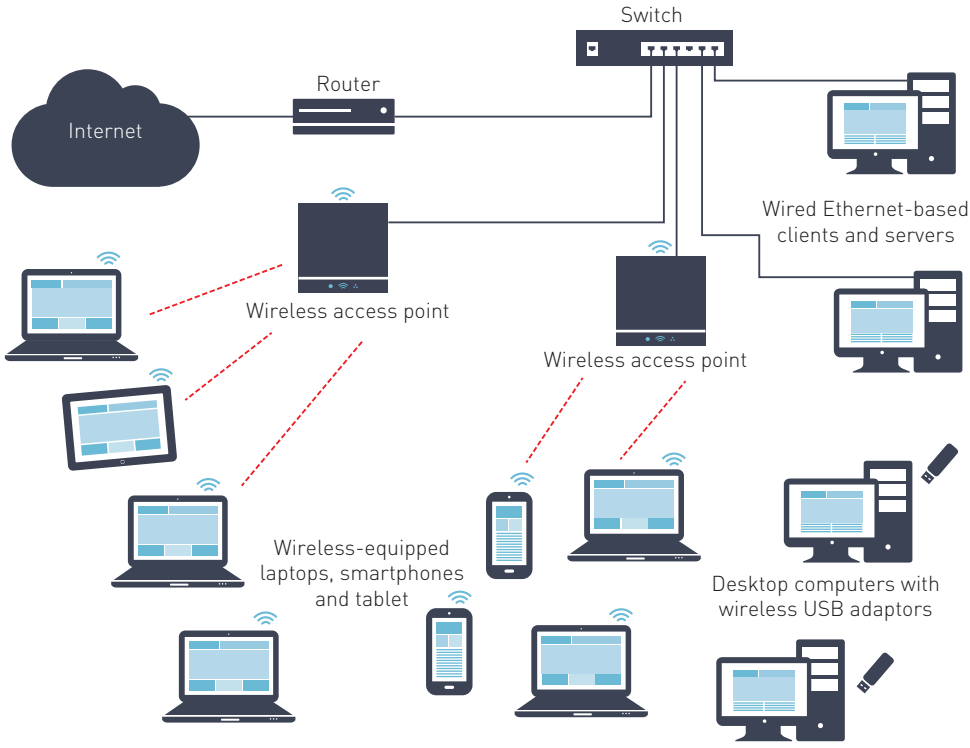


FIGURE 3.14 A business network using multiple wireless access points.

Wireless extender

Wireless extenders (wireless repeaters) increase the area covered by a wireless network (Figure 3.15). They allow users to keep mobile devices, media players and computers connected to a wi-fi network with expanded coverage throughout the home or office. Wireless extenders pick up a wireless signal in the same way as a tablet or notebook computer, then re-broadcast that signal, effectively giving the network a second access point to connect to. This allows users to overcome obstacles that normally would block radio signals and to enhance the signal quality. Extenders typically increase the range of a wireless network by 20 metres (depending on the physical layout and construction materials of the home or office), allowing users to roam anywhere in the home or office and remain connected to the internet.

The location of the extender is critical – too close to the wireless source can cause interference while too far from the source there may not be enough signal to enhance.

FIGURE 3.15 A wireless extender, such as the 5GHz NETGEAR WN2500RP, increases the coverage of a wireless network to all parts of the home. The extender also has four ports for hard-wired devices such as a smart TV, Blu-ray player or game console.

THINK ABOUT COMPUTING 3.4

A concern with a single frequency wireless extender is that they can cause a throughput loss of about 50% of the radio signal, since the extender must receive then re-transmit each packet using the same radio signal on the same channel as the router. An alternative to using a wireless extender is to increase the network's range by using Ethernet-over-power (EOP). Investigate the advantages and disadvantages of EOP over wireless extenders.

Summary of wireless network technologies

Components used in wireless networks have strengths and limitations in terms of data transfer rates, cost, security and reliability. Refer to Table 3.1.

Networks in homes and organisations often use hard-wired technologies to support the wireless connectivity of end users. For example, Ethernet cables may be required to connect a wireless access point to a router, or fibre-optic cable may be used to connect a broadband router to the internet. Storage options in a wireless network can include a hard-wired or wireless connection to a network-attached storage device (NAS), a file server or cloud storage.

Wireless network security options, including WPA2 and WPS, are discussed later in this chapter.

TABLE 3.1 A summary of the strengths and limitations of wireless communications technology; the transfer rates and price data were current in 2015.

Device	Cost estimate	Data transfer rate	Security	Reliability
Wireless broadband router	$150	802.11ac standard, 867–1300 Mbps, dual band (2.4 GHz and 5 GHz frequencies).	High provided • the router identifier is kept secret • encryption is used for broadcasts (WPA2) • Wi-fi Protection Setup (WPS) is used to attach devices to the wireless network	• Throughput on the 5 GHz band can drop quickly as the range increases. • In the early years of 5 GHz connectivity, there is little congestion. • The 2.4 GHz band can be affected by interference from household devices such as automatic garage doors, or from other nearby networks operating on the same band.
Access point	$120			
Wi-fi extender	$140			
Wireless adaptor	$80			

Communications software

Communications software is an application or program designed to pass or support the movement of information over a network. Some communications devices are pre-programmed to accomplish communications tasks; others require a separate communications software program to ensure proper data transmission. Communications software consists of programs that help you establish a connection to another computer or network and manage the transmission of data, instructions and information. For two computers to communicate, they must have compatible communications software.

THINK ABOUT COMPUTING 3.5
Check out the software installed on the network in your school. Find out from the network manager how many users are covered by a site licence for one of the software packages. Compare the cost of the site licence with that of the equivalent number of single-user versions (check the price of a single-user version, such as in the newspaper or online). What is the cost saving for the school by purchasing the network version over the single-user version? What disadvantage is there in purchasing site licence software compared with single-user packages?

Network operating system

A server operating on a client–server network requires operating system and application software that differs from those of a desktop or portable computer. A server is able to share data with multiple users in a secure environment and reduce bottlenecks. A server often is assigned a particular role in a network, with associated software to facilitate that task. A server can be set up to provide email services, internet connectivity, file backup functionality and to manage print requirements.

A server uses specialised software to support its function. For example, an email server may use Windows Mail Server Software. A print server is responsible for sending jobs from multiple clients to a printer in the right order and at the right time. Appropriate software is needed to manage these tasks.

A **network operating system** (also called a **network OS** or **NOS**) is the system software that organises, controls and coordinates the activities on a local area network. A NOS controls the attached computer systems, any peripherals and the communication between them. Some of the tasks performed by a NOS include:

- administration – adding, deleting and organising users and performing maintenance tasks, such as backup
- file management – locating and transferring files

- device management – coordinating print jobs and reports sent to specific printers on the network, ensuring resources are used correctly and efficiently
- security – monitoring and, when necessary, restricting access to network resources.

Network analysis tools

Network administrators use a variety of software packages designed to monitor devices on a network, check the use of different protocols, identify which ports have been accessed on a web server, view event logs and analyse network traffic (Figure 3.16). A **network analysis tool**, or network utility, is software designed to analyse and configure various aspects of computer networks.

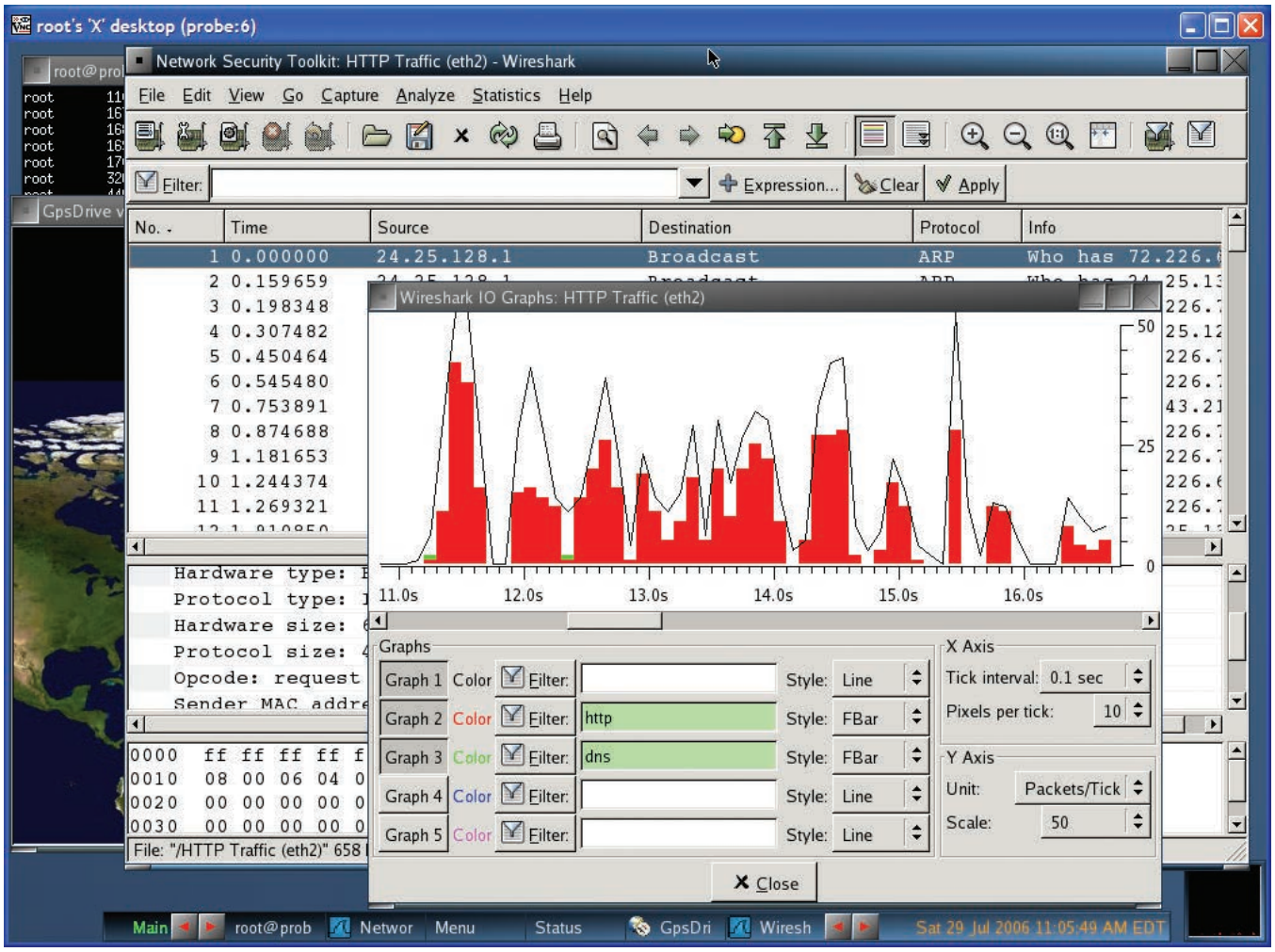


FIGURE 3.16 Analysis of network traffic using Wireshark analysis software.

TABLE 3.2 Popular analysis tools

Angry IP Scanner	Scans a network's open and closed ports. For each IP address found it identifies the hostname, computer name, group name, username and MAC address.
Wireshark	Network protocol analyser and packet sniffer that allows administrators to see what is happening on their network. Used for network troubleshooting and analysis.
Snort	A network intrusion prevention system capable of real-time traffic analysis and packet logging.
NetStumbler	Allows administrators to identify locations that suffer from a weak signal within a WLAN, detect issues of wireless interference and rogue access points. Within a business, it can be used to identify unauthorised wireless LANs that provide access to outside users and thus run the risk of imminent infiltration.
Cain and Abel	Decrypts or recovers lost or forgotten passwords.

Internet services

A significant reason why individuals and organisations purchase computers is to access the internet. The internet is a worldwide collection of networks that links millions of businesses, educational institutions, government departments and individuals. A number of services are available through the internet, with different types of software to support them. **Internet service software** includes web browsers, email, **Voice over Internet Protocol (VoIP)** software and cloud storage.

Web browsers

A **web browser** is an application software package that allows users to access and view webpages. Popular browsers include Internet Explorer, Chrome, Firefox and Safari. When the URL (Uniform Resource Locator) of a site has been entered into the address bar of a browser (Figure 3.17), a webpage from that site is downloaded. The web address consists of a protocol, the domain name, the path to a specific page and the name of the page to be downloaded. A **domain name server (DNS)** identifies the requested site and ensures that data and information are routed to the correct computer.

Search engines such as Google and Yahoo are helpful in locating webpages that contain information of interest. Many search engines maintain a list of words found on the internet. The search engine scans the list for words that have been entered in the search request. The sequence of the list displayed prioritises sites where the search text appears in the page title or descriptor.

Hypertext transfer protocol

Hypertext transfer protocol, better known as **http**, is a set of rules that defines how pages are transferred on the internet. Many browsers and websites do not require the user to enter the 'http://' and 'www' portions of the address.

Hypertext transfer protocol secured

Hypertext transfer protocol secured, or **https**, is a communications protocol for secure transmissions over the internet. The https system provides authentication and encryption communication and is widely used for security-sensitive processing, such as payment transactions and connections to banks. When a secure connection is made, 'https://' is displayed at the start of the URL and a lock symbol appears in the browser window (see Figure 3.18).

When the https protocol is used in an address the computer uses a different TCP port (443) to standard http transfers (TCP port 80). Https uses a protocol called Transport Layer Security (TLS) to transport data safely over the Internet. The predecessor to TLS was called Secure Socket Layer (SSL).

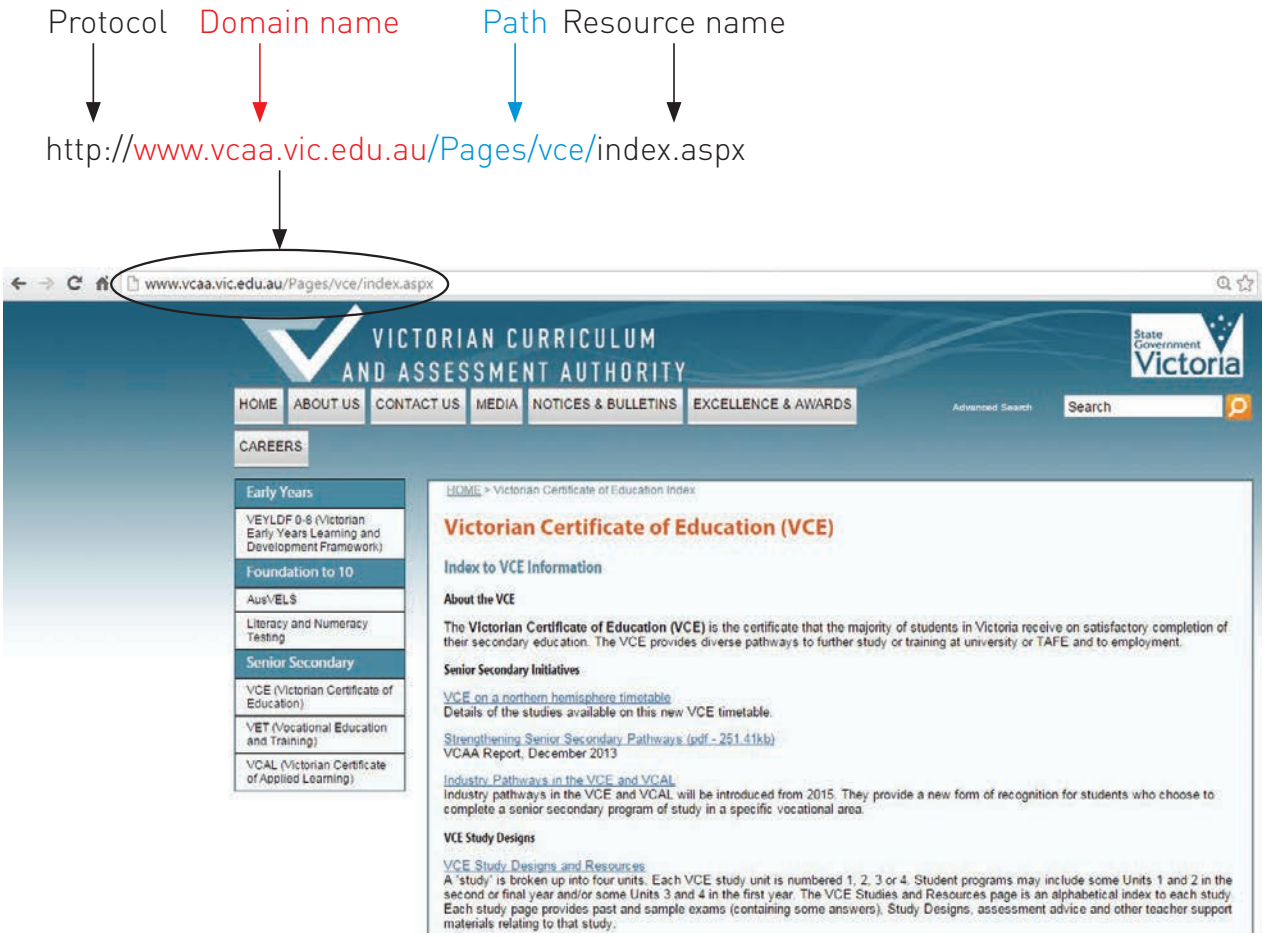


FIGURE 3.17 A URL contains a protocol, domain name, path and resource name

FIGURE 3.18 A secure connection is indicated by 'https://' appearing in the URL and the lock symbol, in this case to the left of the address bar.

Email

Email is the transmission of messages and files via a computer network. Email can be sent over the internet or within a LAN. Email software is used to create, send, read, forward, reply and print emails. A client-server network will have a mail server established to handle the receipt and delivery of mail to and from the LAN over the internet.

File transfer protocol

File transfer protocol (FTP) is an internet standard that allows computers to upload and download files. Uploading is the process of transferring a document, graphic or video from the user's computer to a server on the internet. When someone creates a webpage they need to upload it and associated resources to their ISP's web server using FTP. Files can be downloaded from an internet server to a user's computer also using FTP.

A list of Australian VoIP providers together with plan prices and call rates can be found at VoIP Choice.

Voice over Internet Protocol (VoIP)

Voice over Internet Protocol (VoIP) allows users to speak to other users over the internet. Users need a high-speed internet connection (broadband cable, NBN cable or ADSL 2+), a microphone or alternative audio input device such as existing phones, VoIP adaptor and subscription to a VoIP provider. A VoIP app needs to be installed on a mobile device such as a smartphone.

If calls are mostly within Australia, then a local provider will make billing easier and likely will provide better quality service. If international calls are important, then an overseas provider may offer the best value. There is no reason why users need to limit themselves to just one provider. A local provider for calls within Australia and an overseas provider for international destinations may prove to be the best strategy.

International VoIP providers can allocate users an indial number for use in many countries. This is particularly useful in situations where contacts do not have internet access, such as in some developing countries. An indial number is allocated to the city of the user's choice so that overseas friends and family can call that local indial number to connect with the user. The call is treated as a local call, which saves money.

Cloud storage

Cloud storage refers to saving data to an off-site storage system maintained by a third party. Data is saved to a remote database using the internet rather than on a computer's hard drive (see Figure 3.19). The advantages of cloud storage are that data can be accessed from any location in the world that has internet access and there is no need to carry a storage device or use the same computer to access data that was used to create it. Cloud storage encourages collaborative work practices since members of a project team can be given read and write access rights to the data held in the cloud.

THINK ABOUT COMPUTING 3.6
What are some potential drawbacks of cloud storage?

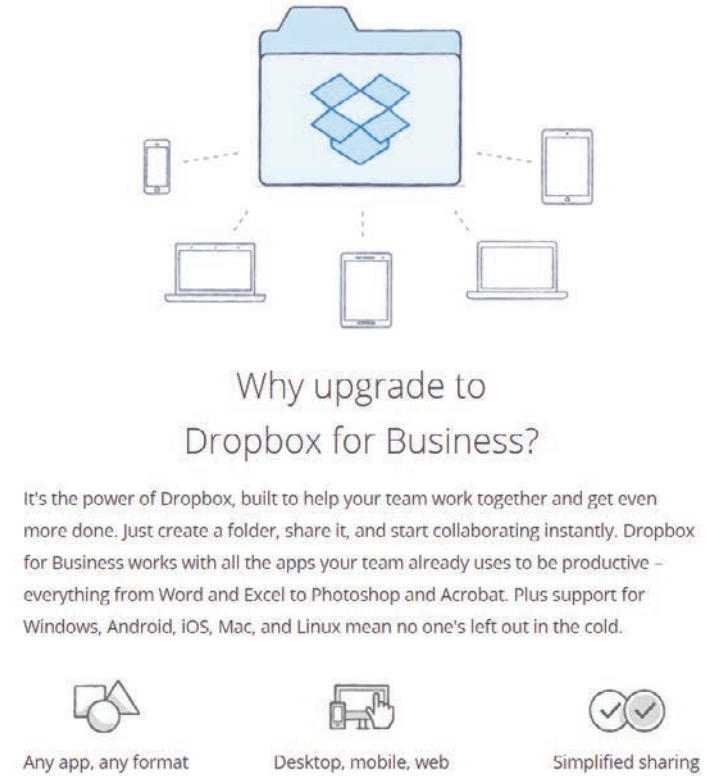


FIGURE 3.19 Dropbox is a file-hosting service that allows users to store and share files over the Internet.

Network communications standards

Today's networks connect terminals, devices and computers from many different manufacturers across many types of networks. These include wide area and local area networks, which use both wireless and wired communication channels. For the different devices on several types of networks to be able to communicate, the networks must use similar techniques for moving data from one application to another.

To avoid problems associated with incompatibility between hardware and software components, organisations such as the Institute of Electrical and Electronics Engineers (IEEE) develop network standards. A **network standard** defines guidelines that specify the way computers access the medium to which they are attached, the type of medium used, the speed at which data flows and the physical technology used. A standard that defines how two network devices communicate is called a **protocol**. The manufacturers of hardware and software must design their products in accordance with the appropriate standard to ensure that their devices can communicate with the network.

The IEEE is a leading authority in technical areas, ranging from computer engineering, biomedical technology and telecommunications, to electric power, aerospace engineering and consumer electronics. The American National Standards Institute (ANSI) is another organisation that has developed globally accepted network standards.

Standards are important in the computer industry because they allow the combination of products from different manufacturers to create a customised system. Without standards, only hardware and software from the same company could be used together. In addition, standard user interfaces make it much easier to learn how to use new applications.

As data flows through a network from one application to another, it may use one or more standards. Some of the widely used network communications standards and protocols for use in both wired and wireless networks are Ethernet and TCP/IP.

Ethernet

Ethernet is a popular network standard that allows personal computers to contend for access to the network.

Ethernet is a popular LAN standard because it is relatively inexpensive and easy to install and maintain. Ethernet networks use cables to transmit data.

The speed at which data is transmitted is usually expressed as bits per second (bps). The original Ethernet standard is not very fast by today's standards – standard Ethernet transfer rate is 10 Mbps. A second Ethernet standard, called Fast Ethernet, transmits data and information at speeds of 100 Mbps, up to 10 times faster than the original standard. Gigabit Ethernet is now in use in many graphic-design studios, and provides an even higher speed of transmission, with speeds up to 10 times faster than Fast Ethernet.

Fast Ethernet uses the same twisted pair cabling as standard Ethernet with upgraded switches and adapters. Signals with Fast Ethernet travel up to 100 Mbps. Gigabit Ethernet (1 Gbps or higher) requires cables capable of handling this faster speed (see Figure 3.32).

Twisted pair cabling is discussed on page 110.

When a workstation on an Ethernet network wishes to transmit, it checks the network to see if it is not busy and then broadcasts a **packet**. A switch ensures that if the packet is addressed to workstation #15, only workstation #15 will read it. If two workstations simultaneously check the network, find it is not busy and send packets at the same time, a collision occurs. When a collision occurs, a special signal travels over the network to indicate that it is 'jammed'. The workstations that sent the messages wait a random amount of time before resending their packets.

An analogy for a **network protocol** would be to consider a computer on a network as a person. The language that person speaks is the network protocol. Only two people who speak the same language can communicate with each other. Similarly, only computers that have the same network protocol can communicate on the network. Just as some people are able to speak several languages, computers on a network may have several different protocols installed and running.

TCP/IP

TCP/IP (transmission control protocol/internet protocol) is a network standard that manages the transmission of data by breaking it up into packets and transmitting the packets over the internet. An important part of the IP protocol is the IP address. The IP addressing standard – four numbers between 0 and 255 separated by full stops – defines a mechanism to provide a unique address for each computer on the network.

A typical TCP/IP packet contains less data than other network protocols. On a WAN, such as the internet, the smaller size allows the packets to travel in multiple paths to the destination. On a LAN there is often only one viable path to the destination, so larger packets are more sensible (less space is allocated to addresses and communications information). Nevertheless, most LANs are now running the TCP/IP protocol because intranets require files to be communicated in this medium.

When a computer sends data over a network, the data is divided into packets. Each packet contains the data, the recipient (destination) information, the sender (origin) information, and the sequence information used to reassemble the data at the destination (Figure 3.20). These packets travel along the fastest available path, avoiding congestion and broken links to the recipient's computer via routers (routers were discussed on page 90). This technique of breaking a message into individual packets, sending the packets along the best route available, and then reassembling the data is called **packet switching**.

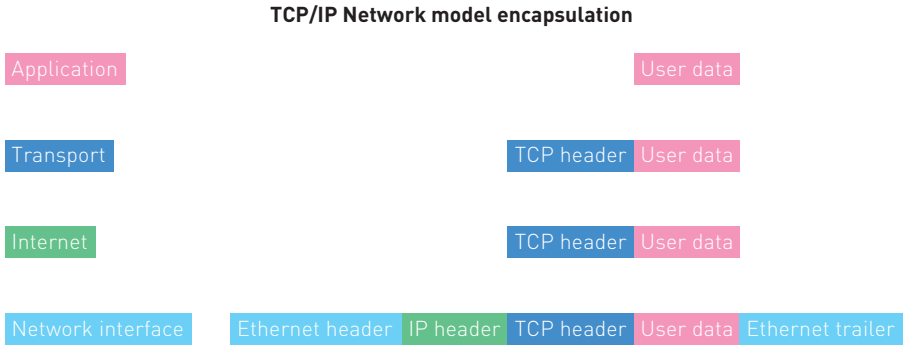


FIGURE 3.20 TCP/IP Network model encapsulation. At each network layer another piece is added to the data packet. The Ethernet frame at the network interface level, consists of a header containing the destination and source address, the middle section, which includes the data and headers for IP and TCP protocols, and a final section that checks for correct transmission.

Some blocks of IP addresses are set aside for internal private use by computers not directly connected to the internet. They are used by business or home networks that need to use TCP/IP but do not want to be directly visible on the internet. The block 10.0.0.0 to 10.255.255.255 is commonly used by school networks and other large organisations with many nodes. Home networks often use the block 192.168.0.0 to 192.168.255.255 to assign addresses to connected devices while IP addresses from 172.16.0.0 to 172.31.255.255 are also for private use. Home network routers often have 192.168.0.1 as their default IP address.

ISSUE Are we running out of IP addresses?

Currently you should not have any trouble connecting your computer to the internet using the Internet Protocol (IP). The version of IP address in common use for many years was Internet Protocol Version 4 (IPv4). An IPv4 address has four groups of numbers that range from 0 to 256 separated by a decimal point (the system is called a dotted decimal number). Each group of numbers is called an octet. The first octet identifies the network on which a computer resides, and the remaining three octets together identify the specific computer or host within that network. The four octets in an IPv4 address allows 2 to the power of 32, or a little over 4 trillion (4×10^9), unique values. Some of these are reserved for special use, leaving more than 3 trillion available addresses.

Three trillion addresses may sound like a lot; however, because so many computers and other devices connected to the internet need a unique address, there is a shortage of IP addresses. To overcome this shortage, a new version of IP addressing was developed. Internet Protocol Version 6 (IPv6) lengthens IP addresses from 32 bits (the octet size) to 128 bits, and increase the available IP addresses to 340×10^{38}.

IPv6 uses eight hexadecimal strings to specify the address. The problem is that adopting the new IP version requires organisations to reconfigure their networks and possibly update their switches to accommodate the new standard.

Home networks and smaller corporate networks do not have enough devices attached to warrant the implementation of IPv6. For the time being, home networks will remain within the 192.168.0.0 to 192.168.255.255 private address block provided by IPv4.

A tablet connected to the Internet on a home wi-fi network can use IPv4. If the tablet is used on a 4G mobile network it will need IPv6.

The 802.11 standard

The **802.11 standard** was developed to specify how two wireless computers or devices communicate with each other via radio waves. This standard uses techniques similar to those of the Ethernet standard and is therefore easy to integrate into existing wired Ethernet networks. The range of wireless communications is up to 300 metres in open areas outside, and about 50 metres inside buildings. Wireless networks are popular in locations where there is no existing wired system, or where the construction of the building makes it difficult to add a wired network. For example, it can be difficult to place wires inside the brick or solid plaster walls that are common in older buildings.

The term wi-fi (wireless fidelity) identifies any network based on the 802.11 series of standards.

The older 802.11b and 802.11g standards use frequencies in the 2.4 GHz band and employ direct sequence spread spectrum technology to avoid interference from microwave ovens, Bluetooth devices and cordless phones. The 802.11n standard uses 2.4 GHz and/or the 5 GHz band. The higher frequency allows the signal to carry more data. The 802.11n standard transmits data at up to 150 Mbps. The newest standard, 802.11ac, operates in the 5 GHz band with data transmission of 1300 Mbps.

THINK ABOUT COMPUTING 3.7

The new IPv6 standard is needed because we are about to run out of IPv4 addresses. Do you think all organisations should be forced to adopt the IPv6 standard to increase the number of addresses available? Do you think the demand for IP addresses will increase over the next few years in the same way that it has over the last 10 years, or will there be an escalation? If an escalation were to occur, what might be the trigger?

The 5 GHz band is not populated by many devices, hence it tends to be quieter, meaning there is less interference from neighbourhood networks or other devices. The 802.11ac standard also makes use of **beamforming**, which means it detects where devices are and intensifies the signal in their direction, rather than using the scattergun approach of earlier standards. The result is the 802.11ac standard allows fibre optic broadband speeds throughout a house, with multiple streams of full high-definition content.

Sending and receiving devices

Sending and receiving devices initiate or accept the transmission of data, instructions and information. Notebook computers, desktop computers, tablets, midrange servers and mainframe computers can all serve as sending and receiving devices. These computers can communicate directly with another computer, with hundreds of computers on a company network or with millions of other computers on the internet.

Web-enabled hand-held computers and devices also serve as sending and receiving devices. A **web-enabled device** provides access to the internet and email from any location. Examples are smartphones and GPS receivers.

Earlier in this chapter we noted that communication over a global network requires:

- a sending device, such as a computer
- a communications device attached to the sending device, such as a cable modem
- a communications channel – either cable, radio, microwave or satellite
- a communications device at the receiving end
- a receiving device that accepts the data or information.

Mobile devices connected to networks

Mobile devices are usually small enough to fit in a pocket. These devices store programs and data permanently on special memory inside the system unit or on a flash memory card. They can usually be connected to personal computers to exchange information or to install applications.

Most mobile devices can be connected wirelessly to the internet. This allows users to chat, send messages, email and access websites. Internet-enabled mobile devices in common use include smartphones, hand-held computers, navigation systems, games consoles and digital cameras. A convergence of technologies has resulted in some devices including functions and features from two or more different types of devices, while other devices like personal digital assistants (PDAs) have become redundant.

Tablets

A **tablet** is a special type of notebook or laptop computer that resembles a letter-sized slate, which allows a user to write on the screen using a digital pen. Tablets are able to connect to wi-fi networks to gain access to the internet and other network resources. Tablets are designed to be held rather than sat on a desk or in the user's lap. Tablet computers have a number of capabilities and some limitations.

TABLE 3.3 Capabilities and limitations of tablet computers

Capabilities of tablets	Limitations of tablets
Allows users to: • surf the internet • send and receive emails • operate a calendar of events • play music • take pictures • store images • use GPS to identify location • run apps.	• No physical keyboard • Difficult to print from • Applications are limited to running on iOS, Android or Windows devices. Some Windows-based software used by businesses is not available on either platform

Smartphones

Besides acting as a phone for voice-based communications, a **smartphone** allows users to send emails and access the web (Figure 3.21). For many businesspeople, smartphones are replacing notebook computers, especially if their work involves frequent travel. A smartphone and its charger weigh about 200 grams, compared with a laptop computer and its related accessories that weigh 2–3 kilograms.

Smartphones communicate wirelessly with other devices or computers. They also function as a portable media player and a digital camera. Smartphones have a number of capabilities and some limitations.

Shutterstock.com/Zeynep Demir

FIGURE 3.21 The Samsung Galaxy S6 is a web-enabled smartphone.

TABLE 3.4 Capabilities and limitations of smartphones

Capabilities of smartphones	Limitations of smartphones
Allows users to: • make and receive phone calls • surf the internet • send and receive emails • operate a calendar of events • play music • take pictures • store images • use GPS to identify location • run apps.	• Screens are smaller than a desktop or tablet, so only a few lines of a document can be displayed • Many webpages are not designed for a small screen • The keyboard is usually a touch screen and smaller than a standard keyboard, so correct data entry can be difficult • Difficult to print from

THINK ABOUT COMPUTING 3.8
Describe how the limitations of smartphones may in part be overcome or addressed.

Most smartphones can connect directly to Bluetooth-enabled devices, such as keyboards and printers. Bluetooth connections are discussed later in this chapter.

Mini-applications software, also called **apps**, are a significant component of smartphones. Smartphones come pre-loaded with apps for social networking, photography, travel and more. Additional apps can be downloaded from Apple's App Store (for iPhone clients), Google's Play Store (for Android devices) or the Microsoft Store (for Windows phones).

The increasing popularity of internet-connected mobile applications among businesspeople, teenagers and students is producing a strain on carrier networks. Applications on portable devices that require continuous connection to a website can place a huge load on wireless networks. As these devices become more popular, the traffic on networks is only going to increase, putting further pressure on the ability of carriers to cope.

Australia's digital economy grew strongly during 2012–13 with increases in internet usage and mobile phone access to the internet, according to the ACMA Communications report tabled in Parliament. The report stated that 7.5 million Australians were accessing the internet via their mobile phones during June 2013, an increase of 33 per cent compared to June 2012. There were 11.19 million smartphone users in Australia in May 2013.

Shutterstock.com/Voznikevich Konstantin

Getty Images/Arno Masse

FIGURE 3.22 A handheld data-collection device is used for high-volume data collection, and features a touch screen, keyboard and number pad.

Handheld data-collection devices

As the name implies, a hand-held data collection device is small enough to fit into your hand. Hand-held devices communicate wirelessly with other computers and devices, and many have miniature or specialised keyboards. Some hand-held devices use a stylus for data entry, including recording a client signature to acknowledge receipt of goods.

Hand-held computers (Figure 3.22) are often used by people who need to record information as they travel; for example, electricity and water meter readers, scientists collecting natural resource data in the field and couriers who deliver parcels. Table 3.5 lists the capabilities and limitations of hand-held data collection devices.

There is concern that drivers relying on GPS are performing illegal or dangerous acts. For example, the rural village of Wedmore, in England, has seen its share of truck drivers wedged between buildings because they blindly followed the route their GPS planned out – regardless of the fact that trucks are prohibited from the village because of their size. There have also been cases in which GPS units have directed drivers onto train tracks or the wrong way down a one-way street.

TABLE 3.5 Capabilities and limitations of hand-held data collection devices

Capabilities of handheld devices	Limitations of handheld devices
• Designed to operate outdoors in variable weather conditions • Shockproof • High visibility screens even in sunlight • Runs apps of spreadsheets, word-processors, presentation software, calendar and email • Allows users to reach a specified location using GPS • Allows the collection, storage and management of large amounts of data • Data collection can be tagged to location using GPS coordinates • Data can be transferred to larger computers or networks using Bluetooth or wi-fi	• Screens are smaller than a desktop or tablet, so only a few lines of a document can be displayed. • The keyboard is smaller than a standard keyboard, so correct data entry can be difficult. • Text written with a stylus may not be easy to read.

Navigation systems

Shutterstock.com/3d brained

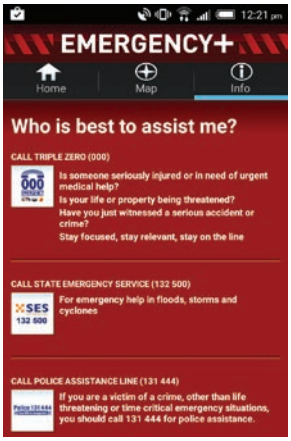
FIGURE 3.23 The GPS navigation device receives three-dimensional data from a satellite network, which enables the receiver to pinpoint the car's location and give travel directions to the driver.

The **global positioning system (GPS)** is a navigation system made up of a group of at least twenty-four satellites that communicate with a fixed or portable device mounted in, for example, a car. The GPS navigation device is able to pinpoint the car's location to within 15 metres (see Figure 3.23). It receives three-dimensional data – latitude, longitude and altitude – as well as precise time. The car's position is then overlaid with digital mapping, and navigation information is stored within the GPS receiver. From this the device is able to give the driver directions, suggest a route around congested traffic or locate a petrol station. The text-to-speech function allows the GPS device to read street addresses rather than simply provide left and right turn directions. This allows drivers to keep their eyes on the road.

Smartphones have GPS technology built in to them. Parents can locate a child's whereabouts through a smartphone with a GPS receiver (Figure 3.24).

Getty Images/Patrick T. Fallon/Bloomberg

FIGURE 3.24 Life360 is an app for a smartphone that allows parents to set up a circle of family members and follow their movements on a map, communicate with them, and receive alerts when members arrive at home, school or work.



National Triple Zero Awareness Work Group © Fire Rescue New South Wales

FIGURE 3.25 Emergency + is an app developed by the NSW Fire Rescue Service.

TABLE 3.6 Capabilities and limitations of GPS devices

Capabilities of GPS devices	Limitations of GPS devices
• Receives latitude, longitude and altitude data from satellites • Plots accurate location on a map • Can be used to plan a route • Provides audio directions • Can provide traffic information and speed limit alerts • Locate places of interest such as petrol stations	• Accuracy may be variable due to triangulation issues and atmospheric conditions • GPS receivers are accurate to within 15 metres on average, though some systems might improve accuracy to 3 metres • Devices do not connect to satellites in thick tree covering or in a gorge • The device may provide a direct route rather than the safest route

The NSW Fire Rescue Service has released an app for mobile phones (see Figure 3.25) that uses a smartphone's GPS function to locate exactly where a person is. Users in an emergency can provide the 000 operator with their precise location – either the address or map coordinates.

THINK ABOUT COMPUTING 3.9

Suna is an automated traffic congestion notification system used by VicRoads. Investigate what information is provided by the system and how the data is collected.

Wearable technology

AFL coaches are keen to know how fast a player runs, track exactly where they go and collect data on the movements of groups of players, such as defenders or the on-ballers. Players wear a wireless position locating system (Figure 3.26) developed by CSIRO in a pouch on their backs. These devices use wireless signals and hence work in areas that GPS satellites do not reach, such as in Etihad Stadium when the roof is closed.

This is a wireless ad-hoc system for positioning (WASP) technology and it monitors player movements on the field. The device is called ClearSky and is used by international sports associations, including the US National Football League (NFL). Rather than using satellites, the WASP system uses fixed reference nodes usually located inside the stadium. The mobile device measures the time it takes signals to travel to each of the fixed nodes and uses triangulation to work out the player's position. The WASP system has accuracy down to 20 cm (compared to 3 metres for GPS systems).

A combination of heart rate measurements and position tracking allows clubs to monitor player fatigue and improve training.

Wearable technology falls into three broad areas – notifiers, trackers and glasses. The notifiers are devices that show off the world around you. Pebble Steel is a smartwatch that can load iPhone and Android apps. Trackers use sensors to record data and include cameras, audio recorders, fitness-tracking devices (see Figure 3.27) and pedometers. Glasses include virtual displays worn on the head. The website www.bluetooth.com/Pages/wearables.aspx describes some uses of wearable technology.

Getty Images/Adam Trafford/AFL Media

FIGURE 3.26 AFL players wear a wireless position locator on their backs so that their movement around the ground can be analysed.

Shutterstock.com/BsWei

FIGURE 3.27 Wearable devices are being used to improve our daily lives, health care and safety; such devices include fitness trackers, health care monitors, smartwatches, heads-up displays and smart clothing.

TABLE 3.7 Capabilities and limitations of wearable technologies

Capabilities of wearable technologies	Limitations of wearable technologies
• Body sensors that track information relating to health and fitness • Synchronise with the user's smartphone using Bluetooth • Smart watches run apps and identify incoming calls on smartphones	• Some of the wearables can be uncomfortable • Contacts can degrade if they get worn or dirty leading to incorrect readings or none at all • Power requirements require regular replacement of batteries (if not rechargeable) • Static electricity (from sports uniform) and electric fields from other devices can cause erroneous results • Most wearables lack a screen so immediate feedback from a tracking app is not possible • Some wearables are limited to connection with a particular smartphone • Devices may have limits to the number of apps they can hold at a time • Portability and size means the devices can be lost or misplaced

Games consoles

A **game console** is a computing device designed for single-player and multi-player video games. A standard game console uses a handheld controller for input, a television screen for output and a hard disk or memory disk for storage. Popular models include Microsoft's Xbox One, Sony's PlayStation 4 and Nintendo's Wii U.

A smaller handheld game console fits in the users hands and includes the controls, screen and speakers in the one portable device (Figure 3.28).

The first home video game was called Magnavox Odyssey and was released in 1972. The console lacked a CPU so cartridges were used for each separate game – tennis, volleyball and chase. The home game industry struggled until the Atari game Pong became popular in the early 1970s.

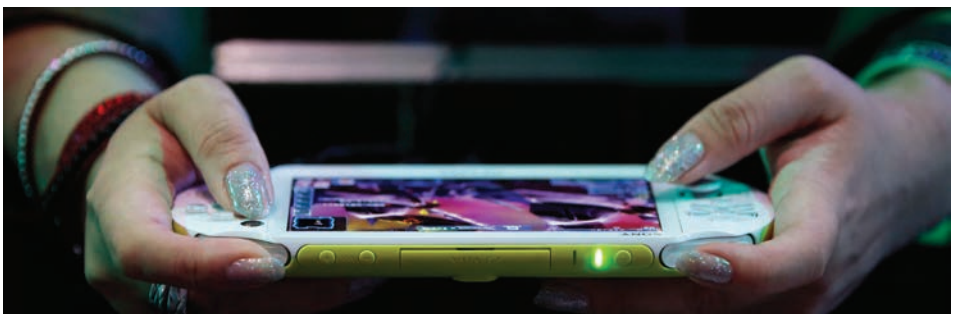

Getty Images/Kiyoshi Ota/Bloomberg

FIGURE 3.28 A portable game console MP5 player

TABLE 3.8 Capabilities and limitations of game consoles

Capabilities of game consoles	Limitations of game consoles
• Play games using Blu-ray disks or digital downloads • Wi-fi connection to internet • Upload saved data to a cloud server • 3D graphics • Stream videos from an online source • Stream gameplay to an online service, such as Ustream • Record videos and images for posting online • Social connectivity – voice chat with online players or group chat with friends	• Tend not to be upgradeable – technological advances require a new model • Old consoles do not play media designed for a newer model • The manufacturer's monopoly over a particular console market forces developers to certify games for that platform. This in turn can force limits on the functionality that can be built into games.

Network-attached storage device

A **network-attached storage (NAS)** device acts as a centralised repository for data. NAS devices (Figure 3.29) share files across a network. In a home environment, the files are often videos, photos or audio files. Most NAS devices can be used as a multimedia server which can share and stream multimedia files to network clients such as computers, tablets, game consoles and phones. A NAS used on a home network is likely to have storage capacity of about 16TB. Most NAS devices require a cable connection to a switch that has Gigabit Ethernet connectivity.

Shutterstock.com/ShaunWilkinson

FIGURE 3.29 A network-attached storage (NAS) device

One terabyte (1 TB) is the same as one thousand gigabytes (1000 GB) or one million megabytes (1 000 000 MB).

Communications channel

An important aspect of communications is the **channel**, which is the communications path between two devices. **Bandwidth** is the width of the communications channel. The higher the bandwidth, the more data and information the channel can transmit.

For transmission of text-based documents, a lower bandwidth delivers acceptable performance. If you transmit music, graphics and photographs, or work with virtual-reality or 3-D games, you need a higher bandwidth. When the bandwidth is too low for the application, you will notice a considerable slowdown in system performance. For example, movies may 'stutter' or pause while buffering more data.

A communications channel consists of one or more **transmission media**. When you send data from your computer to another device, the signal carrying that data most likely travels over a variety of transmission media.

Transmission media are either physical or wireless. **Physical transmission media** use wire, cable or fibre-optics to send communications signals. **Wireless transmission media** send communications signals through the air or space using radio, microwave and infra-red signals.

Physical transmission media

Physical transmission media used in communications include **twisted-pair cables** and fibre-optic cables. These are typically used within buildings or underground. Ethernet LANs often use physical transmission media. Many wireless LANs rely on physical cables to transmit data between a switch or router and a wi-fi device such as a wireless access point. Table 3.9 lists the transfer rates of LANs using various physical transmission media.

TABLE 3.9 The speeds of various physical transmission media when they are used in LANs

Type of cable	Transfer rates
Twisted-pair cable	
10Base-T (Ethernet)	10 Mbps
100Base-TX (Fast Ethernet)	100 Mbps
1000Base-T (Gigabit Ethernet)	1 Gbps
Fibre-optic cable	
10Base-FL (Ethernet)	10 Mbps
100Base-FX (Fast Ethernet)	100 Mbps
Gigabit Ethernet	1 Gbps
10-Gigabit Ethernet	10 Gbps

Noise in a cable can be caused by cross-talk (electric currents between pairs of wires in the same cable) and outside electrical fields, such as power lines, motors and radio transmitters. The greater the cable's ability to resist internal and external noise, the longer the cable that can be used to connect workstations and resources.

Twisted-pair cables

One of the more commonly used transmission media for network cabling and telephone systems is the **twisted-pair cable**. This cable contains one or more twisted-pair wires bundled together (see Figure 3.30). Each **twisted-pair wire** consists of two separate insulated copper wires that are twisted together. The wires are twisted together to reduce noise. Noise is an electrical disturbance that can degrade communications.

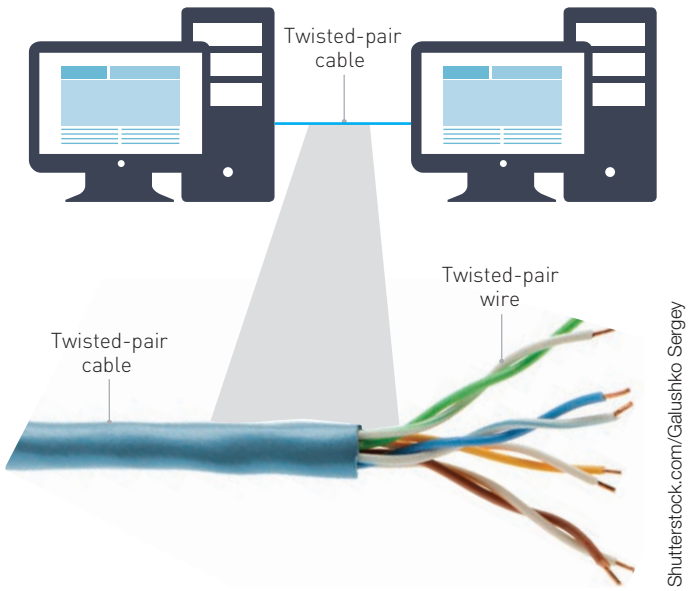


FIGURE 3.30 A twisted-pair cable consists of one or more twisted-pair wires; each twisted pair wire is usually colour-coded for identification. Telephone networks and LANs often use twisted-pair cables.

Fibre-optic cable

The core of a **fibre-optic cable** consists of dozens or hundreds of thin strands of glass or plastic that use light to transmit signals. Each strand, called an **optical fibre**, is as thin as a human hair. Inside the fibre-optic cable, an insulating glass cladding and a protective coating surround each optical fibre (Figure 3.31).

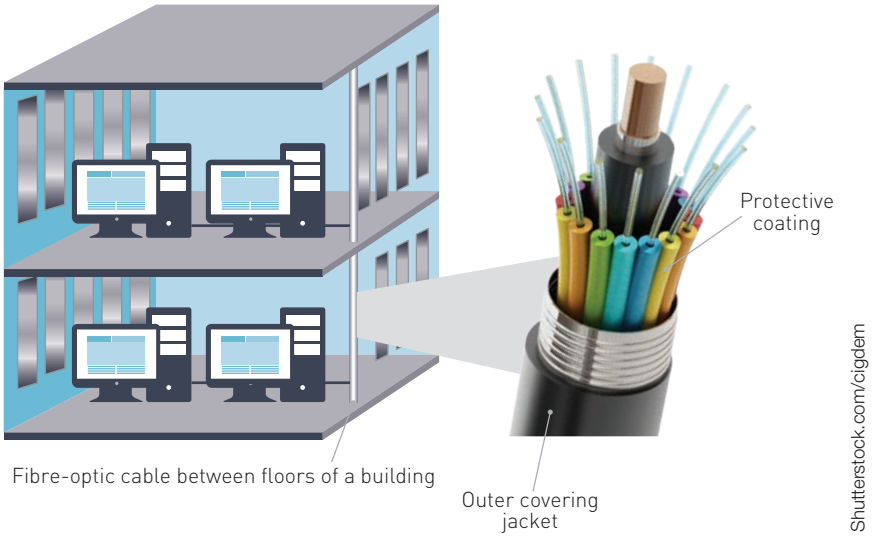


FIGURE 3.31 A fibre-optic cable consists of hair-thin strands of glass or plastic that carry data as pulses of light.

Fibre-optic cables have several advantages over twisted-pair cables. These advantages include:

- the ability to carry significantly more signals than wire cables
- faster data transmission
- less susceptibility to noise (interference) from other devices, such as a copy machine
- better security for signals during transmission as they are less susceptible to noise
- their smaller size (much thinner and lighter).

The material used for fibre-optic cable has optical properties that cause light in the cable to totally internally reflect from the outer surface. This allows all the light to progress down the cable rather than escape to the surrounding air. As fibre-optic cables use pulses of light, they have total immunity from electrical noise. Signals can therefore be sent over much larger distances than with twisted-pair cables.

The disadvantages of fibre-optic cable are that it costs much more than twisted-pair and it can be difficult to install and modify. Despite these limitations, many local and long-distance telephone companies and cable television operators are replacing existing telephone and coaxial cables with fibre-optic cables. Many businesses are also using fibre-optic cables in high-traffic networks or as the main cable in a network.

National Broadband Network (NBN)

The National Broadband Network (NBN) is designed to provide infrastructure for affordable and reliable high speed internet and telephone access to all Australians. The nature and size of Australia means that a variety of technologies are required to deliver the NBN. The original NBN proposal was for communications of 100 Mbps, with fibre-optic cable connecting the internet directly to homes. The scheme has been modified with a change of government, so that the plan

At a connection speed of 25 Mbps the NBN will have a similar maximum speed to ADSL 2+ connections (24 Mbps).

AAP Image/NBNCO

FIGURE 3.32 The roll-out of the NBN

View the progress of the NBN rollout at the NBN website.

THINK ABOUT COMPUTING 3.10

Some future speculators believe that by 2023 households will have more than 100 devices connected to the internet. Devices such as lights, air conditioners, heating systems, door locks, irrigation systems, motion detectors, smoke detectors, home entertainment systems and connections that monitor family members, pets and vehicles all will require bandwidth. Online video chatting will become common practice and requires fast download and upload speeds. Do you think that 25 Mbps provide adequate bandwidth in five to 10 years' time? What devices do you expect to be connected in your home in 10 years' time?

is for fibre-optic cable to reach a node in the street (called the 'street cabinet') from where it will be split and hybrid cables or existing copper telephone networks used to carry signals to homes and businesses. The revised specifications indicate connection speeds of about 25 Mbps. The government argues that this is more than adequate for domestic connections. The estimate for the revised NBN network is A$29.5 billion (the original estimate was A$37.4 billion) with a completion date of 2019 (Figure 3.32).

Households where it is impractical to use fibre cable will connect to the NBN via fixed wireless and satellite technologies. It is estimated that up to 10 per cent of homes will need wireless or satellite connections.

Wireless transmission media

Wireless transmission media are used when it is inconvenient, impractical or impossible to install cables. With the faster speeds afforded by the 802.11ac standard, wireless connections are on a par with wired networks. Wireless transmission media used in communications include broadcast radio, cellular radio, microwaves, communications satellites and infra-red.

TABLE 3.10 Transfer rates of various wireless transmission media

Transmission medium	Transfer rates (maximum)
Bluetooth HomeRF 802.11a 802.11b 802.11g 802.11n 802.11ac	1–2 Mbps 1.6–10 Mbps 54 Mbps 11 Mbps 54 Mbps 108–600 Mbps 867–1300 Mbps
Cellular radio 2G 3G 4G	 9.6–19.2 Kbps 200 Kbps–2 Mbps 2–12Mbps
Microwave radio	150 Mbps
Communications satellite	1 Gbps
Infra-red	115 Kbps–4 Mbps

An 802.11ac wireless router operates on the 5 GHz bandwidth but can still run on the 2.4 GHz network simultaneously. Some vendors may quote a wireless 802.11ac router as a single speed device operating at 1.75 Gbps. This figure is an amalgamation of the 5 GHz and 2.4 GHz capabilities of the router. That is, it is the addition of 1.3 Gbps from the 5 GHz connection with 450 Mbps from the 2.4 GHz network.

Wi-fi communications

For **wi-fi communication** transmissions, you need a transmitter to send the radio signal and a receiver to accept it. To receive the signal, the receiver has an antenna that is located in the range of the signal. Some networks use a transceiver, which both sends and receives signals from wireless devices. Wi-fi communication is slower and more susceptible to noise than physical transmission media, but it provides flexibility and portability.

Wireless computer network components typically use radio signals in either a 2.4 GHz range or a 5 GHz range. A 5 GHz network can carry more data than a 2.4 GHz network; however, the higher the frequency of a radio signal, the shorter its range. So a 2.4 GHz network covers a much larger range than a 5 GHz network. The higher frequency is not as good at penetrating solid obstacles such as walls. On the other hand, there are a number of household devices, such as cordless phones, that operate on the 2.4 GHz band that could interfere with the broadcast transmissions. The 5 GHz band does not compete with other common household devices. A number of components now come with dual band capacity to get the best of both worlds.

Bluetooth

Bluetooth uses short-range radio waves to transmit data among Bluetooth-enabled devices. These devices contain a small chip that allows them to communicate with other Bluetooth-enabled devices. Examples of these devices can include desktop personal computers, notebook computers, hand-held computers, mobile telephones, fax machines and printers. To communicate with one another, they must be within a specified range (about 10 metres, but the range can be extended to 100 metres with additional equipment). A popular use of Bluetooth is to enable hands-free chatting on mobile phones. Most cars are now sold with a built in Bluetooth station that the user can synchronise with their mobile phone. Bluetooth and wi-fi communications technologies use radio signals.

Bluetooth gets its name from a legendary Viking king (one of the main developers was the Norwegian communications company Ericsson). One advantage of Bluetooth is that it can be used to set up a network on the spur of the moment. A group at a meeting can all network their computers to share files, access data from mobile telephones, and send documents to printers and fax machines, all without cables or additional network interface cards. Of course, all of the devices must be Bluetooth-enabled. Data transmission using Bluetooth is fairly slow compared with other wireless transmission media.

Near field communication

AAP/AP Photo/Manu Fernandez

FIGURE 3.33 An example of a near field communication use for short-range contactless communication.

AAP Image/Julian Smith

FIGURE 3.34 The myki public transportation ticketing system uses near field communication to read a commuter's details from their smart card.

Near field communication (NFC) is a form of contactless communication between devices like smartphones or tablets. A user is able to wave their smartphone over a NFC compatible receiver to send information without needing to touch the devices together or set up a formal connection. A customer can simply wave their smartphone near a NFC device to pay for goods purchased in stores or pay for a parking meter (the parking meter can even send messages to the smartphone indicating how much time is left).

An unpowered chip, called a tag, can be used with an NFC device, such as a smartphone with NFC capability. The tag draws its power from the device that reads it using electromagnetic induction. A smartphone can be paired with an NFC tag which can be programmed by apps on the phone to automate tasks. For example, tapping on a smart tag on a poster will transfer information from an embedded chip in the poster onto the smartphone. So a user tapping a movie advertising poster will receive comprehensive details about the film, such as session times, biographies of leading actors, reviews and more, on their smartphone. Tapping on a menu in a restaurant could load the menu into the phone with nutritional information and cooking notes.

NFC tags are small and cheap to produce, so are suited for a range of uses involving mobile payments and creative marketing. More interesting uses will appear as people get more aware of the capabilities of NFC (Figure 3.33).

Pay-wave transactions and Victoria's myki transportation system (Figure 3.34) are examples of cards that use NFC technology. Waving the card near a card reader allows data to pass from the card to the reader and hence the transaction is completed with little effort or time taken.

THINK ABOUT COMPUTING 3.11
Use the Internet to explore creative ways in which people are using NFC tags.

There are four types of NFC tags. Type 1 tags store 96 bytes and operate at 106 Kbps. The biggest and fastest tag is Type 4 which can store up to 32 KB and transmit at 424 Kbps. Type 1 and 2 tags can be written to multiple times, or can use encryption, which permanently locks them so the data cannot be manipulated. Type 3 and 4 tags can be written to once only.

Cellular radio

A **mobile phone** is a telephone device that uses radio signals to transmit voice and digital data messages.

Cellular radio is a form of **broadcast radio** that is used widely for mobile communications, specifically mobile phones (Figure 3.35). A smartphone can be used to communicate with other

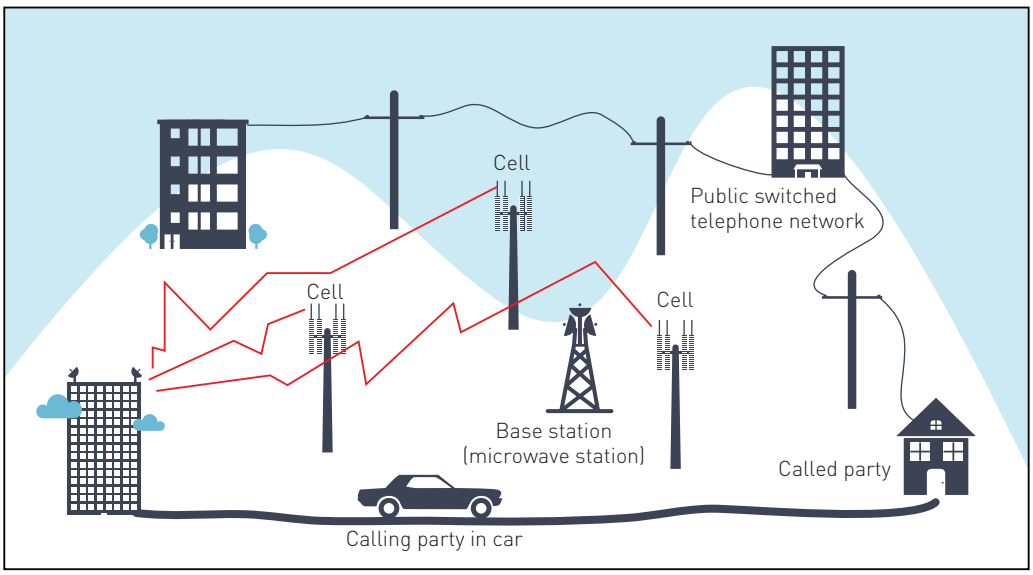


FIGURE 3.35 As a passenger in a car with a mobile telephone travels from one 'cell' to another, the radio signals transfer from the base station (microwave station) in one cell to a base station in another cell.

phones, access the Web, send and receive email, enter a chat room or connect to an office or school network while away from a standard telephone line, such as from a car or a park bench.

Several categories of cellular transmission exist.

- 1G (first generation) – transmitted analog data only
- 2G (second generation) – transmitted digital data at speeds of 9.6–19.2 Kbps
- 3G (third generation) – transmitted digital data at speeds from 200 Kbps to 2 Mbps
- 4G (fourth generation) – transmitted digital data at speeds from 2 Mbps to 12 Mbps

Microwaves

Microwaves are radio waves that provide a high-speed signal transmission. Microwave transmission involves sending signals from one **microwave station** to another (Figure 3.36).

Microwave transmissions are often a viable alternative to cabling where an organisation has premises on different sides of a major road. The cost of digging a tunnel under the road can be expensive and repairs are difficult if there is a break.

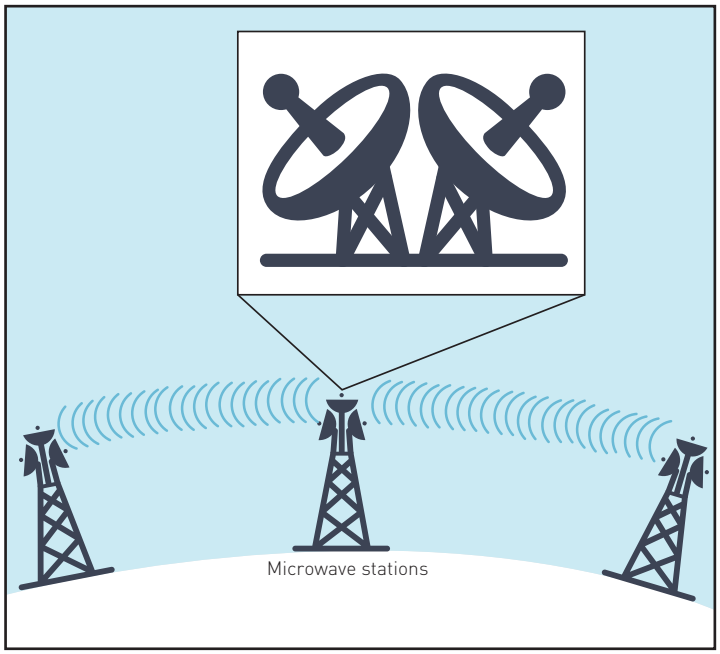


FIGURE 3.36 A microwave station is an Earth-based reflective dish that contains the antenna and other equipment necessary for microwave communications; the dish collects the signals and redirects them to the central collector.

Microwaves use line-of-sight transmission, which means that microwaves must transmit in a straight line with no obstructions between microwave antennas. To avoid possible obstructions, such as buildings or mountains, microwave stations often sit on the tops of buildings, towers or mountains.

Electromagnetic radiation, such as light and radio waves, travels almost as fast through the air as it does through a vacuum (about 300 000 km per second). This means that microwave communication is significantly faster than fibre optic transmissions which send laser light pulses down glass strands. The glass slows the light beam by 30 to 40 per cent. Microwave transmission is used in environments where installing physical transmission media is difficult or impossible, where the organisation occupies a large site and where line-of-sight transmission is available.

A licenced microwave link (operating at frequencies between 7 GHz and 42 GHz) would be used by organisations if the data being transferred is 'mission critical', the connection is long distance or they operate in a high density area where interference is a problem at the lower frequencies.

Communications satellite

A **communications satellite** is a space station that receives microwave signals from an **Earth-based station**, amplifies (strengthens) the signals and broadcasts the signals back over a wide area to any number of Earth-based stations (Figure 3.37). These Earth-based stations are often

Although satellite internet connections are more expensive than cable internet, they are still the only high-speed option in remote areas.

microwave stations. Other devices, such as hand-held computers and GPS receivers, can also function as Earth-based stations. Transmission from an Earth-based station to a satellite is an **uplink**. Transmission from a satellite to an Earth-based station is a **downlink**.

FIGURE 3.37 Communications satellites are placed approximately 37 000 kilometres above Earth's equator.

Network security

Information transmitted over networks has a higher degree of security risk than information kept on a company's premises. Many security techniques – such as usernames, passwords, biometrics and firewalls – are used by network administrators to protect a network. On a vast network with no central administrator, such as the internet, the risk is even greater. Every computer along the path of your data can see what you send and receive.

Security threats

The integrity and security of data and information stored within, and communicated between, information systems can be threatened by a number of actions, devices and events. The threats can be accidental, such as losing a portable storage device containing files; deliberate, such as denial of service and worms; or as a result of an event, such as a power surge.

Accidental threats

Some accidental threats are hard to guard against. For example, people can accidently delete files or send attachments via email to the wrong person.

Portable memory devices are handy for storing files that you can take anywhere with you, but they can easily be lost. For example, losing a memory stick that contains important files would be inconvenient if those files were needed for a meeting or to continue working at a different location. Provided the original files were still available on a computer or network,

however, the situation is retrievable, albeit with time lost and opportunities missed. If the memory stick contained files that were irreplaceable, however, the situation could be dire. Losing a memory stick that contained strategic business information or confidential data could be critical if it fell into the wrong hands.

Deliberate threats

Deliberate threats to data and information occur when someone tries to damage or manipulate the system. This can be through a **hacker** finding a way into a network that bypasses security measures, or an employee sabotaging files or altering data for their own benefit.

A hacker tries to gain access to a network from a remote location using the internet. An authorised client using a password that is predictable or in some way compromised poses a threat to the network. Regardless of the method used, an unauthorised user can seek to damage files stored on the computer or network, steal secret information for their own advantage, or cause mischief in some way.

Networks should also be protected from deliberate attacks from **malware**, such as viruses, worms, Trojans, adware, **spyware**, keyloggers and logic bombs.

Malware is software that is created and used to disrupt computer operation, gather sensitive information or damage the system.

A **virus** is a computer program that can destroy files and alter the performance of the operating system. Once a virus is in a computer, it can spread over a network to other connected computers. The amount of damage caused by a virus depends on the maliciousness of the author. A number of viruses simply spread from file to file without causing any real damage, other than taking up storage space and perhaps the embarrassment of an email attachment being rejected by a colleague's antivirus program. There are many viruses; however, with sinister payloads: some actively destroy files, some overwrite the boot sectors on hard drives to render computers unbootable (unable to load the operating system), alter the directory information so that files cannot be accessed and an increasing number install backdoor programs that allow virus writers to take control of computers remotely. Computers with backdoor software installed are called 'zombies' and are often used for computer crime. Virus-infected files are often transmitted through peer-to-peer network connections.

Spyware is any software that covertly gathers information about a user through an internet connection without the user's knowledge or approval. Spyware is often bundled as a hidden component of free software, but can also be transferred with some software updates. Spyware monitors the user's activity on the internet and transmits that information to a third-party. The intention of spyware is to gather information about email addresses, banking details and credit card numbers.

A number of software developers offer their products as freeware (or adware) until you pay to register. Prior to registering, you may be swamped with reminders and upgrade offers. Other adware products can be embedded in browser software as an 'Add-on Extension'. Cleaning out the 'Add-on' manager regularly will reduce these annoying messages.

A **worm** copies itself repeatedly in memory or over a network, using up system resources and possibly shutting the system down. A computer virus embeds itself in some other executable software, including the operating system, on the target system. When this program is run the virus spreads to other executable files. A worm, on the other hand, actively transmits itself over the network to infect other computers. This suggests that a virus needs the operating system or user to initiate the infection, while a worm will spread by itself.

Trojans are programs that pretend to be one thing, but in reality are performing a quite different and malicious function. For example, a Trojan can pretend to be a game, while in the background it is collecting email addresses stored on your computer and sending them to spammers. Worms and Trojans most commonly use vulnerabilities in email programs to distribute themselves widely and quickly. **Logic bombs** are programs written to do something unexpected, such as deleting all your files, at a triggered event such as a date.

Keylogger software is a type of Trojan that is designed to secretly monitor and log all keystrokes. A number of keylogger programs are used for legitimate purposes, such as parental control that allows parents to track the websites accessed by their children, organisations tracking the use of computers for non-work related activities or businesses tracking keywords and phrases associated with critical commercial information that would be damaging in the wrong hands. While these uses may be legitimate, the majority of keylogger activities are related to stealing people's online payment system data.

Phishing involves sending an email to a user falsely claiming to be an established enterprise in an attempt to scam the user into surrendering private information that will be used for identity theft. Phishing emails typically purport to come from organisations such as banks. Normally, the phishing email will request the user to access a website that will closely resemble a legitimate site. The user will be asked to provide their login details or update their personal information. These are details already recorded by the legitimate organisation, but can be used by fraudsters to plunder the user's bank accounts.

Anti-virus and anti-spyware programs can be installed on computers to guard against malicious attacks. These programs need to be updated regularly to ensure they can recognise the profile of the latest threats.

THINK ABOUT COMPUTING 3.12
Adware, spyware and keyloggers are all considered to be malware. Are they all malicious?

Event-based threats

Event-based threats do not involve accidental or deliberate actions of a human. One example of an event-based threat would be a power surge, which occurs when the incoming electrical power increases more than 5 per cent above the normal system voltage (240 volts). A momentary surge, called a spike, can be caused by a lightning bolt striking power lines. It can result in immediate and permanent damage to a computer or network. Networks can be protected against a power surge with the use of a surge protector. A surge protector limits the voltage supplied to the computer by shorting any excess caused by a spike to ground.

A hard drive crash occurs when the hard drive malfunctions and stored data cannot be retrieved using normal procedures. The cause of a crash can be an impact that forces the read-write head of the device to scratch the disk surface, a magnetic field causing interference or some contamination such as dust or water. Special software can be used to try and retrieve lost files, but success is not guaranteed.

Measures to secure networks

An unprotected network is vulnerable to attack and this can result in serious consequences, such as corruption of files and loss of data. A number of measures can be used to minimise the chances of a security breach, such as usernames and passwords, firewalls, use of wireless security protocols and UPS devices.

Usernames and passwords

Most network operating systems require that you correctly enter a **username** and a **password** before you can access the data, information and programs stored on a computer or network. Some systems assign your username, or user identification (ID). For example, a school may use your student identification number as your user ID. With other systems, you select your own. Many users select a mixture of their first and last names. For instance, a user named Michael Roland might choose 'mroland' as his username.

A two-phase process to control access to a network involves 'identification' to verify that you are a valid user, and 'authentication' to verify you are who you claim to be. A username and password are two methods used to identify and authenticate a user's access rights.

For passwords, most systems require you to select your own. Users typically choose an easy-to-remember word or series of characters for passwords. If your password is too obvious, however, such as your initials or birthday, others can guess it easily. Easy passwords make it simple for hackers to break into a system, so you should select a password carefully.

Longer passwords provide greater security than shorter ones. Each character you add to a password significantly increases the number of possible combinations and the length of time it might take for someone to guess the password. Simply speaking, the more creative you are when selecting a password, the more difficult it is for someone to figure out. Table 3.11 shows the effect of increasing the length of a password that consists of letters and numbers. The longer the password, the more effort required to discover it. Long passwords are more difficult for users to remember, however.

The most common word used as a password is the word 'password'. Some networks require users to change their password each month. A log of the user's previous passwords can be checked to ensure that the password selected has not been used before. This can be an inconvenience to the user, but the increased security can be worthwhile.

Many software programs have guidelines that you must follow when you create your password. Many systems require your password to be at least eight characters long and use a mixture of numbers, upper and lower case letters and special characters.

TABLE 3.11 Password protection

Password		Average time to discover[b]	
Number of characters[a]	Possible combinations[b]	Human[c]	Computer[d]
1	36	3 minutes	0.000 018 second
2	1 300	2 hours	0.000 65 second
3	47 000	3 days	0.02 second
4	1 000 000	3 months	1 second
5	60 000 000	10 years	30 seconds
10	3 700 000 000 000 000	580 million years	59 years

[a] Possible characters include the letters A–Z and numbers 0–9.

[b] Average time assumes the password would be discovered in approximately half the time it would take to try all possible combinations.

[c] Human discovery assumes one try every 10 seconds.

[d] Computer discovery assumes 1 million tries per second.

Banks and other financial institutions concerned about unauthorised transfers of money from online accounts are now using a secure SMS system to validate the transfer. Users are still required to log in to the institution using their login and password; however, if they wish to transfer money to an account at another bank or to an organisation, the institution sends them a randomly generated six-digit code via SMS to their previously registered mobile phone number. The code is completely unique to that transaction and expires within a set time limit, usually five minutes. The user must read the code on their mobile phone and then enter that code on the website to validate the transfer.

THINK ABOUT COMPUTING 3.13

Many websites require passwords. Can a website be secure? Why or why not? Should a website limit the number of password entry attempts to three?

The purpose of a firewall is to securely separate the internet and other external networks from the internal LAN. The firewall examines the content of incoming packets and determines whether they should be allowed to pass through to the LAN. The firewall is usually configured so that the internet connection enters on a separate network interface card, giving total control over the routing of external packets.

THINK ABOUT COMPUTING 3.14

There are many ways that hackers and others with misguided intent may try to access or abuse a network. Use the internet to find out about email bombs, denial of service (website) and spam. What are they? Can they be avoided? How?

Firewalls

As the internet is a public network, anyone with the proper connection can access it. In contrast, a private corporate intranet restricts access to specific authorised users, usually employees, suppliers, vendors and customers. To prevent unauthorised access to data and information, companies protect their intranet with a **firewall** (Figure 3.38). One use of firewalls is to deny network access to outsiders.

To implement a firewall, many companies route all communications through a proxy server. A proxy server is a server outside the company's network that controls which communications pass into the company's network. That is, the firewall carefully screens all incoming and outgoing messages.

A server on a LAN can use one of the 1024 ports available to allow its services to be accessed by external users over the internet. Typical port numbers assigned to network services include:

- 21 – file transfer protocol used for uploading and downloading files (FTP)
- 25 – email (SMTP)
- 80 – web server (http)
- 443 – web server (https).

A firewall can be set to block any or all of these ports to restrict access to outsiders. By blocking the incoming ports, external users cannot use that port to hack into the local network. Holes are opened through the firewall by unblocking a port. This is done to allow legitimate access to the LAN, such as permitting external users to access the web server.

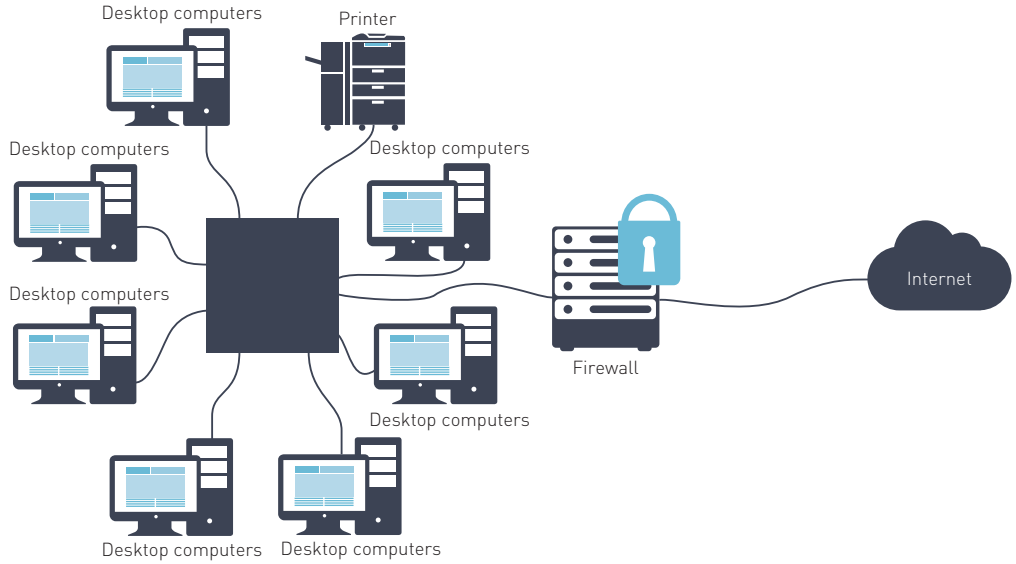


FIGURE 3.38 One use of a firewall is to restrict outsiders from accessing data and information on a network.

Uninterruptible power supplies

Power loss can be caused by storm damage, restricted supply imposed by the power company or heavy demand in the area. An uninterruptible power supply (UPS) provides about 10 minutes of reserve power, which provides sufficient time for the network administrator to shut down the network in an orderly way so that there is no resulting loss of data. To provide the backup power source, the UPS converts its 12-volt DC stored battery charge to a 240-volt AC supply. Most power blackouts last for less than one minute so the UPS can keep the system operating without the need to shut down. Most UPS devices also protect against power surges.

UPS devices come in a range of capacities, with those targeting the home network market operating at 1200 VA or 2200 VA. A volt-ampere (VA) is the unit used for the power in an electronic circuit (the root mean square voltage times the root mean square current).

As a rough guide, a desktop computer requires 200 VA, a 15-inch monitor requires 50 VA, a 21-inch monitor requires 200 VA, a router needs 50 VA and an external hard drive needs 100 VA. So a typical home network might need protection for, say, 550 VA. A UPS rated at 2200 VA will give longer back-up time and extra capacity if more devices are added to the network in the future.

Wireless security

Wireless communications technology has enabled billions of homes and businesses that use notebook computers, tablets and other mobile devices to communicate within a LAN or globally with relative ease. Although wireless provides many conveniences to users, it also poses additional security risks.

A common technique to locate a vulnerable wi-fi network is called 'wardriving'. In this technique, a perpetrator attempts to connect to wireless networks through their notebook computer while driving through areas they suspect might provide easy access.

To avoid unauthorised network access, the wi-fi network should include a firewall and ensure equipment uses one or more wireless security standards. Recommended security strategies include the following.

- The wireless access point should be configured so that it does not broadcast its network name (the service set identifier, or SSID, which is a 32-bit alphanumeric password for the wireless LAN). It should also allow access only to specified devices.
- Use **wi-fi protected access (WPA or WPA2)**, which is a standard that defines how to encrypt data as it travels across wireless networks. Encryption technology scrambles messages sent over the wireless network so that they cannot be easily read. All devices on the wireless network must use the same encryption settings. WPA and WPA2 establish a passphrase that is used to check the identity of all devices on the wireless network.
- Use wi-fi protected setup (WPS) with a push-button connection to attach devices to the wireless network.

The alternative to a push-button WPS is to use a PIN. This is less secure, since many routers do not time-out after an incorrect WPS PIN is used. This leaves the router vulnerable to attack.

Network physical designs

The **physical design** of a network takes into account the hardware and software needed to provide the solution. One way to represent the physical design of a network is to draw a network diagram.

Network diagrams

Networks in medium-to-large organisations can become very complex, with servers, workstations, printers and wireless access points spread widely throughout the premises. Technical support staff need a method of representing both the network and all of its different pathways to provide an overview of the connections and to allow them to identify and locate equipment.

Network diagrams use lines to represent cables, and icons to represent communications devices. Figure 3.39 shows a network diagram for OzDVD, an internet-based DVD sales business. You should be able to identify a number of servers, desktop computers, printers, routers and switches.

The firewall restricts the traffic coming over the internet from directly accessing the internal servers. The web server would be used to hold the static items of the OzDVD webpage.

THINK ABOUT COMPUTING 3.15

To increase the security of a wireless network a number of other options, other than turning on WPA2, should be considered. Investigate the strategies that follow.

- MAC address filtering
- Assignment of static (rather than dynamic) IP addresses to devices on the network

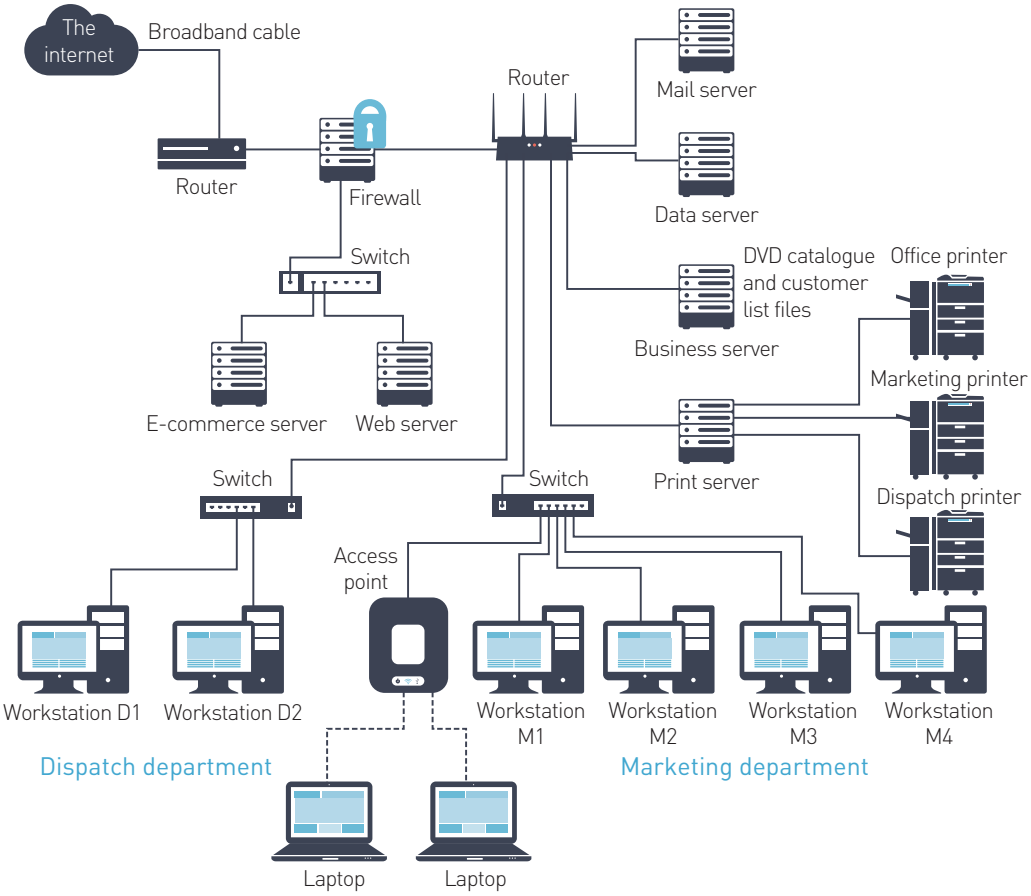


FIGURE 3.39 A network diagram for OzDVD's sales system, showing the communications devices and transmission media used in the network.

The e-commerce server provides the interaction with a customer. It would house the dynamic webpage elements, such as the shopping trolley, checkout facility, DVD-cover graphics and so on. The internal business server contains the catalogue and customer database files. Rules within the firewall allow the business server and the e-commerce server to communicate with each other.

Note that Figure 3.39 is not based on the plans of the building. This would make the diagram cumbersome and difficult to read. The physical buildings are not important in a network diagram, but it is necessary to identify work areas. The network diagram identifies the marketing department and the dispatch department work areas.

The VCE Unit 1 Computing course does not require the students to know the specifics of the laws that cover the legal obligations of network administrators and users with respect to the communication and storage of data and information. The relevant acts are listed below:
The Copyright Act 1968
The Privacy Act 1988
The Information Privacy Act 2000
Health Records Act 2001.

Legal and ethical responsibilities

Network professionals and users of networks have legal and ethical responsibilities with respect to social protocols and the ownership of data and information. Social protocols are a set of rules or behaviours that apply when people use online environments. The protocols cover appropriate behaviour when communicating online, the ownership of intellectual property, the application of digital information security practices and the use of personal security strategies.

Legal responsibilities

Legal responsibilities include those actions to which a person can be held accountable under law. Breaching a legal responsibility has the implication of litigation or facing charges in a court of law. The first legal area to consider relating to the communication and storage of information is ensuring that copyright laws are not infringed. For example, software used on the network should be legally purchased. Text, video and image files loaded on websites, or in files available on the network, should be original works or permission should have been given by the legal owner.

Another legal requirement is to ensure that privacy laws have been complied with. This involves obtaining permission from people for their photos to be used on a website or in documents stored on the network, checking that personal details are not disclosed and information is not available that allows a third party to identify individuals.

Information systems professionals also have a legal responsibility to ensure that data stored on a network or communicated between nodes cannot be accessed by unauthorised users.

The use of cloud computing presents issues for many network managers and users. Moving data into the cloud means the data will move out of the direct control of an organisation or user. That data may be processed and stored outside of Australia. Users need to be aware of their privacy and data security obligations when transferring personal information into any cloud environment.

Ethical responsibilities

Ethical responsibilities relate to following the correct moral path. Not adhering to ethical standards in an online environment has the consequences of a loss of respect, loss of customers and criticism from aggrieved users. Examples are ensuring that sexually explicit material is not stored or accessible; defamatory comments are not posted in email, in messages or on websites; that communications between users cannot be intercepted; and that metadata from tweets and other social media should not allow individuals to be identified.

Resolving legal, ethical and social tensions

Tensions can arise in a workplace if acceptable work practices are not clearly defined. Usually an organisation will adopt a series of policies that outline what workers should or should not do, and how to respond to particular events. If the policies lack clarity, or do not exist, issues can develop that need to be resolved.

There are six steps that an organisation can use to solve a legal, ethical or social tension.

1 Identify the problem: What decision has to be made and what facts are required?
2 Identify the stakeholders: Who are they? What interests do they have? Who is the key player?
3 Identify possible alternatives: What options are available? What are the likely consequences?
4 Identify ethical standards: Are there any applicable laws? Are there any morals or standards that could be applied? Is there a precedent?
5 Evaluate options: Identify strengths and weaknesses. Identify the option that causes least harm. Can the decision be reversed?
6 Make a decision: Select the preferred option. Justify the option. Notify all stakeholders of the decision.

Security practices

Applying appropriate security practices to protect information, particularly sensitive data such as personal details, forms part of the social protocols that users of networks should follow. Network professionals should ensure that networks are safe from accidental, deliberate and event-based threats, as discussed earlier in this chapter. This should include measures to protect the communication and storage of data and information, such as the use of firewalls

and security protocols that apply encryption techniques. Users of networks and members of online sites should be required to use logins and passwords to access a network. Websites that require users to input data online should incorporate a test to establish that the respondent is human rather than a machine.

Personal security strategies

Users of networked information systems should apply safe practices while participating in online environments. This includes users checking their default privacy settings to ensure maximum protection of personal details. Using online filtering techniques to restrict the content that can be communicated over a network is another form of personal security. Filtering can be applied by an employer to personnel within the organisation, by a school to its students and by parents to their children.

Users of social media should activate privacy settings in their accounts to avoid divulging personal data such as photographs, addresses and names.

Responsibilities of network users

Networks reach across societies that have different values and traditions. People using networks have capacities that allow them to do things they could not do before – and do so with anonymity. The norms of society and values can be challenged by the character of human interaction in electronic networks.

Users of networks, including those using social networking sites, must comply with copyright laws and behave ethically. Internet etiquette, also known as netiquette, is a set of guidelines on how users should behave when communicating online. These guidelines provide a set of social online protocols and include:

- avoiding the use of bad language and not saying things to make other people feel bad
- not typing emails in upper case (since it looks like you are shouting)
- not using emoticons in formal emails
- when forwarding an email, removing all personal information relating to the original sender, including their email address
- obeying the rules of online discussion forums
- deciding whether to use your real name or not (using a handle can protect your personal information)
- avoiding running malicious code on a network by not opening emails from unknown sources or opening files that may contain malware
- not making defamatory or discriminatory comments on social media
- not posting text, images, videos or files which infringe on intellectual property rights
- ensuring any sources used or quoted are reliable and authentic
- not uploading or downloading sexually explicit content
- respecting other people's privacy.

Rather than loading material that may be subject to copyright, it is better practice to include links to the website that contains the original source. Any material used should be appropriately cited.

Remember that it is often not possible to remove a comment made on a social media site. Think of the consequences before you post.

Benefits and risks associated with using a network

The use of networks has become widespread as the technology has become more widely available and the use of the internet and social media has become a seamless part of our lives. There are many clear benefits in using a network, from the sharing of hardware and software to accessing the internet. There are also a number of risks associated with the use of networks, particularly within the context of a global environment. In this section, we look at the benefits to individuals and organisations in using a network, then the risks that may be encountered.

Benefits of using a network

Establishing a network has a number of benefits over running standalone computers and resources, including access to peripherals (for example, printers), lower set-up costs (terminals are cheaper than standalone computers) and the speed of communications. Further advantages of using a network are provided below.

Facilitating communications

Using a network, people can communicate efficiently and easily via email, Facebook, instant messaging, chat rooms, VoIP, wireless messaging services and videoconferencing. Sometimes these communications occur within a business's network; at other times they occur globally through the internet.

THINK ABOUT COMPUTING 3.16
List the hardware that is shared on your school's network.

Sharing hardware

Each networked computer can access and use hardware on the network. Suppose several personal computers on a network each require the use of a laser printer. If the personal computers and a laser printer are connected to a network, the personal computer users can each access the laser printer on the network when they need it. Businesses and home users network their hardware for one main reason – it may be too costly to provide each user with the same piece of hardware, such as a printer.

Facebook is an online social network established in 2004. It was originally available only to college students in the United States.

Sharing data and information

In a networked environment, any authorised computer user can access data and information stored on other computers in the network. For example, a large company might have a database of customer information. Any authorised person, including a mobile user using a smartphone to connect to the network, can access this database. The capability of providing access to and storage of data and information on shared storage devices is an important feature of many networks. Project teams can share data, even if they are geographically remote, by using an organisation's virtual private network (VPN), which uses the internet to make global connections. Networks support collaborative work practices through services such as cloud computing, email and file transfer.

Virtual private networks were discussed on page 86.

Sharing software

Users connected to a network can access software (programs) on the network. To support multiple-user access of software, most software vendors sell network versions of their software. In this case, software vendors issue a site licence. A network licence is a legal agreement that allows multiple users to run the software package simultaneously. The site licence fee is usually based on the number of users or the number of computers attached to the network. Sharing software via a network usually costs less than buying individual copies of the software package for each computer.

THINK ABOUT COMPUTING 3.17
What are the advantages and disadvantages of using online programs such as Google Docs and Office 365?

Transferring funds

Electronic funds transfer (EFT) allows users connected to the internet (an example of a wide area network) to transfer money from one bank account to another via transmission media. Consumers can use credit cards or an online payment system like PayPal to make purchases over the internet. Businesses can use the internet to deposit their employees' salaries directly into their bank accounts. Both businesses and consumers pay bills online, which involves instructing their bank to use EFT payment to pay creditors. Global networks are a boon to online retailers, who can effectively trade 24/7.

Risks associated with using a network

The risks associated with using a network relate to inconvenience caused by any fault in the network devices and damage or loss of sensitive data caused by breaches of security.

Breaches of security

Networks with inadequate security systems are liable to be attacked by malware or hackers. These threats can result in valuable information being accessed, stolen, damaged or deliberately altered for fraudulent purposes. Hackers have been able to access customer credit card details by finding an opening in a commercial corporation's network. Viruses and other malware can deliberately sabotage the operation of computers and software.

Wireless networks without suitable encryption security run the risk of outsiders eavesdropping on messages or accessing important files transmitted between users.

User dependence

Network users rely on a network to operate correctly to be able to access files, applications and resources. If a component, such as a file server, develops a fault, users will not be able to run applications or access shared data. This would limit the effectiveness of a worker and impact on the productivity of the business.

Social networks

The popularity of social networking sites such as Facebook and Twitter has increased rapidly in recent years. There are a number of negative effects that overuse of social networking can have on users. These include:

- distracting users from more purposeful tasks such as studying
- reduced learning and research capability as students rely on information easily accessible from social network sites but often unverified
- reduction in traditional communication between people, such as face-to-face conversations
- adoption of poor language skills and underdevelopment of creative writing skills
- negative impact on health due to skipping meals and not participating in physical activity
- reliance on virtual world experience rather than from the real world.

Social media are now important tools for businesses to market their services and keep in touch with their customer base. Customers expect organisations they do business with will be contactable via sites such as Twitter, Facebook and LinkedIn. Social media risks faced by organisations come under three areas: operational, regulatory and reputational.

- Operational risk relates to employees posting material in breach of copyright; the monitoring of employees on social media; ownership of material posted.
- Regulatory risk involves company disclosures – market sensitive data revealed on social media before a public announcement.
- Reputational risk includes what the business or employees say online, or what customers may say about the company.

CHAPTER SUMMARY

ESSENTIAL TERMS

802.11 standard the standard that specifies how two wireless computers or devices communicate via radio waves with each other

app a self-contained program installed on a mobile device that is designed to fulfil a particular purpose; some apps are provided as a standard feature on the mobile device while others can be downloaded from commercial providers

asymmetric digital subscriber line (ADSL) a digital line alternative for the small business or home user. ADSL transmits on existing standard copper telephone wiring. The ADSL2+ technology allows a faster transfer rate than the older ADSL rate – 20 Mbps compared with 8 Mbps.

bandwidth the width of the communications channel measured in bits per second

Bluetooth a protocol that uses short-range radio waves to transmit data along enabled devices, such as notebook computers, mobile telephones and printers

broadband router a basic router that connects a LAN to the internet, also functioning as a switch, a firewall and a wireless access point

broadcast radio a wireless transmission medium that distributes radio signals through the air over long distances (such as between cities, regions and countries) and short distances (such as within an office or home)

cellular radio a form of broadcast radio that is used widely for mobile communications

channel the communications path between two devices

client–server network a network in which one or more computers act as a server (host computer) and the other computers on the network (clients) can request services from the server. A server controls access to the hardware and software on the network, and provides a centralised storage area for programs, data and information.

cloud storage an off-site storage system maintained by a third party and accessed through the internet

communications device any type of hardware capable of transmitting data, instructions or information between a sending device and a receiving device

communications satellites communications facilities that receive microwave signals from Earth, amplify the signals, and retransmit them back to Earth

communications software an application or program designed to pass or support the movement of information over a network

dedicated server a server that performs a specific task, such as file servers, print servers, database servers and network servers

digital signals individual electrical pulses that represent the bits that are grouped together to form characters

domain name server (DNS) residing with the ISP used by the client, the DNS is used by the internet to store domain names and their corresponding IP addresses; when a browser requests that a page be downloaded, the DNS translates the domain name into its associated IP address

downlink a transmission from the satellite to a receiving Earth station

Earth-based stations communications facilities that use large, dish-shaped antennas to transmit and receive data from satellites

Ethernet a standard communications protocol embedded in software and hardware devices that allows computers to operate a LAN; it was developed in 1973 by American electrical engineer Bob Metcalfe, and has become the standard model for LANs worldwide

fibre-optic cable smooth, hair-thin strands of glass or plastic (**optical fibres**) that conduct light with high efficiency; fibre optics are frequently used in new voice and data installations

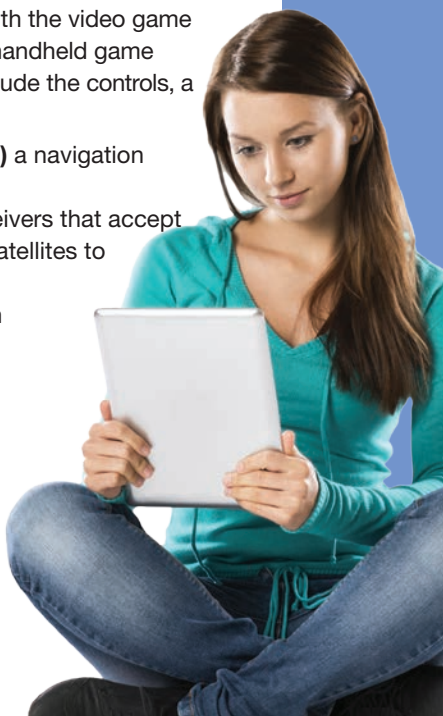
firewall hardware and software that restrict access to data and information on a network

game console an input device with the video game screened on a television set; handheld game consoles are portable and include the controls, a small screen and sound output

global positioning system (GPS) a navigation system that consists of one or more Earth-based receivers that accept and analyse signals sent by satellites to determine the receiver's geographic location

hacker a person who breaks into a computer for profit, from the motivation of a challenge, or to view restricted data

hypertext mark-up language (HTML) a set of special codes that format a file for



CHAPTER SUMMARY

use as a webpage; these codes, or tags, specify how the text and other elements display in a browser and where the links lead

hypertext transfer protocol (http) a set of rules that defines how pages are transferred on the internet

hypertext transfer protocol secured (https) a communications protocol for secure transmissions over the internet; the https system provides authentication and encryption communication and is widely used for security-sensitive processing, such as payment transactions and connections to banks

internet peer-to-peer network an internet network which enables users with the same networking software to connect to one another's hard disk drives and exchange files directly

internet service software web browsers, electronic mail, Voice over internet Protocol (VoIP) software and cloud storage

intranet an internal network within an organisation that uses internet and web technologies

keylogger a type of Trojan designed to secretly monitor and record all keystrokes entered on a keyboard; some are legitimate (such as parental-control programs to allow parents to track their children's internet usage), but others may be used for spying or stealing data

leecher a person downloading a file using a torrent protocol during a P2P session who does not allow their computer to pass pieces of the file to other users

local area network (LAN) a group of neighbouring computers that can share information and resources using a network (such a network is usually within a building or several neighbouring dwellings, but can be up to a few kilometres); an organisation owns the infrastructure used by the network, including the cabling

logic bomb a program that lies dormant until a specific piece of program logic is activated. In this way, it is very analogous to a real-world land mine. The most common activator for a logic bomb is a date. The logic bomb checks the system date and does nothing until a pre-programmed date and time is reached. At that point, the logic bomb activates and executes its code.

malware programs designed to infiltrate and cause harm to a computer or network without the owner's knowledge or consent, such as viruses, worms, Trojans, adware, spyware, logic bombs and keyloggers; the term is short for 'malicious software'.

Mbps short for megabits per second

microwaves radio waves that can be used to provide high-speed transmission of both voice and data; data is transmitted through the air from one microwave station to another in a manner similar to the way radio signals are transmitted

microwave station an Earth-based reflective dish that contains the antenna, transceivers and other equipment necessary for microwave communications

mobile devices compact, lightweight, easily carried and often high-speed wireless broadband enabled devices, such as games consoles

mobile phone a telephone device that uses radio signals to transmit voice and digital data messages

near field communication (NFC) contactless communication between portable devices; a smartphone can be configured to read an NFC tag on a poster or menu to download relevant information (for example, Paywave transactions and the myki transportation system use NFC technology)

network collection of computers and devices connected together via communications devices and wired or wireless transmission media, allowing computers to share resources

network administrator the person who oversees the operations of a client–server network

network analysis tools allow network administrators to monitor devices, check protocols and port activity, view event logs and analyse traffic

network architecture the design of the network; for example, include client–server and peer-to-peer networks

network-attached storage (NAS) a storage device often used to hold video, photo and audio files on a network; a typical NAS has capacity to store 8 TB of data and can be configured as an FTP, web, email and print server

network diagram a schematic method of showing the physical devices and communications lines present in a network

network interface card (NIC) a communications devices that fits in an expansion slot of a computer and enables a computer that does not have in-built network capability to communicate with a network

network operating system (network OS, NOS) software that organises, controls and coordinates the administration, file management, printer management and security activities on a LAN

network protocol rules and conventions for communication between network devices; protocols for networks generally use packet-switching techniques to send and receive messages in the form of packets

network standard guidelines that specify the way computers access the medium to which they are attached, the types of media used, the speed at which data flows and the physical technology used. Examples of network standards are Ethernet, TCP/IP and 802.11

networking software computer programs that establish a connection to another computer or network, and manage the transmission of data, instructions and information

node a network connection point, including desktop or mobile computer, peripheral such as printer or scanner, or portable device such as a smartphone. A node normally is assigned its own IP address.

optical fibre a strand of glass or plastic, as thin as a human hair, that uses light to transmit signals

packet switching breaking a message into packets, sending the packets over a network pathway, and then reassembling the data

packets the small pieces into which data is broken before being transmitted over a network; they will be transmitted independently over the transmission media and, once all the packets have arrived at the receiving computer, are put back together to produce the original, complete message

password a secret combination of characters associated with the username that allows access to certain computer resources

peer someone with equal communication access rights in a network. In relation to torrent downloads, peers are computers on the internet that are downloading a particular file at the same time.

peer-to-peer network (P2P) an internet network that enables users to connect to each other's hard disks and exchange files directly, such as BitTorrent

physical design the communications devices, transmission media and software of a network

physical transmission media media that use wire, cable or fibre-optics to send communications signals

phishing sending an email to a user falsely, claiming to be an established enterprise of some kind, in an attempt to scam the user into giving up private information that will be used for identify theft

protocol a set of rules and procedures for exchanging information between two computers, such as the network transmission control protocol/internet protocol (TCP/IP)

receiving device accepts the data, instructions or information

router an intelligent network-connecting device that can route communications traffic directly to the appropriate network

seed a computer on the internet that has a complete file available for downloading using the torrent protocol

sending device initiates the transmission of data, instructions or information

smartphone an internet-enabled mobile phone that typically includes an address book, calendar and a calculator; a businessperson would require a smartphone that is capable of sending and receiving emails from the company's mail server, as well as opening and editing business applications, such as word processing, spreadsheets, presentations and PDF files, and a web browser.

spyware any software that covertly gathers information about a user through an internet connection without the user's knowledge or approval

switch a device that provides a common connection point for nodes on a network; it enables a packet to reach its destination faster by storing addresses in memory and using logic to direct the transmission

tablet a special type of notebook or laptop computer that resembles a letter-sized slate, which allows a user to write on the screen using a digital pen

TCP/IP (transmission control protocol/internet protocol) a network standard that manages the transmission of data by breaking it up into packets and transmitting the packets over the internet

torrent a small file that holds metadata relating to a file that can be shared on a P2P network; the metadata includes the file's name, size and its location

transmission media the materials or technologies that are used to establish the communications channel; two types of transmission media are physical transmission media, which use some type of physical cabling (twisted-pair wire, coaxial cable and fibre-optic cable); and wireless transmission media, which use wireless technology (microwaves, radio waves or light waves)

CHAPTER SUMMARY

Trojan programs that pretend to be one thing, but that are actually performing another, malicious background function; for example, a Trojan that appears to be a game that is actually collecting email addresses stored on your computer and sending them back to spammers

twisted-pair cable a cable made up of **twisted-pair wires;** twisted-pair cable systems are graded according to categories that describe the quality of the components and the installation technique. The most common twisted-pair cable is graded as category 5, CAT 5e or CAT 6.

twisted-pair wire pairs of copper wires that are twisted together, commonly used for telephone lines and to connect personal computers with one another

uplink a transmission to the satellite

username a unique combination of characters, such as letters of the alphabet or numbers, that identifies one specific user

virtual private network (VPN) a network in an organisation or business that uses the internet to link remote sites and users

virus a computer program that can destroy files and alter the performance of the operating system; once a virus is in a computer, it can spread over a network to other connected computers

Voice over Internet Protocol (VoIP) a high-speed internet connection that allows users to communicate as if they were on a conventional telephone

wearable technology devices that use wireless signals and can be worn by the user, either as clothing or an accessory, such as smart watches and devices that track information relating to health and fitness

web browser an application software package that allows users to access and view webpages

web-enabled device a digital device that provides access to the internet for hand-held devices, such as smartphones

web server a computer that delivers requested webpages to another computer

wide area network (WAN) a network that is geographic in scope (as opposed to local) and uses telephone lines, microwaves, satellites or a combination of communications channels

wi-fi any network based on the 802.11 standard

wi-fi communications a wireless transmission medium that distributes radio signals through the air over long distances

wi-fi protected access (WPA or WPA2) a standard that defines how to encrypt data as it travels across wireless networks

wireless access point a central communications device that allows computers and other mobile devices to transmit data among themselves wirelessly using radio waves

wireless adaptor a device used to connect computers and mobile devices to a wireless network; the adaptor maybe in-built or connected via a USB port on the computer

wireless broadband router a device that combines the functions of a basic router (connecting the LAN to the internet), a switch (for devices, such as a desktop computer, connected by cable), a firewall (security measure) and a wireless access point (to allow wireless connectivity)

wireless extender a device that increases the range of a wireless access point by re-broadcasting the signal it receives, which allows home networks to avoid obstacles and reach further than normal; a dual-band uses both the 2.4 GHz and 5 GHz radio frequencies, so one frequency can be used to communicate with a router while the other frequency can be used to communicate with clients. The device must provide simultaneous transmissions rather than the option of 2.4 GHz or 5 GHz.

wireless transmission media media that send communications signals through the air or space using radio, microwave and infra-red signals

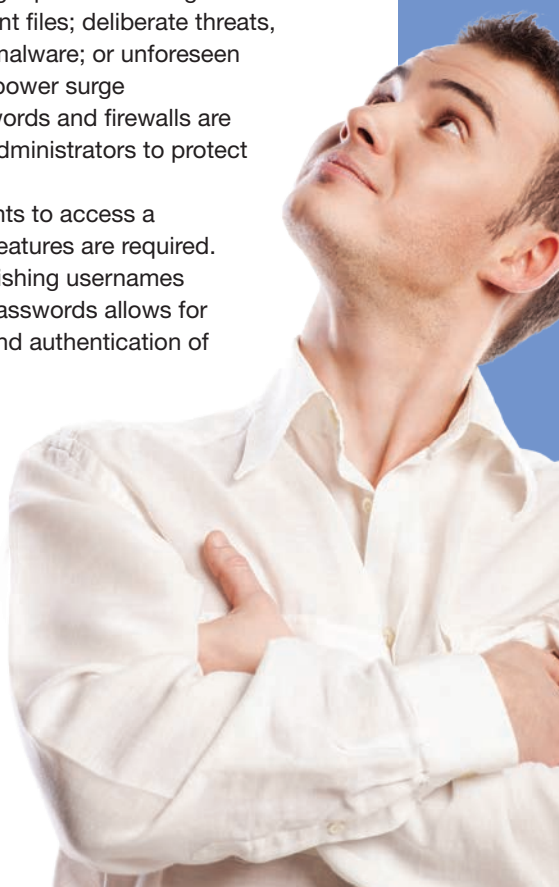
worm a virus that copies itself repeatedly in memory or over a network, using up a system's resources and possibly shutting it down; while a virus will embed in executable software on a computer, a worm will actively transmit over a network to infect other computers

CHAPTER SUMMARY

IMPORTANT FACTS

1 Networks allow users to share resources such as data, information, hardware and software, and to transfer money via electronic funds transfer.
2 To protect wireless networks, security measures are recommended, such as configuring the wireless access point so that it does not broadcast a network name and only allowing access to specified devices. Data should also be encrypted before it travels across a wireless network.
3 Network professionals and users have legal and ethical responsibilities with respect to social protocols and the ownership of data and information. These responsibilities relate to copyright, privacy, socially appropriate material, malware, backup strategies and network access restrictions.
4 Networks allow users to share hardware, software, data and information, to transfer funds globally, attract new customers, entertain and acquire knowledge.
5 The risks associated with using a network in a global environment include breaches in security that allow malware or a hacker to access, steal or damage files and information, users' dependence, and the negative effects of overuse of social media.
6 Businesses need to be aware of the risks associated with communicating on social media. The risks include employees posting information in breach of copyright, issues relating to the monitoring of employee communications, sensitive market information being leaked before public announcements and employees or customers posting negative comments.
7 Software installed on servers, referred to as server software, allows them to provide email services, internet connectivity, file management and print services.
8 The National Broadband Network (NBN) requires a utility box to connect to the fibre-optic cable being rolled out across Australia and a connection box to be installed in the home. The connection box is used for phone connection and to link with the home's network router.
9 The range for a wi-fi network is about 50 metres inside a building.
10 A smart TV can function as a conventional television set as well as link to the internet to screen interactive media.
11 Mobile devices that can be connected to a network include:
   - smartphones
   - hand-held computers
   - netbooks
   - navigation systems
   - digital cameras.

   A convergence of technologies has resulted in one type of device also performing the functions of a different type of device. For example, a smartphone may be capable of running applications software, accessing the internet, playing media and taking digital images beside its main function of voice communications.
12 Video games use a console to input commands, a screen for output and a hard-drive for storage. Popular consoles include XBox One, PlayStation 4 and Wii U.
13 The transmission rate of a communications channel is determined by its bandwidth and its speed. The bandwidth is the range of frequencies that a channel can carry. The speed at which data is transmitted is usually expressed as bits per second (bps), the number of bits that can be transmitted in one second.
14 The National Broadband Network (NBN) uses fibre-optic cable to connect homes and businesses to the internet using connection speeds of 25 Mbps. Fixed wireless and satellite connections are used where cable is impractical.
15 Security threats to networks include accidental loss, such as losing a portable storage device containing important files; deliberate threats, such as installing malware; or unforeseen events, such as a power surge
16 Usernames, passwords and firewalls are used by network administrators to protect networks.
17 To verify users' rights to access a network, security features are required. A system of establishing usernames (or user IDs) and passwords allows for the identification and authentication of each user.



TEST YOUR KNOWLEDGE

Review quiz

COMMUNICATIONS

1 What does the term 'data communications' mean?
2 Identify the devices you use on a day-to-day basis that connect to a network.

NETWORKS

3 Provide an example of a sending device and a receiving device.
4 What is a network? How is a local area network (LAN) different from a wide area network (WAN)?
5 How does a peer-to-peer network differ from a client–server network?
6 Identify the dedicated servers that are used on networks in your school.
7 Describe the type of network a business might use if they have a head office in one city and wish to enable other branches and travelling salespeople to maintain regular contact and access data and files stored on their central servers?
8 What security risks can occur if a user shares files on an internet P2P network?
9 Describe the benefits in downloading a file from a P2P network using torrents rather than from a single source.
10 What do the terms 'peer', 'seed' and 'leech' refer to when referring to torrent downloads?
11 What is an intranet? What benefits to organisations are there in setting up an intranet?
12 Describe the advantages of setting up a home network.
13 There are four types of home network in common use. Which type would you choose to set up a network in your home? Why would you choose that type?

COMMUNICATIONS DEVICES

14 Routers and switches are used to connect separate networks. Explain the circumstances in which each of these would be used.
15 What is the function of a wireless broadband router in a home network?
16 Identify the components that need to be installed in a home or business to enable a LAN to connect to the NBN.
17 What is the intended purpose of the two voice ports on the NBN connection box?
18 What is the purpose of a network interface card? What do mobile computers use in place of a network interface card?
19 A wireless access point can be wired to the fibre-optic backbone of a bus network. What is the purpose of the wireless access point?
20 It may be difficult to pick up a consistent wireless signal in a large home with solid brick walls. Where should the access point be located to provide the strongest signal throughout the house? What solution would you recommend to ensure that some of the more distant parts of the house, such as the back patio, receive a signal?
21 Many home networks include a smart TV and a network-attached storage device. Describe the functions of these devices.
22 A particular network-attached storage device has a capacity of 8 Terabytes. How many megabytes is this equivalent to?

COMMUNICATIONS SOFTWARE

23 Describe four important tasks of a network operating system.
24 Identify the four components of a URL address.
25 When should a website use the https:// protocol rather than the standard http:// protocol?
26 List some benefits for an organisation in using a cloud storage system such as Dropbox.

NETWORK COMMUNICATIONS STANDARDS

27 Why do manufacturers of network hardware and software follow established standards?
28 Briefly describe how an Ethernet network transmits data.
29 In what situations would the TCP/IP protocol be most useful?
30 Explain the term packet switching.
31 What network transmission standard would be useful in a situation where an old, heritage-listed building with solid stone walls needs to be networked?
32 What advantages are there for users if a network moves from the 802.11n standard to 802.11ac?

SENDING AND RECEIVING DEVICES

33 A mobile telephone is an example of a wireless device that can be web-enabled. What does the term web-enabled mean?
34 What functions are likely to be available on a smartphone?
35 What does the expression 'convergence of technologies' mean? Give an example in which convergence of technologies is apparent in portable network devices.
36 Provide three examples of wearable technology.

COMMUNICATIONS CHANNEL AND TRANSMISSION MEDIA

37 What advantage do fibre optics have over wire cables?
38 In what circumstances would a network designer consider using wireless transmission media?
39 Briefly describe the broadcast radio and cellular radio wireless transmission media.
40 Describe how near field communication technology allows a user with a smartphone to read information at a museum exhibit.
41 What is the main limitation of microwave transmission?
42 Many companies use satellite transmission to access the internet. Web satellites, however, will provide faster downlink transmissions than uplink transmissions. Why is the difference in speed not of major concern to these companies?

NETWORK SECURITY

43 Describe how a virus, worm, Trojan and keylogger may threaten a computer network.
44 Describe how verifying the identity of a user can protect a network.
45 What is a firewall?
46 Describe three measures that should be incorporated into a wireless network to restrict access by unauthorised users and to secure the transmission of data.

LEGAL AND ETHICAL RESPONSIBILITIES

47 The manager of a website in an organisation has legal responsibilities related to the ownership of material used on the site. Describe these responsibilities.

48 Use the internet to find an example of how researchers and analysts use Twitter or Facebook entries to study human behaviour.

49 What steps should a network user follow in order to behave ethically when using social media?

BENEFITS AND RISKS ASSOCIATED WITH USING NETWORKS

50 Describe the five main benefits of using a network.

51 Why is user dependence considered a possible disadvantage of networks?

52 What problems could eventuate if a network is not well managed?

53 How might the use of social media be considered as a possible risk for businesses?

APPLY YOUR KNOWLEDGE

XYZ ENGINEERING

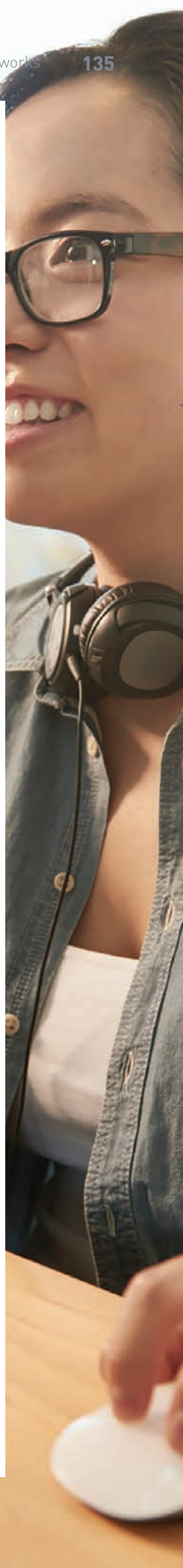
XYZ Engineering has its head office in the northern suburbs of Melbourne. The company designs and develops water systems including channels, culverts and bridges for municipal councils, government authorities and private owners. The head office contains several departments including the civil engineers who design systems for clients, a project team that sub-contracts building works to third parties, an accounts department that bills clients, pays sub-contractors and staff wages, and a small information systems department responsible for establishing and maintaining systems. The company has branch offices in Bendigo, Warrnambool and Bairnsdale. Engineers often need to visit sites to take measurements and make observations. They need to be able to access files stored on the LAN at head office and record their findings.

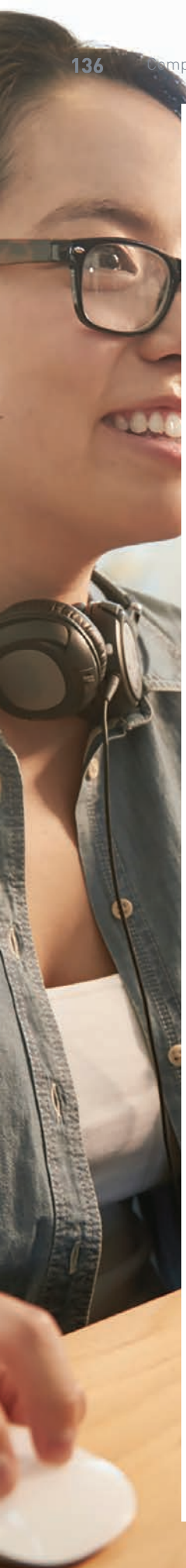
A number of requirements for a network have been identified.

- Desktop computers are used by those staff who work mainly from head office.
- Engineers and project team members who visit sites use notebook computers rather than desktops.
- All departments need access to files including work plans, schedules, job logs, costings and other documents relating to projects.
- Staff need access to printers.
- The company will require a website to promote their business.
- Files that include engineering plans and video clips of existing terrain or proposed developments can be very large and need to be stored centrally.
- Email is the primary means of communication.
- All staff have a company issued smartphone.
- Presentations to clients (involving plans, video clips and costings) need to be made from time-to-time at head office.
- A broadband cable connection to the internet is provided at head office.

Questions

1. What network device is needed to connect to the broadband cable at head office?
2. Employees with portable computers need to be able to use them throughout the head office. What communications technology will these notebooks require? What hardware is necessary to support this technology? What security measures are needed to protect data being communicated using this technology?
3. The desktop computers, printers and other network devices will use wired technology. What type of cabling would be suitable? What devices are needed to link the networks operating in each department?
4. What device would be suitable to store the large media files that the business uses for presentations to clients?
5. The company will operate a client-server network. What servers need to be installed?
6. The rural branches and on-site staff will need to access files from the network. What type of network will link remote sites to a central network via the internet?
7. Describe the legal requirements and ethical responsibilities of the information systems manager with respect to the company's website and files stored on network devices.





8 What risks associated with use of the network does the information systems manager need to guard against? Identify some strategies that will negate these risks.

9 Draw a network diagram of the system at head office. Label all devices and work areas.

10 The company intends to set up a stand at the annual Water Treatment Exhibition. It is hoping that visitors to the exhibition will be able to access information about XYZ Engineering and the services they offer. Rather than providing information in paper form, the company hopes that visitors can download information onto their smartphones. Identify the technology that will be used and describe how it operates.

PREPARING FOR UNIT 1 OUTCOME 2

Design a network with wireless capability that meets an identified need or opportunity, explain its configuration and predict risks and benefits for intended users

For Unit 1, Outcome 2, you are required to design a network for a specific use. The network must have wireless capability, though it may also have some wired components. You must explain its configuration, including identifying network devices and transmission media. You must also predict the risks and benefits for intended users.

OUTCOME MILESTONES

You will be required to:

1. Consider how the information needs of individuals or an organisation could be achieved through the use of a networked information system.
2. Identify the data and information that typically would flow through the information system.
3. Consider how data is to be stored within the network and where it is processed.
4. The network must have wireless capability. Determine whether wired transmissions are also required and identify the communications standard to be used.
5. Determine which network devices are needed.
6. Draw a physical representation of the network.
7. Consider the risks and benefits of using the network for intended users.

STEPS TO FOLLOW

Your teacher may provide a written scenario of a situation in which the implementation of a networked information system would provide benefits to an organisation. Alternatively, you may be asked to identify a system that you have observed and prepare a recommendation for the design of a small network.

1. Identify the information needs of individuals, such as a family, or organisation from the scenario provided or your own observations.
2. Describe the purpose of the proposed networked information system. In most cases, a local area network with wireless connectivity will be required, but users may also need access to the internet. If you are creating a simple network with a small number of users, will you opt for a peer-to-peer network or a client–server network? Why?
3. Identify the data and information that would flow within the information system and the location in which the data would be stored.

4 Identify which portable computers or devices are required to use the network. Determine if a wired connection to some devices is required, or whether a wireless environment on its own is sufficient to meet user's needs?
5 Identify the number of work groups that will operate and determine the number of switches (if any) required. Consider whether a router is needed.
6 Determine whether communications devices are necessary. If there is to be internet access, will you recommend cable, ADSL 2+ or some other form of transmission?
7 Use a software tool to depict the components of the network and its interactions.
8 Explain the function of components and how data and information are transmitted.
9 Predict the risks and benefits for intended users. Risks might include security threats to data and information or user dependence on the network. Benefits might include shared resources, access to the internet and reliable communications via email.

DOCUMENTS REQUIRED FOR ASSESSMENT

1 A brief report that identifies the information needs of the organisation, the purpose of the network and the type of network needed to support the information system
2 A written, oral or visual presentation that outlines your recommendation in terms of the network configuration and a rationale for the decisions made; points to discuss are:
   - LAN, WAN or a combination of both
   - peer-to-peer or client–server LAN
   - number and nature of servers if a client–server network is recommended
   - communications standard(s) to be used
   - communications devices (wireless access points, switches and routers)
   - transmission media (physical and wireless or just wireless)
   - inclusion of a firewall if internet access is recommended
3 A network diagram
4 A report that predicts the risks and benefits for intended users

ASSESSMENT

A set of assessment criteria will be prepared and distributed by your teacher before the start of the task.

CHAPTER 4

ISSUES IN INFORMATION SYSTEMS

Key knowledge

After completing this chapter, you will be able to demonstrate knowledge of:

Interactions and impact

- applications of information systems in a range of settings
- a detailed study in a particular field such as entertainment, agriculture, finance, sport, health, that focuses on:
  - the nature of a contemporary issue associated with the use of information systems
  - legal, social, environmental or ethical reasons for a contentious issue
  - types and capabilities of digital systems associated with the field and issue
  - key stakeholders such as individuals, organisations and governments, and their responsibilities
  - positive and negative opinions of each stakeholder about the issue
- ways in which end-users can express opinions on websites about how information systems are used for particular purposes, such as writing a review in a textbox and a rating system

Data and information

- sources of, and methods and techniques for, acquiring and referencing primary data and secondary data and information
- factors affecting the integrity of data, such as correctness, reasonableness and accuracy

Digital systems

- advantages and disadvantages of using cloud solutions, and using cloud computing for storing, communicating and disposing of data and information
- impact of growth of mobile devices on website design

Approaches to problem solving

- Visualising thinking tools and techniques for supporting reasoning and decision making when analysing issues and ethical dilemmas

For the student

In this chapter, you will learn about a variety of contemporary issues associated with the use of information systems. Techniques and strategies that can be used to investigate information systems issues will also be explained. You are required to collect and analyse data from stakeholders and information from other sources about a specific issue related to the use of information systems. You will present the data you collect on a website you have designed and developed collaboratively. The website should outline the issue, reasons for the issue, key stakeholders involved and their opinions, including how information systems could be used to express opinions about the issue. You will also report your team's point of view on the issue. Both content from this chapter and Chapter 5 will be used to cover all the knowledge required for Unit 1, Outcome 3.

For the teacher

The focus for this chapter, together with Chapter 5, is on preparing students for Unit 1, Outcome 3, in which students design and develop a website, collaboratively. The website presents information related to a contemporary information systems issue in a specific context, which can cause conflict between stakeholders. Students are required to collect data from a variety of sources, then analyse the data to substantiate a point of view.

This chapter discusses a number of contemporary information systems issues, although for the purposes of the Outcome, the information systems issue selected is open ended.

Alamy/Art Directors & TRIP

FIGURE 4.1 Supermarket self-checkout information system

Information systems

There is a significant difference between a digital system and an **information system**. A digital system consists of hardware and software, but on its own a digital system cannot produce any information.

Thus, an information system consists of hardware and software (collectively referred to as a digital system) that also includes the additional components of data, processes and people.

To best illustrate an information system, we will look at the example of a supermarket self-checkout system. It is the additional components that allow the system to produce the output or information required.

TABLE 4.1 Supermarket self-checkout systems are a common sight in Australia today. The information system consists of the following.

Digital systems	Data	Processes	People
• Monitor • Barcode scanner • Weighing machine • Point-of-sale software • Operating system	• Product • Quantity • Price	• Scan item • Make payment • Receive change • Print receipt	• Customer

THINK ABOUT COMPUTING 4.1

An automatic teller machine (ATM) is also an information system. Try to identify digital systems, data, people and processes involved in an ATM system. Make a list of other information systems.

Information systems in action

Shutterstock.com/KPG Ivary

FIGURE 4.2 Conflict between stakeholders

Information systems have provided significant benefits to society over the last 50 years. Systems have been created that allow users to communicate, store and protect data and information and to automate processes that once had to completed manually; for example, writing a letter.

But along with the benefits associated with the use of information systems, issues can also arise from the use of technology that can cause tension conflict between different **stakeholders**.

The following are some examples of contemporary issues associated with the use of information systems.

Information systems and entertainment

The use of information systems for entertainment is widespread. Playing games, watching movies or listening to music are three popular reasons individuals use a digital system. As a result, many websites have been created that support entertainment. A large number of websites are dedicated to reviewing games, movies and music. Some sites allow individuals to play games online or offer online streaming for music and videos, while others facilitate the downloading of files.

For Unit 1, Outcome 3, you will be working collaboratively with other students to design and develop a website that analyses a contemporary issue and presents your team's point of view on that issue.

Issues in entertainment

Illegally downloading movies, music, software and other types of files is the source of much tension and conflict between stakeholders. Australia is at the forefront of this behaviour with sources claiming that more than 30 per cent of Australian adults routinely illegally downloading television shows and movies.

Pirate hunt: is this the end of Australia's love affair with illegal downloading?, Kelsey Munro, Sydney Morning Herald, April 11, 2015

Reasons for the issue

There are a number of reasons why users download content through file sharing and **streaming** websites. First, the cost of downloading a file can be less expensive than legally purchasing the same product. Convenience is also a large factor. Users do not need to leave their own home to access the content and can then enjoy the content whenever and wherever they like. Downloaded content often does not contain advertisements. Being able to access particular content that is not currently available through legal means (for example, a new series of a TV show not yet screening in Australia) is a further significant factor.

So why is downloading or streaming files from some sources illegal?

Downloading or streaming files over the internet is not necessary illegal. There are many websites that offer files to be downloaded or streamed legally. Download.com is an example of a website where users can legally download files for free. For subscribers who are prepared to pay a monthly subscription fee, Netflix provides a legal streaming service of TV shows and movies.

Whether a file is legal or illegal to download comes down to whether the **copyright** holder of the content has given their consent for a user to download their work. Users who download or stream content in Australia without the consent of the copyright holder, are in breach of the *Copyright Act 1968* (Cth) and may face sanctions if found guilty, including being fined or, in more serious cases, being sentenced to a jail term.

THINK ABOUT COMPUTING 4.2

Is watching a video on the streaming service YouTube legal? If someone posted a video that someone else created onto YouTube, would copyright be breached?

P2P is also discussed in Chapter 3.

A summary sheet outlining the *Copyright Act 1968* (Cth) can be found at the Creative Commons website.

Digital systems involved

A common method used, for both illegal and legal downloads, is **peer-to-peer (P2P) file sharing** technology. As discussed in Chapter 3, peer-to-peer file sharing starts with one user offering files they are willing to share. By using the correct software, other users can then download a copy of these files. Instead of downloading a file in one go, P2P software downloads packets of the software at a time. As soon as one packet of data is received, that user can then start offering that packet of data to others. As more users share the same file, it creates more sources from where packets of data can be received, speeding up the downloading process.

File streaming is a data transfer concept where, as a file is received, it can start to be processed or viewed before the entire file is downloaded, meaning the content can be accessed faster.

There are a number of people, or groups of people, who can be affected by illegal downloading and streaming.

In 2015, the company who owned the rights to the movie *Dallas Buyers Club* went to the Federal Court of Australia to demand that a number of internet service providers (ISPs) release the contact details of their customers who were linked, via their IP address, to downloading the movie without the consent of the company, through a file sharing network.

Key stakeholders

One group seriously affected by the illegal downloading or streaming of files are the creators, authors or artists of the original work. Historically, many artists in the music industry earned a large proportion of their income from the sale of their songs including albums and CDs. In many cases, the revenue made from selling music has completely dried up, because of illegal downloading.

Artists suffer as online piracy worsens, Australian Financial Review

On the other hand, some artists have reported an increase in people buying tickets for their live concerts. One reason given for this is that as more people have been able to access the artists' music online, this has helped to increase demand for concert tickets.

The Australian Federal Government is also a stakeholder in the issue. With illegal downloading and streaming becoming such a significant issue in Australia, the federal government may have to act to protect the interests of the creators, authors and artists.

There has been speculation that the government is considering a range of measures to help curb illegal downloading, including blocking websites related to illegal file sharing and streaming and forcing internet service providers to hand over the details of individuals suspected of illegally downloading files.

THINK ABOUT COMPUTING 4.3

Brainstorm some other measures that could put in place that could curb, or stop, the amount of illegal downloads occurring in Australia?

Visualising thinking tools for supporting reasoning and decision making

Lucidchart is a free online tool that allows users to create diagrams

Often, a better understanding of the issues, problems and solutions can be obtained if the process described is presented visually. Visualising thinking tools assist with thinking processes and reflect on the thinking strategies to support understanding. These tools assist with coding and classifying themes, identifying patterns and linking relationships. Tools such as concept maps, sequence charts, mind maps, Venn diagrams assist researchers to document their decisions and clarify their thoughts. In Figure 4.4, a visual analysis of a digital systems issue is presented using Inspiration, a visual mind-mapping software package.

An analysis of an issue can be answered using the following questions.

1 What information system is being used and how is it being used?
2 What is the reason for the use, development of or change in the information system that is causing the concern?
3 Who are the people (the stakeholders) involved? (They can be individuals, institutions, societies, locally or globally, who develop and implement, control and use the information system, or anyone who is affected directly or indirectly.)
4 What are the advantages and disadvantages for the various stakeholders from the use of information system?
5 What are the main issues for the stakeholders? How do the various advantages and disadvantages for the various stakeholders create the issue?
6 What are the ethical principles, laws, policies or rules that apply to the various issues?
7 What are the detailed ethical issues associated with the impact of the information system? Which principles, laws, policies or rules have been violated and how have they been violated?
8 What is the range of feasible solutions (decisions and/or actions) that can reduce or remove the negative impacts? How do they solve the problems?
9 What are the advantages and disadvantages of the various solutions?
10 Which solutions should be chosen? Why? Which solutions should be rejected? Why?
11 What action should be taken? Who should be responsible for performing and monitoring the action? What is the timeline? What resources need to be used?

At the end of the process, all the relevant information will have been presented, various solutions suggested, a solution justified and a plan of action outlined.

ETHICAL DILEMMAS

In Chapter 1 we looked at how competing principles (such as upholding freedom of expression versus protecting children from possible harm) may sometimes present us with a dilemma (for example, whether to permit or ban the sale of violent video games).

Recall the six-step framework presented in Chapter 1, to provide support and guidance for making an ethical decision.

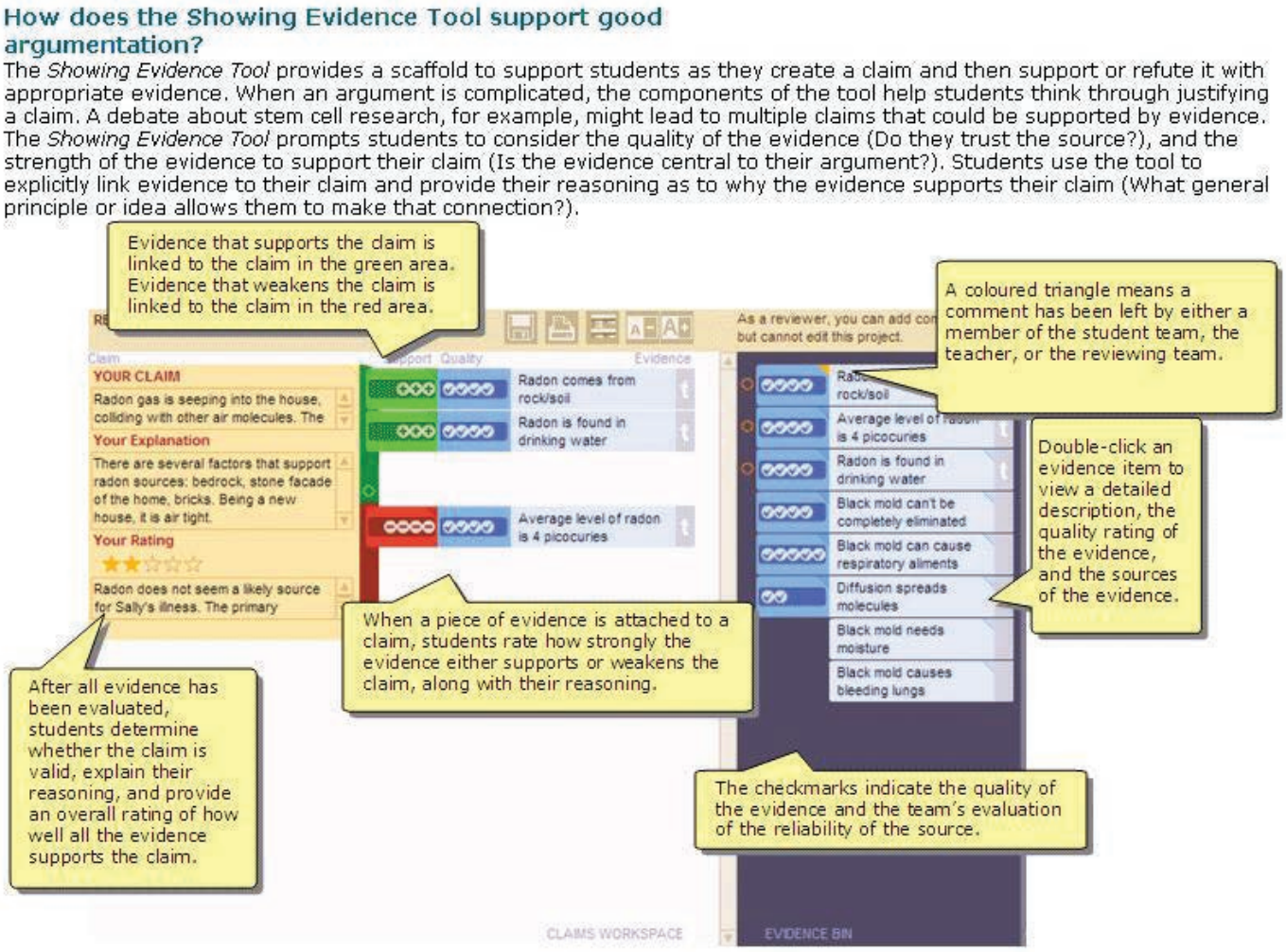
1 Identify the problem: What decision has to be made and what facts are required?
2 Identify the stakeholders: Who are they? What interests do they have? What power do they have? Who is vulnerable? How are the vulnerable to be protected?
3 Identify possible alternatives: What options are available? What are the likely consequences?

4 Identify ethical standards: Are there any applicable laws? Are there any morals or standards that could be applied? Is there a precedent?

5 Evaluate options: Identify strengths and weaknesses. Identify the option that causes least harm. Can the decision be reversed?

6 Make a decision: Select the preferred option. Justify the option. Inform all stakeholders of the decision.

Figures 4.3 and 4.4 provide some further tools for analysing problems and making decisions using visualisation to clarify issues and processes.



Intel Education

FIGURE 4.3 Intel Online Showing Evidence Thinking Tool helps students to construct well-reasoned arguments supported by evidence, using a visual framework

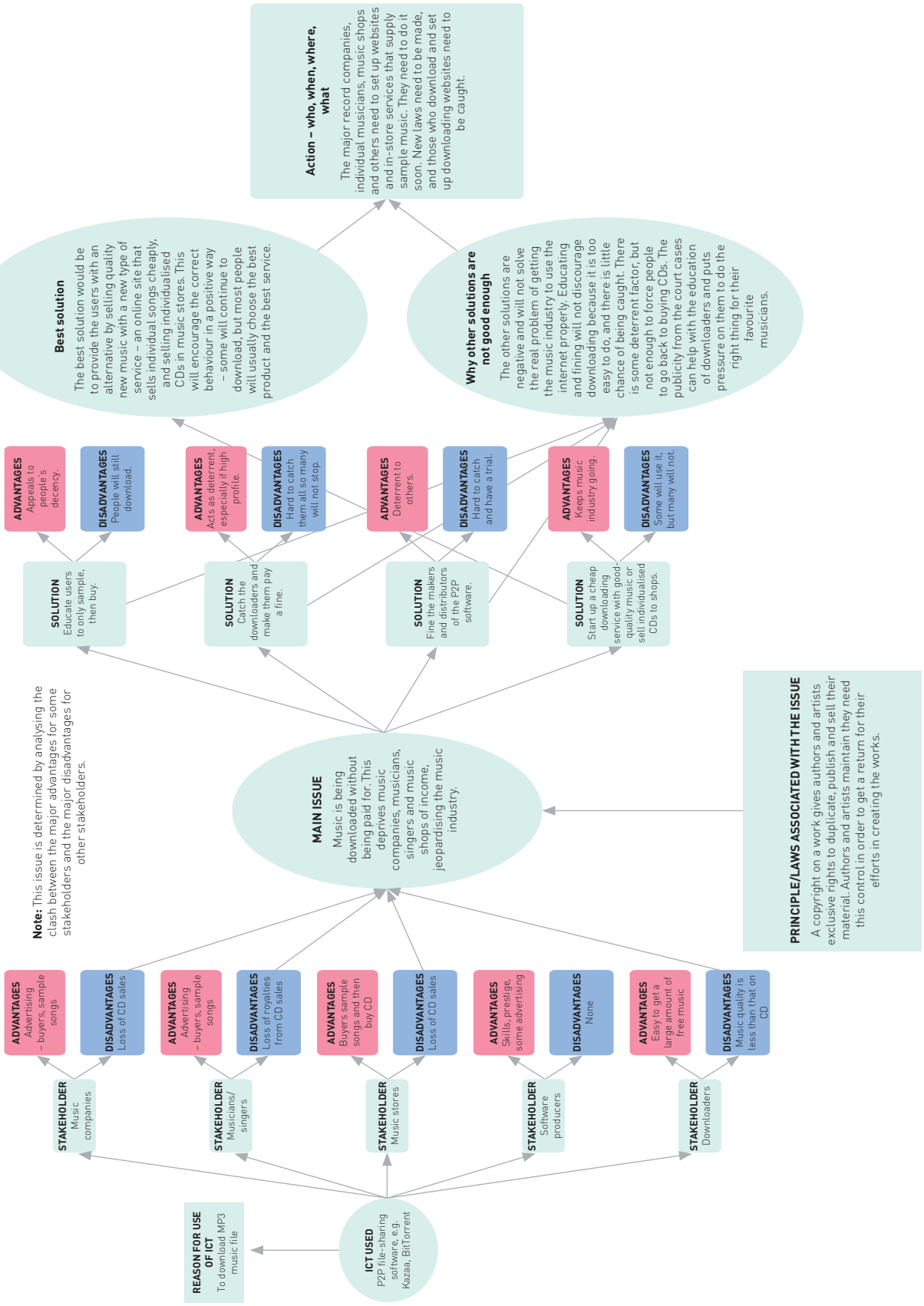


FIGURE 4.4 Visual analysis – the issues surrounding illegally downloading music

Information systems and sport

The growth in the number of people using social networking sites over the last 10 to 20 years has been amazing. One group that harnessed the power of social media are professional athletes.

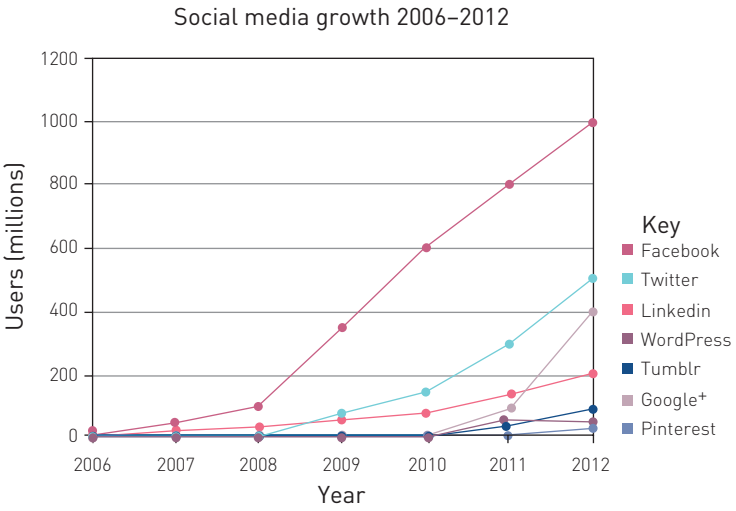


FIGURE 4.5 Growth in social media sites 2006–2012

There are many reasons why social networking sites have become so popular. Interacting with others online is one popular reason. Others reasons include seeking information and entertainment (for example, access to games or movies through social networking sites). Supporting a cause, expressing an opinion, or just for leisure are further reasons why social networking sites have become so popular. Organisations are trying to harness the power of social networking sites to help to promote and market their goods and services, while professional athletes use social media to connect and interact with their fans.

In addition to interacting with their fans, a number of sports stars have made considerable amounts of money by promoting particular products through their social networking accounts – for a fee.

But, along with the benefits that can be achieved through the use of social media, there are also many issues that athletes need to be careful of when broadcasting information to the public.

In June 2015, the account with the most followers on Twitter was Katy Perry with more than 70 million followers. The top sportsperson on the list was footballer Cristiano Ronaldo with more than 35 million followers.

Issues in sports

If a well-known athlete posted something with inappropriate language or information, it could quickly spread through cyberspace and come to the attention of the mainstream media. Not only can this reflect badly on the athlete, but also it may damage the reputation of any individuals or organisations connected with them. Consequently, the athlete may lose supporters and sponsorship deals.

Another issue facing athletes interacting with fans on social media is the potential for online abuse from strangers. Haters, trolls and even fans can leave abusive, racist or violent messages on social networking sites. In some cases, a professional athlete can be attacked relentlessly and consistently by trolls and these attacks can affect the performances of the athlete and even affect their personal or family lives.

Giving away strategic information or making opposition teams angry or upset is another area where athletes need to be careful when posting on social network sites. A seemingly harmless post about a team-mate's injury or clues to team selection may give opposition clubs an advantage.

An article appearing in the *Herald-Sun* newspaper in 2015 referred to the fact that social media was one of the biggest issues facing the AFL at present.

Collingwood star Dane Swan says he is fed up with 6 a.m. drug testing

Furthermore, athletes need to ensure that posts do not get them in trouble with either the sports authorities or the police. AFL footballer Dane Swan landed himself in trouble with the AFL in June 2011, after making a reference to drug testing procedures during the AFL season on Twitter. He later apologised for his outburst.

Reasons for the issues

One reason for the issues professional athletes encounter when they use social media is the instant nature of social media. If an athlete submits a post to a social networking site it will be instantly available for others to view. If the athlete has posted material they should not have, it is too late.

Another cause of athletes having problems when using social media arises from their rather sizeable following, whether friends on Facebook, or followers on Twitter or Instagram. Even if an athlete realises they have posted something they should not have, the chances are that a fan will see the post within minutes (if not seconds) of the post being displayed.

THINK ABOUT COMPUTING 4.4

Can you think of any other issues that sportspeople need to keep in mind when posting on social networking sites?

Digital systems involved

A particularly popular social networking site is Facebook, which allows users to create personal profiles, upload images and video, send messages and keep in touch with friends. There are a number of networking technologies that make up Facebook. One of the most popular components is the Wall, which acts as a bulletin board or forum where messages are left by other users and comments can be added. Status updates are a form of microblogging where a user can broadcast short announcements to their friends. The messaging service allows members to send short messages directly to one (or more) of their contacts without other friends seeing the message. Facebook contains a number of privacy options for its members. An individual can choose to allow anyone to see their information, choose if their information is searchable, keep parts of their profile private and restrict who can see their posts.

THINK ABOUT COMPUTING 4.5

What are your privacy settings on Facebook? Have you ever changed any of the privacy settings? Are you aware of all the privacy setting options?

Twitter is another popular social networking site used by athletes. It is similar in nature to the status update feature in Facebook, in that it is a microblogging site. Members post 'tweets' that other users can see. Unlike Facebook, the default setting for Twitter is public, meaning anyone can view the tweet – as long as the default settings have not been changed. Each tweet can be up to 140 characters and may include hyperlinks. Each tweet is sent to any user who has subscribed to the member's account, and a copy of the tweet will appear on the member's page on the Twitter website for public consumption. A feature of tweets is they can also include a hashtag (#), which acts like a metatag on a webpage to help categorise the tweet. Instagram is another popular social networking site. Like Facebook, it allows members to create a profile, but Instagram focuses on allowing members to upload and share images with others. Images can be posted to a number of social networking sites.

'Friends' in Facebook is a term to describe your user contacts. These contacts may or may not be actually known by the user, but once you add a 'friend' they usually have access to all the data you upload to the site.

Instagram also contains photo editing capabilities, allowing members to enhance the look of their images before posting online. Hashtags are also a feature of Instagram, which helps members to classify each image so others can search for a particular topic.

Details of a BBC study on social media abuse of premier league footballers during 2015

Key stakeholders

An athlete who uses social media appropriately may benefit from increased sponsorship and endorsement deals, just as an athlete that uses social media inappropriately may lose these types of commercial deals.

Fans benefit from being able to interact with their favourite sports stars, but many athletes have stopped using social media as a result of abuse on social media. Missing out on interacting with their favourite athletes means all fans are being punished because of the actions of a few.

A club or team may encourage team members to use social media because the publicity may help to increase membership or sponsorship. However, a decline in membership or sponsorship is also possible if an inappropriate or irresponsible post is made by someone within the organisation.

In addition, companies and organisations who invest significant amounts of money in both individual athletes and teams may quickly see their own brand damaged, as a result of an association that they have with an athlete.

FIGURE 4.6 A wide range of social networking sites

Information systems and agriculture

Agriculture involves crop cultivation, irrigation and harvesting; animal raising, reproducing and culling; as well as handling, transporting, packaging, preservation, processing, value-adding, quality management, storage and marketing, for both crops and animal products. Information systems and digital systems have been developed to assist with most, if not all, of these tasks.

Issues in agriculture

As new technology is being introduced, the amount of training required to be able to use and maintain the new technology is significant. This has caused problems when equipment breaks down and the farmer is not equipped with the skills (or equipment) to fix the problem, resulting in equipment sitting around for days or weeks waiting for a specialist to arrive to solve the problem.

Short- and long-range weather forecasts are another area that can lead to issues with the agricultural sector. Farming is considered one of the riskiest professions in Australia. The type and timing of each crop needs to be carefully planned. Weather forecasts are an important input to the decision-making process for farmers. In the long term, farmers want to know the weather conditions over the next few months to a year, because this will have an impact on their crops. On a daily basis, farmers need reliable information about daily weather conditions so they can make decisions regarding the tasks that will be completed on any given day. Spraying crops and controlled burning are example of jobs where the weather forecast plays a critical role. Getting the forecast wrong can have critical effects on farmers.

The New South Wales Department of Primary Industry publishes a periodical list of apps relevant to the agriculture sector.

Shutterstock.com/My Portfolio

FIGURE 4.7 Modern agricultural equipment – a combine harvester

Reasons for the issues

Many farmers are very skilled at manually fixing equipment but can be very inexperienced when dealing with modern technology and may need to be completely retrained. Also, farmers are often in isolated locations and it can take time for technicians and other professionals to visit farms to fix equipment.

Although technology has assisted in improving the accuracy of weather forecasts in recent years, forecasts, by their nature, are predictions and, like all predictions, there are times when unknown variables may come into effect.

New high-tech farming equipment can be a nightmare for farmers

Digital systems involved

The availability of technology that allows people in remote locations to communicate with others has significantly helped people working in the agriculture industry.

More specifically, global positioning systems are often used in farming, particularly to assist in surveying and fencing property boundaries. Global positioning systems are used to collect data about the topography and contours of properties, where this data is then used to determine the types and timing of crops for a particular property; weather forecasting systems also can be used as inputs to the crop planting process. Most modern farming equipment (for example, tractors and harvesters) come equipped with computing and sensor technology.

Bureau of Meteorology (BoM) forecasts under fire

Key stakeholders

All of these systems have been created to help workers or farmers complete tasks more efficiently, and effectively. This means that the task completed would be better than without technology. Although technology has helped the agriculture sector, there have been some problems along the way.

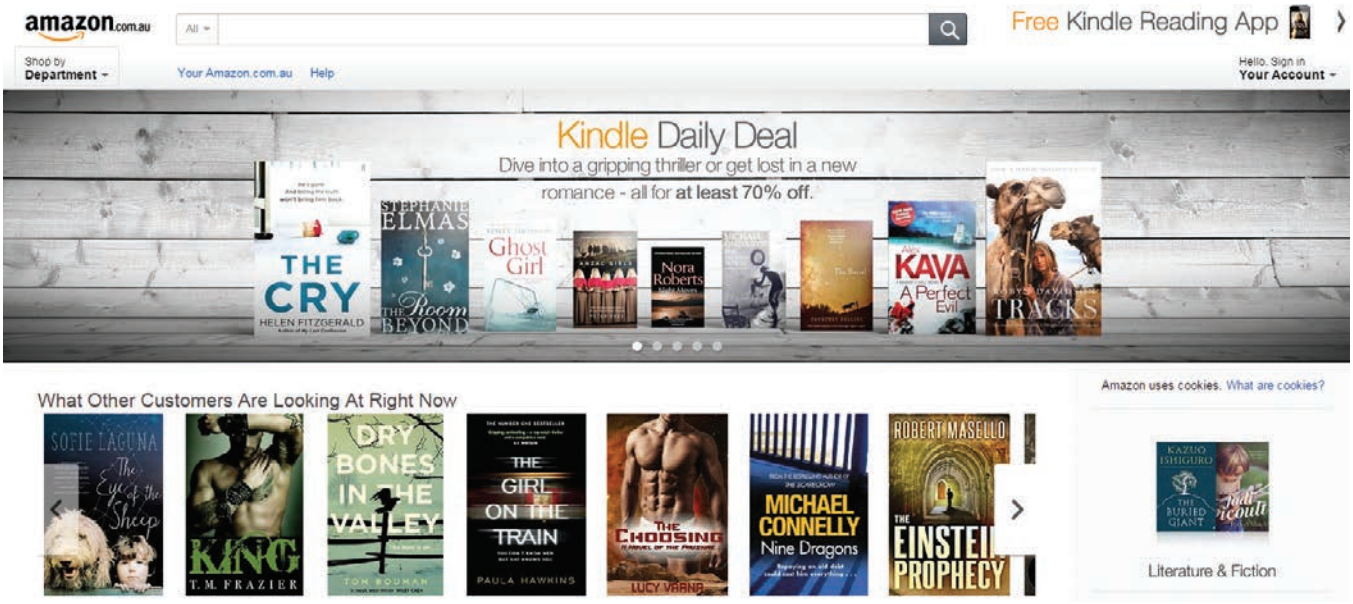
There are many people affected when things go wrong in the agricultural sector besides farmers. Staff, many who work on a casual basis, may have their hours reduced or lose their jobs. The general public can also be affected through the shortage of food or goods available to buy. During floods in Queensland in 2011, the price of bananas increased by more than 200 per cent of the normal selling price as a result of a shortage.

THINK ABOUT COMPUTING 4.6
Identify some other issues where information systems have created issues in the agricultural sector

Information systems and finance

With the advent of the internet during the 1980s and 90s, which helped to create a global computer network, individuals and organisations started to harness the power of the internet to conduct commerce or business transactions online. Many organisations started to set up e-commerce websites that allowed their customers to complete transactions online. Banks, florists, government departments, restaurants, shops and travel agents are all examples of businesses that embraced the chance to conduct business online.

By communicating electronically, it also allowed stores to start accepting payment by credit or debit card instead of cash only, banks introduced automatic teller machines and utility companies (for example, electricity, gas and water) started to accepted online payments of bills.



screen grab of Amazon.com.au at 4 July 2015

FIGURE 4.8 Online retailer Amazon.com was founded in 1994. Amazon.com.au launched in November 2013.

Issues related to e-commerce

With the popularity of conducting transactions electronically has come an enormous amount of issues regarding the security of the data and people involved in these transactions.

Many electronic transactions involve the use of a card, such as a debit or credit card, to complete the transaction. As a result of this, there is a possibility that the card data, including the card number, could fall into the wrong hands and the card could be used for additional transactions not made by the cardholder.

Card skimming often occurs when the customer's card is out of their sight; for example, at a bar or in a restaurant. Another common place for skimming to occur is at an ATM, where a skimming device is hidden behind the panel of the machine and skims the data as the customer inserts the card to complete a transaction. Often, a tiny camera will also be used in conjunction with the skimming device to capture the PIN that goes with the card.

Phishing is another method used to collect card data without the consent of the owner. As discussed in Chapter 3, phishing involves unauthorised people sending emails to individuals pretending to be; for example, a bank or some other type of financial institution, stating something along the lines that the person's account is locked and that to unlock their account the person needs email back their card details and PIN.

Another issue that has developed with the growth of e-commerce transactions is the issue of identity theft. **Identity theft** can be similar to card skimming in that hackers and other unauthorised people start to 'steal' data from individuals conducting e-commerce transactions. But, rather than just steal card details and PINs, they start to collect a wide range of personal and financial details about a person including their name, date of birth, address, work history, banking details and driver's licence number. Some of this data is easily accessible on social networking sites. Criminals have even been known to go through victims' rubbish bins to obtain one or two pieces of data they did not already have.

Once enough information is collected about a person, criminals then may be able to apply online for a credit card or loan, without the knowledge of the victim. Often they change the victim's postal address so no documentation is received and, in some cases, the first the victim hears about these credit cards or loans is when either debt collectors start chasing them for default repayments or they are refused a new credit card or loan.

THINK ABOUT COMPUTING 4.7
What year did the first online shop appear? Was Amazon the first online bookshop?

AAP Image/Queensland Police

FIGURE 4.9 False card reader containing card skimmer added to an ATM

ISSUE Identity theft hits one in five: study

July 5 2012, Alexandra Smith and Dominic Bossi

One in five Australians have had their identities stolen or had their personal or financial data illegally accessed, with credit card crime such as skimming one of the major problems plaguing consumers, a new study suggests.

The Australian Debt Study, released today, shows that Australians aged 35–49 are the most likely group to fall victim to identity fraud while 18–24 year olds are the least likely to report illegal access to their personal or financial data.

The data intelligence group Veda, which surveyed more than 1000 Australian adults, also found that people earning more than $70,000 are much more likely to be targeted for bank account and credit card crime than those earning $40,000 a year or less and cases of identity theft and financial fraud are highest in Western Australia and NSW.

Findings also show that almost one in three Australians suffered some form of credit crime and lost their wallet containing credit cards and identification.

Matthew Strassberg, a Veda senior advisor said: "Identity crime is a thriving industry in Australia, with the Australian Bureau of Statistics estimating the cost of personal fraud to consumers at $1.4 billion dollars a year.

"Whilst credit card fraud is a common form of identity crime, many people do not realise that with only a small amount of personal data, an identify thief could take out a second mortgage on a house, or open up a new line of personal credit and purchase items in their name or under a false identity."

The results of the study come as a 29-year-old Mascot man was yesterday charged with the alleged possession of ATM skimming devices, hundreds of bank cards and seven duplicating machines.

Police spoke to the man and searched his van after he had allegedly parked illegally on Darling Point Road, Darling Point at 12:30pm on Wednesday. He will appear at Downing Centre Local Court on July 25.

[July 5 2012, Alexandra Smith and Dominic Bossi] Sydney Morning Herald

Reasons for the issues

When financial and e-commerce transactions are conducted, sensitive data is transmitted. This attracts criminals who try to use this opportunity for their own personal gain. The techniques and practices that criminals used to try and capture financial data are always changing and being refined.

Another reason that these issues occur is that often the victim is unaware of the dangers that they are putting themselves in by conducting an e-commerce transaction, whether handing over their credit card in a restaurant, purchasing goods online or replying to an email or phone call from the 'bank'.

Digital systems involved

Card skimming involves stealing the data stored on a card's magnetic strip, using a card reading device. The card data is then copied onto a 'clone' card, which is then used for unauthorised transactions, or the card data is used in online transactions where the card does not need to be shown.

Another method used to capture card details is while they are been sent across the network from one location to another. The data is sent across the network in packets and it is possible that these packets can be intercepted by unauthorised people as they are travelling between devices. Once intercepted, the data within each packet can be read and, if useful, then used.

Phishing often involves sending emails pretending to be from a financial institution. The email may include the same font and logo of a real organisation, but often the email address itself may be a giveaway.

Identity theft involves the uses of a variety of digital and physical methods to collect data from the victim.

Key stakeholders

The issues around e-commerce can have far-reaching consequences. For a victim of either card or identity fraud, not only could he or she lose money initially, but any defaults regarding their repayments may be recorded on his or her credit report, which could result in credit applications being declined for years to come. Also, there is always the possibility that the criminals involved may try to use the data again.

Businesses are also affected by identity theft. Banks and financial institutions may cover some or all of the cost of money lost through card or identity theft. This in turn gets passed onto the other customers of the organisation through higher charges or interest rates.

There are a number of precautions individuals and organisations can take to reduce the chances of card or identity theft. Firstly, they can limit the amount of information provided on social networking sites. Many online accounts get users to create security questions – used to reset the password if forgotten. Often the answers to these questions can be found on personal profile sites like Facebook. Cover the keypad when using an ATM. This may not stop the card being skimmed, but may stop the scammers getting the PIN that goes with the card. Do not complete e-commerce transactions on public computers such as those in internet cafes or on

unsecured wi-fi networks, so data cannot be stolen using spyware or keylogging software. Only give personal and banking details on trusted websites and do not respond to emails or SMS messages asking you to give sensitive details such as the username and password of an account – phone the organisation directly if you are unsure. Only give sensitive information on sites that use a secure protocol, where encryption is used to encode the data, so if it is intercepted, it will not be understood.

THINK ABOUT COMPUTING 4.8
Create a list of other issues that may result from completing transactions online.

Shutterstock.com/ Jozsef Bagota

FIGURE 4.10 Secure protocols

Information systems and health

Electronic health records (EHR) are used to collect the medical history of a patient. An EHR holds personal details including age, height and weight, medical history, medications previously prescribed, allergies, blood-test results and more. The data is saved in a digital format that allows the history, or elements of the history, to be shared over a networked environment, the internet and shared between information systems. The system ensures that a complete patient's history is available, in the one location, when required. An EHR helps to reduce duplication of data as only one copy of the file exists. This also reduces the chance of input errors and helps to improve the **integrity of data**.

Dreamstime.com/Pandpstock001

FIGURE 4.11 Electronic health record (EHR)

Issues in health

With all the advantages that an EHR brings there are a number of problems associated with its use. Patient privacy is a large concern related to the growth of EHRs. A significant feature of the electronic records is that, because they are transferable between the multiple health professionals a patient may consult, the chances of the records being sent to the wrong location or accessed by unauthorised staff, either accidently or on purpose, is significantly increased.

The problem of user dependency has also arisen in the use of EHRs. A hospital in Western Australia faced a situation where, because of a network failure, the hospital was without access to any of its digital systems, including EHRs of its patients, for more than 14 hours. The problem was caused by a fault with the hospital's data centre, which was located in Spain.

A study into EHRs found that some medical professionals were reluctant to use a new type of information system.

> Emergency physicians noticed that Electronic Health Records disrupted workflow and was less desirable to use. The health information exchange systems need to adapt to the needs of the end user to be both useful and useable for emergency physicians.
>
> Emergency Physicians' Perspectives on Their Use of Health Information Exchange, Thorn, Shirley A. et al., *Annals of Emergency Medicine*, Volume 63, Issue 3, 329 – 337

Reasons for the issues

The ***Health Records Act 2001*** **(Vic)** is a set of laws that protect the rights of an individual's health records. The act outlines what data can be collected, how the data can be used, who has the right to access the files and even how the data must be stored. The *Health Records Act 2001* (Vic) applies to all medical records in Victoria, in both the public and private sectors. There are 11 Health Privacy Principles (HPPs) contained in the Act, which outline how medical data must be handled.

THINK ABOUT COMPUTING 4.9
Create a list of the 11 Health Privacy Principles contained in the *Health Records Act 2001* (Vic).

Digital systems involved

An EHR stores the data and information regarding an individual's health history on a server in the one location. As the record is stored in the one location, whenever the data in the record is updated or changed it will ensure that the record remains complete and up to date.

The EHR will be accessible via a network connection, so it allows the data and information to be portable. For example, if the patient needs to see a specialist for the first time, the specialist will be able to access the patient's full medical history via a network connection.

Key stakeholders

The patient involved will benefit from an EHR because whenever they need medical attention their full medical history will be available to the health professional they are visiting. Nonetheless because the health records will be stored on a network environment, there is a chance that the records may be accessed by unauthorised users.

Health professionals will also benefit from the use of EHRs as complete patient histories will be available, to assist in providing the best treatment possible. Yet health professionals still need training and support to be able to use the new information system as effectively as possible.

ISSUE Fiona Stanley Hospital systems crash

Daniel Emerson, 16 February, 2015, 5:51 a.m.

Doctors, nurses and administrators at Australia's most technologically advanced hospital were reduced to scribbling on paper when WA Health's computer systems crashed for more than 14 hours.

Fiona Stanley Hospital reverted to 'downtime procedures' during a Statewide outage after the department's main data centre in Malaga, run by Japanese technology giant Fujitsu, failed early this month.

It is understood lightning storms caused the problems on February 1.

A Health Department spokesman said the crash resulted in the loss of clinical and non-clinical computer applications and the IT network, including email, from 2.55pm until 5am the next day.

Staff had to use pens and paper and then enter patient data once the system came back online.

'The health information network staff, together with the vendor (Fujitsu), responded immediately to convene management, resolution and a review of the impact to health care services,' the spokesman said.

'Hospitals also enacted their downtime procedures and hospital business continuity plans to ensure patient safety was never compromised.'

Bankwest, which also uses the data centre, had crashes to its phone and online banking, ATM and EFTPOS services.

[Daniel Emerson, 16 February, 2015] The West Australian

AAP Image/Richard Wainwright

FIGURE 4.12 Fiona Stanley hospital had to respond to an information systems crash.

Expressing opinions

There are many ways a user of an information system can express their opinion about the system online.

THINK ABOUT COMPUTING 4.10

There are many more information systems used in the health sector besides EHR. Can you think of any? Can you think of any issues that may arise from their use?

Social media

Social media are common outlets where users can broadcast their opinions on information systems. Social media sites allow the user, with a few keystrokes, to express an opinion that potentially can spread quickly without much effort. They can then quickly interact with others with opinions on the same system.

There are many social networking sites with sites or topics devoted to the Myki ticketing system used for public transport in Melbourne and the surrounding areas.

When a group of users interact using social media, this may assist in promoting a cause or encouraging action.

Search #Myki to see the social media discussions around this topic.

Forums and bulletin boards

Forums and bulletin boards are also platforms that allow users to express their opinions about information systems.

Many organisations attach a forum page to their website to allow their visitors to share and discuss information and to offer their opinions, while some websites function solely as a forum or discussion board.

Whirlpool is a website created to cater to broadband internet users. It contains one of Australia's biggest internet forums that specialises in information and digital systems.

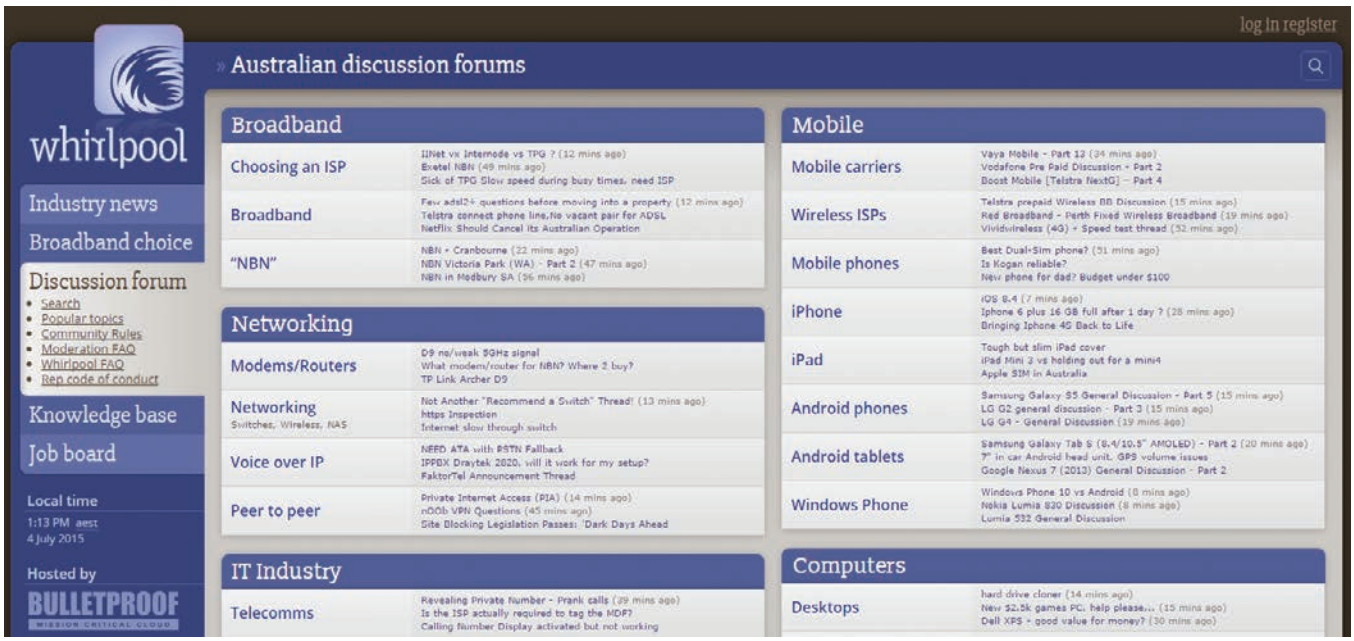


FIGURE 4.13 Whirlpool forum

Ratings systems

Some websites set up rating systems so that users can give feedback about information systems.

Rating system allow the user to give feedback about the information product in a quick and efficient manner. Many sites use a five-star rating system but many variations of rating systems are used on different sites

Also many sites allow a user to give feedback by leaving comments. This offers an advantage over a traditional rating system, because the user can be more specific about what they did or did not like about the system. The comment can be similar to a forum post in that other users can read the opinion and the comment might give more detail about how the system operates. Also, potential users tend to trust the opinions of their peers much more than any organisation-driven marketing or advertising campaign. Finally, online comments allow the organisation to get honest feedback regarding the strengths and weaknesses of their products or services.

CBS Interactive Inc.

FIGURE 4.14 Five-star rating system

CBS Interactive Inc.

FIGURE 4.15 Comment feedback

Methods and techniques to acquire data and information

When investigating issues associated with the use of information systems, you need to use a variety of primary and **secondary data**.

The best research uses both forms of data, as they can be used to explore issues more comprehensively. For example, secondary sources can provide questions to be used in primary research and, in return, primary research can provide ideas that need further secondary research. The good place to start is with a review of secondary sources related to the information system.

Primary data collection methods

Primary sources involve collecting data from the stakeholders involved in the information system issue directly. A number of methods can be used to collect data first-hand. Interviews, observations and surveys are all common methods of collecting **primary data**.

To collect data either by interviewing, observing or surveying can provide different insights and often more in-depth data than information from secondary sources. The data will often be more up to date and can provide unusual and important insights into issues, especially at the immediate local level, than secondary sources can, because secondary sources often present overall conclusions and general summaries.

Interview

An interview involves a conversation between two or more people where questions are asked and the interviewees give their answers to the question. Interviews are useful as the in-depth responses can be elicited from the people involved. A range of questions can be asked and the conversation can change focus at any stage depending on the nature of the responses.

An advantage of the face-to-face interview technique is that both non-verbal signals and emotions can also be picked up by the interviewer. A user's enthusiasm for an information system may not be totally captured by a phone interview technique alone, because body language and facial expression may also paint some of the picture. The same may apply to issues and areas of concern that the user may be reluctant to discuss or only want to give brief generic answers.

Disadvantages of interviews is they can be time consuming as the interviewer obviously needs to be present throughout the process, this, in turn, makes interviewing a relatively expensive data-collection method. As a result, this may limit the number of interviews that take place to only a handful of key stakeholders.

Observations

Observation involves watching users interact with an information system to collect data where and when the activity is taking place.

An advantage of observations is that the observer does not have to rely on other people's responses to or opinions about the system. The observer is seeing for themselves how the system is operating. Observation techniques include making mental and physical notes and taking photographs.

One disadvantages of collecting data through observations is that the people, knowing they are being observed, may act and behave differently to how they would normally. Also, observations may help identify issues with an information system but may not uncover the underlying reasons for them or effects the issues are having and, similar to interviews, observations can be very time consuming and expensive to conduct.

Surveys

Some online surveys, called 'dynamic surveys', are designed to show different questions based on the answers given to previous questions. This is known as question routing.

Surveys are a popular method of collecting primary data. A survey consists of a series of predetermined questions that can then be sent to a participant. The responses to each question might involve selecting from a list of alternatives, similar to a multiple-choice question, may involve writing a comment or a combination of these methods. Techniques for conducting surveys include in person, over the phone, through the post or online.

The biggest advantage of a survey is that it can collect opinions from a wide range of people in an inexpensive and timely manner. Surveys can be sent via email or conducted through a website to get the opinions of hundreds or thousands of users. Some people find completing a survey less intimidating than an interview, so therefore may be prepared to give more honest feedback, particularly if the survey is anonymous.

Shutterstock.com/Andrey_Popov

FIGURE 4.16 Online survey

Disadvantages of surveys include surveys not being completed or only partial completed. Some participants may not take much care or effort in the way they answer the questions on the survey. Another disadvantage is that data collected is based solely on the questions asked. Unlike an interview, in a survey it is difficult to change the course of the questions or ask for in-depth answers, as the structure of the survey is often fixed.

Qualitative and quantitative data

When collecting data, the data can often be grouped into one of two categories, **qualitative** or **quantitative data**.

Preferably, the primary data you collect would be quantitative data, because this is much easier to process into information.

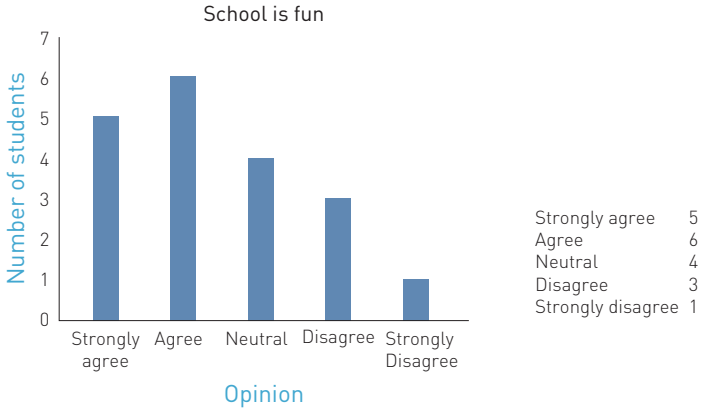


FIGURE 4.17 Example of quantitative data

Unfortunately, not all data required can be collected in this format. Qualitative data is also required, as personal opinions and ideas also make up part of the story. Qualitative data is harder to manipulate into information as the type and number of possible responses make this difficult. Methods are required to manipulate the data into useful information.

Wordclouds are one tool that may be useful in manipulating qualitative data. Each time a word appears in a block of text the font size of that word is increased in the word cloud.

FIGURE 4.18 A wordcloud created using the study design for VCE Computing

A **histogram** is another tool that may help to process qualitative data. Although histograms are used to represent graphically the distribution of numerical data, they can also be used to represent qualitative data. To do this, you first have to sort the data into groups or categories. The illustration below shows an example of a histogram being used to convert qualitative data into a quantitative format. After the opinions were collected about their favourite thing at school, these opinions were classified or grouped into broad categories. The frequency of each group was then converted to a graph.

What is your favourite thing about school?

RAW DATA

Teachers are nice	Home time
My friends	Physical education
French is fun	Range of subjects
Close to my house	Ms Neil my maths teacher
School canteen	English
My science teacher	Playing with friends
I feel safe	Mr Brown

FREQUENCY DATA DRAWN FROM RAW DATA

Subjects	4
Teachers	4
Friends	2
Safety	1
Other	3

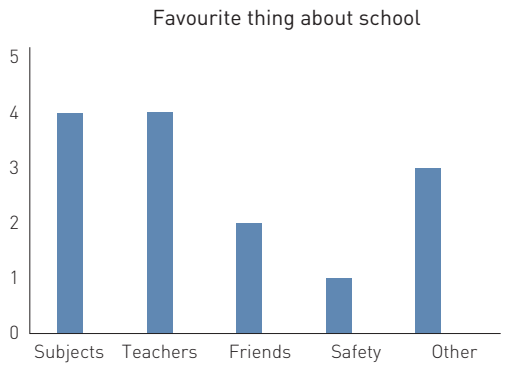


FIGURE 4.19 Example of qualitative data being converted to quantitative data

Secondary sources

Secondary sources were traditionally books, newspapers and journals, and they were often mainly accessed in libraries (Figure 4.20). Nowadays, many secondary sources can be found on the internet, and they come many formats, including text, audio and video. However, despite the easy access to information that the internet affords, do not discount libraries – the reference staff are often able to provide you with better quality, more relevant secondary data.

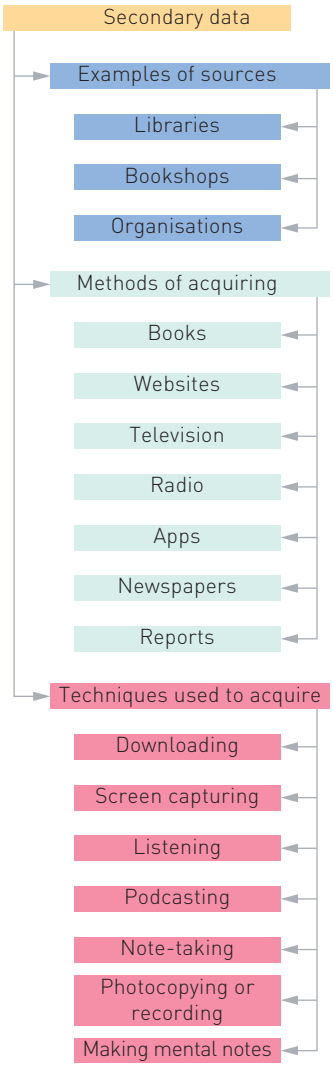


FIGURE 4.20 Secondary data: Sources, methods of acquiring, and techniques used to acquire

Although Wikipedia is a popular and convenient source of data and information, that does not mean the information on it is true.

No matter which secondary sources you use, the quality of the information needs to be evaluated carefully. A good technique when using data from secondary sources is to find two or more sources that confirm the same findings.

Referencing secondary sources of data

Referencing primary sources was discussed in Chapter 1.

There are a number of methods that can be used to reference both primary and secondary sources of data.

The **Harvard referencing style** is also referred to as an author–date style of referencing.

TABLE 4.2 The Harvard referencing system

In-text reference	Reference list
This book looks at how individuals and organisations use, and can be affected by, information systems in their daily lives. *Lawson et al* (2016, p. v)	Lawson, J., Keane, T., Kelly, M., Potts, C., & Sullivan, A., 2016, *Computing VCE Units 1 & 2* 6th edn, Cengage Learning Australia, South Melbourne.

Guide to Harvard referencing method, University of Sydney, 2010

TABLE 4.3 Referencing a website using the Harvard style

In-text reference	Reference list
Usually the author or creator of a work is the copyright owner (University of Sydney 2010, 'Who owns copyright?' para. 1). Note: The heading of the section was 'Who owns copyright?'	University of Sydney 2010, *Guide to copyright*, University of Sydney, viewed 24 January 2011, http://sydney.edu.au/copyright/students/coursework.shtml#who

Many universities publish guides to the Harvard referencing method.

Data integrity

Data integrity refers to the quality of the data held. There are a number of ways that can be used to measure the quality of data. This includes accuracy, correctness and reasonableness.

Accuracy

Measures should be put in place to help reduce the chances of incorrect data being entered. Validation techniques are often used on forms and surveys that may reduce the chances of incorrect data being entered, increasing the chances that the data is accurate. Many online forms include validation techniques such as radio buttons, dropdown lists, checkboxes, date time pickers and so on, to make it easier for the user to enter data and to give them less scope to make an mistake.

THINK ABOUT COMPUTING 4.11
Other referencing methods include the American Psychological Association (Latest Edition), Chicago, and Institute of Electrical and Electronics Engineers (IEEE). Explain any differences between the methods.

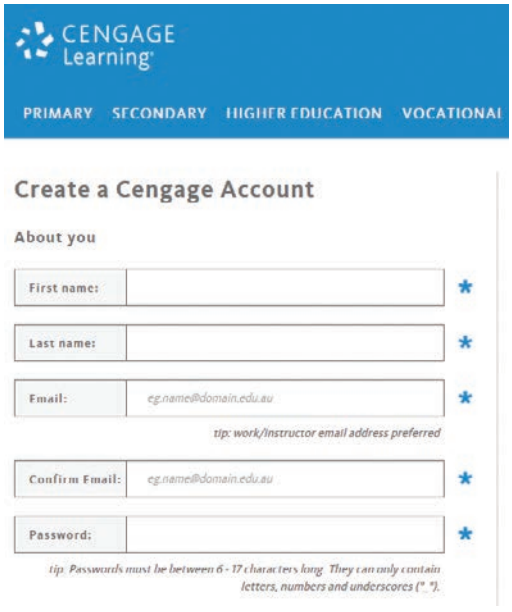


FIGURE 4.21 Input form with validation to reduce the chance of errors

Correctness

The data collected needs to be suitable for its purpose. Is the data collected relevant to the issue being investigated? Is the data collected in a suitable format? Can qualitative data be converted to quantitative data so it can be manipulated? Or is the data collected in a correct data type? Can an audio interview then be converted into text so a wordcloud of the conversation can be created?

THINK ABOUT COMPUTING 4.12
Think of other websites where you have been asked to enter data. What other types of validation techniques are used to help improve the reasonableness and accuracy of the data?

Reasonableness

Reasonableness is related to whether the data collected is normal or acceptable for that type of data item. If collecting data about a student's age, if the data entered were 'apples' this would not be reasonable for this data item (age). Validation techniques can again be used to help to ensure that the data is reasonable.

Storing shared files

In simple terms, **cloud computing** means storing data, information and software on a server located in remote location, rather than on your computer's hard drive. The data, information and software would be accessed via an internet connection.

The 'cloud' is a metaphor for the internet. Traditionally, network diagrams would represent an internet connection as a cloud symbol, so the practice of saving data and information to a remote location, using an Internet connection, was referred to as cloud computing.

The location of the remote servers will depend on the organisation hosting the data. Large organisations, such as Apple and Microsoft, will have numerous data centres set up in different locations around the world. Some smaller cloud providers may only have a data centre in one location.

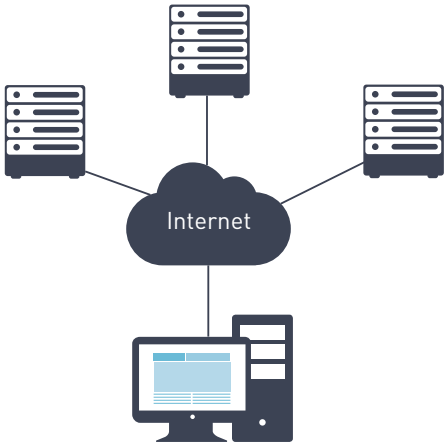
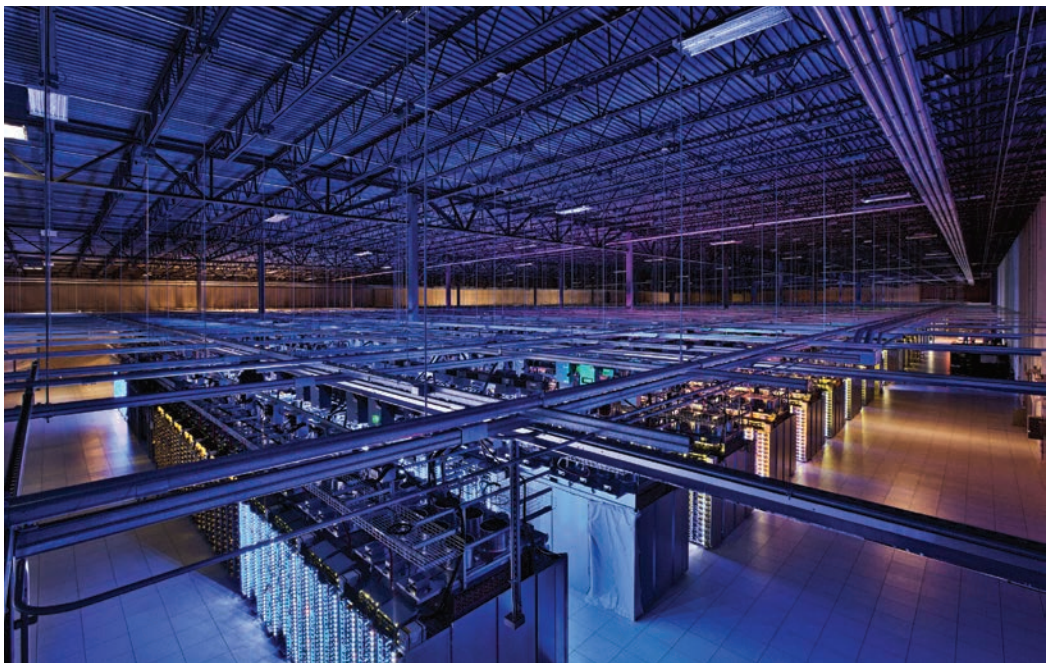


FIGURE 4.22 Network diagram representing the internet as a cloud symbol



Getty Images/Craig Mitchelldyer

FIGURE 4.23 Google's data centre in Council Bluffs, Iowa

Advantages

THINK ABOUT COMPUTING 4.13
Think of some organisations who offer cloud services for free. Why do you think they would be happy to provide this free service?

There are a number of advantages to using a cloud service for hosting or storage.

- Data and information can be accessed using any device.
- Data and information can be accessed from any location.
- It saves storage space and system resources on the user's computer.
- Data is stored in a remote location so in the event of a local disaster, the data will be safe.
- If a copy of the data is stored on the cloud, it automatically creates a backup of the data.
- It can save money as the user does not need to purchase any storage media, such as a file server.
- Data and information can be shared by nominated groups of people.

Disadvantages

There are also a number of disadvantages of using a cloud service for storage or hosting.

- Privacy of data is a concern.
- The level of security at the data centre is unknown.
- If a user's internet connection goes down, then the data cannot be accessed.
- Client data can be lost if the cloud service provider goes out of business.
- The location of the data centre storing the data is often unknown.
- Data could be stored in a country where different laws apply, which could compromise the safety of the data.

Mobile devices and web design

As the number of users using a mobile device to access webpages has rapidly increased, the importance of designing a website for use on a mobile device has become more important.

Many organisations have created separate sites linked to the one URL: One for desktops and notebooks and another site designed for use by mobile devices. When visiting the URL, the system will detect the type of device being used and show the site appropriate for that device.

FIGURE 4.24 Web design for mobile devices

Some of the factors to consider when designing a website for use on a mobile device include the following.

Screen size

The size of the screen will have a significant impact on the design of the webpage. Compared with a website created for use on a desktop or notebook computer, the amount of information that is visible at the one time is significantly reduced when the site is viewed on a mobile device.

Screen resolution

The resolution of the display also needs to be taken into consideration. Resolution is often measured in dots per inch (DPI). The higher the resolution, the higher the quality of the webpage. This may affect the amount of information that can displayed at once or influence the types of input methods used.

Input method

Desktops and notebooks are generally used with a keyboard and mouse. Mobile devices may not have the same options of input methods. Some devices have in-built keyboards, some touch screen capabilities, others use a stylus. The type of input method will need to be considered when designing the website. Should the site have a menu system? Is text entry required? Can all the functions of the site be operated by buttons?

Speed of data transmission

The speed of the internet connection also needs to be considered. With new cellular 4G connections, data download speeds have improved significantly. This has reduced the time taken to download and view larger file types such as audio and video files. But not all devices may be able to access a network with transmission speeds fast enough to download larger file types.

Cost of data transmission

The cost users pay to transfer data also needs to be considered when designing the functions to include in the website. Many mobile devices use plans with data limits or caps, so limiting the amount of data required to view the website would be optimal.

Device resources

Mobile devices may contain components with less capabilities than desktop and notebook devices. CPU, RAM capabilities may be significantly less powerful and the amount of hard disk space significantly less. This may affect the design by reducing the amount of data required or the processing needed to view the site.

Battery life

As mobile devices, due to their nature, generally use a portable power supply (battery) the website should be designed in a way that reduces the drain on the battery. High-resolution videos, animation and bright colours should be avoided.

CHAPTER SUMMARY

ESSENTIAL TERMS

card skimming a practice where the data from a card's magnetic strip is copied, using a card reader device; the card data is then copied onto a 'clone' card, which is then used for unauthorised transactions or the card data is used in online transactions where the card does not need to be shown

cloud computing storage of data at a remote location, rather than on the user's device

copyright a form of intellectual property, copyright states that the creator of the original work (e.g. song, movie or game is the legal owner of that work and has the exclusive rights on how that work can be used and distributed; in Australia, laws relating to copyright are contained in the *Copyright Act 1968* (Cth)

e-commerce the buying and selling of goods and services, or transferring funds over a computer network, commonly the internet

electronic health record (EHR) a digital copy of an individual's medical record

forum an online bulletin board where users can post, read messages, and leave replies; also known as a discussion board

Harvard referencing style a method of referencing primary and secondary data sources

***Health Records Act 2001* (Vic)** a set of laws that protect the rights of an individual's health records in Victoria

histogram a tool that may help to process qualitative data. Although histograms are used to represent graphically the distribution of numerical data; they can be used to represent qualitative data, if it is first sorted into groups or categories

identity theft stealing individual pieces of information about a person, until the thief has enough data to take out a loan or a credit card in the victim's name or otherwise impersonate the victim

information system a structure consisting of a combination of a digital system (hardware and software), data, processes and people that interact to create, control and communicate ideas and digital solutions

integrity of data relates to data quality's three characteristics: accuracy, correctness and reasonableness

peer-to-peer file sharing a method of transferring files often used to transfer files illegally

primary data data collected directly (first-hand) from the stakeholder involved in the issue

qualitative data subjective data about qualities – opinions, comments and personal experiences that are difficult to measure or manipulate

quantitative data objective data about quantities – it is numeric data that is easy to interpret

secondary data data from another source that has already collected the primary data; for example, a newspaper, or a website

stakeholders people affected by a given issue

streaming allows the users to start viewing a file before it is completely downloaded

wordclouds a tool that may be useful in manipulating qualitative data; each time a word appears in a block of text, the font size of that word is increased in the word cloud

CHAPTER SUMMARY

IMPORTANT FACTS

1 Issues arise when an information system causes tension and conflict between stakeholders.
2 The *Copyright Act 1968* (Cth) outlines the laws related to copyright.
3 A breach of the *Copyright Act 1968* (Cth) could result in fines or imprisonment.
4 Many artists are losing out financially as a result of illegal downloading.
5 Many sportspeople use social media to socialise with their fans.
6 Posting of inappropriate information has damaged the reputation of some sportspeople.
7 Online abuse has led some sports stars to stop using social media.
8 Opposition teams have been given an advantage through leaked information on social media.
9 Popular types of social media sites used include Facebook, Twitter and Instagram.
10 Information systems are heavily used in the agriculture sector.
11 Changing practices in the industry are causing conflict between the stakeholders.
12 Incorrect weather forecasts can have significant impact on the farming industry.
13 E-commerce has developed with the development of the internet.
14 Precautions to help avoid card and identity theft include limiting the amount and kind of information posted on social networking sites, covering your hand when entering a PIN at an ATM, and avoiding using e-commerce websites on free or open wi-fi networks.
15 The *Health Records Act* (Vic) includes 11 Health Privacy Principles.
16 Issues related to Electronic Health Records include medical staff resistance to change and user dependency.
17 Social media is a popular way for users to express their opinions of an information system.
18 Rating systems and comments are other ways in which opinions can be expressed.
19 Interviews, observations and surveys are methods of collecting primary data.
20 Word clouds are a method used to visualise qualitative data.
21 Histograms can also be used to visualise qualitative data, if the data can be grouped.
22 When using secondary data it is very important to check the quality of the data.
23 Advantages of cloud computing include accessibility, portability and reducing expenses.
24 Disadvantages of cloud computing include security, privacy and legal concerns
25 Factors to consider when designing a website for use with mobile devices includes screen size, screen resolution, input method, data transfer speed and cost of data transfer.

TEST YOUR KNOWLEDGE

Review quiz

INFORMATION SYSTEMS IN ACTION

1 List the four components of an information system.
2 Explain what the term 'digital system' means.
3 Describe how an issue can occur in relation to an information system.
4 Explain what the term 'stakeholder' means.
5 List three reasons why people illegally download files.
6 Identify a situation where downloading a file may be legal.
7 Define 'copyright'.
8 Explain how the concepts of copyright and intellectual property are related.
9 Name the Act outlining laws about copyright in Australia.
10 Explain how peer-to-peer (P2P) file sharing works.
11 Contrast downloading with streaming.
12 Outline how artists may be disadvantaged by illegal downloading and streaming.
13 Explain why sportspeople might use social media sites.
14 Discuss three ways that social media may become problematic for sportspeople.
15 List three social networking sites often used by sportspeople.
16 Explain the purpose of a hashtag.
17 List some uses of GPS in the agricultural sector.
18 Discuss some of the negatives around the introduction of new technology in the agricultural sector.
19 Explain how incorrect weather forecasts can cause issues for farmers.
20 Explain why e-commerce has become so popular.
21 Define 'card skimming'. List two places card skimming may occur.
22 Explain what 'phishing' means.
23 Explain how identity theft is different to card detail theft.
24 Discuss three precautions people can take to reduce their chances of both card and identity theft.
25 Explain the type of data held in an Electronic Health Record (EHR).
26 Outline two issues arising from the use of an Electronic Health Record (EHR).
27 Name the Act that outlines the laws relating to the use of medical records in Victoria.

EXPRESSING OPINIONS

28 Discuss why social media sites are a popular outlet to express opinions about information systems.
29 Explain an advantage of using a five-star rating system as a method of collecting opinions.
30 Describe the advantages of written comments over a rating system for opinions.
31 Explain how a forum operates.

METHODS AND TECHNIQUES TO ACQUIRE DATA AND INFORMATION

32 Explain the difference between primary and secondary data sources.
33 Discuss when an interview would be preferable to a survey when collecting data.
34 Discuss when a survey would be preferable to an interview when collecting data.
35 Outline one method to assess the reliability of data from a secondary source.
36 Explain how qualitative data differs from quantitative data.
37 Describe how a wordcloud works.
38 Explain how a histogram can be used to process qualitative data.
39 Name two methods used to reference primary and secondary sources.

DATA INTEGRITY

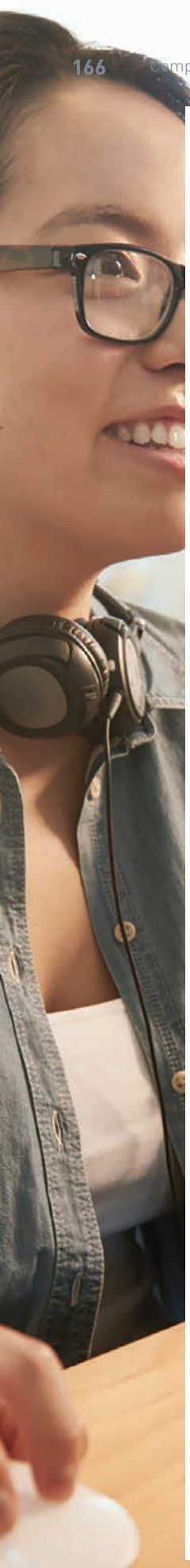
40 List the three characteristics of quality data.

STORING SHARED FILES

41 Define 'cloud computing'.
42 List three advantages of cloud computing.
43 List three disadvantages of cloud computing.

MOBILE DEVICES AND WEB DESIGN

44 Explain why websites used in mobile devices need to be designed differently to websites used on notebooks or desktops.
45 Distinguish between screen size and screen resolution.
46 List three ways that data may be entered into a mobile device.
47 Explain how the speed of transmission differs to the cost of transmission.



APPLY YOUR KNOWLEDGE

Practise analysing information systems issues and problems before attempting Unit 1, Outcome 3.

For each of the following situations:

- outline the nature of the issue
- identify any legal, social, environmental or ethical factors related to the issue
- describe the information system being used
- identify the stakeholders involved
- explain how each stakeholder identified is affected by the issue.

ISSUES

1. An employee uses her computer at work to make comments about the workplace, her employer, her supervisor, fellow workers and customers on her Facebook page. Some of these people have noticed the comments and are not happy about them. They have asked that she be sacked.
2. A family has noticed their son spending a lot of time playing an online computer game with his friends, and it seems to be affecting his school work and his relationships with his family. His parents want to take his computer away from him.
3. The board of a company that wants to save money has decided that most of its hardware and software will be transferred to a cloud-computing company. Many of the digital systems staff do not think it will benefit the company.
4. A country hospital has decided to buy an expert system to help diagnose patient illnesses because it is having problems finding enough medical staff who want to live in a country town. Some of the townspeople are concerned that it might be used by inexperienced doctors and the nurses to diagnose illnesses when the main doctor is too busy or is away.
5. A university student who is short of money has found a website that is selling musical instruments much cheaper than the prices in the local shops. Her friends are not enthusiastic about her using the website, but she insists on going ahead with the purchase using her father's credit card.

CHAPTER 5

APPROACHES TO PROBLEM SOLVING

Key knowledge

After completing this chapter, you will be able to demonstrate knowledge of:
Approaches to problem solving

- key principles of information architecture
- characteristics of effective user interfaces for mobile devices, for example useability, accessibility, tolerance, visibility, legibility, consistency, affordance
- design principles that influence the appearance of websites
- design tools and techniques for representing websites
- formats and conventions suitable for websites
- software functions and techniques for manipulating and validating data, and testing websites
- tools and techniques for coordinating the tasks, people, digital systems resources and time required to create solutions.

For the student

Designing and developing a website collaboratively requires extensive preparation and organisation. The purpose of this chapter is to assist you with designing a website for Unit 1, Outcome 3, paying special attention to the principles of information architecture. The design and development stages of the problem-solving methodology are described in detail. You will explore techniques for representing designs, manipulating data into information, and formats and conventions for websites.

For the teacher

This chapter introduces students to solving information problems through the use of information systems. A problem-solving methodology is discussed, with the emphasis on design and development. The content of this chapter leads students to an awareness of creating digital solutions using various techniques and procedures. At the completion of Chapters 4 and 5, students should be able to demonstrate Outcome 3, which requires students to contribute collaboratively to the design and development of a website that presents an analysis of a contemporary issue and substantiates the team's point of view.

Creating team solutions

Working in a team is a common practice in most workplaces. The benefits of **collaboration**, even with team members in other countries or states, are great. The right team can share work and pool experience to produce more useful solutions far more quickly than any one individual could achieve. With work teams comes the need for organisation. Resources need to be allocated, timeframes determined and personnel briefed on an ongoing basis to ensure that the project or task is completed on time and within budget. In this section of the chapter we will look at some common techniques and tools used to coordinate team-based tasks – that is, to manage a project.

Managing projects

The work of an organisation might consist of routine, day-to-day operations, or it might involve organising unique projects. Managing a retail outlet or a manufacturing business, where sales and production occur continuously, will be different from managing a construction project that is unique and has definite start and completion dates. **Project management** techniques have been developed to help organise a range of unique projects. Examples of projects could include larger-scale projects such as introducing a new brand of toothpaste to self-serve checkouts in supermarkets, and smaller-scale projects such as developing a website.

What is a project?

An idea usually launches a project. An individual or a group of people decide that there is a need for a new solution, a study is commissioned and then project planning begins. A project has the following characteristics.

- A clearly defined purpose
- A starting time
- A finite lifetime
- A number of interdependent tasks

Examples of large projects include the rollout of the National Broadband Network (NBN) across Australia and the Fisherman's Bend Urban Renewal Project in Victoria. Both have a clearly defined purpose and a definite lifetime. Both required a complex set of interdependent tasks. However, in small organisations, projects would not be as complex as these. Organising a school dance, creating a website for a local sports club or creating solutions using databases and spreadsheets are also projects in which project management techniques could be used effectively.

We manage small personal projects every day, and they can range from preparing dinner to buying a car. With larger projects, the number of tasks involved and the complexity of their relationships make it more difficult to manage time and cost. Project management techniques help managers to allocate resources (people and digital systems, for example), schedule tasks, monitor the project timeline and manage costs.

Project management techniques

Project management involves identifying key tasks, allocating resources, scheduling and monitoring those tasks, identifying milestones and determining **contingency plans**. It is therefore important to be aware of the factors you need to consider, as well as the meaning of commonly used terminology.

Tasks

The project should be broken into a series of tasks. Each task should be substantial, but not so large and complex that running over time will seriously affect the completion time of the whole project. The criteria for assessing when the task will be complete must be clearly defined so that a completion time can be recorded. One common method of checking that a task is complete is a checklist. A **milestone** is reached when a particular stage of the project has been completed. A milestone is not a task, so it has no time duration.

Resources

Resources are assigned to each task, and include technical (money and equipment) and human resources. People involved in a task must have input into setting completion times for the task so that timelines are realistic.

Schedule

Tasks are interdependent. A schedule allows you to map the project tasks and display interdependencies. A Gantt chart is one common way of displaying a schedule. A Gantt chart provides a timeline of how the project will be implemented. It visually represents the tasks that have to be completed, when they are to be completed and the time that each task should take. A Gantt chart may also show which tasks are dependent on one another, thus allowing for the identification of the **critical path** of the project.

Because tasks are interdependent, they must be completed in a particular order. Starting some tasks depends directly on the task that is completed before them. For example, when baking a cake, you must first gather the ingredients and materials, then mix the ingredients together, then place the mixture in a tin, and then put it into the oven. To bake the cake before the ingredients have been gathered or mixed together is not possible.

Costing

A charge, or cost, is assigned for the use of resources. Human resources can be charged at hourly rates; other resources may have a fixed charge or a daily rate.

Quality control

Each project has a defined goal, and the project must meet all the specified requirements by completion. **Quality control** involves checking results against identified quality standards. It is just as important to apply quality-control techniques to processes as it is to apply them to products or other end results. Quality control on processes can cut costs and improve timelines.

Tools such as flowcharts, control charts, cause-and-effect diagrams and scatter diagrams can be used to monitor quality control. Of these, the **flowchart** is the most useful, especially for procedures.

The time for some tasks can be extended without affecting the completion time of the project. Other tasks might delay the whole project, even if they are a short time overdue. Such tasks lie on the **critical path** of the project, and extra resources may need to be provided to ensure that they do not run over the time allocated.

Contingency plans

You can create a 'perfect' project management plan; however, if there are some hold-ups with the completion of tasks, your plan can be thrown into chaos. Project managers look carefully to identify possible risks and work out how these potential problems could be overcome with minimal disruption to the overall project timeline; for example, equipment not being available at a critical time, proofreading taking longer than expected or a task proving to be more difficult than expected.

Contingency plans are a set of predetermined actions that a project team will take if some kind of disaster occurs. For example, if the new software is not developed in time for its scheduled implementation, staff may keep using a manual system until the new one is ready. Meanwhile, the software installation task could be delayed while another task, such as hardware delivery, is brought forward.

Tools for managing projects

You can draw on a number of tools to assist you when planning projects. Most are designed to show different aspects of a project's status, such as the time taken as opposed to which tasks depend on one another. Therefore, the project management tools discussed below should be used to complement one another; not in isolation. Project management software does not create tasks and assign resources. You must make informed decisions, and the software will aid the management, documentation, presentation and communication of project information. You can update the documentation easily when changes are made during the course of the project.

Project table

The project manager will brainstorm, on a separate sheet of paper or word processing document, all the possible tasks involved in analysing a problem, and designing, developing and evaluating a solution (Table 5.1). Once **brainstorming** is complete, the project manager can use the task analysis outline shown in Figure 5.1 to order each individual task.

TABLE 5.1 The main headings of a project table for publishing a report

Task number	Task name	Duration (days)	Resources required	Task milestone (Y/N)	Dependent tasks (predecessors)
1	Departmental reports	5		No	N/A
2	CEO's report	3		No	N/A
3	Proofread	2		Yes	1, 2
4	Obtain artwork	5		No	N/A
5	Cover design	3		No	4

Gantt chart

A Gantt chart provides a standard format for displaying project schedule information by listing project tasks and their corresponding start and finish dates in a calendar format (Figure 5.1). Gantt charts not only show a timeline for completion of the project, but they can also highlight

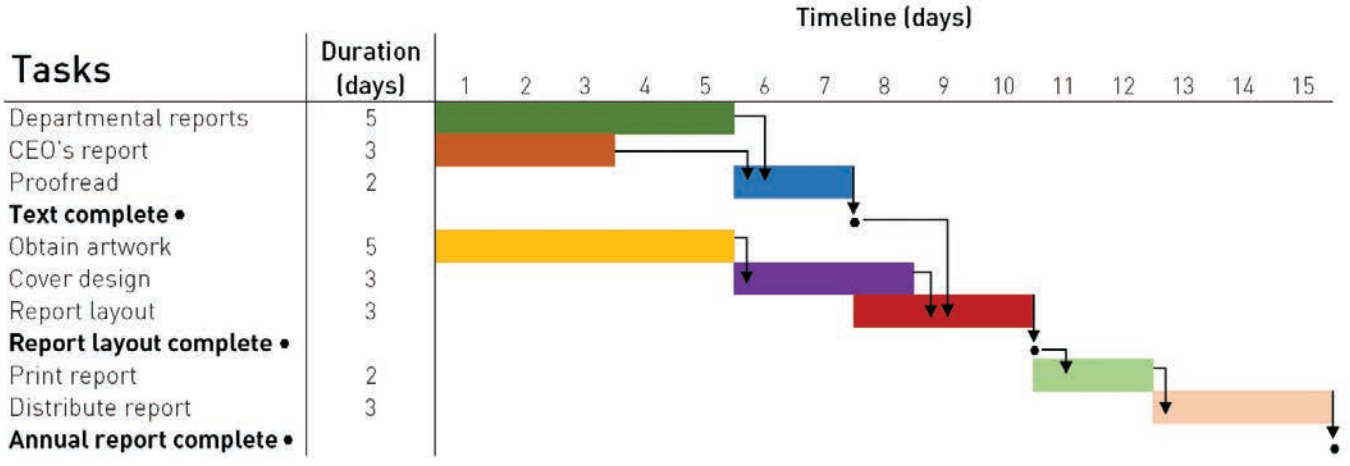


FIGURE 5.1 An example of a Gantt chart created in Microsoft Excel. Simple dependencies have been shown with arrows. Different colours can be used to make it easier to identify individual tasks.

tasks that are critical to the timely completion of a project. Using a Gantt chart makes it easy for the project team members to see when tasks need to start and how long they should take. Many Gantt charts also show milestones and basic task dependencies. Although you can use dedicated project-management software to generate Gantt charts, simple ones can be constructed in spreadsheet programs.

Groupware

When people work together to develop a solution to an information problem, they often do so in sequence. One person may conduct research, the findings of the research are considered for input and manipulation, and may be then passed on to the person performing the data input, and so on to the next person in the work cycle. An example of this might be a team working on a website, such as your group for your Outcome. The team must wait on modifications to the design from each team member before a final design can be agreed on and developing begin. **Groupware** is software that is designed to allow workers to collaborate on tasks, whether they are in the same building or spread across the globe.

A key part of collaboration is file **version control**. Consider the importance of file names that incorporate version information. With many people working on files, it is easy for someone to be working with out-of-date information, unaware that revisions have been made. In addition, workers may accidentally overwrite a newer version of a file with their own, older version.

Groupware applications, such as Microsoft SharePoint, allow project managers to keep control of document versions. Moreover, using a sequential file-naming strategy and storing working files in a central location that all team members can access will help to reduce the problem of working with outdated information.

File-naming conventions are discussed further on in this chapter on page 201.

Overview of Microsoft SharePoint for Office 365

Designing websites

When designing a website, it is best to follow a few key principles for effective design.

- Place most text and images within the dimensions of the screen size to minimise scrolling.
- Make the pages no longer than two average screen lengths. If the content requires more than two screens, divide it into several pages.
- Emphasise important words, not every word, so use bold and italics sparingly.
- Avoid underlining text because it can be confused with **hyperlinks**.
- Navigation buttons should be consistent and every page should link back to the homepage (index page).
- Use thumbnails and small images where possible and appropriate to reduce page loading times to keep visitors happy. If you want large image options, give users the options of clicking on thumbnails and loading images in full size.
- View your site in different browsers and different OS. Different browsers and systems always interpret code a little differently.
- Remember that generally, the colour of hyperlinks changes if a user has clicked it before. Indications of past navigation helps users to understand their current location and consider where to go next.

Lucidchart is a free online tool that allows users to create diagrams

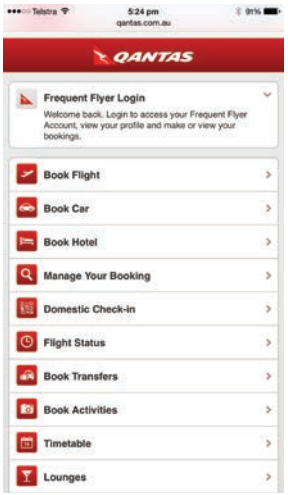


FIGURE 5.2 Qantas website on a smartphone

Designing for mobile devices

Characteristics of effective user-interface design

User-centred design is a methodology that focuses on the needs and characteristics of users, and is applied at the beginning of the design process to ensure that the website is useful and easy to use. The user interface (UI) is the screen that users see when they interact with a device. User interface designs can relate to websites, blogs, computers, mobile devices, ATMs, information kiosks, airport check-ins and any other device that a human interacts with. Mobile devices have become increasingly popular and are a rapidly growing technology. They are transforming our daily lives and providing the means to access the internet without having to use a computer as our primary means for accessing the internet. Given that websites are increasingly accessed through mobile devices, website designers need to carefully consider their interface designs and how they will be viewed on a mobile device (Figures 5.2 and 5.3).

As interactive touchscreen devices such as tablets and smartphones have become prevalent due to their mobility and the low dependence of data entry, these devices present some

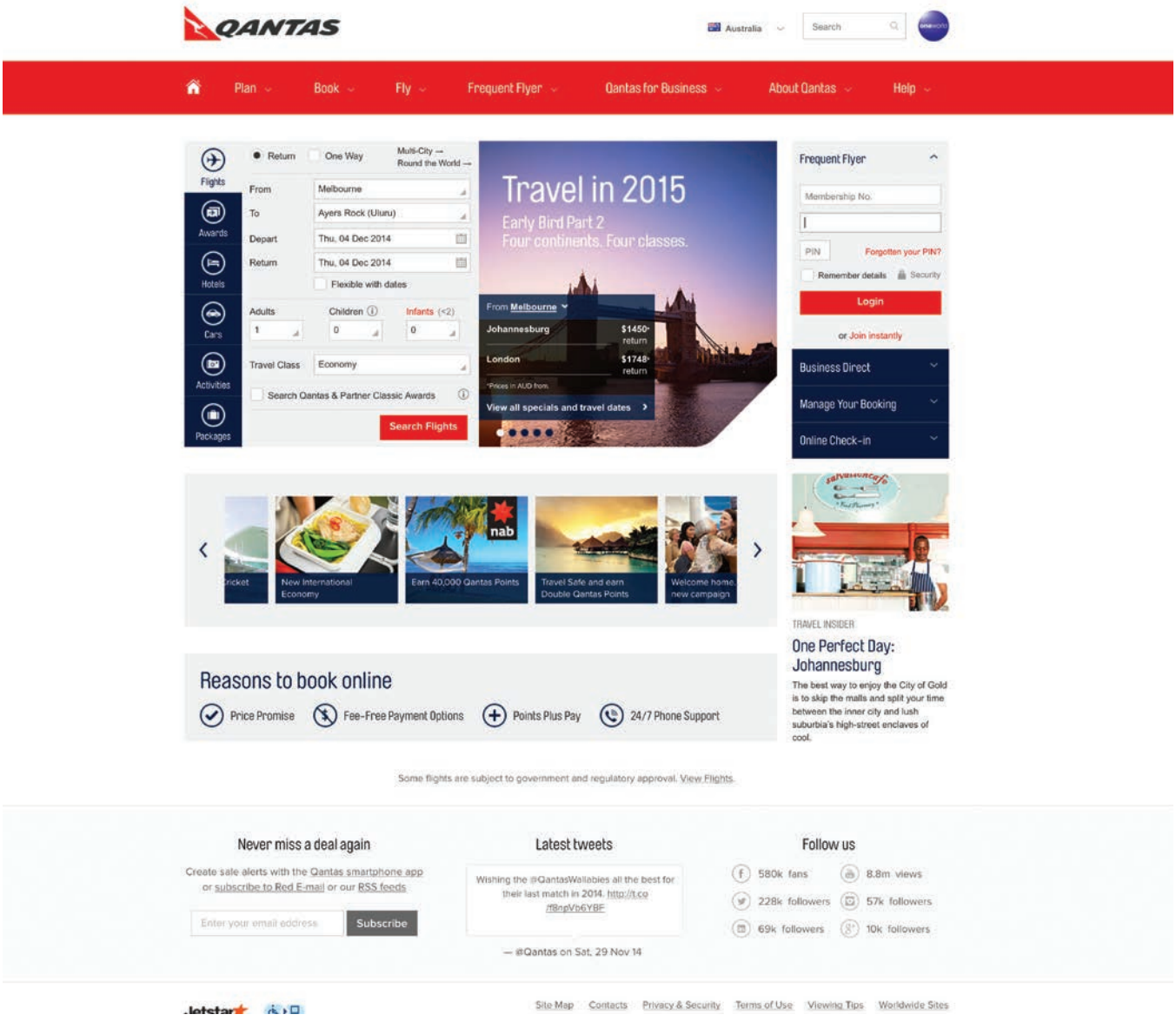


FIGURE 5.3 The Qantas website on a desktop looks very different to the mobile version.

useability challenges. The issues associated with mobile devices include, small, low-resolution screens, limited options for input, unpredictable connectivity issues and lower hardware specifications. Many websites have been designed for full-scale laptops or computers and have not been considered for mobile users.

Useability

Developers of websites need to understand their users, in particular their goals, skills, preferences and propensities, before designing the website.

Useability of a website is a quality or characteristic that represents how easy the website is to learn and to use. It is based on a set of characteristics aimed at designing usable and accessible websites constructed on user-centred design.

When designing the interface, developers need to ensure that it is kept simple. That is, the language is clear on the labels and there are no superfluous elements to distract the users.

Designers need to consider the consistency of common elements throughout the website. Users need to feel comfortable by being able to pick up patterns in language, layout and design throughout the site. For example, if the front page is called 'home' then all references to this page need to be called 'home'. Sometimes, instead of words, a graphic may be used to direct users to a particular page; for example, a picture of a house may be used to indicate the homepage.

Touch zone

When designing mobile interfaces, the touch zones on the website need to be big so that users can easily and accurately tap the targets. Small touch zones or ones grouped closely together make it harder to manoeuvre, as they require more accuracy. Sometimes users have to change the finger used, a thumb, for example, is much larger than the target, and an index finger is slimmer, which can assist with accuracy. The size of a touch zone should not cause users to make errors; therefore, increasing the size will help users avoid mistakes.

Increasingly, users of mobile devices walk while using their device, and therefore context and walking is another constraint. If you need to use your mobile device while walking, then the touch zone has to be bigger to accommodate the possibility of errors. The text of the page needs to be even bigger to allow for the jumping and moving around.

An MIT Touch Lab study found that the average width of the index finger is 1.6 to 2 cm (16–20 mm) for most adults. This equates to approximately 45–57 pixels, which is wider than what most mobile guidelines suggest. There are inconsistencies between Apple and other vendors' guidelines; however, Apple's guidelines are for the touch zone to be approximately 1.55 cm to allow for an average human finger size when interacting with the device.

Read-tap asymmetry

A read-tap asymmetry is where the font is large enough to read, but is too small to touch. While we can read very small fonts, they need to be much larger for us to touch them precisely; therefore balancing this causes everything to increase in size on mobile devices. Using large touch zones on websites will help users of mobile devices to have a positive experience.

Accessibility

People with a wide range of abilities need to be able to use interfaces designed for mobile devices. Interfaces need to be operable, understandable and perceivable for people with all levels of abilities including all disabilities, or limitations such as visual, auditory, physical, speech or cognitive impairments. For example, people who have visual impairments or cognitive disabilities can zoom or magnify the entire screen so that text appears larger.

Accessibility is also a design principle. See page 184 for more information.

The World Wide Web Consortium, sometimes known as WC3, is a widely accepted internal group that works with the public to develop standards and guidelines for the Web. One set of guidelines developed by WC3 refers to web accessibility for disabled users.

Some commercial companies sell software that can check websites against these guidelines. For instance, Bobby software is an application that acts as a spider. It will 'crawl' through a website to check that it meets all accessibility requirements. It then produces a report, which the website owner can act on to improve their site's accessibility.

WC3 standard and guidelines for the web

Accessibility can also refer to technological barriers such as a lack of resources or choice of technologies. For example, internet access can vary even when using broadband, and if users need to download large files, this may not be accessible to those who have a slow connection. Sometimes the location of the house and its distance from the telephone exchange mean that broadband is not available. As a result, the household must rely on mobile wireless data, which is costly.

THINK ABOUT COMPUTING 5.1

1 Visit the World Wide Web Consortium website and look at the five guidelines for Web accessibility. Choose three and write a brief explanation of each.

2 Do you think that these five WC3 guidelines should become law? Discuss.

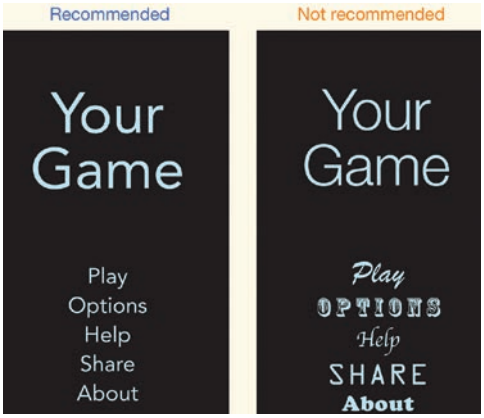
Tolerance

Websites that are well designed should prevent users from making mistakes. Although mistakes will always occur, a tolerant user interface should let users recover from their mistakes. For example, if you forgot your password, how many times will you be allowed to retry it before the system penalises you? Another example of tolerance is when the software auto-corrects spelling mistakes in user messages – the software can anticipate and correct most typing errors. A final example is when conducting a search on Google, the user types in what they are looking for and, even with spelling mistakes, Google refines the search and produces variations.

Legibility

The use of typography, that is choice of fonts, sizes and arrangement of text can assist to create hierarchy and clarity. It can also increase the readability of the site. Developers of user interfaces should take into consideration the following.

- There needs to be sufficient contrast between the use of colours so users can differentiate between contexts. One way to test this is to use the device in varying light conditions such as indoors and then outdoors on a bright sunny day.
- Do not use a combination of red and green buttons to distinguish between two options, as colour-blind people struggle to tell the difference.
- Research the use of colours, as some cultures and countries perceive colours in different ways. For example, in Chinese culture the colour red signifies good luck; whereas white signifies death and mourning.
- Text should be legible. Ideally, body text should never be smaller than 11 points, even when the user chooses the extra-small text size.
- In general, use a single font. Mixing several different fonts can make your website seem fragmented and sloppy. Instead, use one font and just a few styles and sizes.



https://developer.apple.com/library/ios/documentation/UserExperience/Conceptual/MobileHIG/ColorImagesText.html

FIGURE 5.4 Stick to one font throughout the development rather than many different ones

Visibility

By giving thought to the placement of items on the page based on importance, a designer can draw attention to the main pieces of information and can also assist with readability. The use of colour, light, contrast and texture can either attract users or redirect their attention from items. Effective visibility provides prompts and cues, which can assist users through an interaction, guide them through a series of steps, indicate the possible options available to them and communicate the context of the situation (See Figure 5.5).

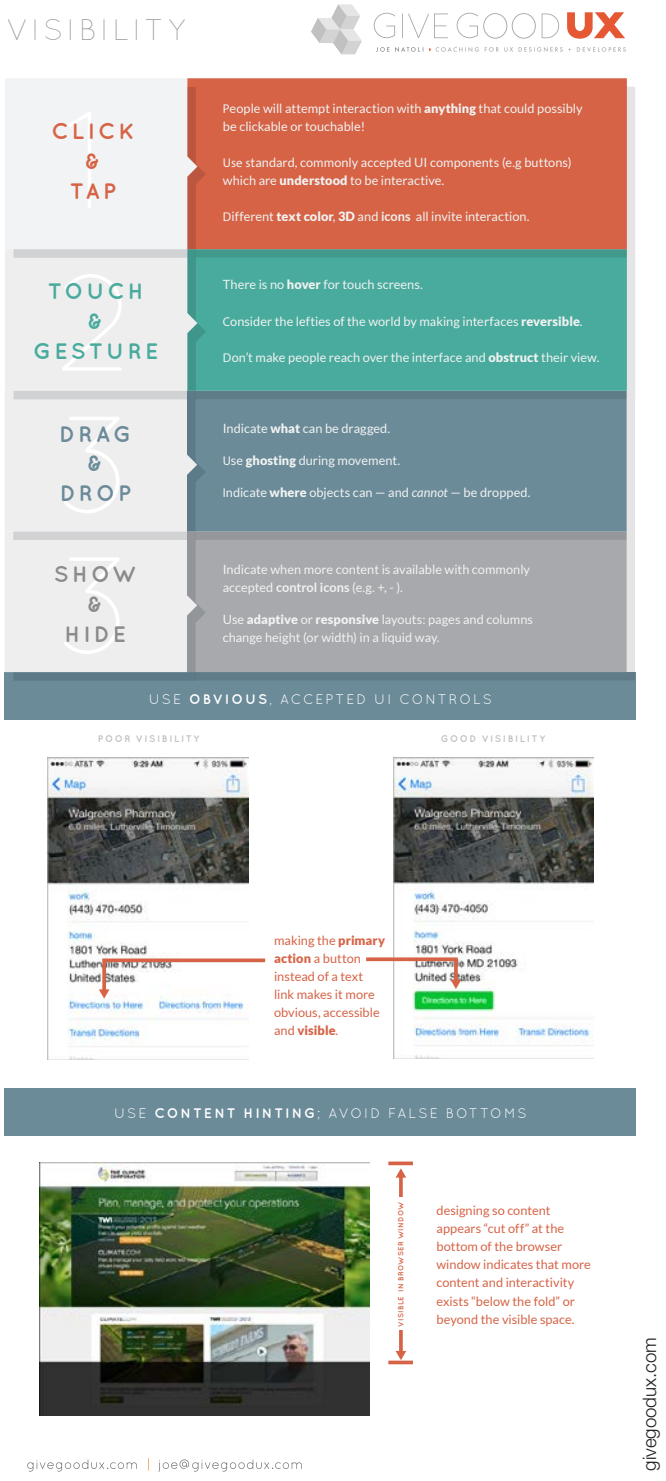


FIGURE 5.5 An infographic on visibility

Useability – US government website guidelines

Consistency

Just as we mentioned that consistency is important in navigation design on page 173, it is also important to use common elements in a user interface to ensure that the site is consistent. By using common elements, users are able to navigate the website confidently and comfortably. Consistency of user interface involves creating patterns in language, layout and design as can be seen in the Useability weblink.

THINK ABOUT COMPUTING 5.2

How many possible actions can you name from the Tumblr home page (Figure 5.6)? For example:

1 Input an email address

2 Input a password

3 Click the privacy link

Affordance

An affordance is a desirable property of a user interface. It logically and naturally leads people to take the appropriate steps to accomplish their goals. Affordances indicate what is achievable and can take many different forms such as shape, texture or size. For example, a toaster slot affords the insertion of bread in the slot; a door knob affords the turning of it. Affordances provide subtle clues on how to interact with objects. Web and mobile interfaces must gain all of their affordance through design. For most designers, this is intuitive and instinctive, based on the thousands of design patterns seen every day. In terms of a website, the solution should provide strong clues to the operation of things. The user should know how to operate a control just by looking at it rather than looking for words or instructions.

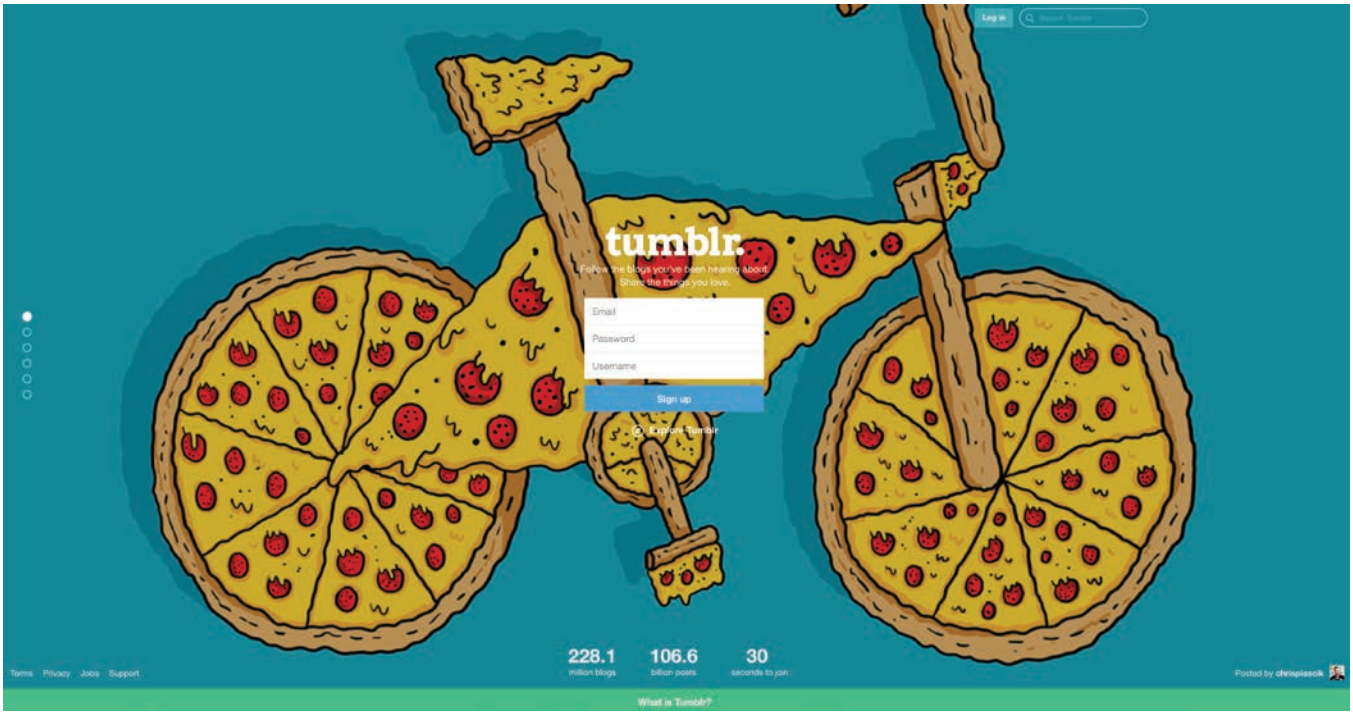


FIGURE 5.6 The homepage of Tumblr has many actions that are affordances.

On the Tumblr home page (Figure 5.6) there are many actions that are affordances, which interact with the way that an interface 'affords' its capabilities to the user. By identifying and recognising patterns, users are able to work their way through the website with little guidance or direction, because of the capabilities of the user interface.

Skeuomorphism is a design principle in which design cues are taken from the physical, real world; for example, the use of folders and files graphics for computer filing systems make them feel recognisable to users. Affordance cues are still evolving; however, we are moving away from skeuomorphic towards digital cues.

Information architecture

When designing your website for Unit 1, Outcome 3, you should consider the site's information architecture. Information architecture refers to the ways that content, which is made up of information and objects, is grouped, labelled and located in online solutions. In other words: how a website is structured and its navigation pathways. If users cannot navigate easily through a website, they will become frustrated and soon leave. Effective website navigation is crucial. If you consider and use the key principles governing information architecture when designing your website for Outcome 3, you will design a website more likely to enable users to intuitively and confidently locate the information they require. The key principles of information architecture include disclosure, classifications, navigation, growth and choices.

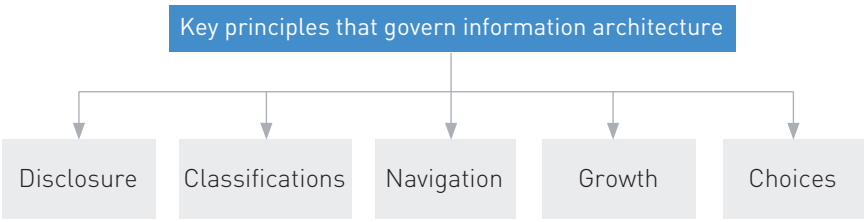


FIGURE 5.7 The key principles that govern information architecture

The user interface is everything designed into an information system with which a user may interact. This can include both hardware and software components. Examples of hardware UI components include the display screen, keyboard, mouse, tablet, stylus and touchscreen. Examples of software UI include the operating system (such as Windows, Linux or Mac), and application software.

See Figure 5.31 on page 199 for an example of a layout diagram.

An interface designer will utilise design tools such as layout diagrams to plan the appearance of onscreen output. Layout diagrams show the placement of important elements on each screen, including headings, graphics, text, hyperlinks and search fields.

Business and government websites would have differing information architecture (Figures 5.8 and Figure 5.9), so it is important that information is structured correctly. You should plan the information architecture of your website carefully using appropriate design tools, such as site maps and multimedia storyboards. These tools are used to plan the way the information is structured, linked and navigated. It is equally important that your website be designed for viewing on mobile devices.

Text-based information, in particular, requires information architecture because navigating through files relies on user interaction. Make sure the structure is logical for the user. You can make video-based information more self-contained and self-guided, but the information architecture must still be well planned and structured. As with the architecture of a house or building, careful forethought and design for information architecture will lead to a solid structure for the solution.

The following section discusses each of the principles of information architecture in turn: disclosure, classifications, navigation, growth and choices.

Disclosure

In information architecture, the principle of disclosure is about presenting only as much information as is necessary to complete a given task. The users of your website will only be able to process so much information at a time.

You have to limit the information, emphasise the important points and present only what your users need. If you go too far and try to give them all the information you have at the same time on the same page, you will create information overload and the importance and emphasis will be lost in the 'noise'.

The key is to show less information, but make sure it is the right information, and offer your users options to access more content if they want to.

Dropdown menus, +/– icons that open and close a hierarchy, alt text and thumbnail images that enlarge to full-size images when moused over are all simple ways of using disclosure to your advantage.

Classifications

Users do not always use the same terms for searching or browsing. Information architecture recognises that human beings are prone to classifying things in different ways and that no single classification scheme will be right for everyone. It is wise to use multiple classification schemes on your website so you can offer users more than one way to find content.

Classification schemes for online content include the following.

- Alphabetical
- Audience
- Formatting
- Geographical
- Organisational
- Subject/topic
- Task
- Combination

UX Booth: Classification schemes – and when to use them

CLASSIFICATION SCHEMES

1 Research classification schemes online. Read the information at the UX Booth weblink as a starting point.
2 Choose two of the classification schemes, other than alphabetical, and find three websites, not from the weblink, which use that classification scheme. Take screenshot examples.
3 Summarise and justify how each website uses the classification scheme indicated. Hint: Annotating the screenshots may help.

Navigation

When you have planned the organisation of your website thoroughly, you should plan your navigation paths accordingly. For example, imagine that you were a shoe retailer and had the following five webpages to add to your new website:

1 Job vacancies
2 Privacy policy
3 Grand opening sale
4 Items $20 and under
5 Mary Jane shoes sizes 8–10

You would never put links to all five of these webpages in the same menu, because they are not of the same category. This principle goes together with the principle of classifications. You need to focus the navigation on your website so that menus group similar items together to make the most sense to users.

As a shoe retailer, you could probably group 'Job vacancies' and 'Privacy policy' in the same menu, because these are often located in the headers or footers of websites and users expect these to be unobtrusive but easy to find if they scroll on the homepage.

'Items $20 and under' and 'Mary Jane shoes sizes 8–10' could be grouped in the same menu – but only at a stretch. One indicates a price ($20 and under) and the other a category (Mary Janes in sizes 8–10) – but they both also indicate a product: shoes.

The only solo item would be the 'Grand opening sale', which would occupy prime real estate on the front page.

By the time you are at the stage of developing the navigation for your website, you will know a great deal about how you have organised its content. Have you organised it alphabetically, by category, time, location or by using some kind of hierarchy? This organisation is relevant to how you develop the navigation pathways.

If you wanted to be able to view both 'Items $20 and under' and 'Mary Jane shoes sizes 8–10' in the same menu (shoes), you could instead use the **principle of disclosure** by adding all products to one page, and apply filters (by size or shoe style) and sorting (by price) to show only the desired items.

Growth

When building a website, you should assume that the content you have today is only a fraction of the content that will eventually be there. Websites keep growing and changing. Most are perpetually under construction. You need to organise content on your website in a way that allows for growth. Consider a simple blog – even this is constantly growing and changing as entries are written, published and commented upon. Blogging software needs to be written to allow for such growth. On the opposite end of the spectrum, websites with massive volumes of sales change rapidly as new products are added, out-of-stock products are removed, promotions are created and new stores-within-stores (such as Amazon's Kindle storefront) go live.

If you do not build a website with the principle of growth in mind, you will be faced with needing to retrofit your design to accommodate it – or you will have a website that will not be easy to update and will slowly stagnate and die.

You should try to make sure you can:

1. Easily add the same type of content to an existing category (for example, add new blog entries to a blog)
2. Add a different type of content to an existing category (for example, add video or audio to a category that only had blog entries)
3. Create a new category of content.

Choices

Sometimes when you are given too many choices for something, it can be harder to make a decision. The longer the dropdown menu is, the less it is will be used. The principle of choices in information architecture is about minimising the number of choices that users need to make at each level, especially the top levels. You do not want to make your users sort through all of the options in the menus to find what they really need, which is what will happen if your menus are too long and are unfocused.

Make shorter, meaningful lists at the top level. Use tabbed, expanding lists, sub-lists and hierarchical levels to show deeper levels of information where needed. Minimise the range of choices that your users have to make at each level.

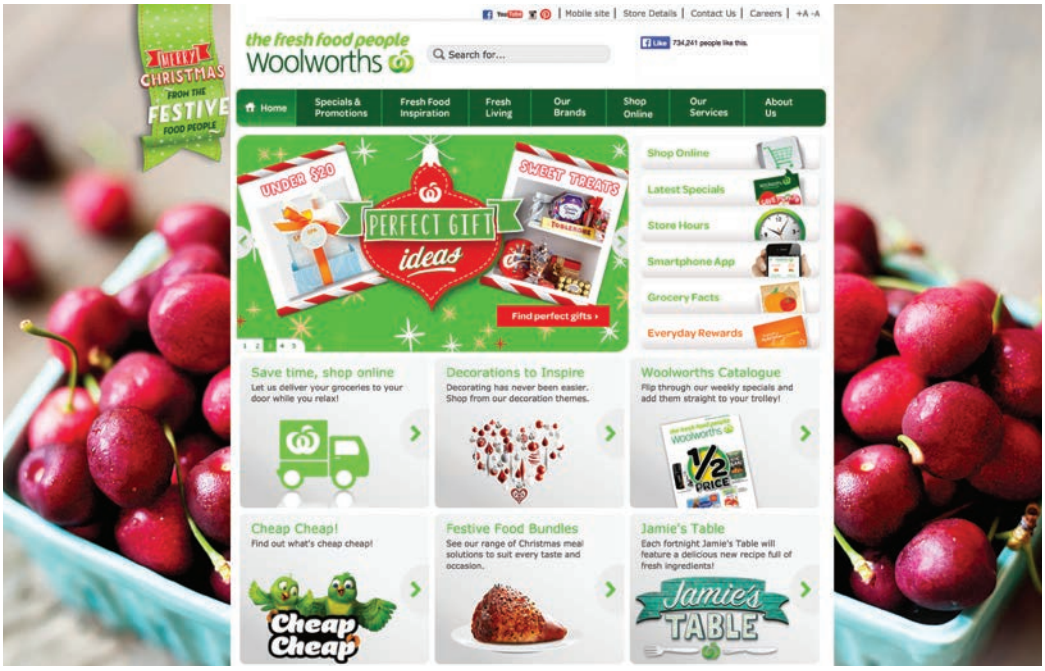


FIGURE 5.8 An example of commercial website architecture

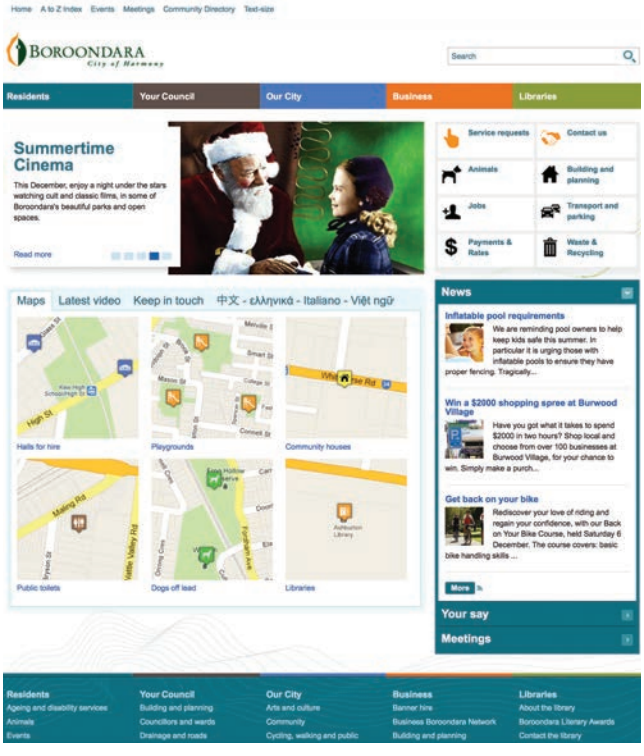


FIGURE 5.9 An example of government website architecture

THINK ABOUT COMPUTING 5.3

Find another example of a government website and see if there are any comparisons between the features in Figure 5.9 and your example. What makes a government website different from a commercial one?

Design principles

Your website for Unit 1, Outcome 3, should be carefully designed. To communicate your issue effectively, it should be easy to access and navigate with minimal time and effort, as well as visually clear and functional.

If your website is not functional, it will frustrate users (Figure 5.10).

FIGURE 5.10 The Liberty Van website is a good example of a website with poor navigation and random links

Design principles are factors that enhance the appearance and functionality of graphic solutions, such as websites. You need to ensure that your issue is obvious and your message is unmistakable. Your website must be carefully designed, taking into account the importance of design principles.

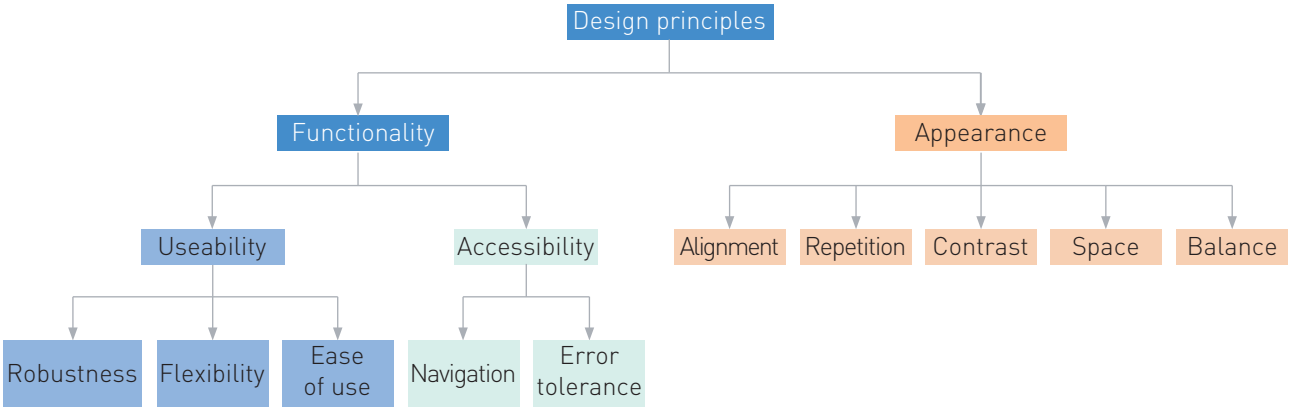


FIGURE 5.11 Design principles are accepted characteristics that contribute to the functionality and appearance of solutions. For VCE Computing, the principles that relate to functionality are useability and accessibility. For appearance, they are alignment, repetition, contrast, space and balance.

Website users want a well constructed website that provides the necessary information and does not make them waste time sifting through pointless material. They also want a website that loads quickly.

The principles shown in Figure 5.11 are discussed in the following section, in terms of how they may apply to the website in Unit 1, Outcome 3.

Functionality

Useability

Useability refers to a website's capability of being used and its availability to a user. It is important that any website incorporates maximum compatibility with a wide range of devices, including future technologies, so that users are not tied to a particular platform or device. This is one reason why websites and apps are designed to work on desktop PCs as well as tablets and smartphones.

If you ensure that users cannot view or access any data structures or important libraries during browsing, you are also abiding by the information architecture **principle of disclosure**.

Robustness is a website's ability to cope with errors during use. In your website, this may mean giving a user the option to correct data or undo an action, step backwards or forwards, and provide accurate and unambiguous error messages. You should also make sure that users cannot access any data structures during browsing because this may introduce bugs into your code and cause further website errors.

Flexibility relates to information architecture's **principles of classifications, navigation** and **choices**.

Flexibility is a website's ability to support multiple ways of performing tasks. For example, many websites were once designed with fixed-width dimensions, such as 800px. These websites only supported a single way of viewing the website – using a browser that supported the fixed-width page size. With so many different browsers, operating systems and devices in use today, it is better to build websites that are flexible, without fixed dimensions, and instead let the user determine the page size.

Scalable type, which is text that can be scaled up or down in size, is also desirable. This is especially important for Outcome 3, because you are creating a website designed for viewing on mobile devices. Device independence and liquid layouts are also desirable. View the Peach Pit weblink for more information on designing for liquid layouts, device independence and scalable type.

Peach Pit: Useability tips you can use: Designing flexible layouts

Ease of use refers to how user-friendly your website is. Some of the user-friendliness aspects overlap slightly with the appearance design principles. However, you should also ask questions when creating, planning and testing the website.

- Is it easy for users to intuit the design of your website and perform basic tasks?
- Can basic tasks be performed quickly?
- Will users remember how your website works when coming back after they have not visited for a long time?

If it is hard to learn the design of your website – *perhaps* because it differs greatly from other websites – it can be very confusing to users and they will be tempted not to return. If accomplishing basic tasks is too hard users will be reluctant to come back: have you ever felt frustrated simply because a website has required far too many clicks on one page just to transition to the next page?

Ask yourself and your group the tough questions about your website. You should make things as easy as possible for users. Too much scrolling, too much clicking, pages that pop-up unexpectedly and websites that are designed in unexpected or unpleasant ways can all reduce ease of use.

Simply being *different* from other websites does not necessarily mean your website will be confusing. If your design is logical or makes sense to users in other obvious ways, it may still work. You never know: your new website design may be a stroke of genius.

FLEXIBILITY

1 Using the Peach Pit weblink as a starting point only, research scalable type, device independence and liquid layouts. Visit at least five other websites.
2 In your Outcome groups, write a summary of:
   a device independence
   b liquid layouts.
3 In your groups, find one website on an issue similar to yours. View it on a mobile device and on a PC. In the table below are the principles of information architecture and design we have so far discussed. In the right-hand column, score each website between 0 for not at all and 5 for very well, in terms of how well you think it is applying each of the principles. You do not have to agree with the rest of your group, but you should discuss your opinions.

	Score (0–5)
Principle of disclosure	
Principle of classifications	
Principle of navigation	
Principle of growth	
Principle of choices	
Useability – robustness	
Useability – flexibility	
Useability – ease of use	

Accessibility

Accessibility as a website design principle really means making your website easier to use for everyone. With so many different versions of browsers and plug-ins available, avoid using special effects created in software such as Flash, Java and JavaScript as the only navigation method. (Particularly because devices such as iPads are unable to play Flash.) Your website needs to be accessible to all visitors, and this is especially true of navigation techniques. Interface designers should design onscreen user documentation that can be read in common software applications, such as Adobe Acrobat Reader. Design your website to be compatible with any browser, and do not require a specific screen resolution or plug-in for correct viewing.

Error tolerance as an aspect of accessibility means helping your users to avoid and correct their mistakes using clear instructions. It also means making it difficult for them to make errors by not allowing them to perform actions that could lead to errors in the first place. Grey out non-selectable options. Ask for confirmation of major actions. This is connected to **robustness**.

Navigation as an aspect of accessibility refers to ensuring that the navigation system is clear, simple and intuitive. For your website, it will also mean making sure that your website can be navigated comfortably by a touchscreen (so the buttons should not be too small and close together, making this difficult). Navigation has a number of other aspects that should be considered.

High-quality UIs are transparent, meaning that your users would not really notice the actual interface because it is so easy to use that they interact with it intuitively. The required information would be found quickly enough that users focus on it rather than the way they found it. The interface is a connection point or navigation tool rather than an end in itself, so it should be unobtrusive but clear.

You should also consider meaning, comprehension, consistency, and appropriateness and relevance in navigation.

Meaning

The labels of icons used for navigation should be meaningful, and the links need to be clear. For example, a school's internal newsletter is known by a particular name; however, visitors to the site who are not familiar with the school may find it hard to find the newsletter if they do not know its name. Someone who has never been to a website before should know immediately where a link will take them.

Comprehension

If images are going to be used for navigation instead of words, these images need to be commonly used and universally understood. For example, when we think of traffic lights, red means stop; amber, caution; and green, go. An image of traffic lights is universally recognised without having to write the words next to each light. Generally, images used as navigation on websites are not universally recognised and can cause problems as visitors to websites do not know how to navigate through the site. It is a good idea to associate text with images so that the links are understandable.

Consistency

To ensure that the website remains consistent and has a similar feel about it on all its associated pages, navigation should appear on every page of the website (Figure 5.12). The basic structure of the navigation in the same location should remain consistent throughout the website, with only minor changes used to indicate location within the hierarchy. It is often desirable to have a link back to the home page. A template or style sheet is often used to provide this consistency in a website. This also relates to **repetition**.

a

FIGURE 5.12 Each page of a website should have a similar feel about it. The personal tab of the ANZ website (a) has a similar feel to the corporate tab (b).

b

FIGURE 5.12 (continued)

Appropriateness and relevance

Your website must meet the needs of the intended audience with appropriate information that is relevant to its purpose.

The website must be able to present the right information for its users in a usable format and in a timely manner. Superfluous information, or distracting but unnecessary extras, should be avoided. A skilled interface designer will produce a streamlined and functional but attractive screen in preference to a screen full of fancy and artistic but useless or irrelevant items (Figure 5.13). Users need efficiency and effectiveness in their interface, not a screen that shows off a designer's bag of tricks.

FIGURE 5.13 The Google search engine is uncluttered with no irrelevant features.

Appearance

The screen layout of your website, or appearance, should not be unnecessarily elaborate and decorative, or contain superfluous animations. Too many buttons, bullets, icons, rulers and flashing graphics will confuse the eye and distract the user. Users should also have the option of turning off any onscreen hints or help features that they do not require.

Long or wide screens that require users to scroll should be avoided when possible, particularly if scrolling is required both across and down the screen (Figure 5.14). This makes the information harder to view. Short screens of information, with links to other screens, are more easily viewed and therefore more effective.

http://www.angelfire.com/super/badwebs/

FIGURE 5.14 An example of a long-screen website

Alignment

Alignment is also discussed in Chapter 2, on page 53.

The alignment of text can be left, right, centre or fully aligned (justified), as shown in Figures 5.15 to 5.18. Choose one alignment for each page and stick to it for the entire page. For paragraphs, left-aligned text is easier to read than centred text because the text begins on the left-hand side every time. The text is in a straight line and readers can follow the text with their eyes starting from the left edge. Centred text makes the eye work harder to locate the start of each line. Unlike left alignment where there is a consistent straight edge for the eye to follow, there is no consistent focus point for eyes to return to once each line is read (Figure 5.16).

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nulla dapibus sit amet orci ut tristique. Mauris vulputate, orci et rutrum tincidunt, neque ante convallis elit, maximus posuere nisi risus quis ligula. Quisque sit amet urna ac leo feugiat mollis in nec purus. Donec fermentum, leo eu rutrum posuere, mi est pharetra est, a cursus orci velit ut magna. Suspendisse scelerisque, ex at luctus vestibulum, enim felis accumsan ligula, non tempor lectus mauris et urna.

FIGURE 5.15 Left-aligned text has a straight left edge and each line begins at the same point on the left edge.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec vestibulum, magna nec dapibus placerat, lectus dui volutpat ex, congue aliquet risus quam ullamcorper nisi. Donec placerat nisl tortor, ultricies faucibus ipsum consequat ac. Aliquam a erat eget enim egestas viverra vitae eget quam. Donec non augue vehicula, congue magna quis, pellentesque nibh. Sed sed ligula nulla. Vestibulum vitae justo et neque fringilla euismod. Sed pulvinar a elit eu dictum.

FIGURE 5.16 Centred text has a jagged edge and each line begins in a different point on the left edge.

Centred text is often used for headlines and short lines of text. Users can read them with ease because the lines are short and the eyes do not have to scan through many lines of text.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec eu est porttitor, fringilla dolor vel, elementum velit. Etiam ipsum justo, pellentesque vitae mauris sed, egestas fermentum mi. Proin posuere, dolor ac ornare tempus, dolor nisi blandit neque, sed blandit magna felis sit amet leo. Curabitur ultrices a sapien vitae dapibus. Morbi tempus lectus ac sem varius, a pretium libero pulvinar.

FIGURE 5.17 Right-aligned text has a straight edge on the right side.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus elementum elementum odio et vehicula. Pellentesque blandit, nibh sed vestibulum pretium, mauris nunc rhoncus elit, quis pellentesque nunc libero nec massa. Donec ut malesuada massa. Nullam sit amet lobortis mi. Maecenas consequat varius elit, vitae laoreet lorem venenatis ac.

FIGURE 5.18 Fully aligned text has a straight left and right edge and each line commences in the same point on the left.

Avoid using all uppercase letters, except in some headings. Long sections of text typed in uppercase are very difficult to read because all the words have a rectangular shape; it is also considered to be bad 'netiquette' as mentioned in Chapter 3.

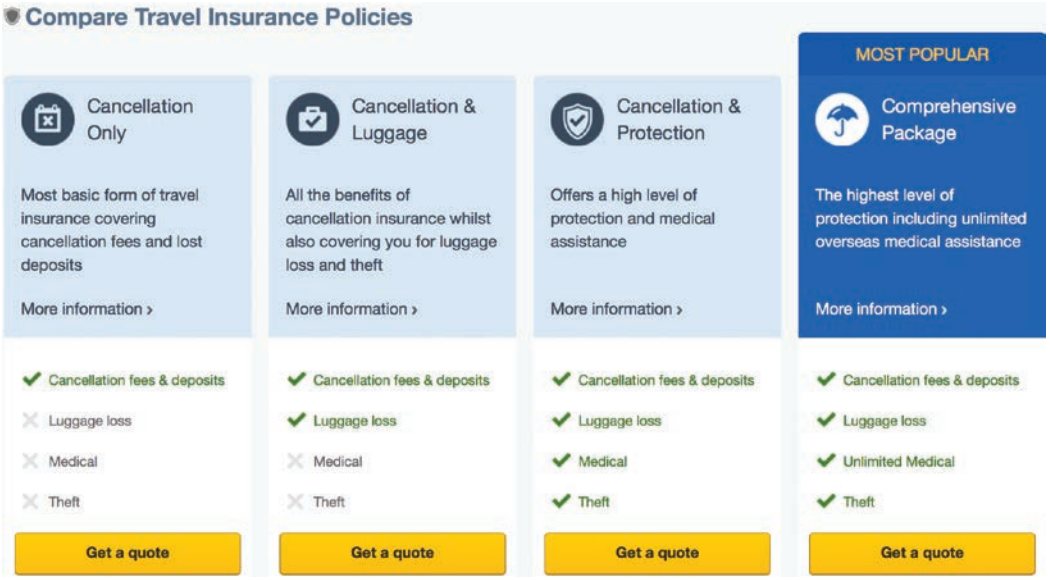


FIGURE 5.19 A website that uses left alignment

Repetition

Repetition is used to unify elements of a layout. This is achieved by repeating patterns, fonts, colours, images and page elements. On a website, use repetition so that users can understand how different pages or menus relate to one another, and to create a strong sense of identity and feeling of consistency.

Repetition is also discussed in Chapter 2, page 53.

The repetitive elements will help to create a sense of organisation to help the user interpret the meaning of what they are seeing. For example, looking at the Qantas image in Figure 5.3, three verbs – Plan, Book and Fly – are repeated in the red navigation bar at the top. Appearing in the navigation bar at the top is one form of repetition. Being in the same font is another. Being in the same form of wording – a short finite verb – is a third. This simple repetition tells the user that these three links lead to similar task-based pages.

Font selection

Most websites maintain the same font selection throughout the entire site. Sometimes different fonts are used in banners or headings, but consistent use of fonts and sizes (usually between 9 and 12 points, with body text at 11 points) is recommended.

Text styles are also discussed in Chapter 2, page 57.

Serif fonts are usually easier to read in printed works than sans-serif fonts. Serif fonts, which have small tails at the end of some letters, are used to make it easier to read, because they can guide the eye and reduce eyestrain when reading large blocks of text. Serif fonts include Times New Roman, Courier, New Century Schoolbook, and Palatino.

The differences between serif and sans serif typefaces

Sans-serif fonts, have no tails (are without serifs) and are better suited to the onscreen reading. Popular sans-serif fonts include Helvetica, Avant Garde, Arial, and Geneva.

Contrast

Contrast refers to the visual difference in colour or tone between objects (both text and images). Greater contrast will make objects stand out more from one another. If there is not enough contrast between two objects, they may blend into each other, making it difficult for the user to see each of them clearly. Contrast between the background of your website and text should make the information clearly visible and legible.

Contrast is also discussed in Chapter 2, page 53.

Colour

Most people do not use a scientific approach to colour schemes: they experiment and choose colours that they think will go together. Trial and error and a fair bit of reflective evaluation are usually required to ensure the colour scheme works. Remember that the colours selected for your website should:

- make it clear, legible and attractive
- enhance important features
- ensure easy navigation.

Use your chosen colour scheme on all pages to be consistent.

Colours can convey a strong message about your website. Table 5.2 shows a summary of the effects that different colours have on the emotions and impressions of users. For websites, what is of concern are how the colours contrast with one another (Figure 5.20). In terms of contrast, it is generally best to use a lighter, paler colour as a page background, with a darker colour for the text. This makes it easier to read the onscreen text, which encourages users to continue to browse the site. It is also important that the actual colours complement one another to achieve a visual harmony. A colour wheel can be used by web designers to find colour schemes that work well (Figures 5.21 and 5.22). Different but complementary colours may then be used for onscreen elements, such as headings and the main body text. Again, an overriding principle is that colour and contrast should be kept as simple as possible.

TABLE 5.2 The information you want to convey can be enhanced with the use of appropriate colours.

Colour	Connotation	Business application
Blue	Conveys the impression of calm, reliability and strength. Dark blue means 'strong and conservative'; light blue, 'youthful'.	In business, blue is associated with fiscal reliability.
Green	Reminder of nature, spring and the countryside. Less appealing traits: poison, decay, illness and jealousy.	Investment bulletins, order forms, training materials, flyers and labels.
Red	Powerful effect on the human eye. Symbolises blood, anger, fire, danger. Deep reds are aristocratic and stylish.	In business, associated with debt and deficit. Banks use red for withdrawal slips, inventory reports and other action documents.
Pink	Has feminine connotations. Pale pinks have calm, restful effects.	Business forms, such as financial statements and invoices.
Brown	Has a sense of practicality, comfort and realism. Sometimes can be seen as dull.	An elegant brown typeface on pale brown paper conveys tradition and quality.
Black	Traditionally in Western cultures, black symbolises the dark side: night, evil, death, despair and mourning.	It also has the modern image of distinction, sophistication and elegance.
Grey	Symbolic of status and authority. Subtle greys are assured, successful and elegant.	Used as background colour, grey is very effective as it provides a contrast to more vibrant colours.
Yellow	The most visible of all the colours, and is often a difficult colour to look at for a long period of time.	It makes shapes appear larger and whiter.
Purple	Sophisticated, mysterious, mystical, meditative, melancholy, dignified.	Traditionally symbolic of sovereignty and justice.

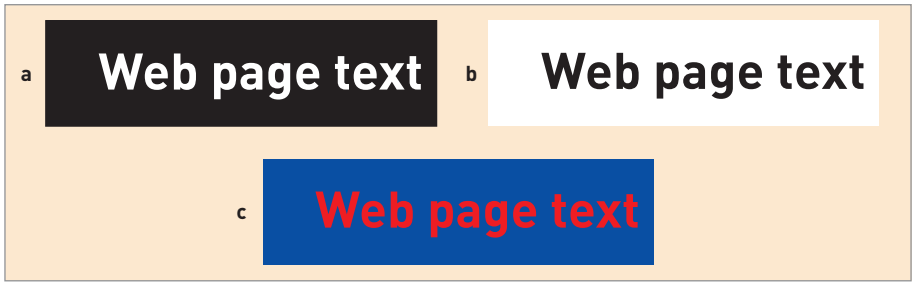


FIGURE 5.20 Illustration of contrast. Note how (b) is easier on the eye than (a) when the colours are reversed; (c) shows the problem of contrast with non-complementary colours – in this case, red text on a blue background.

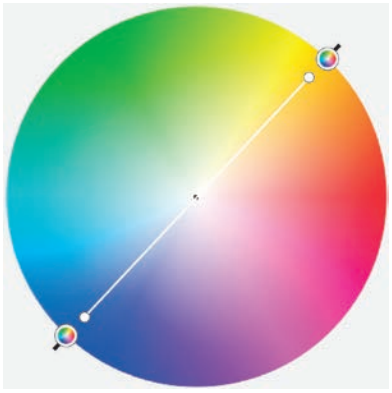


FIGURE 5.21 A colour wheel can show complementary colours. In colour theory, harmonious colour combinations use any two colours opposite each other on the colour wheel, any three colours equally spaced around the colour wheel forming a triangle, or any four colours forming a rectangle (two pairs of colours opposite each other). The harmonious colour combinations are called colour schemes.

FIGURE 5.22 The Adobe Color CC specialises in allowing you to make up attractive and interesting colour schemes for websites.

Conventions for onscreen colours can be helpful in determining colour schemes:

- The most easily readable colours for text are black writing on a white background.
- Avoid using red and green together because colourblind people have difficulty distinguishing between them. Blue and brown together can also be hard to read.
- Limit the variety of different colours on the screen.
- Light or pastel shades (or white) are the best choice for backgrounds. Avoid using bright, vivid colours, except for an area of the screen you wish to emphasise.
- Blue is a standard colour for an unvisited hyperlink, changing to purple when visited. If other colours are chosen for links, they should be consistent.

Traditional primary colours are red, yellow and blue; however, colours used by printers are magenta, yellow and cyan. Some colour printers are unable to reproduce the full range of colours visible to the human eye.

RGB

Combining red, blue and green (RGB) colours creates white, as can be seen in Figure 5.23. Consequently, RGB colours are often known as additive colours. These are mainly used for lighting, optics, video and monitors.

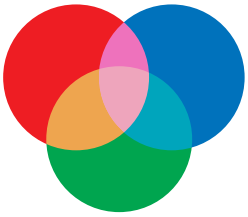


FIGURE 5.23 Red, green and blue (RGB) combine to create white.

To show colour on a screen, a **pixel** must have three coloured phosphor dots. These dots are the additive primary colours, red, green and blue. Each of the three-colour elements making up a single pixel may be set to one of 256 values (ranging from 0 to 255). A value of 0 means that the relevant colour is not visible at that pixel, and a value of 255 would cause the colour to be shown with maximum brightness. There are 256 possible values for each element, so the total number of colours that can be displayed on a (true-colour) screen is 16 777 216 (from 256 × 256 × 256).

Complementary colours are contrasting and stand out against one another. Often it is a good idea to use a complementary colour as the highlight colour. The complementary colour of a primary colour – red, blue or yellow – is the colour you get by mixing the other two: so the complementary colour of yellow is red + blue = purple; for red, it is blue + yellow = green; and for blue, red + yellow = orange. Even though colours are complementary, they aren't necessarily suitable because they might not be a good combination.

RGB (red, green, blue) and CMYK (cyan, magenta, yellow, black) are two major categories of colour representation systems.

Space

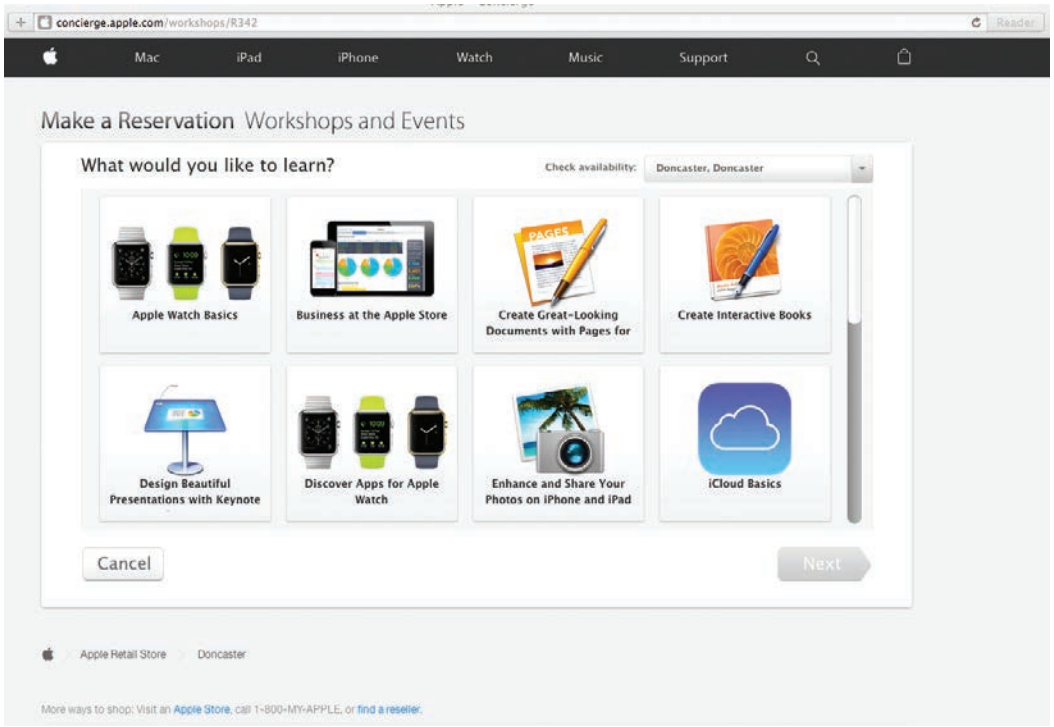
Space refers to the areas around and between objects – text and images. If your website is cluttered, it may be unpleasant to browse. Although you may want to include a lot of information on your website, you still need to put space between objects so they can be individually distinguished and navigated through correctly.

Space is also discussed in Chapter 2, page 54.

You need to space objects on your website so they are easy to perceive but not overlapped and obscured. The screen should not be so crowded with objects and features that the user finds it difficult to see the information they need.

White space can be used as a contrast around a chosen element of the page to draw the user's eye (see Figure 5.24). Within graphics, animations or videos, levels of colour and contrast should make the information clear and attractive. A common convention is also to avoid yellow or other light colours for text on a white background because this can be difficult to read on screen (Figure 5.25).

A large area of white space may be used to balance a section that contains an equally large area of text, because it will be of equal visual 'weight'. Working with space means also working with balance.



screenshot of webpage http://concierge.apple.com/workshops/R342 Copyright © 2015 Apple Inc. All rights reserved.

FIGURE 5.24 An example of a webpage that uses white space effectively. The website emphasises the eight pictures by using white space around them.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Pellentesque blandit nibh nisi, eget blandit sem tincidunt sed. Ut ac dolor at ipsum lobortis consequat. Vestibulum dignissim, eros quis fermentum pellentesque, odio ligula vehicula odio, a vestibulum nibh dolor non tortor. Phasellus vel libero vitae dui aliquam rutrum vel sed sapien. Fusce a diam porta, dictum ante vitae, vehicula purus. Suspendisse vehicula dapibus accumsan. Quisque at tellus nisi. Nulla dapibus ultrices ipsum eleifend dapibus. Nam facilisis pulvinar turpis eget lacinia. Fusce.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Pellentesque blandit nibh nisi, eget blandit sem tincidunt sed. Ut ac dolor at ipsum lobortis consequat. Vestibulum dignissim, eros quis fermentum pellentesque, odio ligula vehicula odio, a vestibulum nibh dolor non tortor. Phasellus vel libero vitae dui aliquam rutrum vel sed sapien. Fusce a diam porta, dictum ante vitae, vehicula purus. Suspendisse vehicula dapibus accumsan. Quisque at tellus nisi. Nulla dapibus ultrices ipsum eleifend dapibus. Nam facilisis pulvinar turpis eget lacinia. Fusce.

FIGURE 5.25 Avoid using light-coloured text on a white background as there is not enough contrast to make it readable.

Balance

A website with a balanced design is visually appealing. Websites with unbalanced designs can lack the appropriate emphasis, can look untidy and they may end up discouraging users from visiting them.

Balance is also discussed in Chapter 2, page 56.

All elements of a layout, websites included, have a visual weight. If the elements on either side or the top and bottom of the screen are of an equal weight, then visual balance is achieved. There are two types of balance: symmetrical and asymmetrical.

With symmetrical balance, the visual elements on each side of an imaginary horizontal or vertical dividing line appear to be exactly the same in terms of visual weight, right down to the proportions and shading.

Asymmetrical balance occurs where visually matched weighting occurs through a combination of objects of differing sizes, shapes and colours.

Specific design considerations

In addition to the design principles and information architecture, there are some other specific design considerations you may find useful.

Proportion

Proportion refers to the visual hierarchy of a page on the screen (Figure 5.26). The designer must ensure that the most important pieces of information have visual prominence on the page, with less vital information taking a less prominent position in the visual hierarchy. Headings, logos or links may be very conspicuous on a company webpage, and therefore obvious to users. Other information that the company does not wish to emphasise as much, such as the cost of an item, may be in a smaller font in a less noticeable part of the screen.

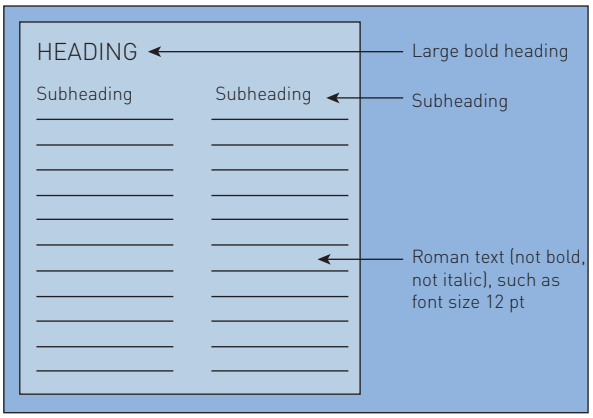


FIGURE 5.26 Visual hierarchy of a page. The main heading is large, bold and at the top of the page. Subheadings are bold. Normal text is roman. The user's attention is visually drawn to the heading and subheadings.

In the graphics industry, the top of a page is sometimes referred to as 'prime real estate' because it is the dominant location where important information will be most noticeable. This concept also applies to onscreen pages, especially for long pages where scrolling is needed. The bottom of a long page is low in the visual hierarchy, so the most vital information a user needs to see should not be placed there.

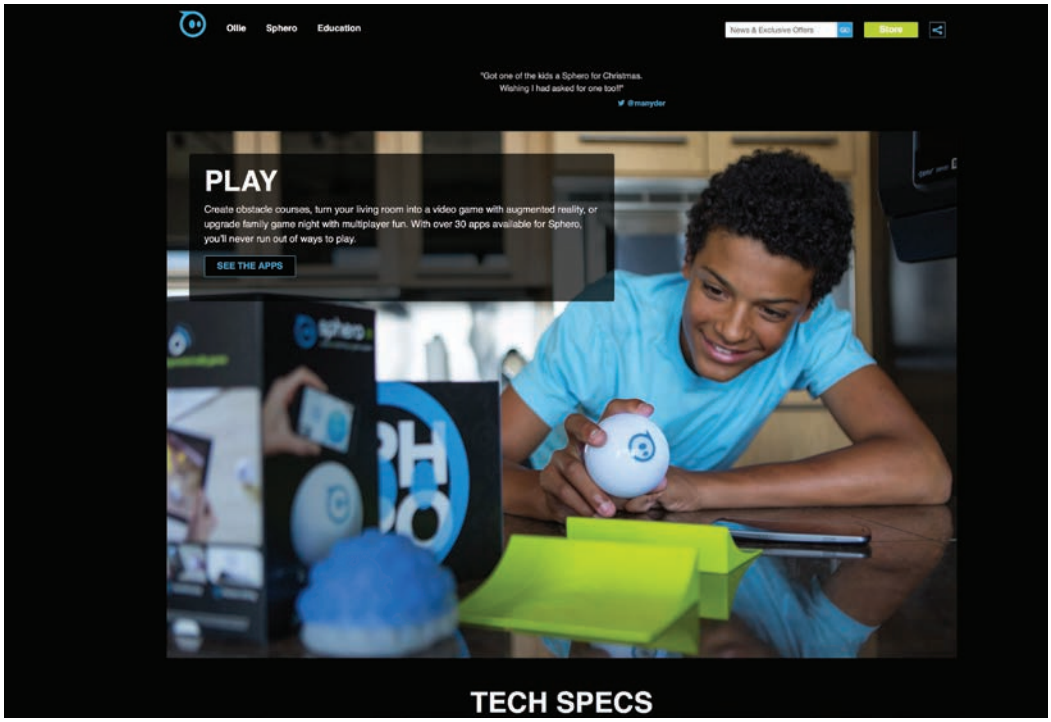
The proportions, or relative sizing, of fonts and graphics affect the visual hierarchy of a screen. The positioning of particular page elements – including white space around objects, the formatting of fonts and the use of animation – also affects the prominence of an element in the visual hierarchy of a page.

Orientation

In designing the layout, the interface designer will consider all elements that are to be included on the screen, taking into account the visual hierarchy or prominence of various items. They will orient objects on the screen and decide the direction that each screen element will face. An important consideration is the aspect of each screen element in relation to other objects on the screen, and to the screen overall. Some typical considerations for **orientation** of screen elements could be the direction that a logo is to face; whether or not a graphic should be flipped

over as a mirror image; whether a frame on a website should be at the top, bottom, left or right of the screen; and how text should be aligned (Figure 5.27).

The size of an onscreen page needs to be considered in the design for orientation of various screen elements. The orientation and **juxtaposition** of screen elements become extremely important.



Sphero

FIGURE 5.27 The designer of this website has used the picture to feature the product and its logo (flipped slightly to the logo on the top left). Text is aligned to the left and in a straight line from the edge of the Sphero logo.

Design tools

A good solution will undergo a rigorous design stage before any attempt is made to implement it. The solution must address the stated problem and provide all the required information. To design a solution to an information problem, you need to choose and justify the choice of your preferred solution, develop detailed design specifications, identify the appropriate procedures and techniques to be used, choose the test data, and correctly apply appropriate **formats** and conventions to the solution.

In the design stage, you focus on creating a plan of the solution. Designing a solution often occurs without using a computer at all. Much time can be wasted if a design has not been properly planned before work starts on developing the solution. As stated above, the two activities associated with this stage are first designing the solution and then designing the evaluation criteria.

In Unit 1, Outcome 3, you are creating a website. Generally, you would follow the steps below.

1 Identify the data you require.
2 Choose an appropriate web-authoring software.
3 Decide how to manipulate the data.
4 Determine the appropriate UIs and apply appropriate information architecture.
5 Develop the site map.
6 Construct storyboards.
7 Choose layout designs.

8 Choose test data.
9 Choose conventions and apply formats.
10 Consider the design principles that influence the appearance of onscreen solutions.

Many of the steps require you to make use of specific design tools. These tools can be considered as addressing either the functionality or appearance design principles. The following sections address each step and the design tools necessary to perform them.

Functionality

Identify the data you require

The PSM is also discussed in Chapter 2, on pages 31–42.

In the first step, you gather the data required, but new material may be need to be produced during the development stage of the problem-solving methodology (PSM).

Choose an appropriate web-authoring software

In the second step, you will be choosing an appropriate web-authoring software. The choice of software will be strongly influenced by the nature of the issue. If, in the analysis stage, it was decided that the creation of a webpage was the most suitable solution, the appropriate software application would be a web-authoring package, such as Dreamweaver. However, sometimes the choice could include custom-made software.

Decide how to manipulate the data

Input–process–output (IPO) charts

IPO charts are also discussed in Chapter 2 on page 66.

For the third step, you will need to decide how to manipulate the data. In problem-solving, certain strategies are adopted to see the 'bigger picture' of the design – or the macro design. These strategies include IPO charts. Also called a defining diagram, an IPO chart identifies a program's inputs, outputs and the processing steps required to transform the inputs into the outputs. Once the overall design has been determined, the finer details of the solution are considered, such as colour, fonts and styles.

Flowcharts

Flowcharts are also discussed in Chapter 2, page 45.

Flowcharts are also useful for the third step. A flowchart can be used to indicate the essential procedures that will be employed to create a solution, and to generate the required information. It uses symbols in a linear sequence to document each procedural step required; a simple set of symbols is shown in Figure 5.28. The flowchart should indicate some detail of the procedures to be followed to produce the solution and output.

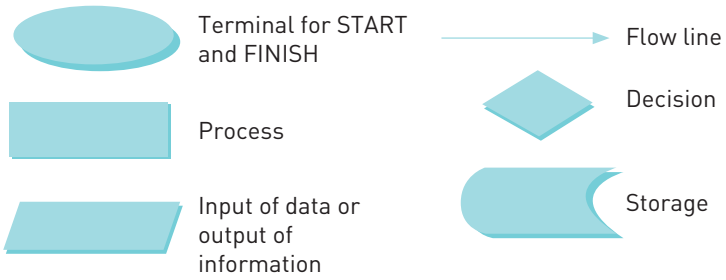


FIGURE 5.28 Simple flowchart symbols

You could create a flowchart to show the steps that would be involved in producing a webpage, based on your previous observations or experience. You would then trial the steps in the flowchart to see how well they work, as a form of quality control. If necessary, you would make changes to your flowchart to make the processes as efficient as possible. Next, you would compare it with another flowchart showing how the same process is actually being done by a web developer. Look for differences between the actual information flow and your 'best possible' flow. These differences may indicate inefficiencies in what the web developer is doing, which can then be rectified. Testing the solution following its development is also considered to be a form of quality control.

Quality control is discussed on page 169.

Develop the site map

A site map, or a linkage plan, is a graphic representation of how each page of a website links together. It is commonly hand-drawn as a series of boxes attached to one another. It gives the designer an overall picture of how large the site will be, and how each page is linked (Figure 5.29).

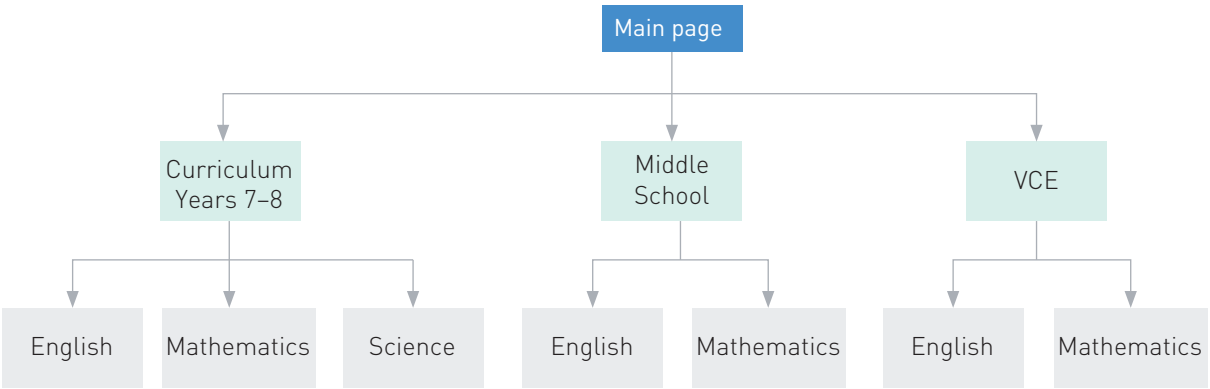


FIGURE 5.29 A site map represents how each page of a school subject selection website links together

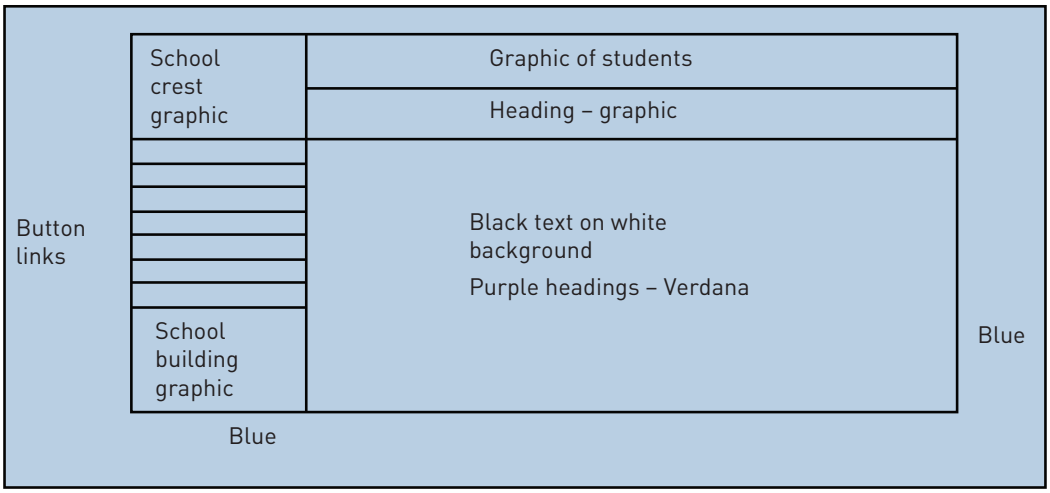
Construct storyboards

A webpage storyboard is a tool used to design the features of each individual page. A series of storyboards provide helpful tools for visual problem solving. Storyboards show general hand-drawn screen designs and the placement of graphics, and describe actions and links to other pages (Figure 5.30). The main part of a storyboard is the drawing of what the site should look like. Other features that should be included in a storyboard for a webpage are:

- title/heading
- page filename
- client's name
- designer's name
- page number
- background colour code/image file name
- the required screen dimensions
- the purpose/description
- a list of required image file names and their memory size
- a list of the required link file names.

To make it easier for a user to navigate through large documents or multiple webpages, the designer can use links, icons, buttons and page numbers. These techniques help the user to find information in which they are interested. It is inappropriate to produce a layout diagram by capturing a screen dump of the solution since the solution has not been created at this point. A storyboard is a common tool for a wide variety of presentations, not just webpages. For example, storyboards are used to help filmmakers design the sequence of scenes needed for their film.

Storyboard for Darby Vale	Page 1 of 43
Client William	Screen dimensions 1024 × 768



Purpose/description		Page title	Middle School
Introduction to Middle School page		File name	m_school
		Links	
		Link name	Link to
Graphics File name	Size (kb)	Subjects	Table of subject
		Year 7 & 8	Yr 7 & 8 page
DVC.jpg	23	VCE	VCE page
building.jpg	27	Policies	Policies page
heading.jpg	4	Contacts	Contacts page
		Site map	Site map page
Background(s)	Blue/white	Home	Home page

FIGURE 5.30 A storyboard is used to plan a website

Appearance

The following sections discuss tools that apply to the appearance of your solution: its layout, testing and formats and conventions, including examples.

Choose layout designs

The tools most commonly used to choose layout designs are layout diagrams and annotated diagrams, or mock-ups. The following sections discuss both of these and provide examples.

Layout diagram

Layout diagrams provide a visual representation of how the final designed product should look. The designs indicate features such as variation in font size, colour and positioning of text. The placement of text and graphics must be planned so that a balanced visual effect is achieved, as can be seen in Figure 5.31.

Annotated diagrams and mock ups

Annotated diagrams are also discussed in Chapter 2, page 67. This includes an example of a *hand-drawn* annotated diagram.

An annotated diagram or mock up provides a visual depiction of how an onscreen solution, such as a website, should look. An annotated diagram or mock up can be hand drawn or mocked up in software using a program such as Balsamiq. It is similar to a layout diagram but it is more explanatory.

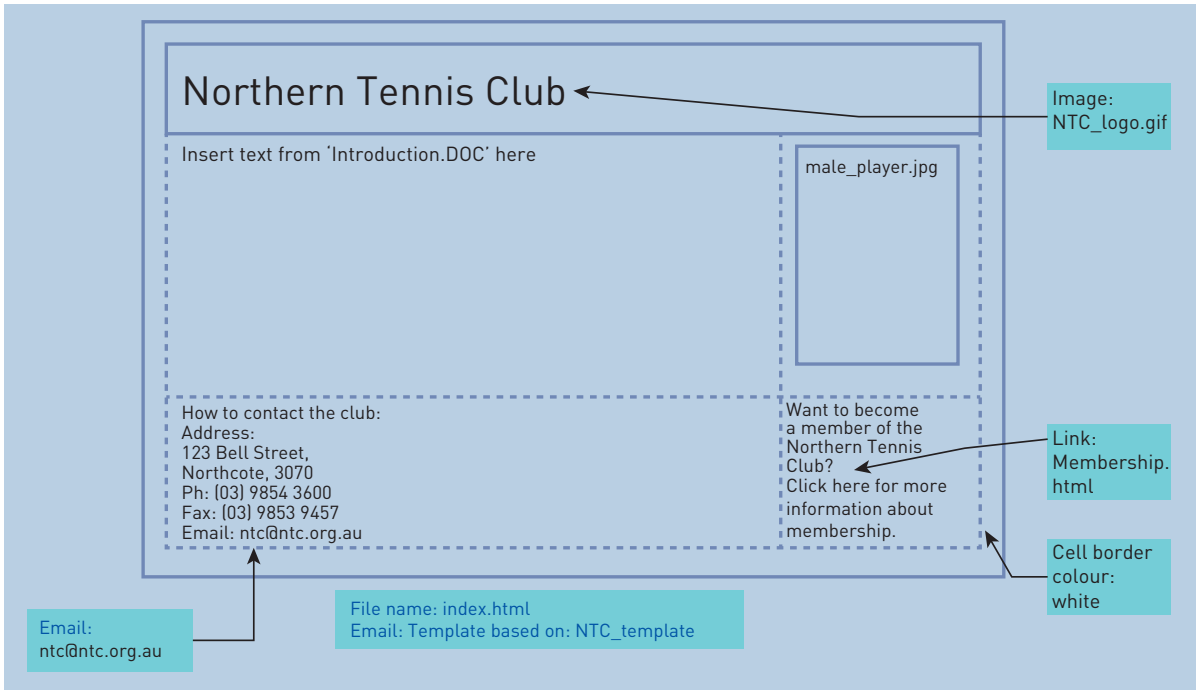


FIGURE 5.31 Sample layout diagram

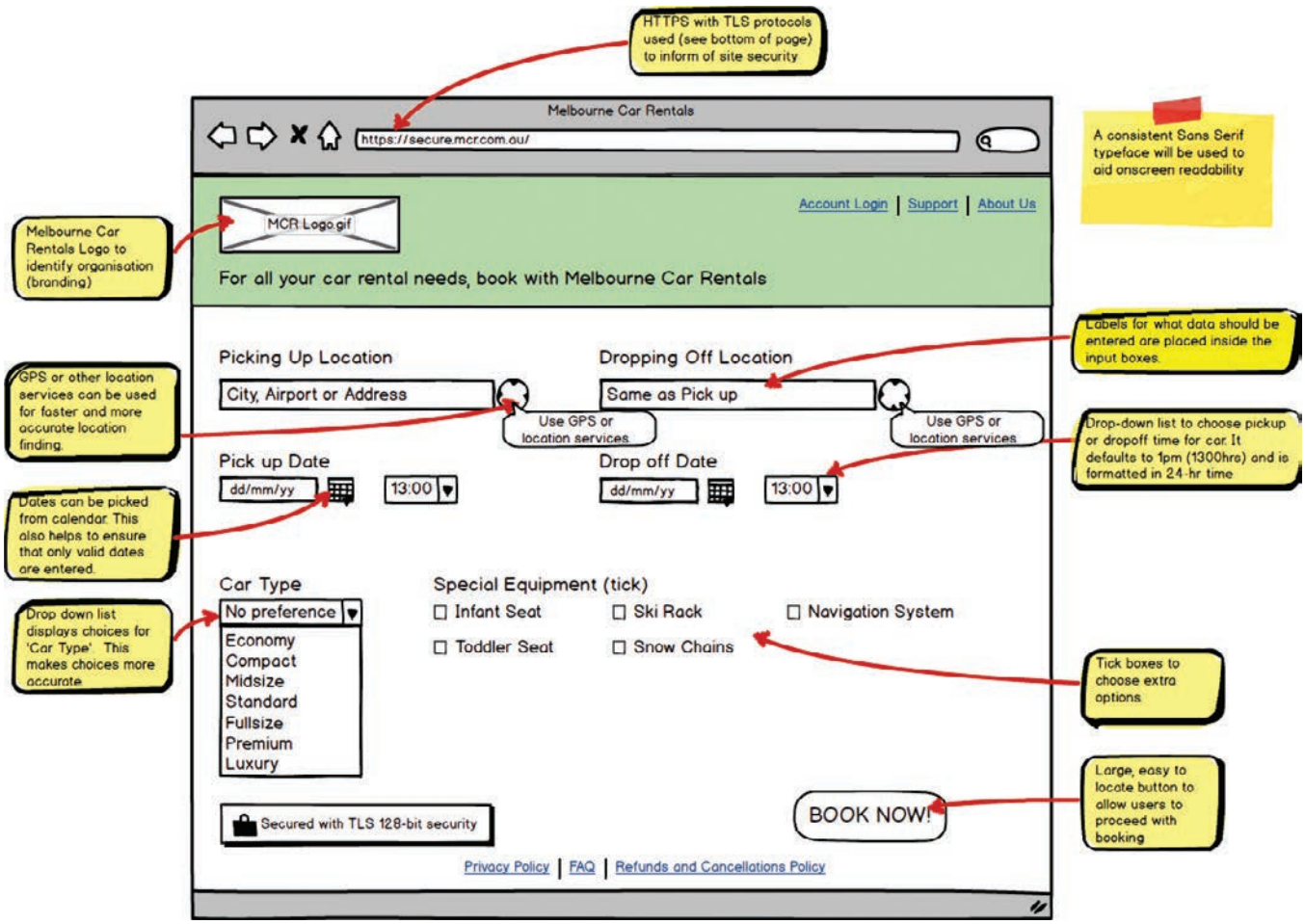


FIGURE 5.32 Sample mock up created with Balsamiq

Choose test data

Testing is a necessary step in problem-solving. Testing the product ensures that it is free of errors. It is not done in an ad hoc way; rather, through the development and use of test (or dummy) data (see Table 5.3). This is produced to test any error-handling or data-validation techniques that may be included in the product. Testing will determine whether the fonts, styles, type sizes and spacing are appropriate for the website. Testing to see that the message is conveyed accurately and clearly is also an important task when creating a website. This ensures that the correct information is communicated to the audience.

In the design stage of the problem-solving methodology, the test data is selected and the types of tests to be conducted are determined. A test table is used to document the tests to be performed. Actual testing takes place after the solution has been developed. Note that the button links need to be tested to ensure all navigation features are working properly. All key features of a webpage need to be tested, including video, audio or animation, if they are used.

TABLE 5.3 An example of test data in a test plan

What is to be tested	Test data	Expected/ required result	Why?	Actual result	Concluding statement
Internal link to 'Senior Singles Results' on 'Results' page	Starting page = results.html Link = Senior Singles Anchor name = `Results.html#SeniorSingle`	When the 'Senior Singles' link is selected, it should hyperlink to the 'Senior Singles Results' anchor.	Because the link is to '`results.html#SeniorSingle`' and there is an anchor called '`#SeniorSingle`' just above the heading '`Senior Single Results`'	See annotated printout.	See annotated printout.
'Mailto' email link to 'ntc@ntc.org.au' on 'Home' page	Starting page = index.html Link = mailto:ntc@ntc.org.au	When the 'ntc@ntc.org.au' link is selected, it should open a new email dialog box with the email address 'ntc@ntc.org.au' in the 'To' field.	Because the link's code is 'mailto:ntc@ntc.org.au'	See annotated printout.	See annotated printout.
Image 'tennis_ball.gif' loads to the left of the cell and in the vertical middle of the text on the 'results.html' page.	Code on 'results.html' is: img src = "Images/tennis_ball.gif" Width = 40 Height = 40 Align = 'absmiddle'	The image 'tennis_ball.gif" must load at the left of the cell and aligned with the vertical middle of the text. Its height and width should be 40.	Because the code links to the 'tennis_ball.gif" file in the images folder, its height and width are set to 40 and its alignment is set to 'absolute middle.' This coding is placed before the text in the table cell.	See annotated printout.	See annotated printout.
User acceptance of functionality	Procedures: Navigate to results page and locate current 'Senior Singles' results. Navigate to membership page, locate information for potential members and download membership form.	Users should be able to perform each of the procedures listed with minimal difficulty.	Interface has been designed to allow for easy navigation to these key areas (navigation panel on left and advertisement on home page).		

Choose conventions and apply formats

Conventions and formats are terms that are applied to the layout and presentation of the webpage. They enhance the appearance of information and make it more readable. Conventions are simply formal ways of displaying information (Figure 5.33). They are rules that people follow when creating webpages. An example is the placement of a navigation menu across the top or down the left-hand side of a webpage.

To format information means to create and/or change the appearance of a document by altering specific features, such as fonts, margins, spacing, columns, tables, graphics, borders, page numbers, headers and footers. Formatting looks at the presentation of the layout and its suitability.

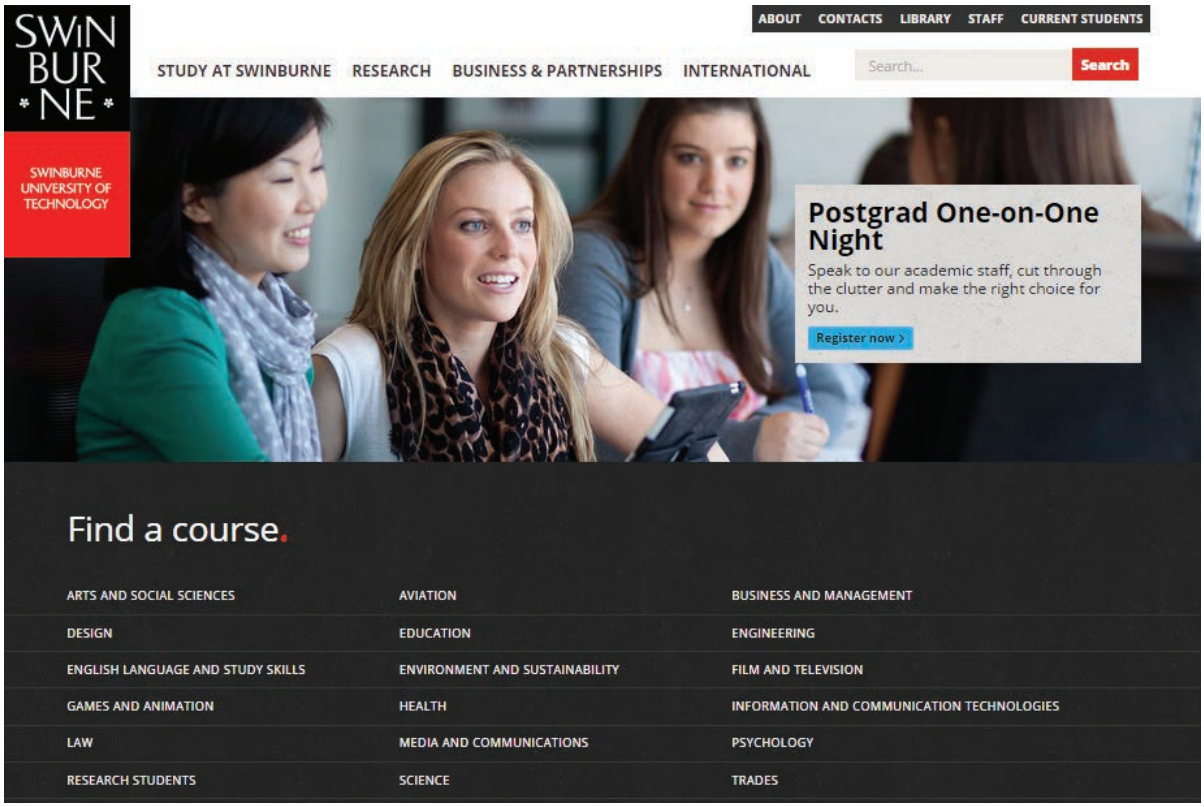


FIGURE 5.33 A website convention involves maintaining a balanced relationship between words and graphics

File naming

The index page, or homepage, is the first page that appears when you open a website. It is sometimes referred to as the default page. The file name of the index page is often `index.htm`, `index.html` or `default.html`, depending on the service that is hosting the website.

File names should be kept short and meaningful and they are easier to read in lower case than in upper case. Every webpage must be saved with a unique file name. File names usually contain letters, numbers or the underscore symbol. In a program such as Dreamweaver, the organisation of resources is assisted by defining a site and creating a resource library.

Most websites contain a large number of files and folders, but keeping them organised and logical is important. Create a set of electronic folders or directories to help locate files quickly. You should name these folders in meaningful ways, such as 'images', 'videos' and 'buttons', to make it easy to find files. File and folder names in websites should never contain spaces or any of the following characters: `/\?%*:|"<>`. It is best to restrict the range of characters in a file

name to alphanumeric characters, hyphens, and underscores; for example, `golden_gate.jpg` or `cable_car2.jpg`. Most file and folder names become part of the webpage's URL, so naming these files clearly is crucial. Keeping the file names short but meaningful makes it easy for users to type into the browser, especially on mobile devices. Long URLs are much harder for users to remember.

The index page

The index page of your website should include important information for users, such as:

- a contact email address or link to a contact form
- the date of the last modification to the website
- the author or company's name and contact details.

Images

A graphic, or graphic image, is a digital representation of information, such as a drawing, chart or photograph. Graphics were the first media used to enhance the text-based internet. The introduction of graphic web browsers allowed webpage developers to incorporate illustrations, logos and other images into webpages.

Webpages use colourful graphic designs and images to convey messages.

To enable graphics files to be used on mobile devices, they must be saved in an appropriate format. There are three universally supported image formats: GIF, PNG, and JPEG.

TABLE 5.4 Graphic formats used on the internet. Some users look at the file extensions to determine the type of file.

Acronym	Name	File extension	Characteristics	Use
GIF	Graphics Interchange Format	.gif	GIF limits the colour palette to, at most, 256 colours, which makes it a poor choice for most images. GIF is recognised as the best format when animation is required. The colour limitations often lead to colouring issues. They appear as white spots or blotches of colour.	• Can be viewed in all browsers • Creates animated images • Used for cartoons
JPEG	Joint Photographics Experts Group	.jpg .jpeg	JPEG uses a combination of lossy and lossless optimisation to reduce the file size of the image. It is best used for optimising photos and screenshots. Although JPEGs look fine from a distance, there is a loss of quality whenever users zoom in on an image.	• Can be viewed in all browsers • Easy to take and send photos via a mobile device as the file sizes are relatively small
PNG	Portable Network Graphic	.png Larger	PNG files are a lossless format so compression does not affect the quality of the image. It produces the highest quality image, but has a significantly higher file size than other formats. PNG files are larger than JPEG files, especially when they are high resolution.	• Can be viewed in all browsers • Not suitable for photographs

JPEG is the standard for photographic images and uses compression techniques to reduce the file size. These smaller sizes result in faster downloading of webpages. The more compressed the file, the smaller the file, but the lower the quality. The goal with JPEG files is to reach a balance between image quality and file size.

A graphic image saved as a GIF also uses compression techniques to reduce file sizes. The GIF format works best for images with only a few distinct colours, such as line drawings, single-colour borders and simple cartoons. GIF formats are limited to 256 colours and therefore not recommended for large blocks of solid colour in simple graphics. One of the disadvantages of using GIF is that it produces jagged edges when supporting transparency.

A PNG format was created in response to the limitations with the GIF format. Compression does not affect the quality of PNG files. Unlike JPEGs, which blur images at a certain point, a PNG file will always look as sharp as the original image. PNG files tend to be slightly larger than JPEG files, especially when they are high resolution.

JPEG, PNG, and GIF formats have benefits and limitations. Some formats are better suited for specific purposes; for example, when the image needs to be animated, you should use a GIF format. If file size is a consideration, and a smaller file size is required with some loss of image quality then JPEG would be the best choice. PNGs are best used when image size is not as important because they are better for capturing for **lossless** images.

To ensure that images are manageable when used with most applications, they must be compressed, which reduces the size of the photos.

THINK ABOUT COMPUTING 5.4
Take an image and convert it into different formats and evaluate the differences.

Video and audio files

As images, audio and video files take up a large amount of file space, it is worth investigating methods to make their file size smaller. You should limit the use of movie footage or sound on a webpage. Some users object to downloading large files, so clearly label the size of the file on the page so that users can estimate the time needed for download. Users with slow internet connections may have difficulty getting data fast enough and therefore the image quality will be poor. Sound files can be stored in the formats MP3 and movie files as MP4.

MP4 video formats are widely used for computers and mobile devices, and for downloading and streaming video. It is the default choice for smartphones and tablets and watching movies on computers and TVs. MP4s are compatible with Android and Apple mobile devices. MP4s support the H.264/AVC (Advanced Video Coding) codec, which processes the original large media file into a small package that can produce high-quality video at very low bit rates. The H.264 codec is optimal both for streaming internet video such as videos on YouTube and iTunes. H.264 is perhaps best known as being one of the video encoding standards for Blu-ray. Blu-ray discs handle high bit rates.

MPEG-4 is a file format that is commonly used to store media types defined by the Moving Picture Experts Group (MPEG), though it can store other media types as well. The files of this format usually have the extension .mp4. Apple's QuickTime file format was the foundation of the new MPEG-4 standard. MPEG-4 allows streaming over the internet, combining of multiple video and audio streams in one file (multiplexing), variable frame and bit rates, subtitles and still images.

MP3 (known as MPEG-1 or MPEG-2 Audio Layer III), is an audio coding format. It is commonly used to download digital music or audio streaming. Many digital audio players play MP3s. Even though it uses **lossy** compression, and the file is reduced in size, the sound reproduction is similar to the original uncompressed audio. Audiophiles would disagree and notice the quality loss; however, most people would not notice the difference.

MP3 was designed by the Moving Picture Experts Group as part of its MPEG-1 standard and later extended in the MPEG-2 standard.

Style guides

Style guides provide instructions to website developers about where to use different type sizes and fonts, and whether these should be in bold or italic. They can also contain information on colours and patterns to be used as backgrounds on webpages (Figure 5.34). Formats and conventions are also documented in the guides.

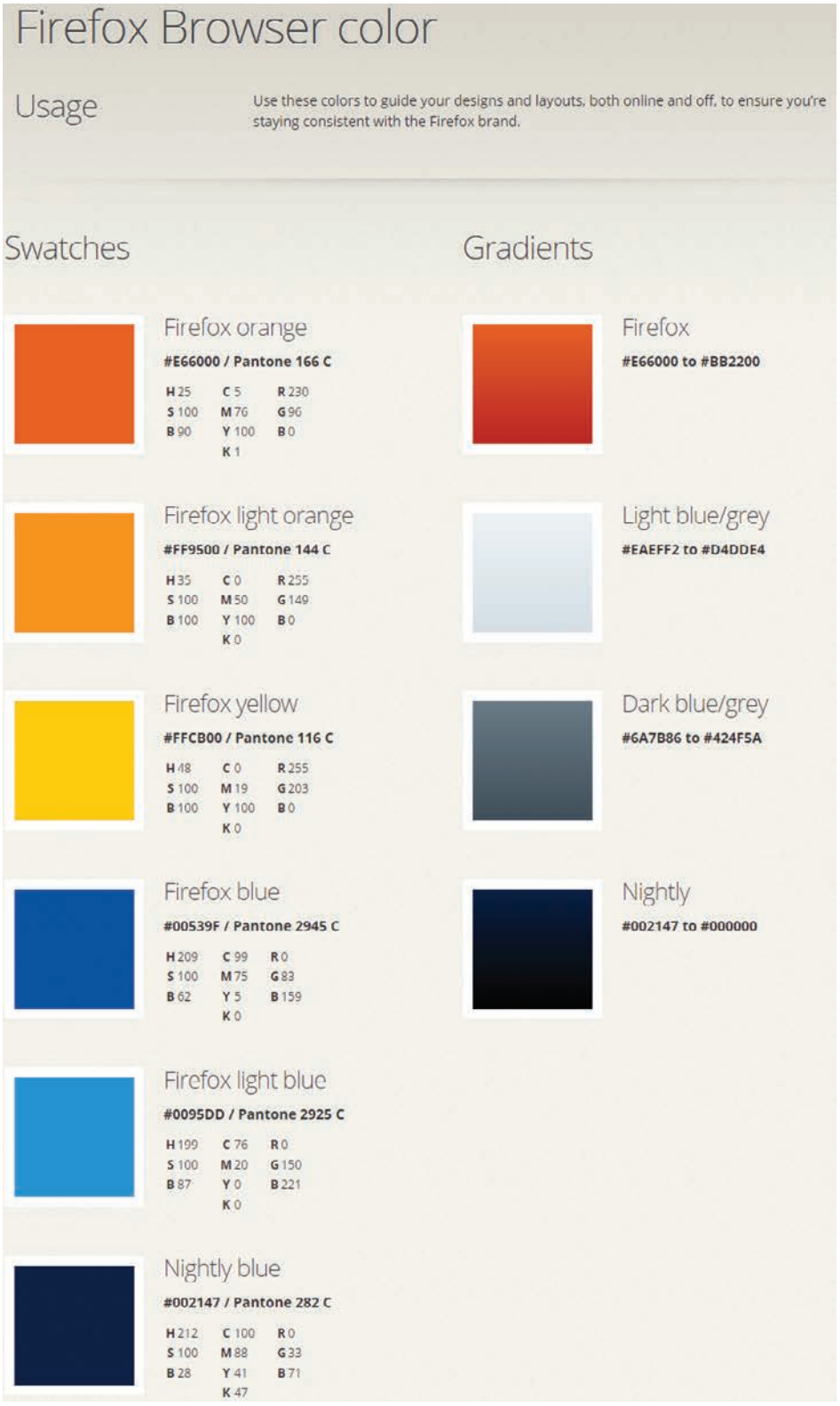


FIGURE 5.34 An extract from Mozilla Firefox's style guide

Developing websites

The development stage of a problem-solving methodology involves using the appropriate software to produce a solution as outlined in the design stage. There are four activities associated with this stage: manipulation, validation, testing and documentation.

During the design stage, planning techniques such as a site map and storyboards were used to conceptualise the solution. The development stage involves using appropriate software to apply the designs and build the solution.

Manipulation

Manipulation occurs when the data is transformed into information; it is the process of making sense of the data. Before computers existed, manipulation took place by hand and was subject to human error and interpretation. With the aid of computers, however, data manipulation now requires far less effort.

The following web-authoring functions and processes allow the user to manipulate data to create a webpage.

Buttons

A clickable image or a button is created by the website developer to allow users to navigate within a webpage, or to a different page or an external URL. Buttons need to be clearly identifiable to users. In some cases, buttons may have to be created in another program, for example Adobe Fireworks, and then imported back into the web authoring tool.

Hyperlinks

A hyperlink generally has two ends: one is called an anchor and the second is the direction. The anchor is the starting point of the link (source) often found in the original webpage and the other point is where the link leads to (target). The target point can be located on the same page as the source, or it can be found on another page in the same website, or it can link to an object, a document or URL.

Tagging

Meta tags assist webpage developers to provide search engines with information about their website. The type of text to be inserted into the source code of the webpage include important keywords about the content of the website. Often the title, keywords and description tags are used as meta tags. Meta tags assist with search engine optimisation and how the website is ranked within the search engine.

Sound editing

When sound grabs are used in a website, given the large file size, only portions can be used. Sound needs to be edited so that only the parts that are essential to the website are retained. Sounds, similar to images, need to be 'cropped'. That is, all superfluous parts of the sound file will be removed with only the portion that is needed kept. Applications such as Audacity are excellent for sound editing.

Text editing

Changing the text size or fonts, aligning headings and justifying text are all ways in which text can be manipulated. Copying data from a word-processed document, reformatting it and placing it on a webpage is another example of manipulating data into information.

Image editing

The inclusion of images into a website needs careful consideration. Images may need to be scaled or cropped when placed on a webpage so that they fit into the allocated space. If images need more attention, such as further editing, then a dedicated image editing package such as Photoshop or GIMP can be used.

Storage and retrieval

The user in the retrieval stage is the webpage creator.

Both data and information need to be stored so that they can be retrieved for later processing or communication. The best storage device to use is determined by the amount of information or data being stored, and the amount of time the data needs to be stored. In relation to websites, writeable CDs and DVDs, USB flash drives, and internal and external hard drives are the most common forms of storing data. It is important that files are stored using appropriate file name conventions.

A document can be retrieved easily if the user knows the file name and where the file is stored.

Retrieval is the process of accessing stored data or information. The retrieval process involves transferring a copy of the data or information from its place of storage – for example, the hard drive, or the cloud – to the computer's main memory.

Validation

Before data is validated, it needs to be entered so that it can be processed. In a web-based solution, the storyboards created during the design phase will indicate where elements are to be placed and the links that need to be made.

Many spellcheckers favour American spellings and cannot be relied on to choose the correct word, especially if one word has several different meanings and spellings, such as bare and bear.

Validation can be done either manually or electronically:

- **Manual validation** refers to a person checking the data to ensure it is correct and/or reasonable by proofreading the information. This includes checking both for correct spelling, grammar and punctuation, and for accuracy of transcription and reasonableness.
- **Electronic validation** refers to the use of software features, such as spell- and grammar-checking, to verify accuracy.

Testing

Testing is a necessary activity in problem solving: It is done to ensure that the solution is free of errors. Testing the solution is done through the development and the use of test data (also called **dummy data**.)

Test data is produced to test any error-handling, data-validation techniques or formulas that may be included in the product. Testing determines whether:

- the solution works within the scope of expected data and produces the required outcomes
- wrong or unexpected data needs to be rejected
- variations in graphics images fit templates, such as pictures fitting a particular space
- links work correctly
- fonts, sizes, styles and spacing are appropriate
- the message is clear. This particular test applies only to a small number of software tools, such as desktop publishing, word processing, webpage authoring and presentation software. Testing requires the creator of the product to ensure that the message they are conveying is not lost among other parts of the document and that the message is clear and concise.

There are a number of attributes, or properties, of a web solution that need to be tested once the solution has been built. Some of the attributes that should be focused on include:

- functionality
- presentation
- useability
- accessibility
- communication of message.

See Table 5.3 for examples of testing.

Functionality

The functionality of a website relates to the activities or actions it was designed to carry out. When testing functionality, it is important to look at the original problem and determine if the solution meets the organisation's needs.

When testing the functionality of a webpage it is wise to continually preview your site on a browser and test the links. This involves systematic testing using a variety of different browsers to see how the website functions. The testing of the website's functionality on a number of different browsers, such as Internet Explorer, Firefox and Apple's Safari, will reveal whether any of the popular browsers have problems supporting the website. Any difficulties identified with a browser accessing the website need to be addressed.

Presentation

During the design stage of the problem-solving methodology, a decision would be made about the appropriate format of the solution; for example, whether the solution should be a multimedia presentation, a website or a brochure. One of the most important considerations should be the intended audience: there is no point creating a web-based solution if the audience is unlikely to have easy access to the internet.

No matter what format is chosen, the presentation of the solution can be a difficult attribute to test – what looks good to one person may appear ugly to someone else.

A sample of formats and conventions appropriate to webpage presentation that can be tested include:

- ensuring that the webpage fits within the screen dimensions
- users of the website do not have to scroll any page more than two average screen lengths
- text styles (bolding and italics) are used with restraint
- hyperlinks are not confused with underlining
- the combination of colours used as backgrounds and text allows for easy reading
- each page in the website has consistent navigation buttons.

Useability

All presentations, whether they are brochures, posters or websites, need to be user-friendly. The information being conveyed should be easily accessible.

When testing the useability of a website, it is worth asking some of the following questions:

- Can the user navigate throughout the website without getting lost?
- Is it easy to get back to the home page?
- Are navigation buttons always visible and placed in a consistent location?
- Do all the hyperlinks work?

Accessibility

A web solution needs to be easily accessible via a search engine. It must also be easy to load, with small images that do not take too long to open.

Communication of message

Whatever the format, the important information presented in the solution should be clear and obvious. An advertisement on a brochure, poster or website that intends to advise the date, time and venue for a meeting must convey those details, without the readers getting lost among other less vital information. The purpose of the website should be obvious to users, and the message it is conveying should also be clear and simple.

CHAPTER SUMMARY

ESSENTIAL TERMS

accessibility a functionality design principle comprising navigation and error tolerance; accessibility of websites is primarily about making them easy for users to access, navigate and use, even in the face of errors or inexperience

brainstorming the collection of many people's ideas to devise or find a solution to a problem

collaboration where several members of a team work on parts of the same solution at the same time; team members may be separated geographically from one another

contingency plans a set of predetermined actions that a project team will take if some kind of disaster occurs

convention an accepted and standard way of formatting data; the way it is displayed

critical path the line that runs from the beginning of a project until its conclusion, and is the longest such path through the sequence of events constituting the project

dummy data a set of 'pretend' data used to establish how a product will respond

flowchart a chart used to indicate the essential procedures that are to be employed to create a solution, and to generate the required information; it uses symbols in a linear sequence to document each procedural step required

format altering the appearance of a document by changing features such as fonts, margins, spacing, columns, tables, graphics, borders, page numbers, headers and footers; also refers to the actual font, margin, spacing, etc., chosen

groupware a type of application software that enables workers to collaborate by enabling file sharing and 'real time' conferencing

hyperlink a connection/link to a website or webpage

juxtaposition being placed side by side

lossless a type of image compression that does not affect the quality of the image because all original data can be recovered when the file is uncompressed

lossy a type of image and audio compression in which bits of information deemed unnecessary are eliminated so the data file is much smaller

manipulation the process of making sense of data so that it is transformed into information

meta tag an information tag used to specify a page's title, description and keywords; essential in helping search engines locate the page

milestone the achievement of a significant stage in a project and has zero time duration; for example, completing the printing of an annual report so it can be distributed to shareholders would be a task of zero time and represents a milestone. This follows tasks in which the report has to be researched, written and proofread, all of which take time

orientation the direction and aspect of elements of an onscreen page; for instance, portrait (vertical) or landscape (horizontal), the positioning of frames at left or right on a website, text justification or the direction a graphic is facing

pixel short for picture element, a small dot of colour that forms part of a picture on a screen

plug-in a program that enhances the capability of an internet browser; used to enhance multimedia, such as Flash and Quicktime

project management the process of planning, organising and monitoring a project in order for it to be completed on time and within budget; another way of describing project management is that it is a type of collaborative problem-solving involving the coordination of tasks, people, technical resources and time

quality control checking results against identified quality standards; it is just as important to apply quality-control techniques to processes as it is to apply them to products or other end results

skeuomorphism where design cues are taken from the physical, real world; for example, using folder and file graphics for computer filing systems to make them appear recognisable to users

useability a functionality design principle comprising robustness, flexibility and ease of use; useability of websites is about maximising compatibility for a range of devices, allowing for them to continue functioning even if errors occur, supporting multiple ways of performing tasks, and providing a user-friendly experience

validation checking data for accuracy and completeness

version control managing changes to documents, applications, websites, and other collections of information; Wikipedia has version control to provide a history of the edits to articles

CHAPTER SUMMARY

IMPORTANT FACTS

1 An idea usually launches a project. An individual or a group of people decide that there is a need for a new product or service, a study is commissioned and then project planning begins. A project has the following characteristics:
   - a clearly defined purpose
   - a start time
   - a finite lifetime
   - a number of interdependent tasks.
2 A project manager must be skilled in human resource management, communication, quality control, time management, costing, accounting and contract management.
3 A project is broken down into a series of tasks. Each task should be substantial, but not so large and complex that it will affect the completion time of the whole project.
4 Resources are assigned to each task and include technical and human resources.
5 A schedule allows you to map the project tasks and display interdependencies. A Gantt chart is one common way of displaying a schedule.
6 Tasks that might cause a project to be delayed lie on the critical path of the project, and extra resources may need to be provided to ensure that they do not run over the time allocated.
7 Project management software, such as Timeline or Microsoft Project, enables the project manager to schedule and monitor complex projects. Information relating to starting dates, completion times, resources, costs and dependencies is entered and then displayed in graphic form.
8 Common tools used in project management include project tables and Gantt charts.
9 A project table enables a project manager to brainstorm key tasks and work out their dependencies before putting them into a formal chart.
10 Gantt charts provide a standard format for displaying project schedule information by listing project activities and their corresponding start and finish dates in a calendar format.
11 Visualising thinking tools assist with thinking processes and reflect on the thinking strategies to support understanding.
12 Information architecture refers to the structure of the website and its navigation pathways.
13 When planning navigation design, the following aspects should be considered: accessibility, meaning, comprehension and consistency.
14 User-centred design is a methodology that focuses on the needs and characteristics of users, and is applied at the beginning of the design process of the website to ensure that the website is useful and easy to use. The user interface is the screen that users see when they interact with a device.
15 Useability of a website is how easy the website is to learn and to use. Useability assessment is based on a set of characteristics aimed at designing usable and accessible websites constructed on user-centred design.
16 When designing mobile interfaces, the touch zones on the website need to be big so that users can easily and accurately tap the targets. Small touch zones or ones grouped closely together make it harder to manoeuvre.
17 Websites that are well designed should prevent users from making mistakes. Although mistakes will always occur, a tolerant user interface should let users recover from their mistakes.
18 Effective visibility provides prompts and cues that can assist users through an interaction, or guide them through a series of steps, indicate the possible options available to them and communicate the context of the situation.

CHAPTER SUMMARY

19 Consistency of user interface involves creating patterns in language, layout and design.

20 An affordance is a desirable property of a user interface. It logically and naturally leads people to take the appropriate steps to accomplish their goals. Affordances indicate what is achievable and can take many different forms such as shape, texture or size.

21 The problem-solving methodology provides a structured approach to problem solving and consists of the following stages: analyse the problem, design the solution, develop the solution and evaluate the solution.

22 Designing the solution involves describing the process of solving the problem and considering the layout of the output. Input data needs to produce the information required, and test data is needed to test that the solution produces the desired results.

23 Common design tools used to plan websites include flowcharts, IPO charts, site maps, storyboards and layout diagrams.

24 A flowchart might be used to show the process or procedure that the user needs to go through to create a website.

25 Also called a defining diagram, an IPO chart identifies a program's inputs, outputs and the processing steps required to transform the inputs into the outputs.

26 Layout diagrams provide a visual representation of how the final designed product should look. The designs indicate features such as variation in font size, colour and positioning of text. The placement of text and graphics must be planned so that a balanced visual effect is achieved.

27 Storyboards show general hand-drawn screen designs and the placement of graphics, and describe actions and links to other pages.

28 Testing of the solution involves the use of dummy or fictitious data to check that the solution is producing the expected output. The data, although fictitious, is similar to what the user will input. Test data includes both reasonable and unreasonable values.

29 A test plan shows all the elements that will need to be tested to ensure correct functionality and user acceptance.

30 File names should be kept short and meaningful. They are easier to read in lower case than in upper case. Every webpage must be saved with a unique file name. File names usually contain letters, numbers or the underscore symbol.

31 Alignment of text is when the text on a page is positioned to the left or right side of the margin, centred within the page or spaced out evenly between the left and right margins.

32 Serif fonts are usually easier to read in printed works than sans-serif fonts. Serif fonts, which have small tails at the end of some letters, are used to make it easier for the viewer to read, as they can guide the eye and reduce eyestrain when reading large blocks of text.

33 A graphic, or graphic image, is a digital representation of information, such as a drawing, chart or photograph. There are three universally supported image formats: GIF, PNG, and JPEG.

34 A graphic image saved as a GIF also uses compression techniques to reduce file sizes. The GIF format works best for images with only a few distinct colours, such as line drawings, single-colour borders and simple cartoons. GIF formats are limited to 256 colours and therefore not recommended for large blocks of solid colour in simple graphics. One of the disadvantages of using GIF is that it produces jagged edges when supporting transparency.

35 JPEG is the standard for photographic images and uses compression techniques to reduce the file size. These smaller sizes result in faster downloading of webpages. The more compressed the file, the smaller the file, but the lower the quality. The goal with JPEG files is to reach a balance between image quality and file size.

36 Compression doesn't affect the quality of PNG files. Unlike JPEGs, which blur images at a certain point, a PNG file will always look as sharp as the original image. PNG files tend to be slightly larger than JPG files, especially when they are high resolution.

37 Complementary colours are contrasting and stand out against one another. Often it is a good idea to use a complementary colour as the highlight colour.

38 Combining red, blue and green (RGB) colours creates white. Consequently, RGB colours are often known as additive colours. These are mainly used for lighting, optics, video and monitors.

39 Contrast in an onscreen product refers to the visual difference in colour or tone between items on the screen. Greater contrast will make items appear to stand out more from one another. If there is not enough contrast between two items, they may appear to blend into each other, making it difficult for a user to see them clearly.

CHAPTER SUMMARY

40 MP4 video formats are widely used for computers and mobile devices, and for downloading and streaming video. It is the default choice for smartphones and tablets and watching movies on computer.

41 MP3 (known as MPEG-1 or MPEG-2 Audio Layer III), is an audio coding format. It is commonly used to download digital music or audio streaming. Many digital audio players playback MP3s.

42 Style guides provide instructions to website developers about where to use different type sizes and fonts, and whether these should be in bold or italic. They can also contain information on colours and patterns to be used as backgrounds on webpages. Formats and conventions are also documented in the guides.

43 Common design elements to consider when planning a website include: proportion, orientation, clarity and consistency, colour and contrast, useability and accessibility, and appropriateness and relevance.

44 Proportion, or visual hierarchy, refers to the prominence of various elements on a screen. The most important elements, such as headings, should stand out visually on the screen.

45 Clarity is important for onscreen products. All elements on the screen should be able to be seen clearly to convey the message effectively to the user.

46 Consistency of navigation links, colour schemes and other repeatable features allow users to navigate an onscreen product comfortably with minimal confusion.

Review quiz

TEST YOUR KNOWLEDGE

APPROACHES TO PROBLEM-SOLVING METHODOLOGY

1. List the characteristics of a project.
2. List the important skills of a project manager.
3. If a task is said to be on the critical path, describe what will happen if the completion of this task is delayed.
4. Explain how a project table and a Gantt chart assist a project manager to plan and manage a project?
5. Why is it important to keep track of file versions when collaborating with others on a project?
6. How does groupware assist workers to collaborate?
7. List and describe the activities involved in the design and development stages of the problem-solving methodology.
8. Create a storyboard for the development of your school's website.
9. Create a layout diagram for the homepage of your school website so that it can be viewed on a mobile device.
10. What is the difference between a storyboard and a layout diagram?
11. What is the purpose of using a site map when creating a large website?
12. Why do solutions need to be tested using test data?
13. Why is it important to consider the elements of design when constructing a webpage on a website?
14. Describe and explain the purpose of at least one method of manual validation and one method of electronic validation when producing a website solution.
15. What is the difference between formats and conventions?
16. Why does particular attention need to be given to naming files?
17. Why are sans-serif fonts easier to read than serif fonts on websites?
18. Why is it important to ensure that light background colours are used in webpages?
19. Why is it necessary to give careful consideration to website navigation?

APPLY YOUR KNOWLEDGE

GANTT CHARTS

1 Examine the Project Table in Table 5.5. Create a Gantt chart to represent this project. Be sure to include simple dependencies and milestones.
2 Answer these questions based on the Gantt chart created in Question 1.
   a Which tasks are on the critical path?
   b How long is the critical path (in days)?
   c How is the critical path affected if Task C takes an extra three days?
   d What impact is there on the critical path if Task E takes four more days than planned?

TABLE 5.5

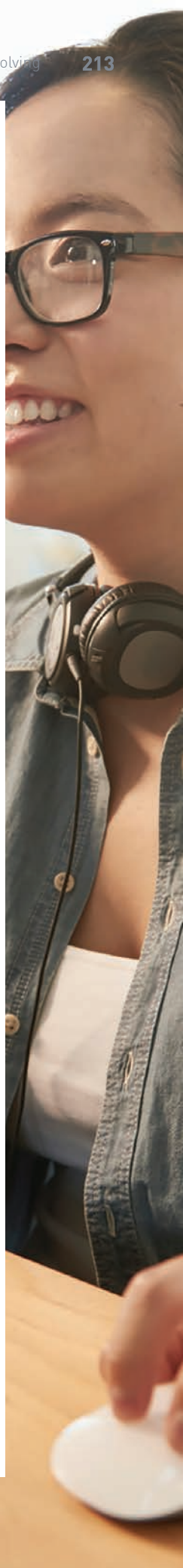
Task name	Duration (days)	Task milestone?	Predecessors
A	5		
B	5		A
C	4		B
D	3		A
E	5		D
F	2		E
G	4		F
M1	0	YES	C, G
H	2		M1
I	3		M1
J	2		H
M2	0	YES	I, J
K	4		M2
M3	0	YES	K

CHARLOTTE – ONLINE SAFETY PREFECT

Charlotte works alongside students at her school providing guidance and leadership in online safety. Her school has an iPad program where all students have access to their own device. She wants to set up a website for students to assist with issues they might encounter and places to get help, while at school and outside of school. She wants the website to be viewed on mobile devices.

Questions about the development of the student-support website

1 Create a layout diagram for the first three pages of the website.
2 Create a detailed storyboard for the first three pages of the website.
3 Create a site map for the entire website.
4 Draw a flowchart and an IPO chart to help plan the solution.
5 List five formats and conventions that will be applied to the website.
6 Identify the file names that will be used for the first three pages of this website.
7 Nominate the colour scheme that could be used in this website, and justify the choice of colours for the background and text.



PREPARING FOR UNIT 1 OUTCOME 3

Design and develop a website collaboratively with others that presents an analysis of a contemporary issue and the team's point of view on the issue

Working in teams (virtual and face-to-face), you will use web-authoring software to create a website that is designed to be viewed on mobile devices. The website will present an overview of a contemporary issue associated with a particular field, such as entertainment, agriculture, finance, sport or health. Your website should be informed by your knowledge of information architecture. You should focus on:

- the nature of the issue associated with the use of information systems
- legal, social, environmental or ethical reasons for a contentious issue
- types and capabilities of digital systems associated with the field and issue
- key stakeholders such as individuals, organisations and governments, and their responsibilities
- positive and negative opinions of each stakeholder about the issue.

OUTCOME MILESTONES

1. Select and apply appropriate methods and techniques to acquire and reference data and information.
2. Use digital systems to document and monitor project plans when creating team solutions.
3. Analyse the causes and effects of issues using visualising thinking tools.
4. Synthesise viewpoints to formulate your team's point of view.
5. Evaluate cloud computing as a data storage solution.
6. Select and use digital system components appropriate to your team's needs.
7. Select appropriate design tools and represent the appearance and functionality of solutions, taking into account user interactions.
8. Recommend online techniques for encouraging end-users' support of published viewpoints.
9. Use web-authoring software and select and apply functions and techniques to manipulate data and create solutions.

STEPS TO FOLLOW

1. Create a team to work collaboratively to research an issue.
2. Conduct preliminary research into issues that are affected by the use of information systems.
3. Select one particular field to study in detail.
4. Develop a project plan to document and monitor the project.
5. Acquire primary and/or secondary data and information about the issue.
6. Reference the data and information collected.
7. Use visualisation thinking tools to analyse the causes and effects of the issue.
8. Blend individual team members' opinions to formulate a team point of view on the issue.
9. Evaluate the use of cloud computing as a data storage solution.
10. Select appropriate design tools to represent the appearance and functionality of the solution.
11. Recommend techniques to allow end-users to express their opinions on websites.
12. Select and use the appropriate digital system components.
13. Use web-authoring software to manipulate the data and create the solution.

DOCUMENTS REQUIRED FOR ASSESSMENT

1. Project plan annotated with progress made and any changes that occurred during the project
2. List of tasks completed by each member of the team
3. Primary and secondary data collected regarding the study
4. Visualising thinking tool used to analyse the issue
5. Documentation used to formulate the team's point of view
6. Evaluation of cloud computing as a data-storage solution
7. List of design tools used for the appearance and functionality of the solution
8. Recommendations of techniques used to allow users to express opinions
9. List of the digital system components used to complete the solution
10. Completed website solution

ASSESSMENT

Assessment for the task will be based on the quality of the analysis of the subject involving the use of information systems, the quality of the website solution developed and other required documentation. A set of assessment criteria will be prepared and distributed by your teacher prior to the commencement of the task.

UNIT 2

INTRODUCTION

Unit 2 of VCE Computing focuses on data and how computational, design and systems thinking skills are applied to support the creation of a range of solutions.

Throughout Unit 2, students will be required to apply the analysis, design, development and evaluation stages of the problem-solving methodology outlined in Unit 1. In Area of Study 1: Programming, students develop a range of knowledge and skills while using programming and scripting languages, and associated software, to create solutions. In Area of Study 2: Data analysis and visualisation, students expand on their knowledge of data and the various tools that are used to extract it, reduce its complexity and manipulate it to create clear, attractive and useful visualisations. In Area of Study 3: Data management, students use database management software to create a solution that applies all stages of the problem-solving methodology.

AREA OF STUDY 1: PROGRAMMING

Outcome 1

You are required to design working modules in response to solution requirements, and use a programming or scripting language to develop the modules.

To achieve this Outcome, the student will draw on key knowledge and key skills outlined in Area of Study 1.

AREA OF STUDY 2: DATA ANALYSIS AND VISUALISATION

Outcome 2

You are required to apply the problem-solving methodology and use appropriate software tools to extract relevant data and create a data visualisation that meets a specified user's needs.

To achieve this Outcome, the student will draw on key knowledge and key skills outlined in Area of Study 2.

AREA OF STUDY 3: DATA MANAGEMENT

Outcome 3

You are required to apply the problem-solving methodology to create a solution using database management software, and explain the personal benefits and risks of interacting with a database.

To achieve this Outcome, the student will draw on key knowledge and key skills outlined in Area of Study 3.

CHAPTER 6

PROGRAMMING

Key knowledge

After completing this chapter, you will be able to demonstrate knowledge of:

Data and information

- characteristics of data types and methods of representing and storing text, sound and images

Digital systems

- functions and capabilities of key hardware and software components of digital systems required for processing, storing and communicating data and information

Approaches to problem solving

- functional requirements of solutions
- methods for creating algorithms such as identifying the required output, the input needed to produce the output, and the processing steps necessary to achieve the transformation from a design to a solution
- suitable methods of representing solution designs such as data dictionaries, data structure diagrams, object descriptions and pseudocode
- characteristics of effective user interfaces, for example useability, accessibility, structure, visibility, legibility, consistency, tolerance, affordance
- techniques for manipulating data and information
- naming conventions for files and objects
- testing and debugging techniques, including construction of test data

Key skills

- interpret solution requirements
- select and use appropriate methods for expressing solution designs, including user interfaces
- apply techniques for manipulating data and information using a programming or scripting language
- devise meaningful naming conventions for files and objects
- apply testing techniques using appropriate test data.

For the student

This chapter relates to VCE Computing Unit 2, Area of Study 1: Programming. It introduces basic programming concepts such as software development tools, storage and control structures, the software development process, design tools, types of programming languages, and universal programming ideas such as pseudocode, modules, loops, debugging and testing. You will also be learning a specific programming language that you will use to develop a series of small programming tasks for Unit 2, Outcome 1.

For the teacher

This chapter introduces students to the universal theoretical concepts behind programming that are required for Unit 2, Outcome 1. It does not assume knowledge of, or use source code from any actual language. Instead, pseudocode is used. Little previous programming experience can be expected from many students, so it is important to introduce the chosen language to them early, and give them time to train up. Unit 2, Outcome 1 should consist of a few (perhaps five) small, independent tasks that include basic programming concepts, such as storage, logic, loops and calculations.

Information systems in programming

This chapter deals with Unit 2, Outcome 1 of Computing. Throughout this Area of Study, you will focus on using a programming or scripting language that is capable of supporting object-oriented programming (OOP) to create working software modules. Programming and scripting languages provide more flexibility than applications do, because you can insert specific instructions to create a purpose-designed solution.

Information systems are explained in Chapter 3.

The specific language that you study is flexible. You will also develop and hone your skills in interpreting solution requirements that come from your teacher, and in designing working modules. During this Area of Study, you will apply methods and techniques for completing a series of small discrete tasks or working modules that use features of a programming or scripting language, including predefined classes.

You will also apply knowledge and skills associated with the design and development stages of the problem-solving methodology (PSM).

Information systems comprise people, data, processes and digital systems. In the context of programming, the key parts are:

1. **people**, who interact with systems according to their needs, such as programmers, data entry operators, system managers, technicians and end users
2. **data**, which is composed of raw, unprocessed facts and figures, such as someone's date of birth, that is used as input to be processed into meaningful information as output, such as someone's age
3. **processes**, which are the manual and automated ways of achieving a result, such as a manual data backup or an automated hard disk error scan
4. **digital systems**, which are made up of the hardware and software needed to support programming and software use.

   Digital systems are made up of the following components.
   - Networks exchange data between computers.
   - Protocols are rules used to coordinate and standardise communication between devices.
   - Application architecture patterns are sets of principles used to provide a framework for structuring solutions to recurring problems; for example, **thin client** is the philosophy that, rather than use powerful computers it is better to use 'dumb' workstations connected to a powerful central computer that does all the processing work for them.
   - Software comes in three types: systems, applications and utilities.
   - Hardware is physical equipment for input, output, storage, processing and communication.

See Chapter 3 for more information on networks and protocols.

The following section will discuss in more detail the hardware and software components of digital systems.

Hardware

The physical components of digital systems are known as hardware. They include familiar items such as the monitor, mouse, hard disk drive (HDD), motherboard, graphics card, sound card and so on.

Hardware requires software instructions to control it; software requires hardware to carry out its instructions. They work together to form a useable digital system.

Hardware falls under a number of categories, including:

- **input devices**, which are instruments and peripherals, such as keyboards, that enable users to send data and commands to software and the operating system

- output devices, which are instruments and peripherals such as printers and monitors that display information from a computer in human-readable form
- processing hardware
- storage hardware
- communication hardware.

The following sections cover processing, storage and communication hardware in greater depth.

Processing hardware

The key element of programming hardware is the processing hardware – the digital processor that converts data into information and controls all of the other hardware in the system.

CPU

THINK ABOUT COMPUTING 6.1
Research the CPUs of a mobile phone, a laptop and a desktop gaming machine. How do they differ, and how is their performance measured?

The central processing unit (CPU) is often thought of as the 'brain' of a digital system and it handles most of a system's data manipulation. The CPU is helped by other processors, such as those in the video card, hard disk drives and audio controller chips. Major CPU designers include Intel, AMD, ARM and IBM.

Reduced instruction set computing (RISC) CPUs, such as ARM, have smaller instruction sets than **complex instruction set computing (CISC)** CPUs, such as Intel's i7. Being cheaper and smaller and therefore drawing less power and producing less heat makes CISC CPUs ideal for use in smartphones and tablets.

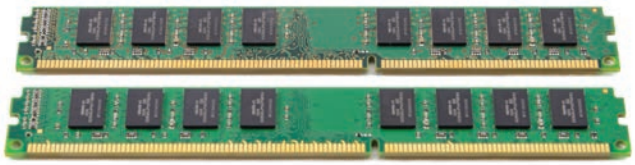
GPU

As CPUs get smaller, the laws of physics start to be bent. When electrons travel in time, or disappear and reappear elsewhere in the universe, CPUs will become unreliable. Quantum computing hopes to fix this problem.

The graphics processor unit (GPU) is a very fast and expensive processor specifically designed for high-speed image processing in graphics cards. Application software, such as Adobe Photoshop, video editors and 3D games, exploit GPU power to accelerate processor-intensive calculations.

Storage hardware

Storage hardware retains data and software for both immediate and later use. It comes in two main types: primary storage and secondary storage.



Shutterstock.com/ Jiggo_thekop

FIGURE 6.1 RAM modules

Primary storage

Primary storage is a computer's random-access memory (RAM). It has billions of storage locations in silicon chips. RAM stores instructions and values including variables, arrays and other storage structures when programs are running or being created. RAM chips are volatile because they lose their data when electricity is turned off. Dynamic RAM (DRAM) is used as the main memory in computers; high-speed (and expensive) static RAM (SRAM) is used in graphics cards and CPUs.

Secondary storage

Permanent secondary storage stores data, information and applications when they are not actively used. Secondary storage includes hard disk drives (HDD), solid state drives (SSD) and network-attached storage devices (NAS).

Hard disk drives (HDD) are aluminium disks densely crammed with magnetically recorded bits of 1 and 0. Spinning at up to 10 000 RPM, they store and retrieve data at incredible speed, with breathtaking accuracy and reliability. They are very cheap per megabyte of capacity, and

still the biggest, and most reliable long-term storage medium you can find. In 2015, a 4TB (approximately 4000GB) HDD cost approximately $200, which works out to about 20GB of storage per dollar.

Solid state drives (SSD) store data in non-volatile NAND RAM (similar to that used in Flash drives and SD cards). They have no motors to age and fail, run silently, start up instantly, consume less electricity, generate less heat, and may access data faster than a HDD. Unfortunately, NAND RAM eventually loses its ability to be written to, stores less data per square centimetre of storage space, and is expensive. In 2015, a 128GB SSD cost approximately $105, or around 1GB of storage per dollar.

A network-attached storage (NAS) device is a networked team of HDDs. Using a NAS offers more speed, capacity (e.g. 12TB), data protection (e.g. hot-swap disks), convenience and reliability than a simple USB hard disk alone.

Network-attached storage (NAS) is discussed in more detail in Chapter 3 on page 109.

TABLE 6.1 Storage units

Unit	Symbol	Equivalent to	
		RAM	Data storage
Byte	B	8 bits (1 or 0), the basic unit of storage	8 bits
Kilobyte	KB	1024 bytes	1000 bytes
Megabyte	MB	1024KB (roughly 1 million bytes – the size of two average novels)	1000KB
Gigabyte	GB	1024MB (PCs have gigabytes of RAM)	1000MB
Terabyte	TB	1024GB (hard disks have terabytes of storage)	1000GB
Petabyte	PB	1024TB	1 000 000GB (1×10^6GB)
Exabyte	EB	1024PB	1 000 000 000GB (1×10^9GB)
Zettabyte	ZB	1024EB	1 000 000 000 000GB (1×10^{12}GB)
Yottabyte	YB	1024ZB	1 000 000 000 000 000GB 1×10^{15}GB)

Storing 1YB would take 2 500 000 cubic metres of 64GB microSD cards – the equivalent of the volume of the Great Pyramid of Giza

THINK ABOUT COMPUTING 6.2
Get online Australian prices for various sizes of SSD and HDD. Graph their costs against their capacity.

Communication hardware

Communication hardware is used for sending and receiving data and information.

Ports are physical sockets or connectors that carry data between a computer and external devices, often referred to as peripherals. Universal serial bus (USB) is a standardised high-speed way to connect many devices, including Flash drives, printers, modems, keyboards, mouses, speakers and smartphones.

In the days before USB, many manufacturers invented their own type of port. Computers were jam-packed with ports to accommodate individual makes of modem, printer, mouse, keyboard, monitor, joystick, etc.

As a programmer of a high-level language you will not need to worry about directly controlling devices such as printers or disk drives. Your programming language will issue commands such as 'display this' or 'save this data' and the OS will negotiate with the hardware to fulfil your requests. The OS knows how to talk to hardware because each device comes with a software **driver**, which is like a dictionary that tells the OS the commands that the hardware understands. The OS gives a generic command, such as 'print this', and the driver translates the command into language that the specific piece of hardware understands. Hardware misbehaviour is often caused by using an incorrect or outdated driver.

Networking hardware – modems, routers, switches, cables and wi-fi are explained in Chapter 3.

Software

Software is used to control computing devices to process data. There are many types of software programs used to:

- calculate, such as spreadsheets
- store and organise data, such as databases
- entertain, such as games
- communicate, such as web browsers, email and instant messaging
- control devices, such as embedded software in TVs, toasters and car engines.

Hundreds of other programs exist. These pieces of software are all created by programmers.

Types of software

System software tools are used by a computer to manage hardware and run the user's programs; for example, the operating system (OS), device drives and communication protocols.

Applications are used to perform work or complete larger tasks. Popular examples of applications include Microsoft Word and Excel, Adobe Photoshop and Mozilla Firefox browser.

Utilities are usually small, single-purpose software tools that do a specific job or add functionality to an operating system. They include text editors, audio format converters and DVD burners.

The OS (operating system)

An OS such as Windows, Mac OSX, Linux or Android is system software that controls a computer's hardware and runs the user's application software. Operating systems are usually incompatible with one another, but they all perform similar functions.

Windows users can run *services.msc* to see the dozens of tasks the OS is managing in the background.

- Loading and saving data and programs
- Displaying output and printing
- Processing sound and music
- Allocating memory for user programs
- Watching the user's keyboard and mouse activity
- Controlling network and internet access
- Encrypting, decrypting, compressing and decompressing data
- Caching downloads
- Controlling user logins and maintaining security over accounts, files and access to resources
- Running background programs to keep the system working efficiently; for example, disk defragmenters, virus scanners and checking for upgrades

THINK ABOUT COMPUTING 6.3

Linux is a free, open source OS. Where is Linux used, and what benefits and drawbacks does it have compared with Windows and Mac OS?

Programming and scripting languages

There are more than 700 programming languages, but you will not have to learn them all.

Programming languages are used to give instructions to computer processors so they can calculate useful information or carry out tasks for humans. Whether your phone is playing an MP3, your car is turning on its anti-skid braking, or McDonald's is calculating staff wages, programming languages are needed.

Scripting languages conveniently store sequences of instructions that, alternatively, could be entered one at a time. Like human languages, there are many programming languages, each with distinctive grammar, punctuation and vocabulary. Most programming languages have special abilities or strengths that make them more useful than other languages for a particular task.

Professional programmers know a handful of languages and choose the best language for each job based on its strengths and weaknesses. Choosing the languages to learn is a big decision, but remember that learning one language makes it easier to learn others. The most popular programming languages include C (C++ or C#), Python, Java, JavaScript, Perl and PHP, SQL and Visual Basic.

- C, C++ or C# is used for writing low-level utilities and fast applications.
- Python is a scripting language used widely across the internet and to control devices.
- Java is used for server-side website programming and for Android apps.
- JavaScript is a client-side scripting language for websites.
- Perl and PHP are also used for websites.
- SQL, or structured query language, is a scripting language for database programming.
- Visual Basic is a good first programming language to learn and it is good for **prototypes**.

While programming languages may differ, they all do basically the same job: they control a digital system such as a computer, tablet or smartphone.

Programming languages differ in the amount of direct control they give over a computer's hardware and operating system. With a **high-level** language such as Visual Basic or Python, programmers avoid having to worry about complex details of the structure of actual disk files or where data is stored in memory. High-level languages are simpler to use, but lack the control of complex but more difficult to learn low-level languages. Conversely, a **low-level** language such as C or machine code requires more skill and knowledge from the programmer, but allows more direct control of the workings of a computer.

High- and low-level programming languages each have their uses. To write a simple alarm clock program, a high-level language is fine. To write a device driver to control a printer, only a low-level language will do.

SQL is covered in more depth in Chapter 8, beginning on page 314.

THINK ABOUT COMPUTING 6.4

Research three popular languages to discover their origins. What did existing languages lack that led to the need for the new languages?

C has both high-level features (for example, FOR loops and arrays) and low-level features (for example, memory pointers and byte-level operators) making it popular and suitable for many occasions.

Software development tools

To develop software, you need a number of basic, essential tools, including an **editor**, **compiler**, **linker** and **debugger**. The following section discusses these in more detail.

An editor is a specialised word processor that is used for creating human-readable **source code**, or rather, human-readable programming instructions. Code editors come with specialist features designed to make programming easier, such as highlighting programming keywords, detecting unbalanced parentheses and adding line numbering. Programming editors, such as the one shown in Figure 6.2, show colour-coding, indenting, collapsible text and line numbering.

A compiler converts source code into executable programs that a computer can carry out; that is, that a particular **central processing unit (CPU)** and operating system, such as Windows or Mac OSX, can understand.

```
321
322 function getState(  stsDOC, stsDSC, stsDJS, stsDBS, stsDWS )
345
346 function getStateReason( stsDWS, stsDOC, stsDSC )
347 {
348         var stateReason = '';
349
350         if( typeof stsDWS !== 'string' || stsDWS === '' ||
351             typeof stsDOC !== 'string' || stsDOC === '' ||
352         typeof stsDSC !== 'string' || stsDSC === ''
353         ) {
354                 return '';
355         }
356
357         if (stsDSC !== 'NO' || stsDOC !== 'NO') {
358             stateReason = 'AttentionRequired';
359         } else if (stsDSC === 'NO' && stsDOC === 'NO' && stsDWS === '1900') {
360             stateReason = 'Paused';
361         } else if (stsDSC === 'NO' && stsDOC === 'NO' && stsDWS === 'NO') {
362             stateReason = 'None';
363         } else {
364             stateReason = '';
365         }
366
367         return stateReason;
368 }
369
```

FIGURE 6.2 A programming editor

THINK ABOUT COMPUTING 6.5
Research online to choose a popular code editor. Examples are Notepad++ and Programmer's Notepad. What features does it possess that make it better than a plain text editor for creating source code?

Executable code compiled for one **platform** will not work on another without being ported (re-compiled for another platform). Porting is hard and expensive work; this is one reason why many apps are available for Windows but not Mac or vice versa.

Part of the difficulty of porting is that programs must often be substantially changed to work under a different operating system.

A linker loads information that the executable code will need; for example, how to read a keyboard or how to calculate square roots. Linkers come with useful pre-written code libraries.

A debugger helps programmers to find bugs – or programming errors. Sometimes debugging can take as long as the original programming time, or even longer if the program has been poorly designed. Debuggers may:

- highlight incorrect **syntax** (programming expression) and show how statements should be expressed
- allow programmers to set **break points** in code, where the compiler will stop and let the programmer inspect the current values of variables
- allow line-by-line **stepping** through code so developers can find exactly where a problem arises.

Few programmers today use separate editors, compilers, linkers and debuggers. Most use an **Integrated Development Environment (IDE)** such as Microsoft's Visual Studio to combine the development tools into a single package.

Figure 6.3 shows:

1. the **toolbox** of **graphical user interface (GUI)** objects the programmer can insert into the program
2. the **code window,** which the programmer uses to instruct the program how to act when an event takes place, such as the clicking of a button
3. the **form**, which will be the visible **interface** for the program's user
4. the **properties** of the selected **object**, which let the programmer modify an object's characteristics or behaviour
5. the **project manager** with which the programmer can manage the various files and components related to the program
6. **onscreen help** to give brief reminders of what the currently selected object is like.

Refer to page 253 for more information on GUI objects.

Refer to page 235 for more information on interfaces.

Refer to page 232 for more information on objects.

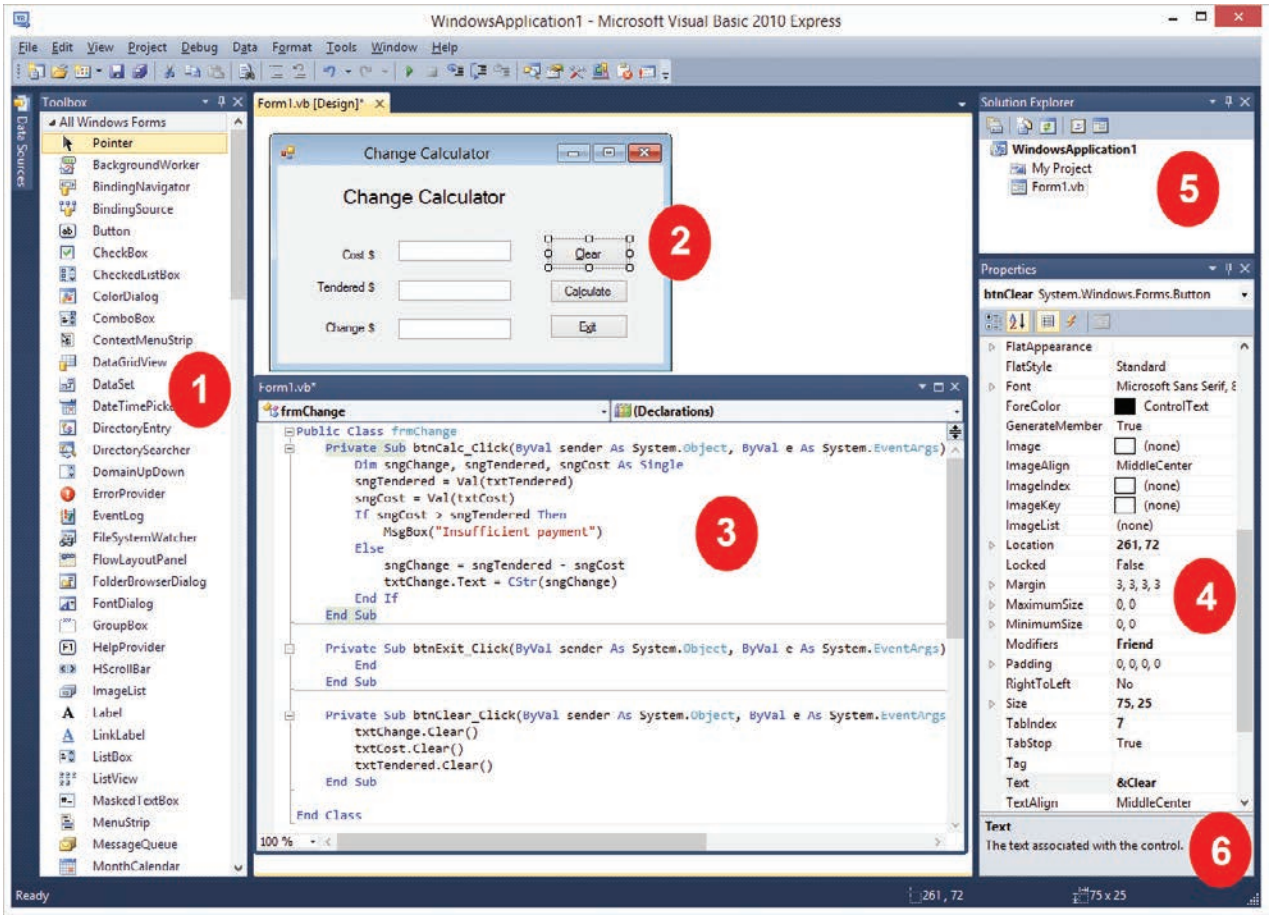


FIGURE 6.3 An example of an IDE

Storage structures

A storage structure is a location in RAM where data is stored during the execution of a program. The two main data storage structures are **variables** and **constants**. Variables are so-named because the value stored in a variable can *vary*, or be changed by a program. Constants, on the other hand, have fixed (unchanging) values during a program's execution, such as the value of pi, or the number of the Australian states and territories.

Arrays can store many values in numbered 'slots'. For example, to store 12 monthly rainfall figures, create an array called `intRain[12]` to house 12 values. To address (refer to) an individual value, give the name of the array and the desired **index** (slot number); for example:

```
IF intRain[1] < intRain[12] THEN
  DISPLAY "January was drier than December!"
END IF
```

Using variables and arrays is discussed later in this chapter.

Data types

Most languages want programmers to declare the *type* of data that needs to be stored, so they can most efficiently store the data. After all, if someone asks you for a box, do they want to store a pair of shoes, a wild cat or a widescreen TV? Knowing the intended contents lets you choose the container of the best size and type.

Chapters 7 and 8 discuss data types in greater depth.

All programming languages support various data types for variables and arrays, but they differ in what types they support. Most languages support text (string), integers, floating point (real) numbers, date/time (timestamp), and Boolean data types, as described in Table 6.2.

TABLE 6.2 Common data types

Data type	Description
Text (string)	Alphanumeric characters and punctuation – any group of characters that can be typed, such as `Tom Smith, *Hello*` or `123abc`. Numbers can be stored as string, but cannot be used for calculations in that form. Use strings to store phone numbers like `(03) 3945 2394`, because strings can hold parentheses, leading zeroes and spaces while numeric types cannot.
Integers	Whole numbers with no fractional part, such as `1, 0, 3, 67` and `341567` Fractional data is lost when stored as integers. Integers require little RAM.
Floating point (real) numbers	Numbers with fractional parts, such as `2.42`. Floating point numbers use more RAM than integers. Types include single precision and double precision.
Date/time (timestamp)	Calendar date and/or time of day. Using date/time data type allows programming languages to perform complex time and date calculations.
Boolean	Stores only two values: true or false. Although this may not sound very useful, computers spend a lot of time making true/false decisions and the dedicated, concise Boolean data type saves a lot of RAM.
Character	A single character only, such as `M`, `F` or `$`
Byte	An integer between 0 and 255
Currency	Numeric, for storing monetary values only; it has a large number of decimal places to prevent rounding errors
Pointer	C uses this special data type to store pointers to the locations of items in RAM

Languages differ in how they treat floating point numbers. A single precision variable may use 4 bytes of memory to store a value up to about 38 digits long, with 7 decimal places. Double precision is much bigger!

When you select a data type, you should choose wisely so that you will not lose vital information. For example, if you chose to store a floating point number in an integer variable, it would lose its decimal places because, as described in Table 6.2 above, integer data types do not store decimal places. Conversely, small integers do not always need to be stored as floating point numbers, particularly very large ones such as double precision, because they are *small* and using large data types will just waste RAM and slow down your execution. Thus, when choosing data types, consider the form of data and its size.

THINK ABOUT COMPUTING 6.6
Look into the specifications of the programming or scripting language you are using and find the minimum and maximum values that it can store in an integer.

You should also think about both current and future needs. For example, Kelly is programming an employee database for a company that has 30 employees. She chooses 'byte' data type for the variable to hold the number of employees. Six years later, the company has grown considerably and now has 256 employees. Kelly's program crashes as soon as employee number 256 is entered into the system, because byte variables can only store numbers up to 255. Kelly's conservative choice of data type killed her software.

THINK ABOUT COMPUTING 6.7
List the data types your language supports. Identify their purposes.

Media data types

A **medium** is a channel or means through which data and information are sent or stored. In information and digital systems, multimedia means the use of several media, such as video, audio, text and photographs, to convey information. Storage media are ways of storing data, and include Blu-ray, SSDs, HDDs and the cloud.

Media are data-hungry. An 30-minute audio file uses 6–30 MB of storage. Video can consume 20 times that amount. A digital photo can be anywhere from a few kilobytes (KB) to several megabytes depending on its quality.

Byte measurements are confusing. Since 1998, official IEC standards say that a kilobyte is 1000 bytes and 1024 bytes is now a 'kibibyte'. Many ignore that and use the traditional 1024 byte kilobyte.

Data storage is measured in bytes. One byte is made up of 8 binary bits (1 or 0). A kilobyte (KB) is approximately 1000 bytes. A megabyte (MB) is approximately 1 million bytes. Transmission speeds are usually measured in bits per second, not bytes. In abbreviations, byte is represented by the uppercase 'B', and bits use 'b', so 56 Mbps means 56 million bits per second (7 million bytes per second). Do not confuse bits and bytes! Refer back to Table 6.1 on page 221 for more information on data storage units.

The quality of stored media involves their resolution and compression, which is discussed in the next section.

Resolution

Resolution is a measure of how much detail is in:

- images (dots per inch when printed or pixels per inch when onscreen and colour depth)
- video (frames per second, frame size and bitrate)
- audio (sample rate or frequency, bit depth and number of channels).

The higher the resolution, the more realistic the reproduction will be, and the bigger the media file… and the longer it will take to transmit. Choosing a level of resolution for media usually involves a compromise between quality, cost, storage requirements and the time needed to transmit.

THINK ABOUT COMPUTING 6.8
Research the resolutions used for laser printing, phones and monitor screens. How do the differences affect file size?

When saving JPEG files, the 'Quality' compression number is not a percentage. It just means 'a little bit' or 'a lot'.

Compression

Bulky media are usually compressed before being stored or transmitted.

Lossy compression shrinks media by throwing away details. Common lossy media formats are used to shrink:

- photos, such as JPEG
- video, such as DivX, MKV and WMV
- audio, such as MP3 and WMA.

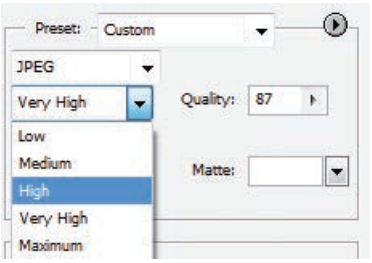


FIGURE 6.4 Setting the compression level when using the 'Save for Web' option and saving as a JPEG in Photoshop.

JPEG collapses similar colours into a single shade. MKV may reduce the number of frames per second. MP3 removes low or high musical notes that most people cannot hear. When users choose a level of lossy compression, they must balance quality against size.

THINK ABOUT COMPUTING 6.9
Using free software such as Audacity take an audio file and save it with different sample rates (e.g. 8 KHz and 44.1 KHz) and sample sizes (e.g. 16-bit or 32-bit) in mono and stereo. Compare the sizes of the files and sound qualities.

Detail lost during lossy compression can *never* be restored. Always keep original, uncompressed master copies of media for later use.

Lossless compression file formats (such as GIF, PNG, TIFF images, and FLAC audio) reduce media size as much as possible without losing data. They work by summarising data. For example, instead of recording 500 blue pixels individually, it says, 'The next 500 pixels are blue'.

JPEG format is designed to store photos with subtle colour transitions. GIF is designed to store images with big blocks of solid colour, using a palette of only 256 colours. This explains why JPG is a poor choice for storing logos, as GIF is for storing photos.

Text is never stored in a lossy format; throwing away little words or changing big words into similar smaller words to save space would not be not wise! Text formats include:

- plain text (txt), which stores nothing but text characters
- comma-separated values (CSV) or tab-separated data files
- rich-text format (RTF), which includes formatting tags, such as HTML
- portable document format (PDF), which is a searchable, compressed format that can be viewed in its original and perfect formatting without using the obscure or expensive software that created it
- proprietary formats such as DOCX and WPD that have detailed formatting, images and metadata (data about the text).

Image colour information is stored in two different ways. Images to be viewed on monitors are stored as RGB with three bytes representing the brightness of a **r**ed, a **g**reen and a **b**lue pixel between 0 (off) and 255 (fully on). Images intended for printing use CMYK to describe the strengths of the four ink colours – **c**yan, **m**agenta, **y**ellow and blac**k**.

THINK ABOUT COMPUTING 6.10
Find a digital logo with a couple of solid colours, and a digital photo of a landscape. Using an image editor like Photoshop or GIMP, load both images and save them under both JPEG and GIF with new names. Compare the file sizes and image qualities. Now try saving the JPEG image with increasing compression levels. Graph the sizes and subjective assessments of the image quality.

FIGURE 6.5 A document in print layout when using word processing software

```
{\plain \fs36 \b\f1\fs36 {\tc {This is a main heading}{\plain \
fs36 \b\fs36 .}}\par
}\pard \fs24
{\plain \fs24 \b This text is in }{\plain \fs32 \b\fs32 bold}{\
plain \fs24 \b .}{\plain \fs24 \par
}{\plain \fs24 \i Italics here.}{\plain \fs20 \fs20 \par
```

FIGURE 6.6 Part of the same document, saved as RTF

Problem-solving methodology is explained further in Chapter 2.

Developing software

The software development used in VCE Computing can be applied either as the single stage-by-stage process of the problem-solving methodology (PSM) or to each iteration of an agile problem-solving process. An agile process is a flexible and responsive approach that allows a return to earlier production steps when needs change; for example, if changes in technology force a radical redesign of the product, the client wants to add new functionality, or market pressures force a change of direction.

PSM stage: Analysis

The task of this stage is to develop the software requirements, both functional and non-functional.

Functional requirements

Functional requirements describe *the tasks* that a program should be able to perform. In the simplest terms, these are the things a program must be able to do – the main reason for creating it. For example, a program's functional requirements may specify that it must be able to:

- edit, crop and touch-up illustrations and photos
- create illustrations
- create handwriting typefaces.

The functional requirement or requirements are usually achieved in a specific and identifiable place in a program or a solution, such as a particular formula or a piece of programming.

Non-functional requirements

Non-functional requirements describe the *attributes* or *qualities* that your solution should have. Using the design program as an example again, its non-functional requirements may require it to be the following: precise, flexible, fast and easy to use.

A non-functional requirement will probably not be achieved in one specific place in a program. It usually requires a combination of factors across an entire program. For example, achieving ease of use in a design program may involve building in simple menus, shortcut keys, context-sensitive help, using clear language and making sure the interfaces are user-friendly in many different places.

The results of the analysis stage, comprising solution requirements, constraints and scope, are recorded in a document called solution requirement specifications, or SRS. In this Outcome, your teacher will give you the analysis.

Constraints

Constraints are limiting factors or conditions that you need to consider when you are designing a program. A constraint will usually reduce your freedom of design choice. Constraints generally fall into five categories: economic, technical, social, legal and useability.

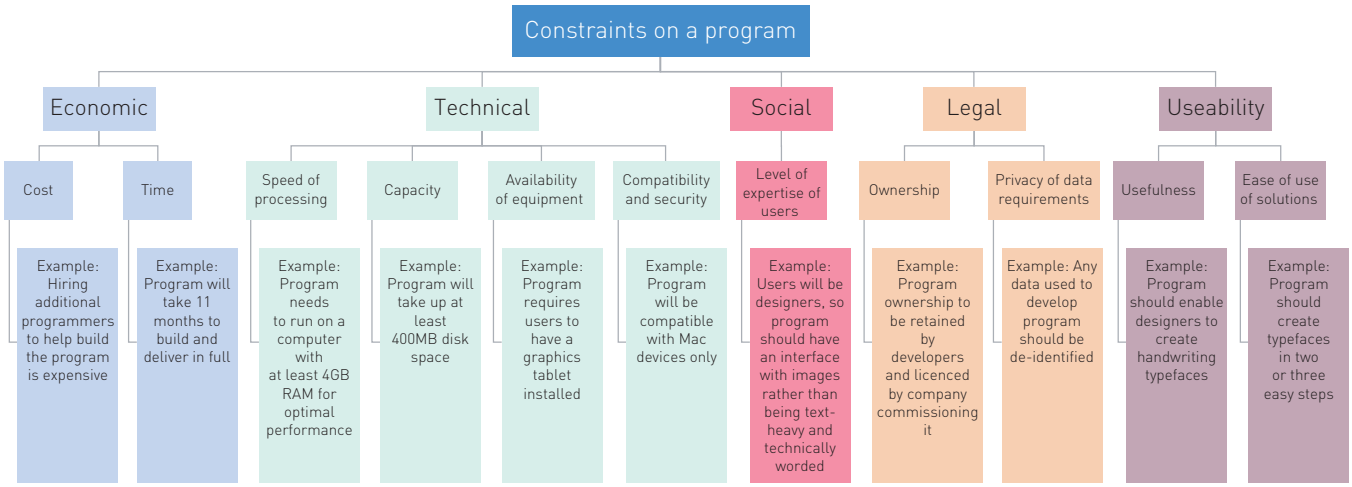


FIGURE 6.7 Constraints on a program

PSM stage: Design

Design is a vital stage in starting to create a good program, and there are several acknowledged good practices to observe.

Software design uses various tools to plan a program's architecture (how it will be constructed) and its appearance. The following sections discuss several types of software tools that can be used to help plan a program's architecture: data dictionaries, data structure diagrams, input-process-output charts, pseudocode, object description and interface mock-ups. You may find these tools useful during the development of your Outcome. Each design tool has its own specific purpose and reason for existing. If you ever find yourself using a design tool that gives no more information than that already provided by other design tools, you are probably using it incorrectly.

TABLE 6.3 Tools to plan a program's architecture and appearance

Design tool	Designs
Data dictionary	Data types, names, formatting, validation
Input-process-output (IPO) chart	Output and data requirements, calculation strategies (algorithms)
Data structure diagram	Structure and relationships within and between data items
Object description	The behaviour and properties of components
Pseudocode	The logic behind processing
Mock-up	The appearance of output and the user interface

Data dictionaries

A data dictionary is used to plan storage structures including variables, arrays, and GUI objects such as textboxes and radio buttons. The data dictionary should list every structure's name and data type. It may also include the data's purpose, source, size, description, formatting, and validation.

For more information on data dictionaries, refer to Chapter 8, page 312.

TABLE 6.4 Data dictionary

Name	Type	Format	Size	Purpose	Example
txtCustomerID	Text	XXX99	5	Customer ID	SMO40
dateDOB	Date	YYYY-MM-DD	Fixed	Date of birth	1992-12-28
sngSales	Single precision	$##,###.##	Fixed	Total amount spent	$12,456.78
boolClubMember	Boolean	Yes/No	Fixed	Is a member of the buyer's club?	Yes
txtFamilyName	Text	Xxxxxxxxxxx	25	Customer family name	De Silva
txtFirstName	Text	Xxxxxxxxxxx	15	Customer given name	Horatio
intAge	Integer	999	Fixed	Age in years	34
intMemYears	Integer	99	Fixed	Years a member	12

Data dictionaries are valuable when code needs to be modified later by other programmers and the purpose of a variable or array is unclear.

Object naming

Creating clear and obvious names for your program's variables, arrays, GUI controls, forms and windows makes your source code more readable and maintainable over time. It also makes it more easily understood by programming colleagues, and will save you some time after you have finished the code.

Descriptive names make it easier to know what an object is for, and how it should be treated. For example, the name 'Temperature' is much more informative than 'T'.

There are two industry-standard naming techniques.

Hungarian notation involves adding an object's type as a prefix to its name, such as `intTemperature` for an integer, or `lblHeading` for a label. This practice is particularly common in databases and programming. The prefix reminds programmers how the object should be handled; for example, you would not accidentally try to change the Multiline property of `lblHeading` because the `lbl` party reminds you that it is not a textbox and does not have a Multiline property.

CamelCase involves the use of capital letters to mark the start of new words in a file or object name. Multi-word names can be hard to read in a name that has no spaces, such as `inttemperaturecelsius`, but programming languages forbid spaces in names. In situations like this, CamelCase helps by capitalising the initial letter of each word in a name, so `inttemperaturecelsius` becomes `intTemperatureCelsius`. Using another example, in the filename `annualcompanyauditreport2017.docx`, the initial capitals system of CamelCase would make things much easier: `AnnualCompanyAuditReport2017.docx`.

The way that files are named is very important. Get started with best practices early.

- Use underscores (underlining) instead of spaces; for example, `z_dev_MediaGrid.fmp` instead of `z dev MediaGrid.fmp`.
- Keep names short but meaningful; for example, some of the most common programs have short but meaningful names: WINWORD.EXE, Acrobat.exe, and so on.
- Be consistent; for example:

  intTemperatureCelsius.xlsx, intTemperatureFahrenheit.xlsx, intTemperatureKelvin.xlsx

  inttemperaturecelsius.xlsx, intTemperatureFahrenheit.xlsx, int Temperature Kelvin.xlsx

- You will be unable to use characters forbidden by the operating system, which include the following punctuation marks in Windows:
  * ? < > : / \ " |

Data structure diagrams

A data structure diagram shows the structure and relationships within and between the data items in the data dictionary. A data structure diagram does not repeat information already present in the data dictionary; that would be a waste of time.

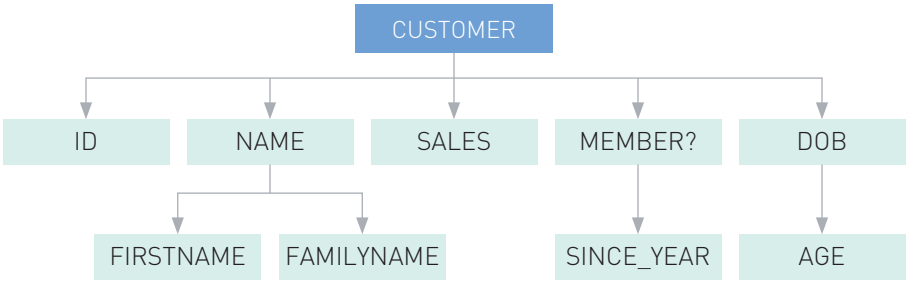


FIGURE 6.8 Example of a data structure diagram

Input–process–output (IPO) charts

IPO charts help programmers to design formulas and **algorithms**. An algorithm is the strategy for a calculation. A few steps must be followed to create and complete a chart correctly. First, enter the information required, such as a person's age, into the **Output** column.

Next, ask what data is needed to calculate that output. To calculate a person's age, we need to know two things: Date of birth (DOB) and the current date. Enter these into the **Input** column.

Finally, work out what kind of processing (algorithm) needs to be done on the input to calculate the desired output. Enter this algorithm into the **Process** column as pseudocode – do not write source code.

The second row of Table 6.5 below uses our example of calculating a person's age in years. The algorithm in the Process column describes a technique for calculating the answer. Find and then subtract the number of days between the birth date and now, and divide that by 365 to get years. The IPO chart can now be completed.

Row three of Table 6.5 calculates a subtotal by multiplying quantity by cost per item. Row four calculates total cost by taking a subtotal amount and adding tax (if payable) by multiplying the tax rate percentage by the subtotal. Row five uses a condition: if age is greater than 60 years, then total cost should be total cost minus the total cost divided by discount rate percentage. This becomes the senior citizen cost, the output.

A bad algorithm can make a program slow, fat or unreliable. A good algorithm can make a program responsive, small and trustworthy.

Every calculation – even the really simple ones – should be designed and put in the IPO chart. Later programmers might need that information.

TABLE 6.5 An IPO chart

Input (data)	Process (algorithm)	Output (information)
DOB Current date	(Current date – DOB) / 365	Age in years
Quantity Cost per item	Quantity * Cost per item	Subtotal
Is tax payable? Tax rate %	Subtotal + (If tax is payable, subtotal * tax rate %)	Total cost
Age in years Total cost Discount rate %	If age in years >= 60, Total cost – (total cost * discount rate %)	Senior citizen cost

See Chapter 2 for more information on IPO charts.

Notice how the information from previous calculations is often used in later calculations.

The IPO chart can then be given to a programmer and the algorithm converted into source code for any chosen programming language.

Object description

In programming, an object is any item that a program can inspect and/or change in terms of its appearance, behaviour or data. Today, object-oriented programming (OOP) is a common practice. OOP focuses on objects, such as icons, menus, buttons and listboxes, in a GUI that a user manipulates to issue commands and display information. OOP objects have properties, **methods** and events.

1 Properties are the attributes of an object, such as width, colour, size, name and visibility. For example:
`listbox1.` **`width`** `= 200` sets a property.
2 Methods are the actions that an object can carry out, such as move, refresh, setfocus or hide. For example:
`mainwindow.` **`refresh`** uses a method of the window.
3 Events are actions or occurrences that an object can detect and respond to accordingly, such as a mouse click, key press or a timer going off. Each event usually has its own procedure, which describes what will happen when the event occurs. For example:
`txtFamilyName.` **`keypress`** responds to a user's typing into the `FamilyName` textbox object.

An object description is a way of describing all of the relevant properties, methods and events of an object.

```
OBJECT: txtName
PROPERTIES

Class: textbox
Left position: 300
Width: 500
Font: Arial
Justification: left
Visible: yes
Font colour: black

METHODS
Cut: save cut text to disk

EVENTS
Keypress: if key is CTRL+[ set text justification to left.
```

FIGURE 6.9 Example of an object description

Pseudocode is probably the most important tool for students of Software Development VCE Units 3 & 4.

Pseudocode

Writing an algorithm in source code is slow. An algorithm written in source code also limits itself to use in only one compiler. **Pseudocode**, also known as Structured English, is a quick, flexible, and language-independent way of describing a calculation strategy – halfway between

English and source code. Once the algorithm is sketched out in pseudocode, it can be converted into source code for any desired programming language.

'Pseudocode' literally means 'false code'.

A good algorithm can be extremely valuable and bring forth great change. A clever strategy can make software run twice as quickly or use half the amount of RAM. An ingenious idea can lead to the development of a program that was once considered impossible. For example, Google's PageRank completely changed the way the world searched the internet, and made billions of dollars for its inventor in the process. The invention of **public key encryption** finally cracked the age-old problem of how to encode and transmit secrets without having to also send an unlocking key that could be intercepted.

Google PageRank checker

This pseudocode determines if a year is a leap year:

```
if (year is divisible by 4 and not divisible by 100)
or (year is divisible by 4 and 100 and 400) then
    it's a leap year
else
    it's not a leap year
end if
```

THINK ABOUT COMPUTING 6.11

Use a testing table to check whether the leap year algorithm works for the years 1999, 2000 (which *was* a leap year), 2020 (also a leap year) and the current year.

The rules of pseudocode

What are the rules of pseudocode? Easy: there are none. As long as the intention of the calculation is clear, it is good pseudocode. If not, it is bad.

However, ensure that you specify assignment (the storage of a value) using the ↗● symbol rather than the equals sign (=) that is used in algebra and in most real programming languages; for example:

```
IsLeapYear ↗● True
```

The equals sign is reserved for logical comparisons, such as:

```
IF B=0 THEN CALL SoundAlarm
```

Common features found in pseudocode include:

- loops, such as `WHILE/ENDWHILE` and `FOR/NEXT`
- control structures, especially `IF/ELSE/ENDIF` blocks
- logical operators – `AND`, `OR`, `NOT`, `TRUE` and `FALSE`
- arrays, such as `Expenses[31]`
- arithmetic operators `(+ - * /)` and the familiar order of operations, as used in Year 7 Mathematics and Microsoft Excel spreadsheet formulas.

Pseudocode punctuation and the names of key words are largely up to you if it is clear what you mean; for example, it does not really matter if you prefer `WHILE/WEND` or `WHILE/ENDWHILE`.

To 'Get data from keyboard', you could use `INPUT`, `GET`, `FETCH`, or another keyword. To read data from a disk file, you could choose `INPUT`, `GET`, `READ` or something else. To avoid ambiguity, you could explain your pseudocode's conventions using comments, as shown in the example on page 234.

```
# The hash symbol precedes a comment
# GET reads the keyboard.
# READ loads data from a disk file.
# DISPLAY shows output on screen.
# WRITE saves output to a file.
DISPLAY "What is your name?"
GET UserName
OPEN FILE "Users.txt"
READ data for UserName
IF new data exists THEN
    WRITE new data to file
END IF
```

Interface mock-up

If software will be used directly by people (rather than running hidden deep in the OS), it needs an **interface** – a place where people can control the program, enter data and receive output. A successful interface cannot be cobbled together. It must be carefully designed to make it usable and clear.

To design an interface, use a **mock-up**, which is a sketch showing how a screen or printout will look. A mock-up should typically include the following features.

- The position and sizes of controls such as buttons and scroll bars
- The positions, sizes, colours and styles of text such as headings and labels
- Menus, status bars and scrollbars
- Borders, frames, lines, shapes, images, decoration and colour schemes
- Vertical and horizontal object alignments
- The contents of headers and footers

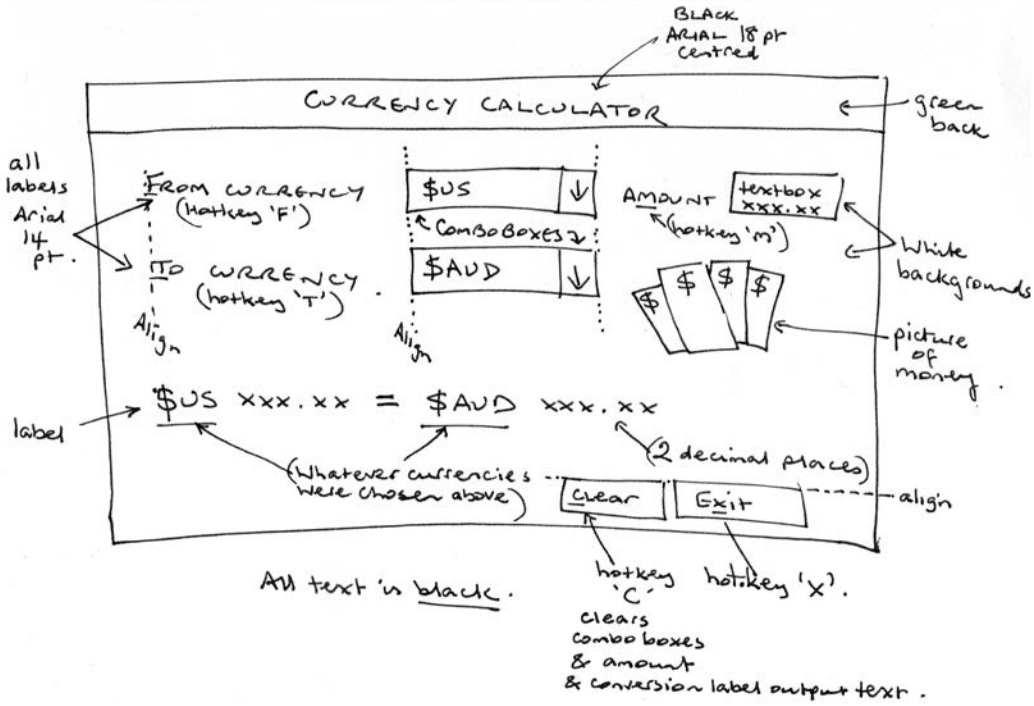


FIGURE 6.10 A mock-up of a screen interface

In VCE Computing it is not mandated that you use software to create your mock-ups. You may use software such as Balsamiq Mockups if you wish, but you may also create them by hand using pen and paper.

A mock-up can be considered successful if you can give it to another person and they can create the interface without needing to ask you questions.

Creating effective user interfaces

Very good interfaces are difficult to make because human beings are individuals. We have our own preferences and operate differently depending on gender, experience and cultural background. However, to create usable interfaces, the same several factors must be always considered and applied: useability, accessibility, structure, visibility, legibility, consistency, tolerance and affordance.

Useability

Software needs to do more than just create accurate output. It must let users work efficiently and require minimal learning, memorisation and stress. You should make commonly used features the quickest and easiest to find. Do not hide basic functions deep in a menu, because users will find it frustrating.

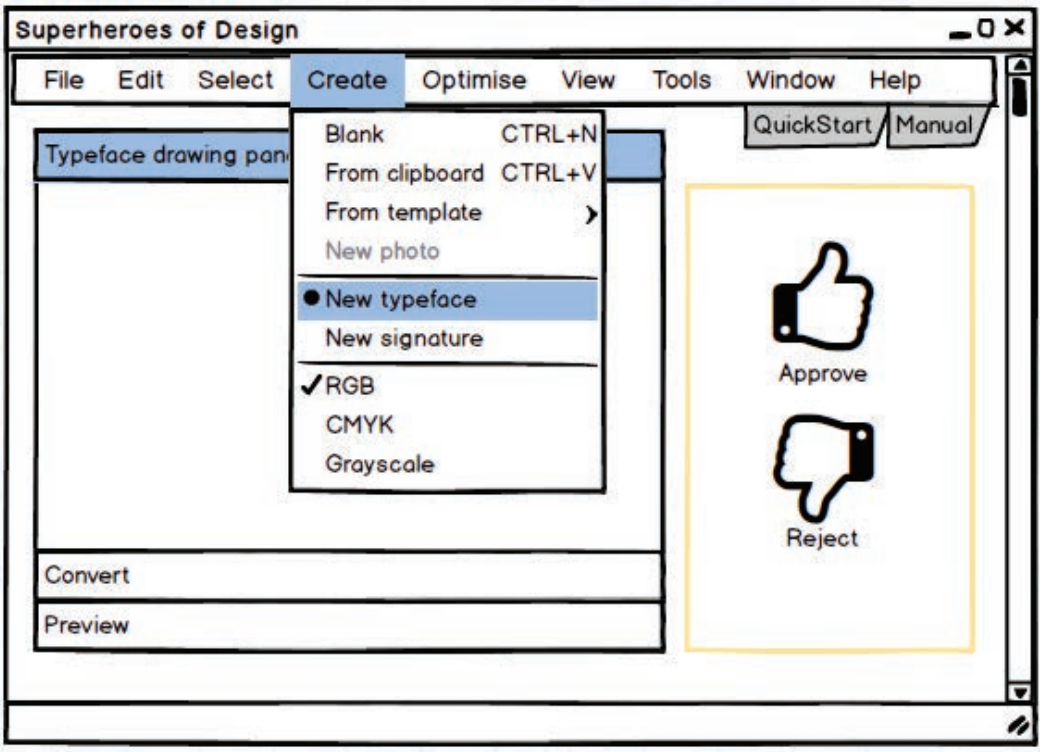


FIGURE 6.12 Ease of use: Although this mock-up looks easy to use, users would normally expect that creating a new document of any kind, such as the new typeface, is usually a menu option contained within the 'File' menu at the far left, which may make this interface *less* useable. This is something to consider.

There are many ways to provide help for your users. Consider providing printed manuals, quick start guides, internet help guides, context-sensitive help, onscreen instructions, pop-up tool tips and/or examples.

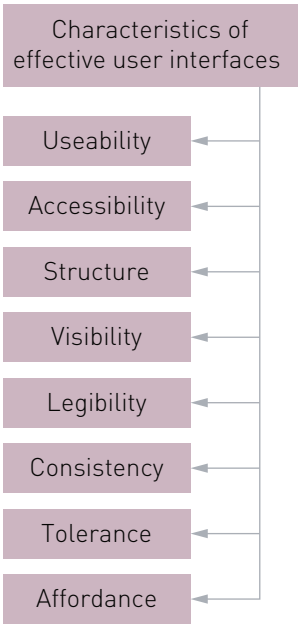


FIGURE 6.11 Characteristics of effective user interfaces

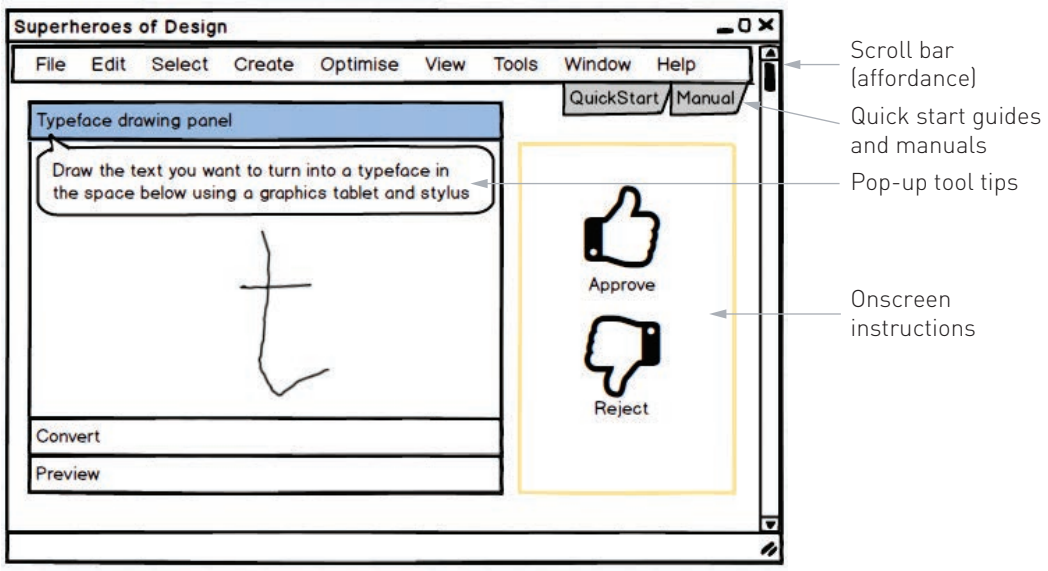


FIGURE 6.13 Tool tips, quick start guides, manuals and onscreen instructions make this interface somewhat easy to use. What else would make this interface even easier to use?

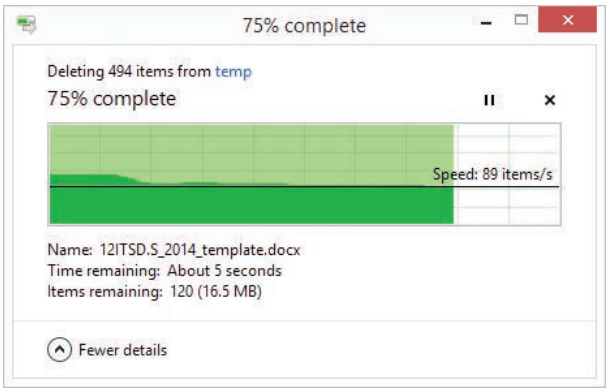


FIGURE 6.14 Showing progress during a long operation

Another useability factor to consider is showing progress. Few things cause users more anxiety than when a program gives no indication about what it is doing. Has it frozen? Is it deleting everything on the computer? I need to leave in five minutes – what is this machine doing? Computer users panic easily when programs stop communicating with them. Give the user some reassuring feedback about the operation's status with a progress bar, an estimate of time remaining, a spinner – anything to stop users reaching for the unresponsive computer's reset button (Figure 6.14).

Affordance

Affordance refers to the concept that objects on your interface should immediately suggest what they do and how to use them. An interface with good affordance naturally leads people to use it accurately, efficiently and intuitively to accomplish their goals.

- The shadow effect on a button suggests it should be clicked with a mouse.
- A scroll bar looks like the natural thing to do would be to drag it up and down.
- A flashing red icon instinctively suggests a problem.

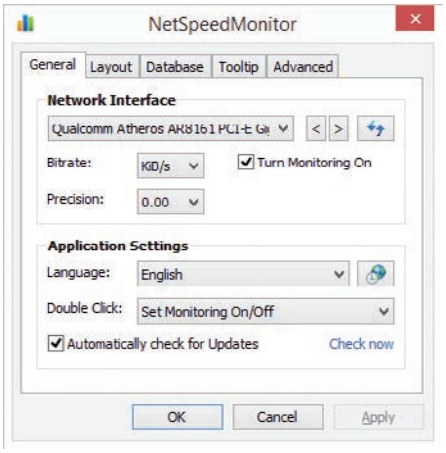


FIGURE 6.15 A small form using tabs and dropdown lists to prevent overcrowding

THINK ABOUT COMPUTING 6.12
What forms of colour blindness exist?

Accessibility

In software development, **accessibility** refers to catering for the disabilities or special needs that your software's users may have.

- To cater for colour blindness, avoid putting green text on a red background.
- To cater for poor eyesight or low vision, do not make text too small.
- To cater for those who have limited hand coordination, make buttons larger and space them further apart.
- To cater for those who have limited reading ability, use short words and avoid colloquialisms and jargon.

Structure

You should organise any user interface that you design quite deliberately so that it makes sense to end users. You could base it on existing models that users are already familiar with, rather than changing fundamentals. For example, Figure 6.12 moves the creation of a new typeface document from the 'File' menu to a new menu 'Create' away from the top left, and this is a needless shift. It differs from a working model that is what users already expect. Part of a working structure that you could use would move this grouping back to the 'File' menu, as shown in Figure 6.16.

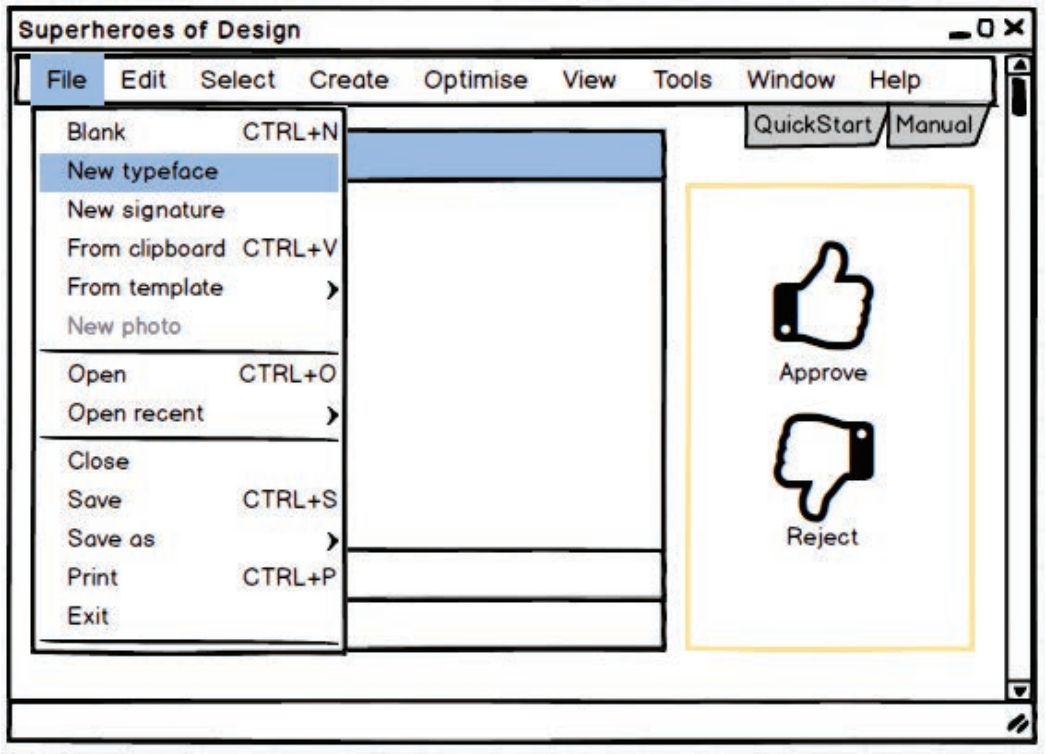


FIGURE 6.16 Structure based on a consistent, recognisable model

You should also put related items together so they are easier to find, and separate distinct items in clear, sensible groupings. Ensure that users do not need to hunt across several forms, menus or screens to carry out related actions. See the example of a GIMP interface shown in Figure 6.18.

Software users do not want to have to learn each programmer's personal stylistic preferences. They want to start a program and use standard techniques, knowing instinctively where things are and how they work. For example, in Windows, the software version is found under the Help menu > About, and the Help menu is always the right-most menu item. Do not put it elsewhere just to be different!

Software consistency is no accident. All major software developers publish **style guides** giving programmers instructions about how to design software for their platforms. To ensure software is predictable and easy to learn, some major software companies will not certify software unless it obeys their guidelines.

THINK ABOUT COMPUTING 6.13
Download interface design style guides from Microsoft and Apple (search for 'windows style guide' or 'apple style guide'. How do their styles differ?

Visibility

Visibility in a user interface means that the tools and options that the user needs to perform a specific task should be visible to them without them being distracted by superfluous information. If a user does not need to see extra information to make a decision to approve or reject a character

they have drawn in their typeface, then do not show it to them. For example, the thumbs up or thumbs down Approve/Reject step shown in Figures 6.12, 6.13 and 6.16 will suffice.

You should only show as much information as a user needs to make a decision or proceed to the next step, and no more. Overwhelming visual detail can make an interface confusing and undesirable.

Instead of squeezing many objects onto a small form, use multiple forms, tabbed controls, pull-down menus and combo boxes that collapse when not needed (Figure 6.15).

An uncrowded interface is even more important when programming for mobile devices with small touchscreens that will be operated with big fingers.

Legibility

Promote ease of use and reading comprehension by ensuring that the information on the user interface is noticeable and clearly distinguishable, and that any text that appears on the interface is readable. (Refer to Chapter 2 for more information on design principles and formats and conventions, which cover the contrast aspect of legibility in more detail.) Legibility deals with aspects including contrast, leading, kerning, line length and font size.

For interfaces:

- avoid using text in all uppercase
- use underlined text only to indicate hyperlinks
- use familiar, plain typefaces designed for reading onscreen, because decorative typefaces can be hard to read
- beware of overusing bold, italic and other font effects – not everything needs to be called out
- left-align text
- apply a hierarchy, with important things larger than less important things
- use appropriate contrast between text colour and background colour to maximise readability
- make smart use of white space; an interface crowded with controls is ugly and can lead to errors if users accidentally click the wrong item.

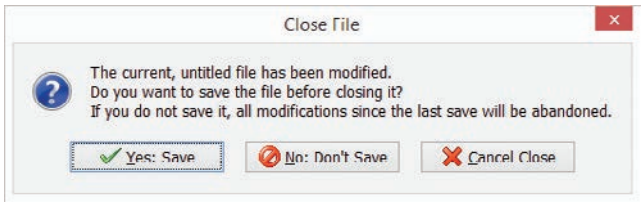
Tolerance

Tolerance is the capacity of software and interfaces to compensate for a user's errors and cope with people's natural differences in how they carry out tasks. An interface that forces users to obey its rigid expectations and is unforgiving of individual variation will be unpleasant and difficult to use.

Here are some examples of tolerant practice.

- Allow users to cancel or undo actions, and do not lock them onto a path from which they have no escape, such as printing 300 pages with no option to cancel.
- Provide settings and preferences so users can adjust a program's behaviour.
- Use the `MouseButton_Up` event rather than `MouseButton_Down` to trigger a button click. If users click a wrong button, a tolerant interface allows them to slide the mouse pointer off the button and safely let it go, while an intolerant interface would trigger the mouse event immediately, permitting no escape after the click.
- Warn users when they are about to do something dangerous or costly, such as deleting an account.
- Compensate for users' poor choices, such as by backing up data before deleting it – the Recycle Bin utility in an OS does this.
- Anticipate common errors and handle them gracefully. For example, Microsoft Word's Autocorrect quietly fixes typos like 'ACcidentally' caused by holding the SHIFT key down for too long at the beginning of a word.

- Make the **default action** the least harmful one possible. **Default values** should be the most popular and commonly chosen values. Courteous software lets users change default values.
- Make 'Cancel' the default button on a form so that no harm is done if the user carelessly hits ENTER.
- Provide users with choice. People use computers and software in different ways, and good software lets them choose how they do things, such as printing a document from the menu bar, from shortcut keys on their keyboard (such as CTRL+P in Windows), or from a button on a quick-access toolbar.



In Figure 6.17, by default, simply hitting ENTER triggers the 'Yes: Save' button rather than closing the file and losing data. The default button can be identified by its thicker border line.

FIGURE 6.17 The default button

Consistency

An interface should look and behave consistently from start to finish. Consistency should be applied in as many aspects as realistically possible. Refer to the GIMP interface shown in Figure 6.18 for an example of consistency. This includes:

- icons
- body text and heading styles
- text and background colours
- margins, borders, headers and footers
- navigation and menus.

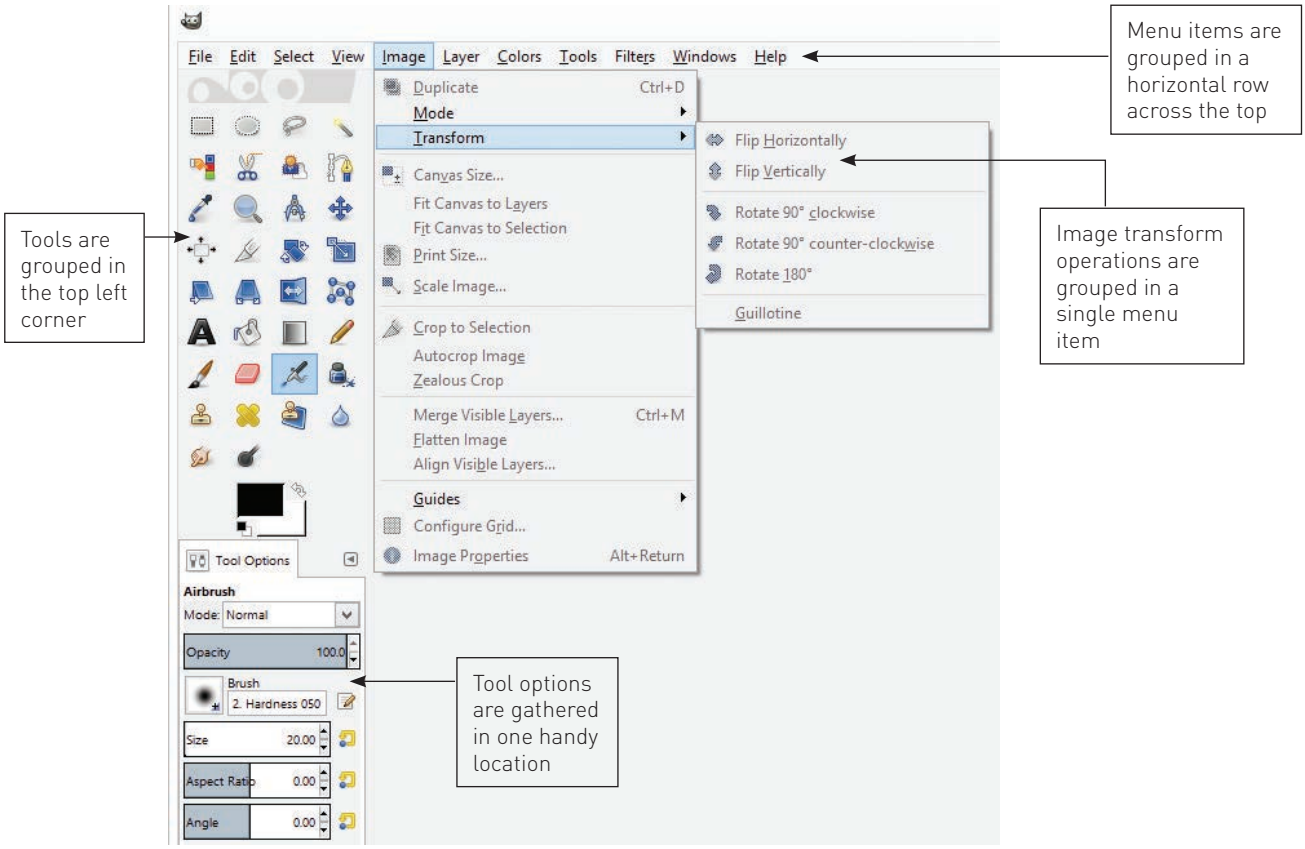


FIGURE 6.18 GIMP interface

Fundamental programming concepts

Despite the differences between platforms and languages, many concepts are universal to all programming environments.

Compiled and interpreted languages

Many programming languages, such as the C family, are compiled, meaning that source code is converted once by a compiler into executable code, such as an EXE file in Windows, to be run many times under a particular OS.

Scripting and interpreted languages, such as Python, PHP, Perl and JavaScript use a different approach. The source code is compiled *every time* it is run, instead of being compiled once by a compiler into a stand-alone executable program. This process is slower than once-off compiling, but:

- programs can be easily, swiftly and repeatedly modified by the programmer or end user without a compiler
- the code is human-readable and no viruses can be hidden in the source code
- source code only needs to be written once for all computers on all platforms that have the interpreter to run it; this is important for server-side internet programming.

All OSs, and some applications, support scripted batch files to automate tasks. Windows Powershell, Applescript, and Unix shell scripts are such scripting languages. They make complex or often-repeated tasks simple, especially for unskilled users.

THINK ABOUT COMPUTING 6.14
Python is a widely used scripting language, and well worth learning. You can download a Python interpreter and investigate it.

```
@echo off
echo Moving torrent files from c:\down and p:\torrents to Revo (Y:)
if exist c:\down\*.torrent     copy c:\down\*.torrent     y:\down
if exist p:\torrents\*.torrent copy p:\torrents\*.torrent y:\down
timeout 4
echo.
echo Done
```

FIGURE 6.19 A Microsoft DOS batch file

Modular programming

Modular programming involves breaking programs into small sections of code. Large programs are typically created as a collection of small, self-contained **code modules** (also known as subroutines or subprograms) for these reasons.

1. It is easier to find a bug in a small module than in a massive chunk of code.
2. Software development is faster when several programmers can work simultaneously on different modules.
3. A useful module can be re-used in other programs. This saves time and effort.
4. A program often needs to carry out the same action in different places. Rather than repeating the code, it appears once and is called upon multiple times.

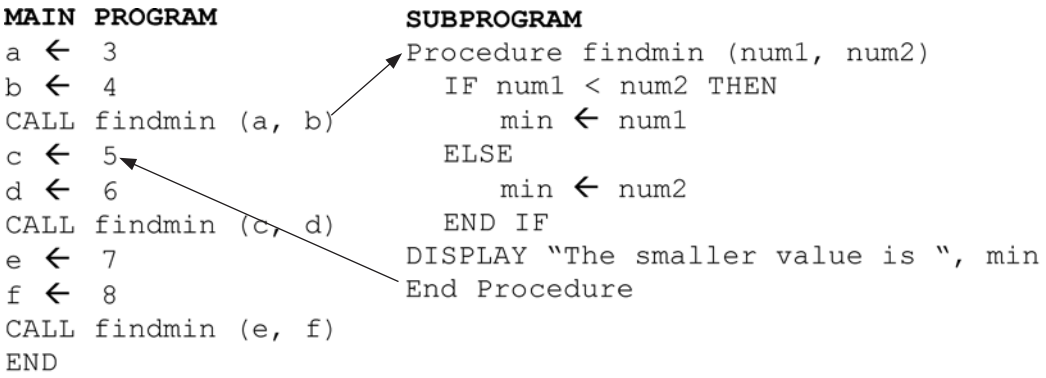


FIGURE 6.20 A main program calls a subprogram

There are a few notable features in Figure 6.20.

- The same code is needed in three places in the main program (left). Each time, the main program **calls** subprogram `FindMin`.
- For each call, the main program **passes** the subprogram two values it needs to calculate with. Values passed to a subprogram are called **parameters**.
- The values sent to the subprogram are copied to variables (`num1`, `num2`) that are **local** to the subprogram; that is, the main program cannot see or change them. In this way, a subprogram can use variables without worrying about variable names in the main program or other subprograms. **Global** variables are visible to – and changeable by – the main program and other subprograms. For safety, avoid using global variables unless absolutely necessary.
- When the subprogram finishes, it passes control back to the main program. Execution continues with the statement following the one that called the subprogram.

Functions

Functions are procedures that calculate and return a value. Function calls usually have parentheses after them, to contain parameters; for example:

```
answer ↗● SQRT(num)
```

Here, the `SQRT` function calculates the square root of parameter `num` and returns the answer to the main program where it is assigned to `answer`. Commonly used functions like `SQRT` come with the compiler for programmers to use. Typical function libraries include:

- mathematical: absolute, ceiling, cosine
- string: get left/right/middle characters, convert to lowercase, find substrings
- conversion: convert to Boolean/integer/string, convert string to number
- miscellaneous: time and date, logical (e.g. `isDigit`, `isUppercase`), random number.

Functions are often nested; for example:

```
PrintName ↗● Upper(Left(Firstname,1)) & "." &
Upper(Left(FamilyName,1)) & Lower(Right(FamilyName,Length
(FamilyName)-1))
```

Data validation

Validation checks that input data are reasonable. Validation does not and cannot check that inputs are accurate.

A **range check** checks that data are within acceptable limits or come from a range of acceptable values. For example, students enrolling in kindergarten must be between the ages of

THINK ABOUT COMPUTING 6.15
Visit a few websites that have data-entry forms. What **validation rules** are used on the data, and why? Are any of them unreasonable (e.g. insisting on a 5-digit zip code)?

3 and 6 years (acceptable limits). As another example, the product size must be small, medium or large (acceptable values).

A **type check** is a useful way of confirming that the values entered into fields are of the expected type. It will confirm if values are entered in the wrong fields, such as if numbers are entered into fields that expect only text values.

An **existence check** checks whether a value has been entered at all.

Internal documentation

Internal documentation explains the functioning and purpose of source code to programmers to make code more meaningful.

Useful comments add information that is not already obvious in the code:

```
// IntTemp - temperature is in Celsius
IntTemp ↗● 0
```

It should not be trivial or contain obvious information, like this:

```
// Set temperature to zero
IntTemp ↗● 0
```

What to include in internal documentation

- The purpose of a module
- The author's name
- Date of last modification
- Version number, to keep track of the latest version of the code
- Information, if any, about further work that is needed
- Problems that still need to be fixed
- Assumptions; for example, the customer file already exists
- Constraints; for example, it must work on a screen of only 300 × 500 pixels
- External code libraries or resources required by the module.

The contents of internal documentation may be dictated by a programming team's or organisation's style guide. It promotes consistency in a team's work, making it easier for programmers to collaborate and work with the code of others.

Adding internal documentation takes extra time and effort, but it is easier than studying obscure code.

There are no rules regarding how comments should be marked in pseudocode. You could use any of the following:

```
/* comment */
// comment
' comment
# comment
```

Note: Programmers are encouraged to write internal documentation, but for the purposes of this Area of Study, internal documentation is *not* required.

THINK ABOUT COMPUTING 6.16
Search online for 'google style guide'. Why do you think those rules were made?

Loops

Much of the power of software comes from the ability to automate repetitive actions. Using a loop – doing something 100 000 times – is just as easy as doing it twice.

If the number of required repetitions is not known in advance, use an **uncounted loop**; for example, when the number of accounts to process in a database might change every few minutes. It keeps testing whether it should continue looping.

If the number of required loops is known, use a **counted loop**; for example, the number of accounts has already been counted.

Counted loops

The classic counted loop is FOR/NEXT, which uses a variable as a counter (the **index**) to keep track of its progress as it loops from its starting point to its ending point.

This C code uses x as its index while looping from values 0 to 9 as it repeats the code within the curly { } braces to produce the output:

```
0   1   2   3   4   5   6   7   8   9
```

```
for ( x = 0; x < 10; x++ ) {
    printf( "%d\n", x );
    }
```

If you break down the typically concise C syntax, you can see the following.

- `x = 0` **initialises** the counter to value zero.
- `x < 10` looping continues while the value of x is less than 10.
- x++ increments x after each **iteration** (loop).

Inside the {loop}:

- printf is a function to display a formatted value.
- %d\n formats the output as a decimal followed by a new line.
- The semicolon tells the C compiler that the statement is finished.

This example in Basic is more typical of a real program because it uses variables rather than constants for the loop's limits.

```
FOR x = StartingValue TO EndingValue
    PRINT x
NEXT x
```

Uncounted loops

Uncounted loops keep cycling while a logical test is true; for example, *while* the temperature is less than 54 degrees Celsius, or *until* we reach the last **record** in the file.

Top-driven (test at top) loops carry out their continuation test *before* the loop begins. In contrast, **bottom-driven** (test at bottom) loops carry out their instructions at least once and then test to see if they should loop again.

TABLE 6.6 Uncounted loops

Top-driven loop (Test at top)	Bottom-driven loop (Test at bottom)

READ strName
Found ↗● FALSE
Pointer ↗● 1
WHILE NOT Found
  IF lstNames[Pointer] = strName THEN
  Found ↗● TRUE
  END IF
END WHILE
	READ strName
	Found ↗● FALSE
	Pointer ↗● 1
	DO
	  IF lstNames[Pointer] = strName THEN
	  Found ↗● TRUE
	  END IF
	LOOP WHILE NOT Found

Table 6.6 shows a number of classical programming features.

The spaces at the start of some lines are called **code indentation**. Code indentation makes it easier to see where loops and tests begin and end. You are expected to use indentation in your pseudocode.

Note also the **initialisation** of the `Found` and `Pointer` variables. Some compilers automatically initialise variables, but not all. Give variables explicit starting values, just to be sure. Close files explicitly and free up reserved memory before ending programs.

The `IF/THEN/END/IF` control structure is a classic example of programming logic.

Don't mix up '<' and '>'. It's embarrassing, and causes serious logical errors!

Logic

Digital logic is used to control a program's behaviour under different conditions. For example:

```
IF B > 0 THEN
    B ↗● B + 1
END IF
```

The logical test (`IF B > 0`) must result in a true or false (Boolean) answer. If true, the code following `THEN` is executed (`B ↗● B + 1`). If false, execution skips to the line after `END IF`.

The other logical comparisons in pseudocode are shown in Table 6.7.

TABLE 6.7 Logical operators

Symbol	Meaning
=	is equal to
<	is less than
>	is greater than
<=	is less than or equal to
>=	is greater than or equal to
<>	is not equal to

Some logical decisions are more complicated than a single test.

Other logical operators can be used to allow powerful, intelligent decision making. Here are some examples.

- **AND** adds another condition that must also be true for the result of the test to be true.
  IF (ID length=5) **AND** (first character is alphabetical) the ID is valid.
  Both conditions need to be true, otherwise the whole **IF** statement becomes false.
- **OR** adds a condition that, if true, would make the whole test true.
  If (destination is far away) **OR** (time is short) then travel by jet.
  If *any* condition is true, the test result is true.
- **ELSE** describes what happens when the result of the **IF** test is false.
- **Parentheses** can be used to group tests into related logical bundles.

Take the example, 'An ID is valid if it is five characters long and starts with a letter, or it is seven characters long and starts with a digit'. The pseudocode for this might be as follows.

```
IF
            ((ID length=5) AND (first character is alphabetical))
            OR
            ((ID length=7) AND (first character is a digit))
THEN
            ID is valid
ELSE
            Display a warning.
END IF
```

- **CASE** is a handy structure used by many languages when there are many possible conditions, each with its own appropriate action.

Debugging

The best-laid plans of programmers can go astray in three main ways:

1 syntax errors
2 logical errors
3 runtime errors.

Syntax errors

Compilers expect precise instructions in a strict format with no ambiguity. If any source code cannot be understood by the compiler because the syntax does not match what the compiler expects, it will stop working until the code is fixed.

Such syntax errors are caused by incorrect punctuation, spelling and grammar. Incorrect punctuation may result from simple mistakes, such as using a square bracket instead of a parenthesis.

Incorrect spelling in a syntax error is not the same as incorrect spelling in regular communication. A compiler has a small dictionary of key words that it recognises. If a source code instruction is not in the dictionary, even if it is a real word, the compiler will return an error message. While a person may know that the word 'colour' has the same meaning as 'color', a compiler does not, so it returns an error until 'color' is used and 'colour' removed.

Commands in source code must follow a precise format to avoid incorrect grammar errors. For example, a language might expect the following syntax:

```
INPUT "prompt"; variablename
```

If the source code said the following instead:

```
INPUT variablename; "prompt"
```

the compiler would not understand the source code any more than you would understand a person who greeted you by saying, 'Like I hat your, hello!'

Thankfully, syntax errors are easily found by compilers, and easily fixed by programmers. In fact, many modern source code editors pop up helpful syntax tips when they detect the use of a key word, as shown in Figure 6.21.

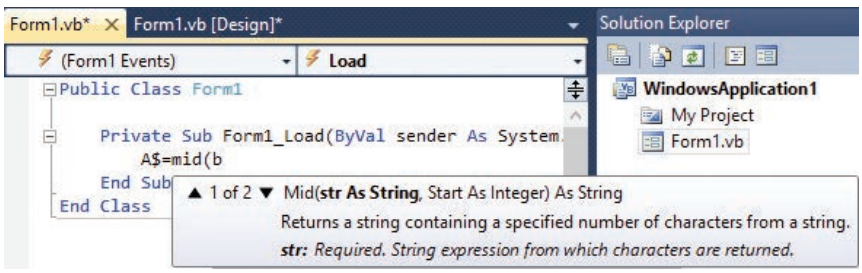


FIGURE 6.21 Visual Basic editor pops up help when you type a command

Logical errors

Logical errors occur when a programmer uses a wrong strategy. To the compiler, nothing is wrong with the syntax. The problem is that the instructions are just plain *wrong*. For example, to add 10 per cent tax to a price, the following statement would give an answer, but it would be wrong.

```
TotalCost = Price + 10%
```

A `Price` of 30 would yield a `TotalCost` of 30.10 instead of 33.00. What went wrong? The **algorithm**.

To calculate a price plus tax, the correct algorithm is:

1 Calculate 10 per cent of the `Price`.
2 Add that amount to the `Price` and store it as `TotalCost`.

The faulty algorithm simply added 10 per cent (0.10) to the price. It should have been:

```
TotalCost = Price + (10% of Price)
```

Logical errors are the hardest to fix because the compiler cannot detect faulty logic any more than a car knows when you are driving in the wrong direction. The only way to find logical errors is to create test data and manually calculate the correct answers for that data. Compare the algorithm's answers with the expected answers. If they do not match, fix your algorithm.

It is important to be aware of the potential for introducing incorrect spelling syntax errors because compilers use American English and Australian students write in Australian English.

Testing is discussed on page 248.

THINK ABOUT COMPUTING 6.17
Write an algorithm to convert a fraction like $\frac{3}{4}$ to a percentage.

Runtime errors

Runtime errors are caused by factors during the execution of a program, such as:

- the computer running out of memory
- hard disk errors
- operating system failure
- a problem with network connectivity
- incompatibility with another program running on the computer, such as antivirus software
- incorrect or outdated device drivers.

An OS may detect and handle unexpected system states to protect a program from crashing, but you should try to anticipate possible runtime problems and make allowances for them; for example, by making your code check for free space *before* saving data to disk.

The loop pseudocode (Table 6.6) introduces a common but unwelcome programming feature: a logical error. See if you can identify it before reading on.

The problem is that the program's loop uses the `Pointer` variable to count its progress through the array `lstNames`, but `Pointer` never changes and `Found` will never become `True` so the looping will never end. Such an **endless loop** will force the user to shut down the program to regain control of the computer.

To fix the logic, `Pointer` needs to be **incremented** (increased by one) with `Pointer ↗● Pointer + 1`. But where should the statement go?

Let's put the increment statement here:

```
Pointer ↗● 1
WHILE NOT Found
    IF lstNames[Pointer] = strName THEN
            Found ↗● TRUE
            Pointer ↗● Pointer + 1
    END IF
END WHILE
```

Is the code now debugged? We can find out by doing a desk check.

Desk checking

Desk checking is a technique used to check the logic of an algorithm manually. Essentially, the programmer imitates a compiler and manually tests pseudocode logic by stepping through the lines of code to check that the values are as they should be at each point. Throughout the check, you must adhere to specified logic. In the table below, assume that the name being searched for is 'Ted' and the `lstNames` array contains the following test data.

Array index	Value
1	Bob
2	Carol
3	Ted
4	Alice

Use your brain as a compiler to step through the code using our test data to calculate actual values. Beside the lines of code, draw a table where you record the values of variables.

This may help you remember the errors:
1 Syntax error: Trying to board a bus through the exhaust pipe.
2 Logical error: Getting onto the wrong bus.
3 Runtime error: The bus breaking down.

Remember that a logical error is a fault in the program's logic and will not be detected by a compiler or debugger.

When they first see statements like `Pointer = Pointer + 1`, some young programmers exclaim, 'How can something equal itself plus one?' In most languages, '=' indicates assignment, and means 'Evaluate `Pointer + 1` and store the result back into `Pointer`.' Unfortunately, most languages also use '=' for logical comparisons like `IF X=1 THEN X=0` The first '=' means 'is equal to'. The second means 'is assigned the value'. This is why, to avoid confusion, VCE pseudocode indicates assignment with '↗●' `IF X=1 THEN X ↗● 0`

THINK ABOUT COMPUTING 6.18
List each loop structure supported by your chosen language, and categorise its type.

The first desk check in Table 6.8 uncovers a new problem: When the `IF` test fails, execution skips to the `END IF` line, which bypasses the new increment statement, and causes that endless loop again! In the second desk check, we can move the increment statement again.

TABLE 6.8 First desk check

Code		Loop 1
READ strName	Strname = 'Ted'	
Found ↗● FALSE	Found = False	
Pointer ↗● 1	Pointer = 1	
WHILE NOT Found	Found = False, so test is True, so enter loop.	
IF lstNames[Pointer] = strName THEN		Pointer = 1. lstNames[1] = 'Bob'. 'Bob' <> 'Ted' so skip to END IF
Found ↗● TRUE		Skip
Pointer ↗● Pointer + 1		Skip
END IF		What? Hang on! That's NOT RIGHT.
END WHILE		

TABLE 6.9 Second desk check

Code		Loop 1	Loop 2	Loop 3	Loop 4
READ strName	Strname='Ted'				
Found ↗● FALSE	Found = False				
Pointer ↗● 1	Pointer = 1				
WHILE NOT Found	Found = False, so enter loop.		Found=False so loop again	Found still =False, so loop again	Found = TRUE! Skip to line after END WHILE
IF lstNames[Pointer] = strName **THEN**		Pointer=1. lstNames[1] = 'Bob'. 'Bob' <>' Ted' so skip to END IF	Pointer=2. lstNames[2] = 'Carol'. 'Carol' <> 'Ted' so skip to END IF	Pointer=3. lstNames[3] = 'Ted'. 'Ted' = 'Ted' so drop to next line	skip
Found ↗● TRUE		Skip (Found=False)	Skip (Found=False)	Found ↗● True!	skip
END IF					skip
Pointer ↗● Pointer + 1		Pointer=2.	Pointer=3	Pointer = 4	skip
END WHILE		Back to the top	Do it again	Back to the loop test again	skip

THINK ABOUT COMPUTING 6.19
Use the leap year pseudocode from earlier to desk check the years 1400, 1700 and 2100.

We expected the pseudocode to find 'Ted' in slot 3 of `lstNames`. The second desk check verified it.

Good test data caters for all possible circumstances, so you should also test that the code works when the name being sought is *not* in the array; for example, `strName` is 'Humphrey'.

Process testing

You must always test the software you have developed, whether the program is a game, a shopping cart for a website, or an aeroplane's autopilot navigation system. If the game you have developed fails, it may simply annoy the users. If the shopping cart fails, it could prevent purchases or overcharge customers. However, if the software in an aeroplane's autopilot system fails, people could die.

In addition to all of this, buggy, poorly tested code can ruin a freelance programmer's reputation and make it difficult to find future work.

Several forms of testing can be conducted.

Alpha testing, which is sometimes called informal testing, is when programmers test their own code during software development.

Beta testing is usually the first time that software is tested by future 'end users' and specially chosen reviewers, using live data that is more random in nature. It usually happens once the program has been completed and the goal is to weed out useability problems – you are actually trying to crash the software.

Validation testing verifies that the code properly validates input data. For example:

```
IF (Sex <> "M") AND (Sex <> "F") THEN
    DISPLAY "Invalid Gender!"
END IF
```

In **component testing**, individual modules within the software are tested in isolation. In contrast, **integration testing** is used to test whether modules work together; that is, that they can exchange parameters properly. **System testing** checks that the program works as a whole.

Finally, in the **formal testing** phase, the client who is paying for the software is shown how the finished program meets all of the functional and non-functional requirements specified during analysis.

Test data

Never trust a program's output. It may look perfectly authoritative but still be 100% wrong.

To prove the accuracy of a program's output, you need to feed the program sample data to work on, and compare the program's answer with one you know is guaranteed to be correct. Choosing this test data is not as easy as it sounds.

Good test data should include the following.

- Valid data, which is data that is perfectly acceptable, reasonable and fit to be processed
- Valid but unusual data, which is data that should *not* be rejected even though it seems odd; for example, a gifted 12-year-old child may enrol in university so that age number should not necessarily be processed as an error
- Invalid data, to test the code's validation routines; for example, if people must be 18 years of age to be granted a credit card, the test data should include people under 18 so they can be rejected
- Boundary condition data, which are on the borderline of some critical value where the behaviour of the code should change; such 'tipping point' errors are a frequent cause of logical errors in programming

Testing tables

Part of software design is to write a checklist of all the input, processing and output the software should be able to do, based on the design specifications. This list is used throughout the development of the application to check and test that the application meets those specifications. This list is called a testing table.

As an example, an online club allows members aged between 6 and 16. The validation pseudocode says:

```
IF age > 6 AND age < 16 THEN accept member
```

Is there a problem with this logic? You can use a testing table to use the test data to calculate a result manually and compare it with the output of the pseudocode. It is a good way to show evidence of testing in your Outcome.

In this case, good test data would be 5, 6, 7, 15, 16 and 17, because they cover every possible type of input: below the lower limit, on the lower limit, within the limits, on the upper limit, and above it. The test data set is as small as it can be.

TABLE 6.10 Testing table 1

Testing table			
Data	**Expected result**	**Actual result**	**Fix**
5	Don't accept	Don't accept	
6	Accept	Don't accept	
7	Accept	Accept	
15	Accept	Accept	
16	Accept	Don't accept	
17	Don't accept	Don't accept	

The pseudocode behaves accurately most of the time, but on occasion it fails spectacularly. Why? It only accepts members who are *over* 6 and *less than* 16 and rejects applicants who are *exactly* 6 or 16.

Once these logical failures have been highlighted, the cause of the errors need to be found. The rule effectively says 'between 6 and 16' not 'older than 6 and younger than 16'. The logical operators > and < are wrong. Now we need to devise a fix and fill in the last column of the testing table.

TABLE 6.11 Complete testing table

Testing table			
Data	**Expected result**	**Actual result**	**Fix**
5	Don't accept	Don't accept	🕑
6	Accept	Don't accept	Change Age>6 to Age >=6
7	Accept	Accept	🕑
15	Accept	Accept	🕑
16	Accept	Don't accept	Change Age<16 to Age <=16
17	Don't accept	Don't accept	🕑

Using the completed testing table in Table 6.11 as a guide, the pseudocode can be corrected as follows.

```
IF age >= 6 AND age <= 16 THEN accept member
```

Handy tip: Often when you are asked to identify a bug in code, the problem is a result of boundary condition errors.

Arrays

A variable can only hold a single value, which is a major limitation when processing a large amount of data. Adding up annual rainfall using variables looks like this:

```
GET Rain01
GET Rain02
GET Rain03
GET Rain04
GET Rain05
GET Rain06
GET Rain07
GET Rain08
GET Rain09
GET Rain10
GET Rain11
GET Rain12
TotalRain = Rain01 + Rain02 + Rain03 + Rain03 + Rain04 + Rain05 +
Rain06 + Rain07 + Rain08 + Rain09 + Rain10 + Rain11 + Rain12
```

Elegant code is effective, short and clever (also known as 'a neat hack'). Especially good algorithms are sometimes described as 'beautiful', and are more art than science.

Imagine the pain involved if there were 1000 rainfall figures. The solution is to use an array: a storage structure with multiple, numbered storage slots. Arrays used with loops are a powerful programming tool.

Using loops and arrays, the rainfall calculation could become far more elegant:

```
nMonths ↗● 12
DECLARE ARRAY Rain[nMonths]
FOR monthnum = 1 to nMonths
          GET Rain[monthnum]
          TotalRain ↗● TotalRain + Rain[monthnum]
NEXT monthnum
```

Using `nMonths` avoids repeating '12' throughout the code. The code is easier to maintain when key values can be changed in a single location.

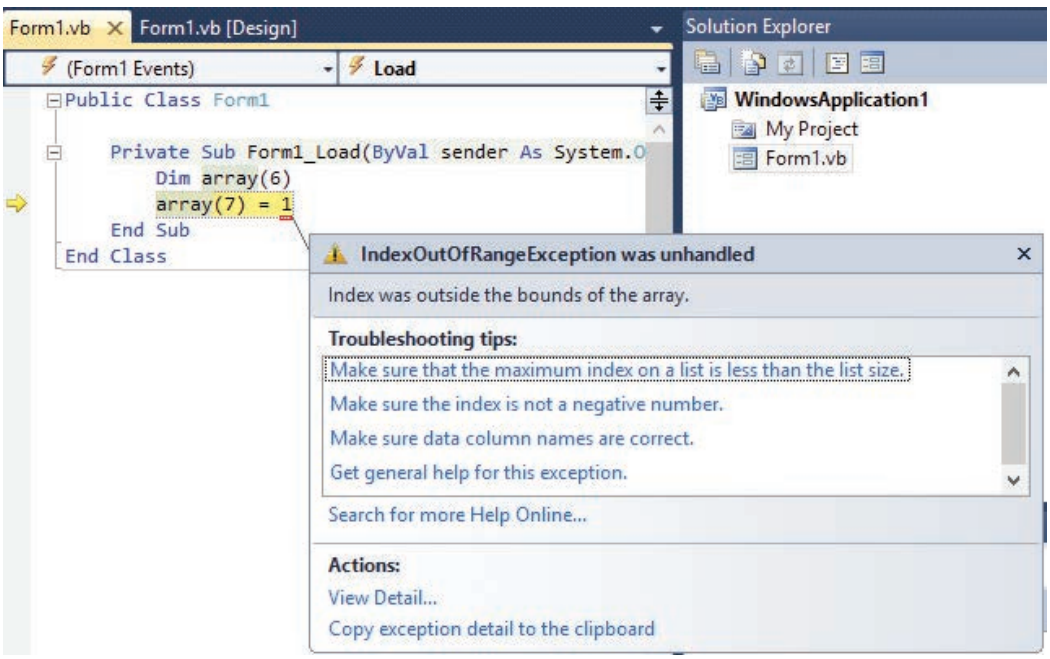


FIGURE 6.22 An attempt is made to refer to slot 7 of a 6-slot array

To cover 1000 months instead of just 12, simply change '12' to '1000' in the first line. It is that easy.

To create (or **declare**) an array, a programmer must usually specify the data type it will hold and the number of slots (or **elements**) it has. Referring to indexes outside of the defined limits will cause a runtime error.

The `Rain` array is a **one-dimensional (1D) array** because it is a single column of data, similar to a list. To store data in a table with rows and columns like a spreadsheet, use a two-dimensional (2D) array.

If the rainfall data needed to be stored for 100 years, you would create the array `Rain[100,12]` with 100 rows, each with 12 columns – 1200 slots all together.

As a table, the rainfall data may look similar to Table 6.12.

Some languages use [square brackets] to enclose array index numbers. Others use (parentheses). In pseudocode you can use either, but be consistent.

Arrays are much like tables. With tables, it often does not matter if you create it like this

	A	B
1		
2		
3		

or like this.

	1	2	3
A			
B			

Arrays are similar. We could have created the array as `Rain[12,100]`.

TABLE 6.12 A 2D array seen as a table

Array table								
	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Etc.	Month 12
	Jan	Feb	Mar	Apr	May	Jun		Dec
Year 1	67	34	67	45	34	45	...	67
Year 2	56	45	75	41	75	32	...	61
Year 3	41	63	82	46	56	31	...	86
Year 4	59	51	74	31	56	78	...	78
Etc.	...	...	...	...	...	...	...	...
Year 100	45	34	67	50	56	45	...	89

Having two dimensions requires two loops: one to loop through the 100 rows and one to loop through each row's 12 columns:

```
nMonths ↗● 12
nYears ↗● 100
DECLARE ARRAY Rain[nYears , nMonths]
// remember that years are the FIRST dimension, Months are the SECOND.
FOR yearnum = 1 to nYears
    FOR monthnum = 1 to nMonths
        GET Rain[yearnum , monthnum]
        TotalRain ↗● TotalRain + Rain[yearnum , monthnum]
    NEXT monthnum
NEXT yearnum
```

For each repeat of the outer year loop, the nested month loop carries out its full 12 cycles – like the hours and minutes on a digital clock. When the inner loop has run its 12 cycles, the outer loop increments its counter and the inner loop runs its full course again.

So the data is processed in the order shown in Table 6.13.

TABLE 6.13 How nested loops work

Year	Month	Comment
1	1	Inner and outer loops begin with their starting values
1	2	Outer loop's counter is constant until inner loop finishes its full count
1	3	Monthnum = 3
1	4	

Cont.

Year	Month	Comment
1	Etc.	
1	12	Inner loop finishes. Outer loop ticks over.
2	1	Inner loop starts its next full cycle.
2	2	Yearnum = 2. Monthnum = 2.
Etc.	Etc.	
100	10	
100	11	
100	12	Inner loop finishes. Outer loop finishes.

Important – an inner loop must be *completely* enclosed within the outer loop!

What if each of these rainfall figures for 12 months over 100 years were recorded in five different locations? Use a 3D array, `Rain[5,100,12]`, and three nested loops.

THINK ABOUT COMPUTING 6.20
Can you suggest a case when a 4-dimensional array would be needed?

Stacks

A **stack** is a simple, basic method of temporarily storing data, with the PUSH command, and releasing it as needed, with the POP command. Like a stack of pancakes or parked shopping trolleys, the most recently added item is always the first to be removed, so stacks are referred to as **first-in last-out (FILO)** or **last-in first-out (LIFO)** structures.

Stacks and queues are required knowledge for Software Development VCE Units 3 & 4.

Queues

Queues are often used when some code or a device such as a printer cannot keep up with incoming data or commands. Instead of discarding the code or device, they are stored in a queue in the order in which they arrived. When the processor or device becomes available, queued jobs are processed in turn. A queue is a **first-in first-out (FIFO)** stack.

Files and records

Primary storage, or RAM, holds data during a program's execution. Between runtimes, data and instructions are stored in files on non-volatile secondary storage, such as hard disk drives.

The main file types you should become familiar with are **sequential** and **random**.

Sequential files

Sequential (or serial) files are plain text documents that contain human-readable text data. Sequential files are easy to create but can be slow to search through because data items can be of any length. This means that the location of a particular data record may not be easily found.

A record is a complete set of data **fields** relating to a single item, person, transaction or event. An employee's record may include fields containing her given name, family name, date of birth, ID, department and so on. The concept of records and fields is also used in database theory, so it is worth remembering.

Sequential files are often used to store logs – ongoing histories of events. New events are easily appended (added to the end of) sequential files. OSs continuously maintain many logs.

A common type of sequential file is called CSV, which stands for comma-separated values. In a CSV file, each line is a record, composed of comma-separated fields:

```
Susan , Wasib , 1990/12/31 , WAS0001, Accounting
Spiro, Papadopolous, 1966/02/13, PAP0023, Management
Alphonse, Capone, 1924/11/01, CAP0003, Security
```

Random files

Random files are made up of records of identical length, with fields that must be rigidly defined in advance. Table 6.14 shows an example of a random file for an employee record.

Every record is the same length, so the location of any record in the file can be calculated precisely with:

```
StartingPoint ↗● (RecordLength * (RecordNumber - 1)) + 1
```

For example, the length of an employee record is 52 bytes. The starting point of record 3 in that random file would be (52 × 2) + 1 = 105. Using this calculation, the program can instantly seek position 105 in the file and load record 3.

Thus, any record can be accessed instantly without having to step through every record from the start of the file, as with sequential files.

Therefore, the benefit of random files is fast access. However, there are also some drawbacks to consider.

Data that is longer than the reserved file size is simply truncated (chopped off) and lost. 'Marybelle-ChristieAnne' would forever be known in the employee file records as 'Marybelle-Chris'.

A field value that is shorter than the allotted length still uses the allotted space. Thus, employee 'Su' would have 13 wasted bytes in her `FirstName` field on disk. With millions of records like these, a great deal of disk space can be wasted.

Try to imagine random files as books where every individual paragraph is allotted one exact full page of space. If you want to find paragraph 99 it is very easy: go to page 99. If paragraph 99 should be longer than a single page, the extra information is truncated and lost forever. If paragraph 99 is very short, most of the page is wasted.

Locating paragraph 99 in a normal book is much more like locating a sequential file. You must start at page 1 and count through 99 paragraphs. It takes more time but no information has been lost or space wasted.

Choosing between random and sequential files is not always easy, and it depends how much data there is, and how important it is to be able to access data quickly.

TABLE 6.14 A random file

Employee record	
Field	**Length (in bytes)**
`FirstName`	15
`FamilyName`	25
`DateOfBirth`	3
`ID`	7
`SickDays`	2

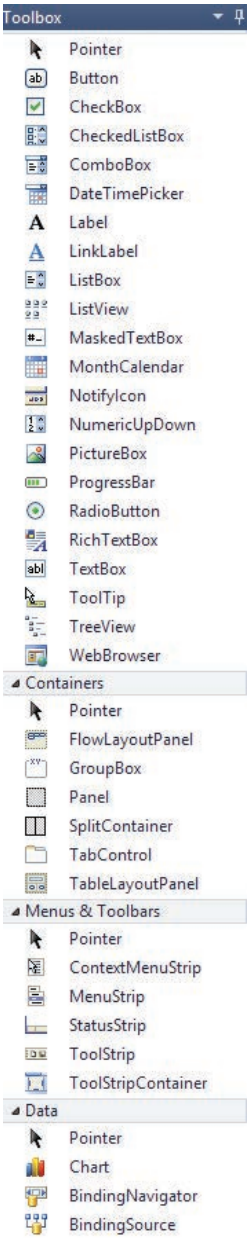


FIGURE 6.23 Visual Basic GUI Objects

GUI controls and structures

Most high-level languages have graphical user interfaces to make programs more intuitive for users.

These languages provide pre-packaged classes of objects for programmers to use, such as list boxes, menus, radio buttons and text boxes. With the GUI responsible for managing these objects, the programmer's work is made much quicker and easier.

Programmers can use GUI objects to avoid mastering arrays, variables, stacks and other programming basics, but this is short-sighted, because their programs will be tend to be fat and slow.

GUI structures are designed for ease of use, not raw power or optimal speed, and they consume far more memory and processor time than arrays, stacks and variables. The programmer also has little fine control over them, and essentially they are like cheap automatic cameras that are capable of producing acceptable, but rarely exceptional, photographs.

As a programmer, you must learn basic concepts, practise your skills, and master your tools to become successful.

When RAM was rare and very expensive, software had to be very basic, so it was difficult to use and difficult to create. Complex commands had to be memorised and usually typed. Software did not have the resources (RAM and hard-disk space) for all of the user-friendly features taken for granted now, such as easy menus, dropdown lists, cut and paste, undo, icons, dialogue boxes and scrollbars.

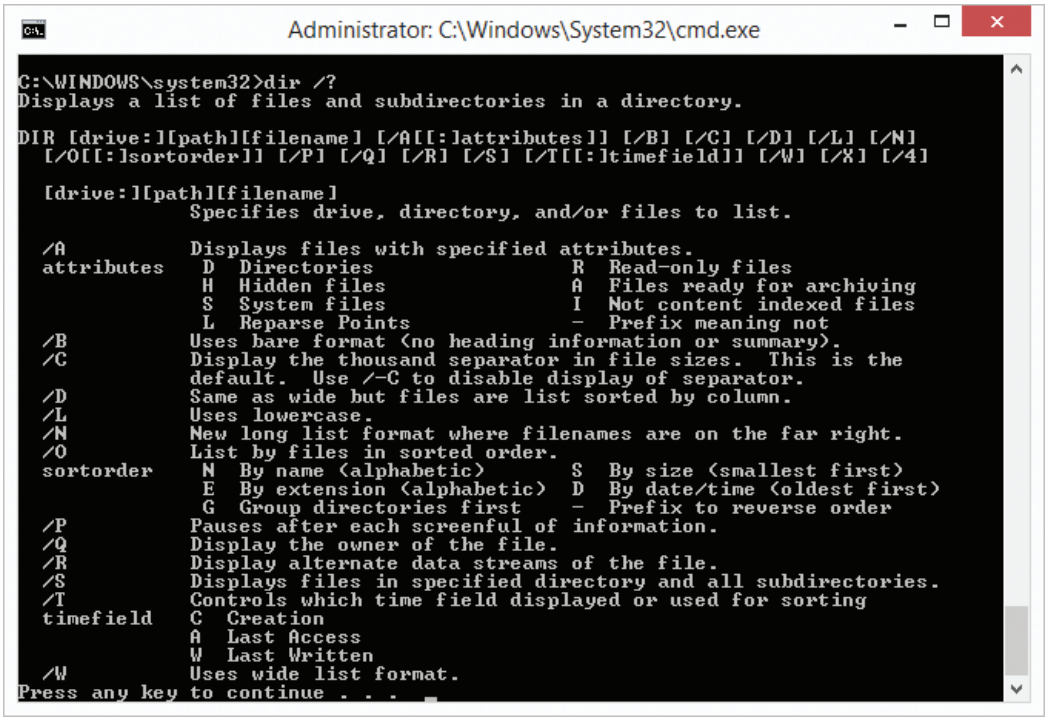


FIGURE 6.24 The unfriendly but powerful Windows command-line interface (CLI)

Eventually, researchers developed the concept of a graphical user interface where text commands could be replaced by intuitive actions such as clicking on icons and dragging objects around on screen. Credit for this 'WIMP' (Windows, Icons, Menus, Pointer) concept must go to Xerox PARC researchers from Palo Alto, California, who developed it in 1973. Apple later popularised WIMP ideas with the Macintosh in 1984. Microsoft and others soon followed.

GUI make software easier for end-users not interested in becoming power users. Power users are generally experts who have learned various efficient techniques, such as keyboard shortcuts and command-line text-based instructions.

GUI programming requires an object-oriented language that provides predefined classes (objects), such as windows, buttons and listboxes, with offer properties and methods, and which can respond to events.

The programmer writes code that can:

- read properties; for example, current mouse position
- change properties; for example, increase window height
- respond to events, such as a timer or a mouse drag
- cause an object to carry out a method, such as close, or refresh.

Object-oriented programming (OOP) lets programmers hand off a lot of work to the OS, such as detecting events like mouse clicks, sorting listbox items and scrolling windows. The programmer creates **event handlers**, which are procedures to handle events, such as, 'When the user clicks this button, take this action ...'

This type of programming is considered **event-driven** because the OS detects and reports events and the programmer provides responses to them. A drawback of GUI programming is that it creates programs with large appetites for RAM, disk storage and CPU power. Such programs may run poorly, or not at all, on mobile devices.

Searching

Finding a single item among billions, like a database of eBay auctions, is the sort of thing programmers constantly need to do, and need to do efficiently; that is, without wasting time, money and effort.

A **linear search** checks every individual item in turn to see if it matches the item for which you are searching. It can be very slow when there is a lot of data to search, but it is easy to program.

The faster, more clever search is the **binary search**, which is preferable when searching large data sets, especially when speed is important. The drawback is that the data set must be sorted, and sorting is itself a slow, difficult task.

There are many other types of searches, each with benefits under certain circumstances. Entire treatises, such as *The Art of Computer Programming* by legendary Donald Knuth, have been written to study the theory of searching.

CASE STUDY ETHICAL DILEMMA

To help you tackle this dilemma, search online for 'ACS code professional conduct' and look at the ethical requirements, particularly in Section 1, that the Australian Computer Society expects its members to follow.

Jean, a statistical database programmer, is trying to write a large statistical program needed by her company. Programmers in this company are encouraged to write about their work and to publish their algorithms in professional journals. After months of tedious programming, Jean has found herself stuck on several parts of the program. Her manager, not recognising the complexity of the problem, wants the job completed within the next few days. Not knowing how to solve the problems, Jean remembers that a co-worker had given her source listings from his current work and from an early version of a commercial software package developed at another company. On studying these programs, she sees two areas of code that could be directly incorporated into her own program. She uses segments of code from both her co-worker and the commercial software, but does not tell anyone or mention it in the documentation. She completes the project and turns it in a day ahead of time.

Source: a case study from Australian Computer Society: ACS Code of Professional Conduct Case Studies, April 2014.
More case studies: https://www.acs.org.au/__data/assets/pdf_file/0004/30964/ACS_Ethics_Case_Studies_v2.1.pdf
Code of professional conduct:
https://www.acs.org.au/__data/assets/pdf_file/0014/4901/Code-of-Professional-Conduct_v2.1.pdf

THINK ABOUT COMPUTING 6.21

Without reference to copyright or other laws, identify the following.

- What is Jean's ethical dilemma?
- What are her options?
- For each of these options, what are the likely consequences of her choosing it?
- How could she get out of this dilemma?
- What relevant clauses in the 'Australian Computer Society's Code of Professional Conduct' cover this situation?

Ethical dilemmas are discussed in Chapters 1 and 5.

CHAPTER SUMMARY

ESSENTIAL TERMS

accessibility ease of use by people with disabilities or special needs

algorithm strategy behind a calculation or procedure

arrays storage structures with many 'slots' (elements) that are addressed by number; arrays are managed with loops

central processing unit (CPU) the 'brain' of a digital system; the handler of most of a system's data manipulation

code modules subroutines or subprograms; a function is a special type of module because it sends back (returns) a value; programmers must keep track of storage structures in multiple modules, making data dictionaries even more valuable

communication hardware hardware that transmits data between computers and networked devices

compiler programs that convert source code into executable programs

debugger a program that helps to remove programming errors (bugs)

default action the operation that the software will carry out when the user does not give more detailed instructions

default value used if the user does not provide an alternative value; a word processor may default to using Arial typeface, 12pt, with single-spacing

editor a specialised word processor for creating human-readable programming instructions

field a single data item in a record; e.g. FamilyName

graphics processor unit (GPU) a very fast and expensive processor specifically designed for high-speed image processing in graphics cards

initialise to give a starting value to a variable

input devices instruments and peripherals, such as keyboards, that allow users to give data and commands to software and the OS

integrated development environment (IDE) a unified programming tool

interface within software, the place where people control the program, enter data and receive output

iteration looping or repeating

linker a program used to load information that the executable code will need, such as read a keyboard or calculate square roots

methods actions a GUI object can carry out; e.g. window.refresh.

mock-up a sketch showing how a screen or printout will look, which is used to aid in the design of an interface

modular programming breaking programs into small sections of code to simplify debugging and allow the re-use of modules in other programs

object any item that a program can inspect and/or change, in terms of appearance, behaviour or data

operating system (OS) software programs that manage a computer's hardware and run programs

output devices instruments and peripherals, such as printers and monitors, that display information from a computer in human-readable form

platform a combination of OS and CPU

ports physical connectors (sockets) for cables

primary storage random-access memory (RAM), which provides storage for data, information and software during program execution

processing hardware (CPU, GPU) hardware that runs the operating system, utilities and applications

properties characteristics such as width, colour, visibility

prototype a demonstration product that looks and feels like a finished program, but may be incomplete or not fully functional

pseudocode code that designs algorithms in a clear, human-readable, language-independent format

random-access memory (RAM) the primary and most common form of hardware storage; it can be accessed randomly, meaning that any byte of memory is accessible without touching the preceding bytes

random files records of identical length with fields that must be rigidly defined in advance

reduced instruction set computing (RISC) CPUs (like ARM), which have smaller instruction sets than complex instruction set computing (CISC) CPUs

queue a 'First In First Out' stack, storing incoming data or jobs to be processed in order

record a complete set of fields relating to an entity, such as a person

secondary storage (HDD, SSD) permanent storage

sequential file plain text with variable field lengths, such as CSV; large files are slow to search

source code human-written and human-readable version of a program

CHAPTER SUMMARY

stack a simple, temporary 'first in last out' (FILO) storage structure

storage structures places in memory holding data that is being used by a program; includes variables, arrays, textboxes and radio buttons

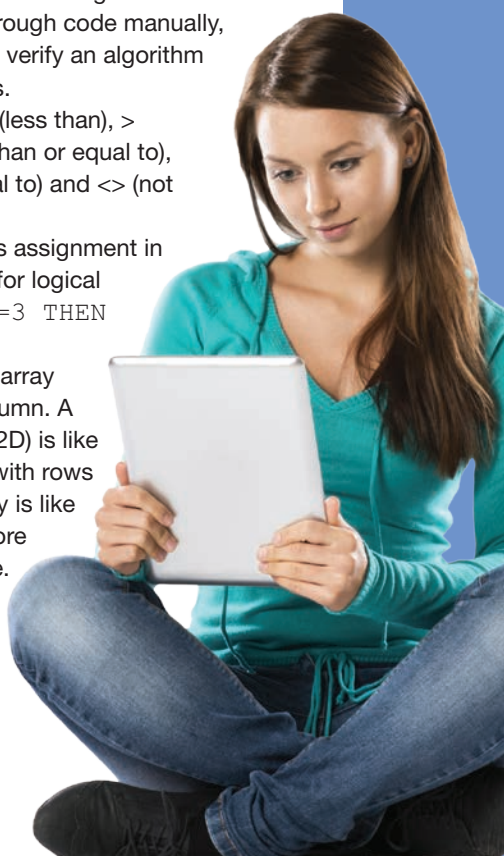
style guide in software development, a set of instructions for programmers about how to design software for the developer's platforms

thin client the belief that it is better to use 'dumb' workstations connected to a powerful central computer that does all the processing work, rather than use many powerful computers

validation rules rules that check the reasonableness (not accuracy) of data's range, type, and existence

IMPORTANT FACTS

1. The three types of software are system (OS), applications and utilities.
2. Interpreted languages convert source code to executable code only when a program runs.
3. Information systems are made up of people, data, processes and equipment (digital systems).
4. Digital systems comprise hardware, software, networks, protocols and application architecture patterns.
5. Transmission media such as twisted-pair cables, fibre-optic cables and WAPs each have their strengths and weaknesses.
6. The problem-solving methodology (PSM) is used to guide software development.
   - Analysis determines the software's requirements, scope and constraints.
   - Design devises a method of solving the problem, including its appearance and architecture. Tools include pseudocode, IPO charts, data dictionaries and mock-ups.
   - Development is when software and hardware are created, assembled and tested. Training and documentation are created.
7. Good naming practices include Hungarian notation (e.g. intTemp) and CamelCase.
8. A data dictionary lists all storage structure requirements, including data types and names, size, validation rules etc.
9. Data types include integer, floating point, string, character, date/time and Boolean (true/false).
10. IPO charts plan the data and processing needed to produce output.
11. Data structure diagrams describe the structure and relationships within complex data structures.
12. Object descriptions completely describe the properties of objects; e.g. textbox.width.
13. Internal documentation in source code explains the workings of the code to programmers.
14. Mock-ups show what screens and printouts should look like, including sizes and positions of items, colours, alignment and fonts.
15. Onscreen images are saved with RGB colour information. Printed images use CMYK.
16. Modules are called subprograms, procedures or subroutines. Functions are modules that return a value.
17. Global variables are visible to and changeable by the main program and all subprograms.
18. Local variables only exist within a single module.
19. Counted loops (`FOR/NEXT`) repeat a known number of times.
20. Uncounted loops (`DO`, `WHILE`) loop while a condition is true.
21. Uncounted loops can be top-driven (test at top) or bottom-driven (test at bottom), depending on where they test for continuation.
22. Syntax errors occur when a compiler cannot understand instructions in source code.
23. Logical errors occur when a faulty calculation strategy produces incorrect output.
24. Runtime errors occur when problems arise during execution, such as running out of RAM.
25. Desk checking steps through code manually, acting like a compiler to verify an algorithm and detect logical errors.
26. Logical operators are < (less than), > (greater than), <= (less than or equal to), >= (greater than or equal to) and <> (not equal to).
27. The operator ↗● means assignment in pseudocode. = is used for logical comparison (e.g. `IF A=3 THEN B ↗● 4`)
28. A one-dimensional (1D) array is like a list with one column. A two-dimensional array (2D) is like a table or spreadsheet with rows and columns. A 3D array is like a stack of 2D arrays. More dimensions are possible.



CHAPTER SUMMARY

29 Data can be saved to secondary storage as sequential or random files.
30 Random files have fixed-length records that are quickly accessed.
31 Testing ensures that programs work properly and generate accurate information.
32 Good test data checks software behaviour under all possible circumstances.
33 Prove your algorithm and show evidence of testing by using a testing table.
34 Events include mouse clicks, timers and so on.
35 Object-oriented languages use classes of pre-defined objects (e.g. listboxes) that programmers can manipulate.

TEST YOUR KNOWLEDGE

Review quiz

HARDWARE

1 Describe two differences between primary and secondary storage.

SOFTWARE

2 Explain the difference between system and application software.

THE OS (OPERATING SYSTEM)

3 Summarise the role of an operating system.

4 Describe the roles of people, data, processes and digital systems in creating an information system.

PROGRAMMING AND SCRIPTING LANGUAGES

5 'Programming languages may differ in syntax, but they are all basically alike.' What does this mean?

SOFTWARE DEVELOPMENT TOOLS

6 How are source code and executable code related?

STORAGE STRUCTURES

7 Name and give examples of five data types.

8 Describe how lossy and lossless compression reduce file sizes.

THE SOFTWARE DEVELOPMENT PROCESS

9 Why is the PSM analysis stage important?

10 Show examples of three different software design tools.

11 What is an algorithm, and how does pseudocode relate to it?

12 Suggest two possible good names for a variable used to hold an integer value describing men's average shoe size.

CREATING EFFECTIVE USER INTERFACES

13 Describe affordance and tolerance and give examples of each.

14 List five tips for creating an effective user interface.

FUNDAMENTAL PROGRAMMING CONCEPTS

15 List examples of syntax, logical and runtime errors.

16 How do compiled and interpreted languages differ?

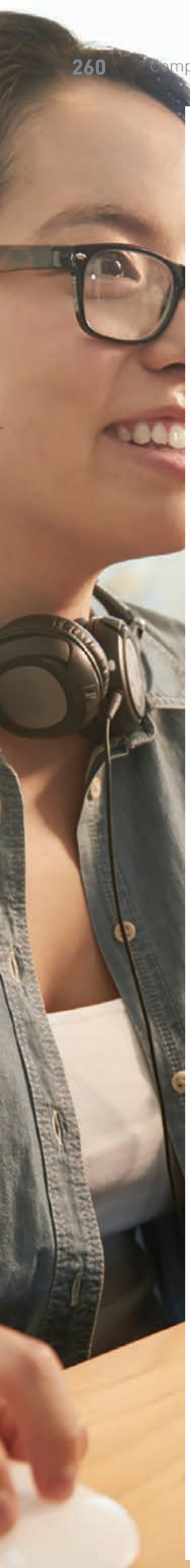
17 Justify the use of internal documentation.

18 'Arrays and loops were built for each other.' Explain.

19 When is desk checking used?

20 Why is it important to use test data that focuses on boundary conditions?

21 When would you recommend the use of a random file?



APPLY YOUR KNOWLEDGE

GUESSING GAME

You will design and develop a program that will play a game with a human. The human thinks of an integer between 1 and 100. Your program will use the most efficient strategy possible to guess the number. For each guess, the human must confirm if the guess was correct, higher or lower than the secret number.

Tasks

1. Write pseudocode to describe the processing strategies involved in developing the program.
2. Create an IPO chart to describe the data and information requirements.
3. Design the program's interface with a mock-up.
4. Create test data to fully exercise the validation and the algorithms.
5. Use your test data to desk check the pseudocode.
6. Develop the solution, using good object naming and creating useful internal documentation along the way.
7. Test the solution with your test data. Fix all bugs.
8. Get a classmate to play your game and give you written feedback on its appearance and ease of use.

PREPARING FOR

UNIT 2 OUTCOME 1

Design working modules in response to solution requirements, and use a programming or scripting language to develop the modules

STEPS TO FOLLOW

The Outcome will require approximately 10 classes to complete. For each programming module, you will be given solution requirements (the analysis part of the PSM), and you will select appropriate design tools to plan a solution, and then develop and test a working model of the solution. Use pen and paper for the design to avoid prematurely beginning development.

Your teacher may choose to give you the tasks one at a time after periods of relevant theory instruction, or as a group after all of the theory has been covered. To encourage serious attention to design, your teacher may choose to enforce design-only periods of time where no computers may be used.

Keep all of your designs, even failed attempts. A change of mind during design can actually be a very desirable event!

DOCUMENTS REQUIRED FOR ASSESSMENT

- Screen mock-up
- Data dictionary
- IPO chart
- Internal documentation (must be used within the source code)
- All solutions must show evidence of testing using appropriate test data.

ASSESSMENT

You will be assessed on the following measures.

- Whether your chosen design tools are appropriate and useful
- Your choice of data types and storage structures
- Object and file naming
- The accuracy of calculations
- The appropriateness of software functions you have used
- Functionality: Does the program do everything it is meant to do?
- Appearance of both the interface and the output
- Ease of use
- Internal documentation
- Thoroughness of testing

CHAPTER 7

DATA ANALYSIS AND VISUALISATION

Key knowledge

After completing this chapter, you will be able to demonstrate knowledge of:

Data and information

- sources of authentic data in large repositories
- factors influencing the integrity of data, for example accuracy, timeliness, authenticity, relevance
- characteristics of data types and data structures relevant to selected software tools

Approaches to problem solving

- types and purposes of data visualisations
- problem-solving activities related to analysing needs: functional and non-functional requirements and constraints
- characteristics of file formats and their suitability to be converted to other formats
- design tools for representing data visualisations
- formats and conventions applied to visualisations to improve their effectiveness for intended users
- functions of appropriate software tools to extract targeted data and to manipulate data when developing visualisations
- criteria and techniques for evaluating visualisations.

For the student

If you can imagine that the amount of data that is generated every day just by using social-networking tools, you might also be able to imagine that there is someone, somewhere, who is looking through a mountain of data looking for meaning.

Data visualisation is the process by which we take large amounts of data and process it into effective graphical representations that will meet the needs of users or clients.

These representations can take the form of charts, graphs, spatial relationships and network diagrams. In some cases, the data visualisation might involve interactivity and the inclusion of dynamic data that allows the user to deduce further meaning from the visualisation.

For the teacher

This chapter introduces students to the knowledge and skills needed to use software tools to access authentic data from repositories and present the information in a visual form.

The key knowledge and skills are based on Unit 2, Area of Study 2. If a data visualisation is effective, it reduces the effort needed by readers to interpret information. This chapter takes students through the different types of visualisations, and then uses a case study to explore some of the tools available to process data from repositories, such as the Bureau of Meteorology and the Australian Bureau of Statistics.

Information needs and data visualisations

When clients or users require particular information, and no system currently exists that provides the information, then an information need has been identified.

This could be due to an existing **information problem** (an organisation is worried about declining sales), an **identified need** (park rangers need a method to communicate weather conditions on total fire ban days), or an **opportunity** (currently no list of driving instructors in Victoria exists).

When an information need has been identified, one process used to help to create a solution that will meet the needs of the clients or users is the **problem-solving methodology (PSM)**.

The problem-solving methodology is discussed in Chapter 2.

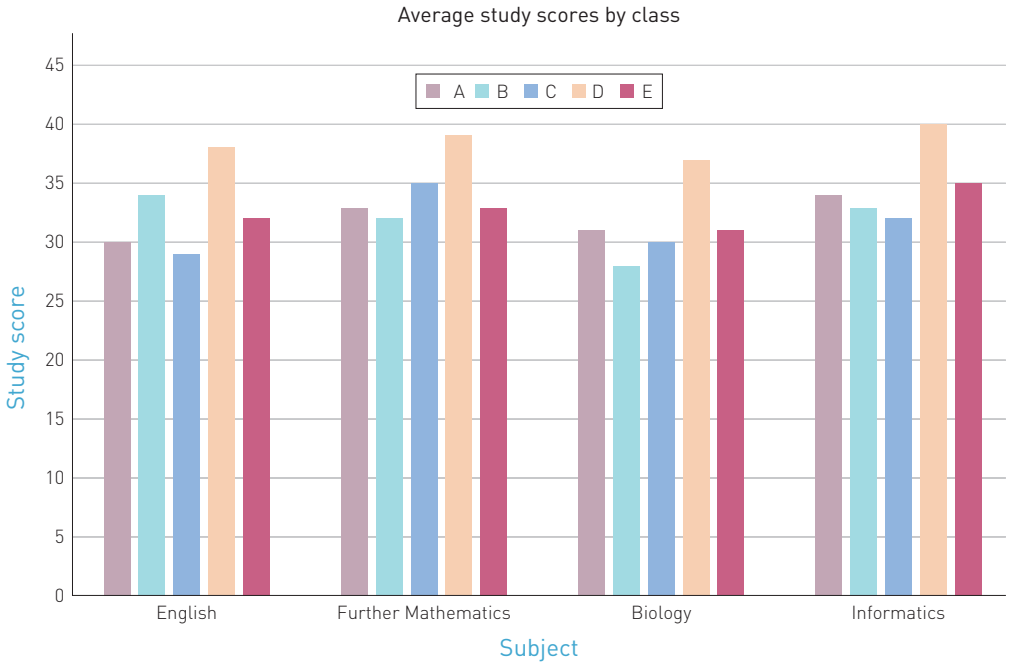


FIGURE 7.1 Visual representation of average VCE subject study scores by class

Sometimes the information need might revolve around finding a method to represent a data set. Looking at data in a table does not provide much meaning for the user. Table 7.1 displays the subject results for students in different classes.

TABLE 7.1 Student test results

Class	English	Further Mathematics	Biology	Informatics
A	30	33	31	34
B	34	32	28	33
C	29	35	30	32
D	38	39	37	40
E	32	33	31	35

At first glance, the data does not really have much meaning, nor are any relationships or patterns in the data identified. By following the four stages of the problem-solving methodology, the requirements needed from the solution will be identified, possible solutions designed, then

developed, and after the solution is created, the user can evaluate whether the solution is meeting their needs (the requirements identified in the analysis stage).

Figure 7.1 on the previous page is a solution that visually represents the data shown in Table 7.1.

The solution presented in Figure 7.1 used a column chart to represent the data. A column chart is useful for comparing different data items. In this situation the average study score of each class is compared with the other classes. When looking at the data in the table it is difficult to identify any patterns or relationships. Once the data is converted into a visual format it quickly becomes apparent that Class D performed the best in each of the subjects.

With the use of data visualisations, further investigation can be undertaken to understand why Class D performed better in the VCE than the other classes. Did Class D consist of all the high achievers at the school? Did they have the best teachers? Did the students in Class D complete more homework than the other classes, or was it just a coincidence?

THINK ABOUT COMPUTING 7.1

TV weather reports use data visualisations to present weather conditions. Think of other situations from day to day, where data is presented in a visual format that is more meaningful and is quick and easy to understand.

Sources of authentic data

As discussed in earlier chapters, there are two sources of data: primary and secondary.

Primary sources

Primary data sources involve collecting the data directly or firsthand from stakeholders. Methods used to collect primary data include **interviews**, **observations** and **surveys**. By its nature, data collected from primary sources should be considered authentic. Measures, though, need to be put in place when collecting from primary data sources to ensure the data is accurate, free from bias, reliable and relevant.

Secondary sources

Collecting data from **secondary data sources** involves using data that other users have collected firsthand. The range of sources holding secondary data include newspapers, books, magazines and reports, to name a few, with a wide range of organisations collecting data for their own particular information needs.

One important question to ask when using data from a secondary source concerns how authentic it is. Finding similar data from two or more sources helps to authenticate the data. Also, collecting data from organisations that are considered reliable and trustworthy helps to increase the chance that the data is authentic. **Government departments** collect and store vast amounts of data and take many precautions to ensure that the data they collect is authentic. They are increasingly making these data sets available for public use. As a result, government departments are considered a good source of authentic data. Some of these sources are discussed on pages 265–67.

Bureau of Meteorology

The Bureau of Meteorology (BOM) is Australia's national weather, climate and water agency. They collect a wide range of climatic data to be able to make regular forecasts, issue warnings and offer advice. Much of the data that they collect is available for public use.

Bureau of Meteorology

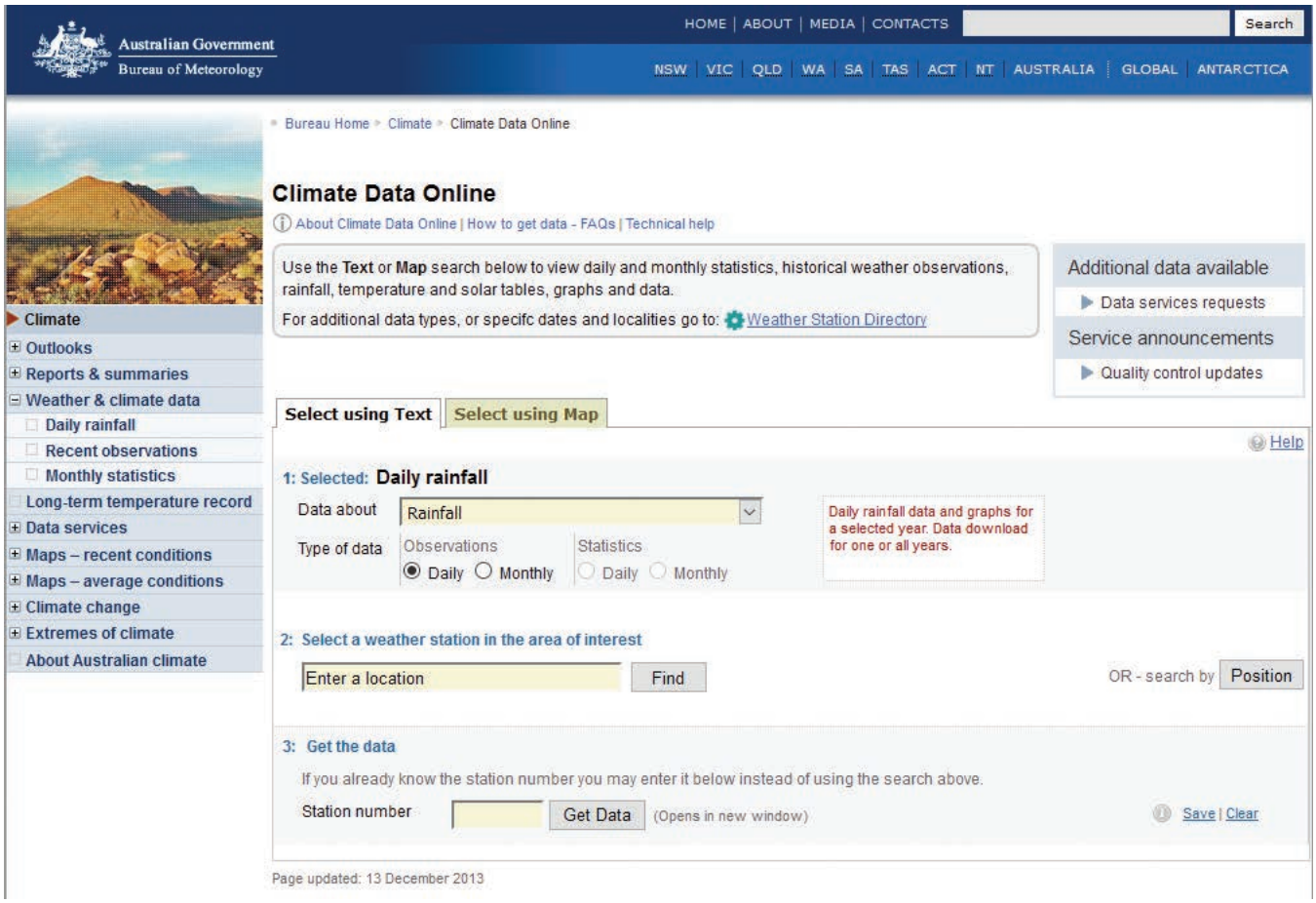


FIGURE 7.2 A screenshot from the BOM's website

Victorian government data directory

The Victorian government data directory contains a range of data sets from Victorian government departments. There are data sets available on a diverse range of subject areas including education, finance, health and town planning. The data sets offered are in a range of file formats, including spreadsheets, audio and video, comma separated values, geospatial and API tools.

Victorian government data directory

https://www.data.vic.gov.au

FIGURE 7.3 A screenshot from the Victorian government data directory website

Australian Bureau of Statistics

The **Australian Bureau of Statistics (ABS)** is the statistical agency of the federal government. The ABS provides statistics on a wide range of economic, environmental and social issues, for use by governments and the community. Data sets are available on a diverse range of topics from foreign trade, agriculture, sporting facilities to crime statistics. In fact, the ABS website provides an endless supply of open data.

ABS

FIGURE 7.4 A screenshot from the ABS website

Data.gov.au

Data.gov.au is similar to the Victorian government data directory in that it provides publically available data sets that have been collect by federal government departments.

Many other organisations offer data sets to the public.

Data.gov.au

FIGURE 7.5 A screenshot from Data.gov.au

Data integrity

Data integrity refers to the quality of the data. An information system's purpose is to transform data into useful information. It therefore is important that the data inputted has integrity, because the information produced is based on the data itself. For example, if a customer database contains the wrong postal details for a customer this is incorrect data. When a data set contains these types of errors, it loses its integrity. The more errors the data set contains, the lower its integrity.

There are many ways the quality or integrity of the data can be measured. When using data from either primary or secondary sources, factors to consider include accuracy, timeliness, authenticity and relevance.

THINK ABOUT COMPUTING 7.2

Create a list of four other organisations in Australia who provide data sets for public use. Each organisation can be a government department or private organisation.

Accuracy

Accuracy involves ensuring that the data collected is correct and does not contain any errors.

When using primary methods to collect the data, validation may be able to be used to reduce the chances of incorrect data being entered. Data validation often involves restricting the data that can be entered into a particular field and by restricting what can be entered. The chances of incorrect data being input are therefore reduced.

Many online forms contain a number of validation methods that help reduce the chances of errors being input. Validation techniques include dropdown lists, radio buttons, predictive text, check boxes and required fields.

THINK ABOUT COMPUTING 7.3

How many validation techniques are used in the online checkout form from cafe press (Figure 7.6)?

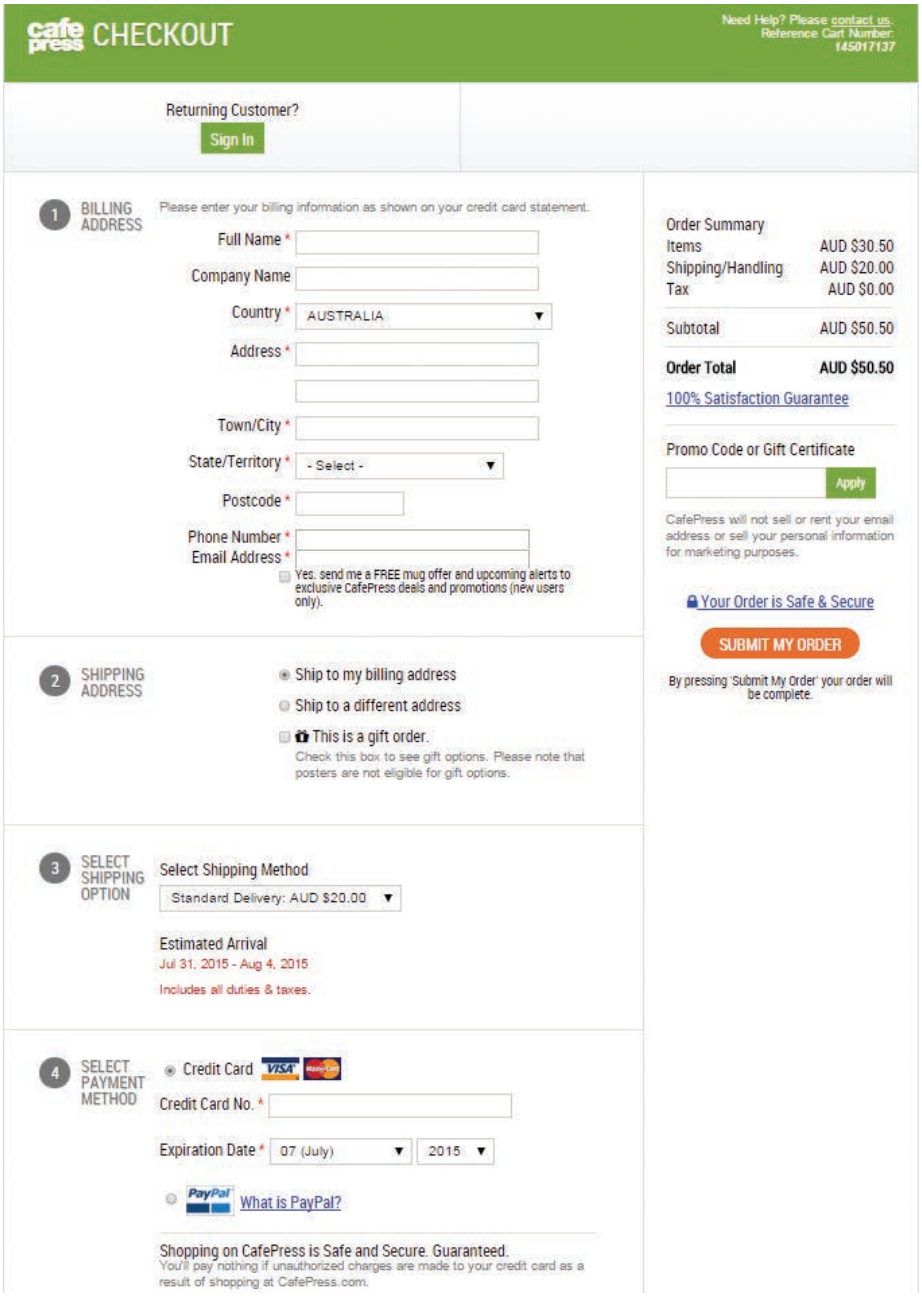


FIGURE 7.6 Online form containing validation techniques

Timeliness

Timeliness relates to the age of the data – how old is it? The data used should be relevant for the time period. For example, using Melbourne's population data from the 1990s to help plan the location for new primary schools would result in a plan that does not match the city's current needs. It is important that the data input into the information system is timely to match what is needed, and is not collected too early (or too late).

Timeliness can also relate to the information provided by the information system. If the information produced is not provided in a suitable timeframe it would be useless. Imagine a school produces a daily bulletin outlining all the events that occur each day, but the bulletin is always published at the end of the school day, and not the start. The information is not being received in a timely fashion and, therefore, is useless to the organisation.

Authenticity

Authenticity relates to how genuine the data set is. Is the data genuine, original, accurate, reliable and trustworthy? The authenticity of primary data is easier to confirm as it has been collected firsthand, although precautions should be put into place to ensure it is accurate and reliable. Also, when collecting primary data it is important that the sample size is not too small, as this may lead to inaccurate results.

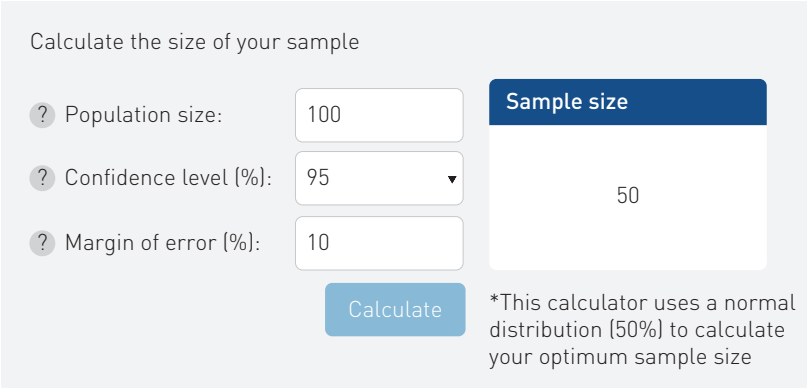


FIGURE 7.7 Online calculator to calculate optimum sample size

Relevance

Relevance is related to whether the data helps to produce the information required to meet the information need of the user. There is no point using the data that does not contribute to producing the information required. Collecting data about the weather is useless if the purpose of the system is to identify where new hospitals should be located in Australia. There are many attributes of relevant data but ultimately the data should help achieve the needs of the user.

Data types and data structures

Data types are particular forms that an item of data can take, including numeric, character and Boolean. The number of data types can vary between software applications. Each piece of data can be classified into one of the data-type groups. Common data types include the following.

Character

A **character** data type holds a single letter, number or symbol and is usually stored in 2 bytes of memory. Numbers stored as a character data type cannot be used in mathematical calculations. Examples of characters include `a`, `&`, and `5`.

Text (string)

A **text** data type consists of a 'string of characters'. The terms 'text' and 'string' are sometimes used interchangeably to refer to the same data type. Within software applications, the term 'text' is often used. If the data consists of 2 or more characters then a string data type is used. The amount of memory used to store the string will depend on the length (2 bytes × length of string). Again, numbers stored as a string cannot be used in mathematical calculations. Often a phone number is saved as a string. Examples of strings are `Hello`, `0406000000` and `ABC123`.

Integer

An **integer** is a number without a fractional or decimal component, e.g. a whole number. An integer is usually stored in 4 bytes of memory. Numbers stored as integers can be used in mathematical calculations. Examples of an integer include `-345`, `0`, and `1234567890`.

Floating point (or decimal)

A **floating point** number is a number with a fractional or decimal component, such as a decimal number. A floating point number is split into two parts: The mantissa and the exponent. The mantissa holds all of the digits with a decimal point after the first digit. The exponent holds the power of 10 by which the mantissa must be multiplied to regain the original number. A floating point can take up to 12 bytes of memory. Examples of floating point data types include `3.17`, `-6.890` and `0.001`.

Boolean

A **Boolean data type** can only hold two possible values, usually true or false. A Boolean data type is known as a logical data type. It is often used in conditional statements to test if a condition or value is true or not. Boolean data types are efficient in terms of storage space as they only require 1 byte of memory for storage.

THINK ABOUT COMPUTING 7.4

In theory, a Boolean data type only requires 1 *bit* of memory space. Find out why it needs to take up 1 byte of memory instead.

Data structure

A data structure is a particular way of organising a collection of data items. Using data structures helps to improve the efficiency of the data handling process. Three common data structures are arrays, records and files.

Array

An **array** is a collection of data items generally of the same data type. Each item in the array is allocated an address.

Table 7.2 illustrates an array data structure. The structure is named 'Number' and holds six data items – all integers. The value of Number (0) is 40 and Number (5) is 67.

TABLE 7.2 Array data structure

Number	
Index	**Value**
0	40
1	34
2	56
3	21
4	12
5	67

Record

A record is a collection of data items of different data types. Each element of a record is generally called a field. The record in Table 7.3 consists of six fields and involves data of different data types.

TABLE 7.3 Record data structure

Customer ID	Name	Phone	VIP	Quantity	Cost
01901	Karen	0406348790	Y	5	3.99

THINK ABOUT COMPUTING 7.5
Referring to Table 7.3, how many different data types can you identify?

File

A file data structure can hold numerous data items, arrays or records. A file is saved separately from the software program that utilises the file. Large data sets can be read from, or written to, a file for storage purposes. The file to the right shows a collection of records related to student results.

Figure 7.8 is an example of a comma-separated value (CSV) file. A range of file types will be discussed later in the chapter.

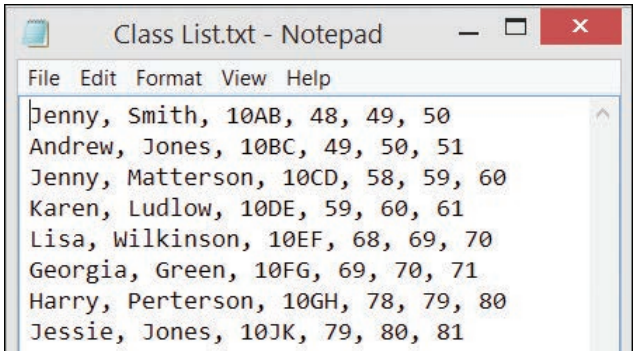


FIGURE 7.8 File data structure

Types and purposes of data visualisations

Data visualisation is a relatively new field of computing. As more and more data sets are being made available, the number of individuals and organisations performing analysis on this data is also increasing. The types of visualisations being used is endless. Data is being manipulated and processed into new forms of visualisations every day.

Figure 7.9 shows different types of charts used for different purposes, but these only scratch the surface in terms of types of data visualisations.

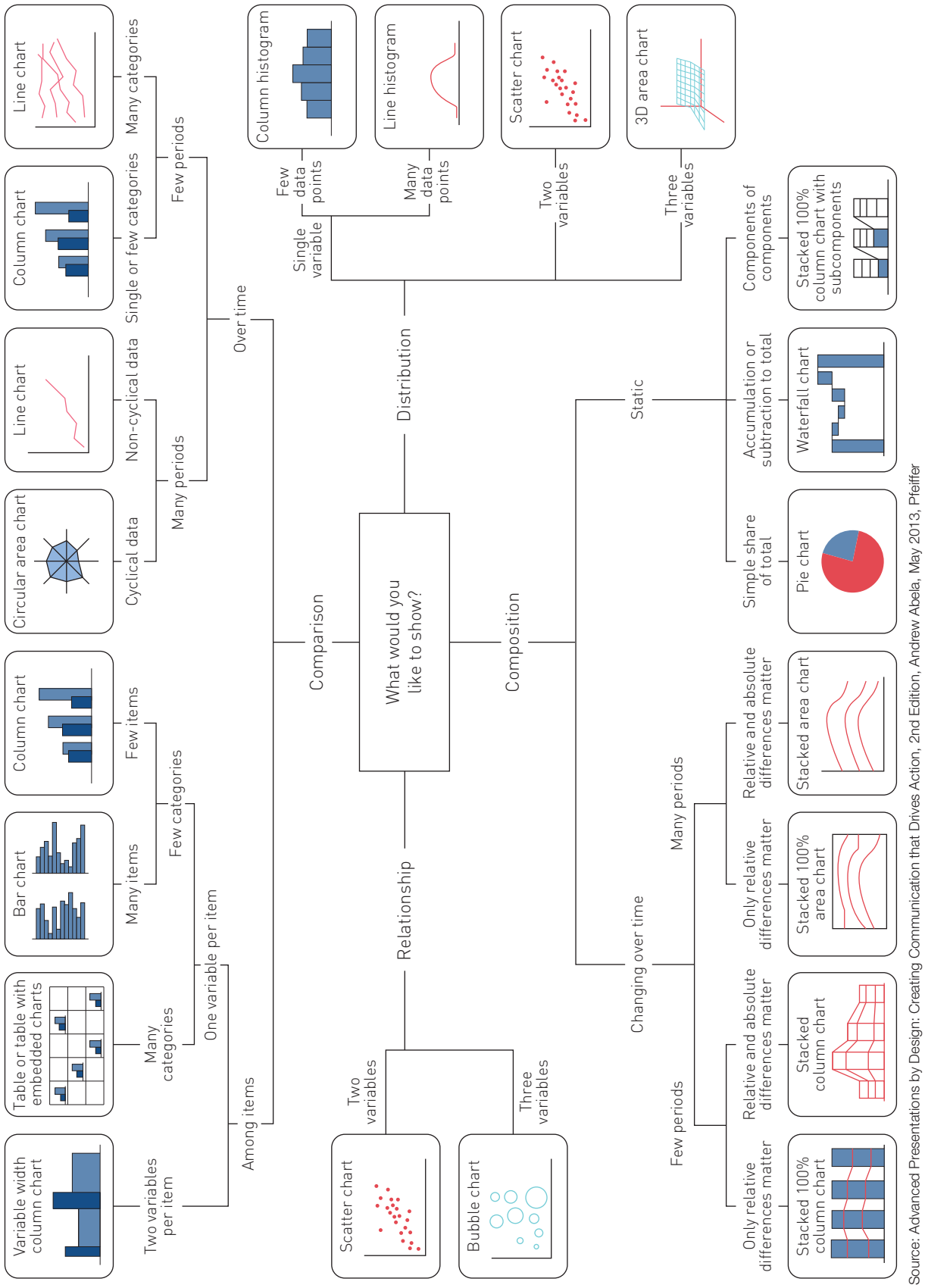


FIGURE 7.9 Some common types of data visualisations; it does not include network diagrams or maps

Charts

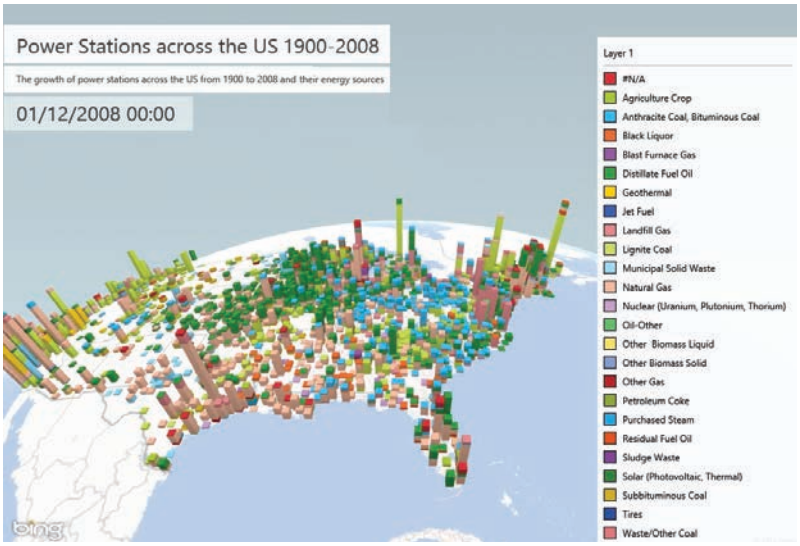
A **chart** (also known as a graph) is a method of displaying data visually, where the data set is represented as symbols in the chart. Charts are a popular way of presenting data visually and many spreadsheet applications have charting (or graphing) capabilities. Features of a chart can include a title, axis, scale or grid, data labels and a legend.

Charts are often used to visualise numerical data. There are a range of chart types, where each type is used for different purposes. A bar chart can be used to compare different items, a pie chart to show each data item as a proportion of the population, line charts are useful for showing the trend in a data item over time and histograms are useful for grouping data then showing the frequency of each group.

Shutterstock.com/vasabii

FIGURE 7.10 Types of charts

Maps



© 2015 Microsoft

FIGURE 7.11 Geospatial visualisation

A popular method to display geographical data is by using **map-based visualisations**. These types of visualisation are often called **geospatial visualisations**. Geospatial data is data that is related to the geographical location covered. Data could be related to population, roads, rivers, climate, mobile phone towers or any other characteristic of the area. Many geospatial visualisations are dynamic and allow the user to zoom in or out or navigate over an area. Geospatial visualisations are becoming more popular because they are a very powerful tool that allows the data to be brought to life through visualisation. As a range of data can be overlayed with a geographical location the uses of these types of visualisations are enormous. Common uses have been for agricultural, environmental, mining and urban planning purposes, but the list is endless.

Networks

Network visualisations show relationships between different data items and relationships between different data sets. Finding relationships within and between data sets has been an increasing area of interest in recent years as more data has become publically available from both government and private organisations around the world.

A network visualisation might show the frequency with which individual players might pass the ball to each other in a football game or the number of people who travel on a public transport system each day. Network visualisations are also used to represent the layout of computer networks, or public transport systems. Figure 7.12 shows an example of a network visualisation.

NelsonNet additional resource: Figure 7.11 Geospatial visualisation, p. 274

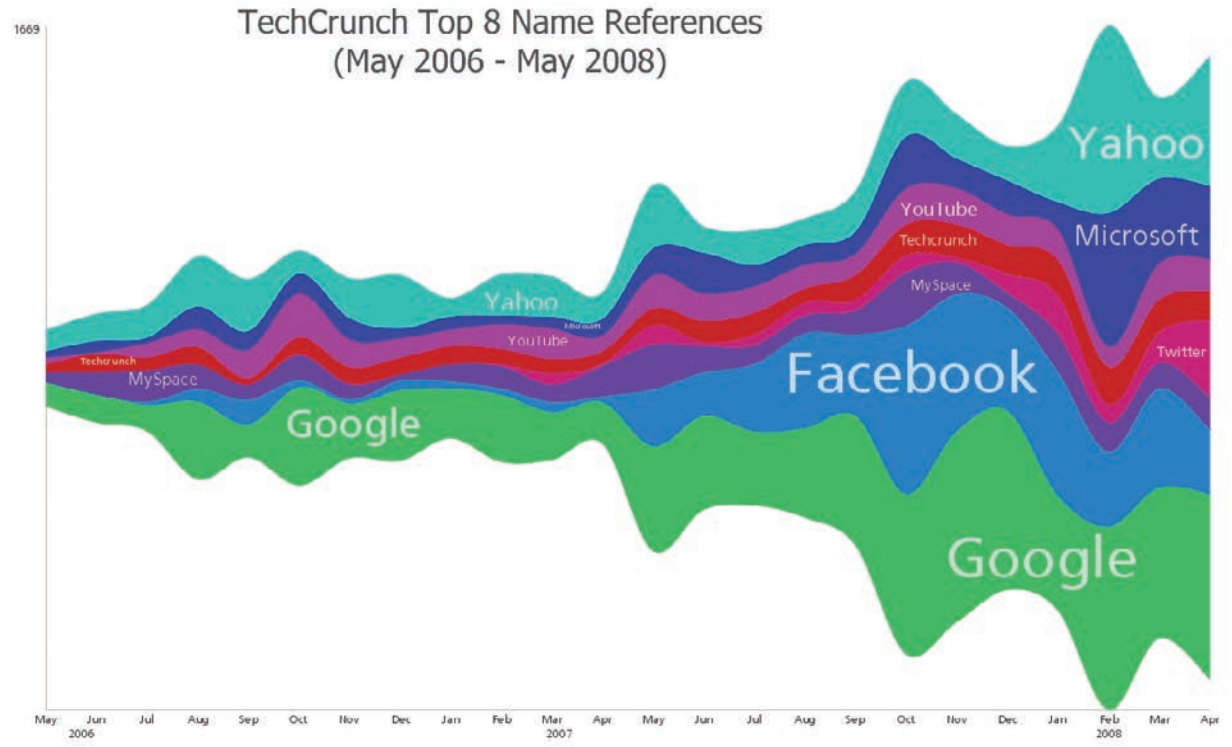
Alamy/Route55

FIGURE 7.12 Network visualisation

Time series

Time visualisation represents a data item or data set over a period of time. Some time-based visualisations will show historical data, while others capture live data to provide real-time information. It is also possible to display the dimension of time by adding motion or animation to create a dynamic data representation.

The data could also be related to timeline or time series. Timeline data may relate to individual items or events and show the order in which the items or events occurred over a time period, while time series data may relate to the same data item and show the variations or changes in the item over a time period. The visualisation shown in Figure 7.13 is an example of a time series representation, showing the top eight name references to social networking sites on the internet from 2006 to 2008.



Jeff Clark, Neoformix.com

FIGURE 7.13 Time visualisation

Hierarchy

Hierarchy visualisations show the relationships and structure between data items. In a hierarchy, data items are represented as being above, equal or below other items in the data set. Types of hierarchy visualisations can include organisational, structure and tree charts and mind maps.

They can be used to show the relationship between items in a data set or they can illustrate the breaking down of items into smaller components. Hierarchy visualisations are also useful for representing non-numerical data types. Figure 7.14 shows a word map representing Chapter 3 of this book. The subject networks have been broken down into sections and then the components of each section branch out from the centre.

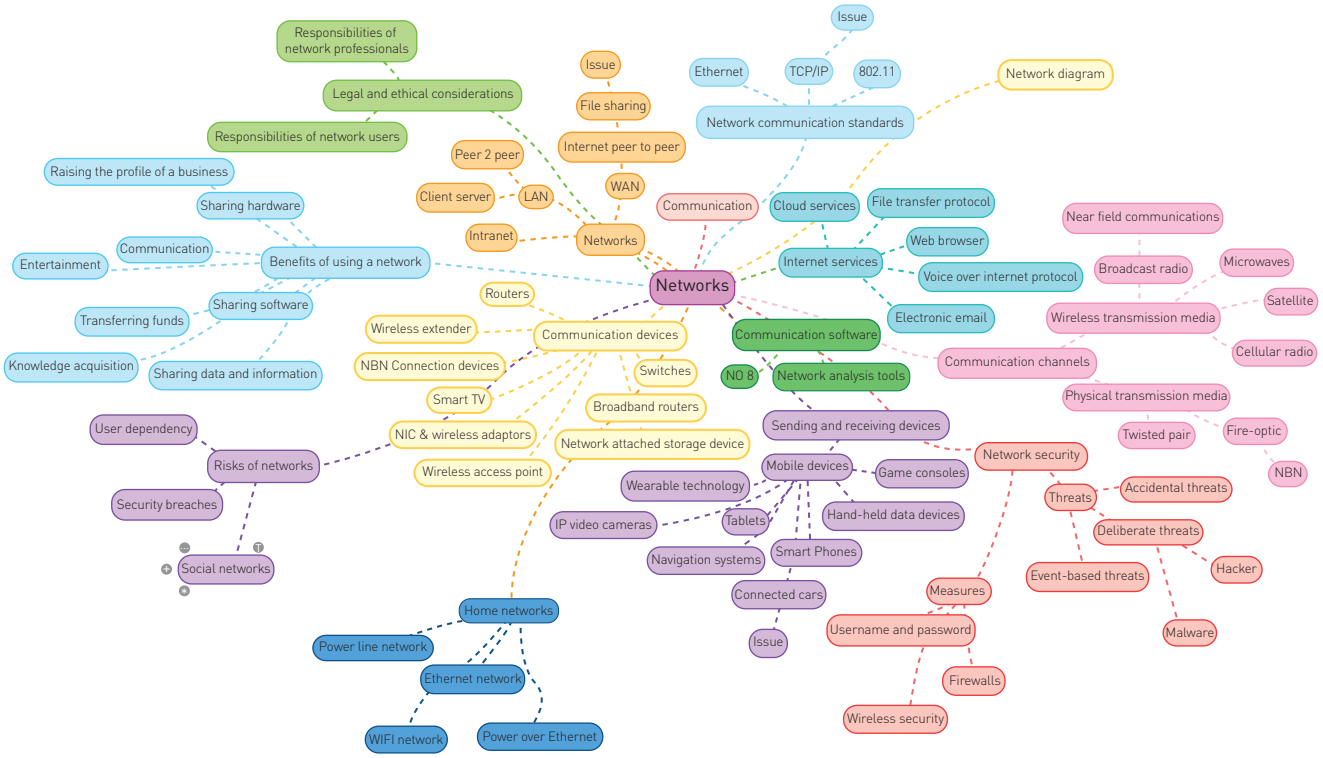
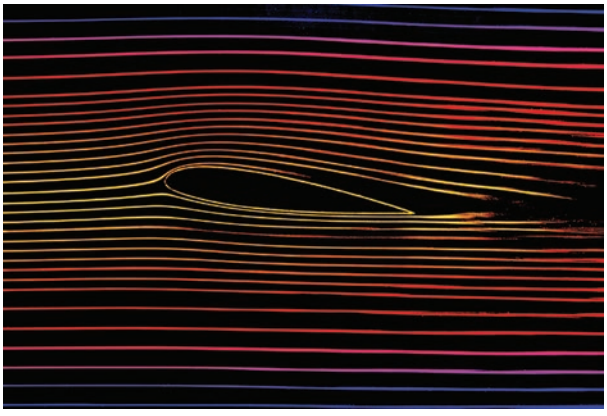


FIGURE 7.14 Hierarchy visualisation

Flow

Flow visualisations involve representing data that illustrates the flow pattern of a data item or items. This could be the pattern of customer movements through a supermarket or the series of pages a user would visit on a website to complete a transaction (user-flow diagram).

THINK ABOUT COMPUTING 7.6
User-flow diagrams are studied in Unit 3, Outcome 1 Informatics. Using a search engine, find three examples of user-flow diagrams.



Science Photo Library/Omikron

FIGURE 7.15 Flow visualisation

Flow visualisations are also used for scientific purposes to visualise the flow patterns of objects that are normally invisible, including air and water. Figure 7.15 represents the effect of an aircraft wing on the airflow passing the wing. The data for this visualisation was collected during testing using a wind tunnel, and the data converted to a visualisation.

Matrix

Matrix visualisations can be used to show the composition of individual items in the sample size. In this regard, they can be considered similar to pie charts. Matrix visualisations often divides the display area up into grids (similar to cells in a spreadsheet). Different sections of the display area are then used to represent the proportion of individual (or groups) of data items.

Matrix diagrams can be used to compare different data items, similar to a scatter chart. Again, the display area is divided in a grid or table format. Figure 7.16 shows a number of matrix visualisations with each visualisation representing a county in the United Kingdom. The colours represent the percentage of people who voted for a particular political party in an election.

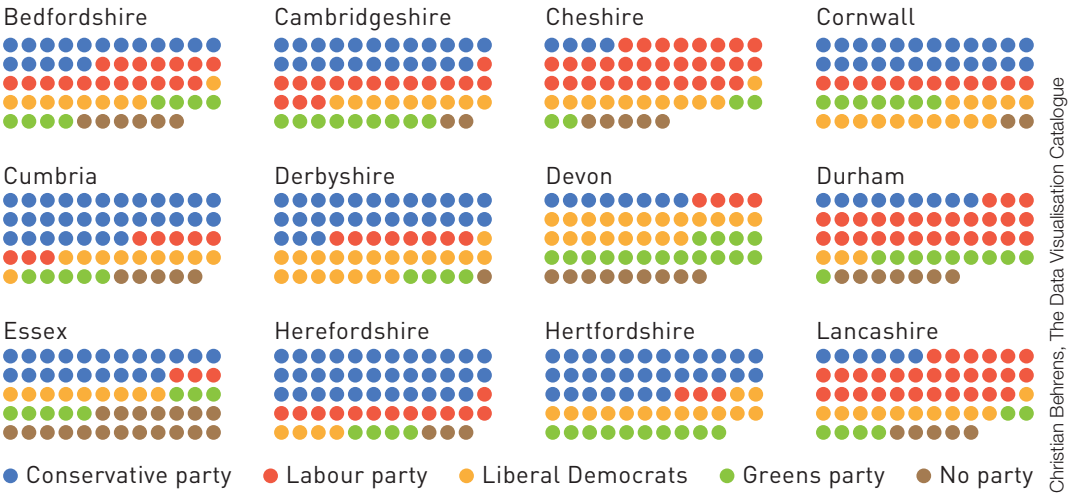


FIGURE 7.16 Matrix visualisation

PSM: Analysis

As mentioned earlier, when an information need is identified, one method to create a solution that will meet the need is to use the problem-solving methodology.

The first stage of the methodology, the analysis stage, is often considered the most crucial. Many organisations invest more time, effort and money for this stage, as they consider getting the analysis stage correct and having a clear picture of what is required will save time and money in the latter stages.

The analysis stage is typically about 'what' questions. What is the current information need? What is required to meet the current information need? What constraints may restrict the requirements? What is the scope of the solution?

The analysis stage consists of three activities: determining solution requirements, determining solution constraints and deciding upon the scope of the solution.

THINK ABOUT COMPUTING 7.7
Explain how completing a thorough analysis stage may save time and money in the latter stages of the methodology.

Solution requirements

Solution requirements are what the client needs from the solution. What features do they want in the solution? Solution requirements can be broken down into functional and non-functional requirements.

Functional requirements are directly related to what the solution will do. Some examples of functional requirements for data visualisations include:

- display population data of towns in Victoria in a visual format
- represent the distances between these towns
- allow the user to zoom in on a particular region or town
- allow the user to select an individual town to read more detailed information.

Non-functional requirements are other requirements that the user or client would like the solution to have but that do not affect what the solution does. Examples include:

- user friendly
- not display any personal details
- compatible with different web browsers and operating systems
- portable so it can operate on devices of different sizes.

When buying a car the amount of money available for the purchase may restrict the user's requirements. The user may not be able to purchase the car they really want. Instead they may need to re-evaluate their needs.

Solution constraints

Solution constraints are factors that may limit or restrict the solution requirements.

Like requirements, constraints can be broken down into groups: economic, technical, social, legal and useability. Chapter 6, Figure 6.7, page 229, shows an example of each type of constraint.

Economic

Economic constraints include time and cost.

The deadline by which the user or client needs to have the solution operational will define the time available to design and develop the solution. The longer the time available, the more time there is to complete an in-depth analysis, detailed designs and develop advanced features of the solution. The shorter the timeframe, the faster that each stage in the problem-solving methodology needs to be completed.

Meanwhile, the funds (money) available to complete the project may affect the hardware and software (digital systems) available for use, the number and range of staff who are available to work on the solution and even the data used as input, if the data sets required need to be purchased.

Both a lack of time or money may result in a re-evaluation of the user's requirements, or a re-evaluation of how the requirements can be achieved.

Technical

Technical constraints are constraints related to the hardware and software available for the project. Available hardware and software, memory and storage capacity, processing and transmission speeds, and security concerns are all examples of possible technical constraints.

THINK ABOUT COMPUTING 7.8
List three other technical constraints developers of smartphone apps need to consider when developing a product.

For example, developers need to keep in mind that smartphone users may not always have access to a high-speed network connection, so they need to ensure that any animated data visualisation solution does not require a large amount of bandwidth to download and view.

Social, legal and useability

Non-technical constraints relate to areas other than hardware and software. Useability and the user's level of expertise (social) are examples. If a solution is being developed for users with little digital systems expertise, this may restrict some of the requirements that would involve complex manoeuvres to complete. Creating a solution for a child audience may restrict the method used to input data into the solution.

Legal requirements are another type of non-technical constraint. Privacy laws may restrict features linked to displaying personal data in the solution, or to collecting data from the devices of someone using your solution. Copyright laws may restrict features that allow other users to upload content to the solution.

Scope of solution

The scope outlines the boundaries or parameters of the solution so all stakeholders are aware of exactly what the solution will contain. The scope of the solution consists of two elements. Scope is not required for your solution for this Area of Study.

Many house and land package contracts state exactly what is included with the package and what exactly is not included in the package for the price. For example, tiles on the floor of the kitchen are included, but the garden will not be landscaped.

What the solution will do

What the solution will do is a list of all the solution requirements (both functional and non-functional) that will be included in the solution.

What the solution will not do

This is a list of all the solution requirements that will not be included in the solution.

Usually these are solution requirements initially sought by the client, but because of constraints they have been left out of the solution project.

At the start of the project, it helps to outline what will and will not be included in the solution to prevent arguments later in the project between the client and the developer.

THINK ABOUT COMPUTING 7.9
What problems do you think a clear scope of solution can avoid later in the project?

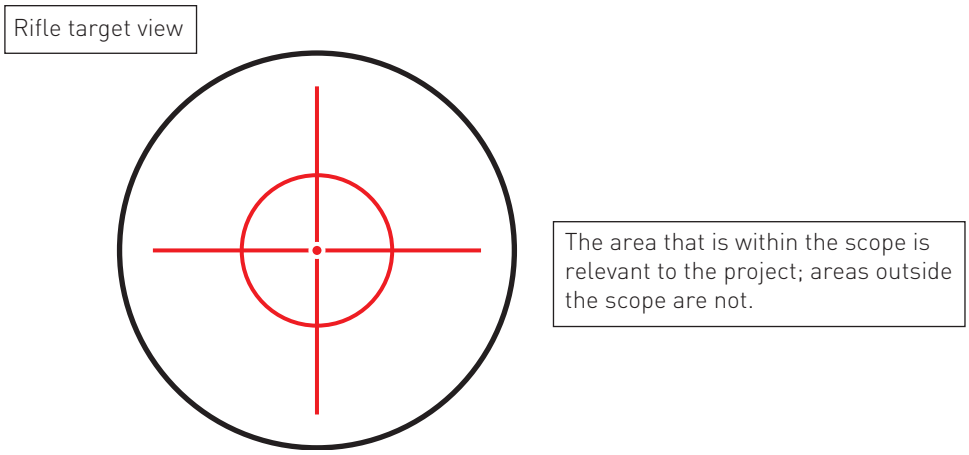


FIGURE 7.17 Scope of solution

An example of a scope of solution would be as follows: The solution will display population data of towns in Victoria in a visual format and graphically represent the distances between those towns. It will be created to be user friendly and for privacy reasons, it will not display any personal details.

The requirement that allows the user to zoom in on a particular region or town will not be included in the project because of economic factors, but may be added at a later stage.

Design tools

The second stage of the problem-solving methodology involves designing the solution. The analysis stage was all about 'what' the solution required. The design stage is all about 'how' the solution will be achieved.

The two activities involved in the design stage are the solution design and evaluation criteria. We will look at evaluation criteria in a later section of the chapter.

Solution design

The solution design involves planning how the solution will appear and function. To assist with planning, a range of design tools can be used to represent the solution's appearance and functionality.

Appearance design tool – layout diagram

A layout diagram can be used to plan how the visualisation will appear. The diagram can be hand drawn or computer generated, but if computer generated, should not use the same software as the solution. The purpose of a layout diagram is to illustrate how the final solution will appear. Once completed, the layout diagram can be inspected to see if the type of visualisation chosen is appropriate for the data set.

A number of alternative layout diagrams can be created to represent the data set in a variety of types of visualisations, to help decide which type to use in the final solution.

The layout diagram allows the developer and the client to give feedback on other areas of the solution, including formats and conventions used, colour schemes and other design principles. It is much easier to make changes to the solution in the design stage than once development begins.

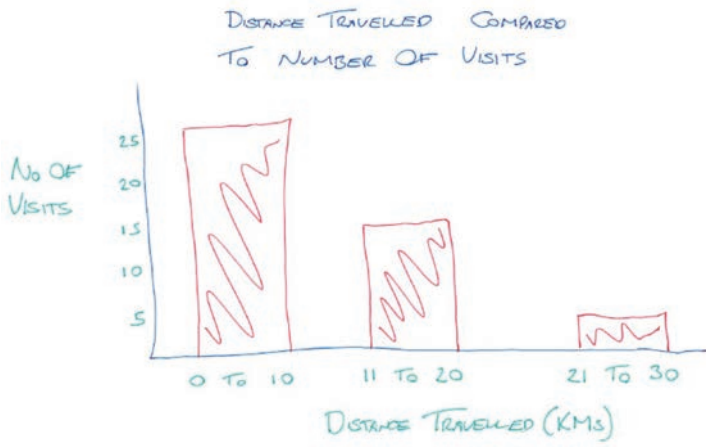


FIGURE 7.18 A hand-drawn layout diagram set out as a column graph

Nisha Samuel Design
Scroll to the section at the bottom of the page and click to view the 'Crowdfunding' storyboard.

Appearance design tool – storyboard

Storyboards can be used to show how the data-visualisation animation might work. For example, if you are using animation in the data visualisation, the storyboard might represent different stages of the solution. View Nisha Samuel's website for an example of a storyboard showing the sequence of scenes in a visualisation used to promote the understanding of crowdfunding communities. The website also shows other useful storyboards.

IPO charts are discussed in Chapters 2 and 3.

Functionality design tool – IPO chart

An input-process-output (IPO) chart is a design tool that helps to plan how a solution will function. The IPO assists in identifying the data required to produce the information (or output) required by the solution. Often by reverse engineering from the output, you can identify what processing steps are needed for the input data.

TABLE 7.4 Input process output design tool

Input	Process	Output
Distance travelled Number of visits	Select data Select insert tab Select column chart Add title and labels	Column chart

Functionality design tool – flowchart

Flowcharts can be used to show the procedure users follow to create a data visualisation.

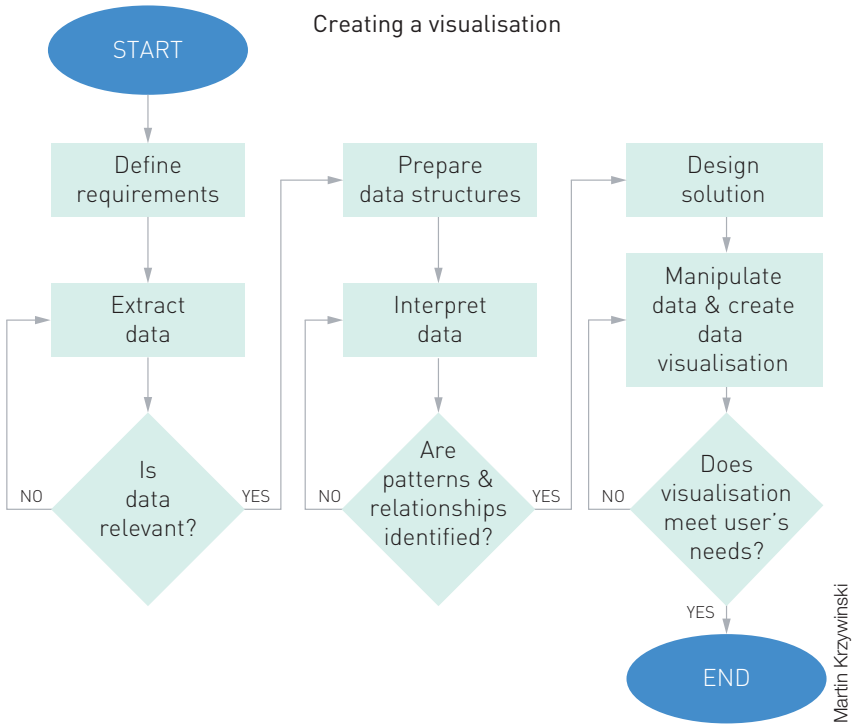


FIGURE 7.19 Flowchart on how to make a visualisation

Formats and conventions

Formats are related to the font type and size selected, background colour used and any other change of appearance. Conventions are general rules that are followed when using a particular format. When we present data in a table, often the column headings are in bold, the text left aligned and the numbers right aligned. When presenting data in a website, we use clearly named hyperlinks to make it easy for the user to find the required information. When addressing a letter, we often follow standard rules (or conventions). For example, the stamp goes in the upper right-hand corner, the return address goes in the upper left-hand corner while the 'to' address is centred on the envelope.

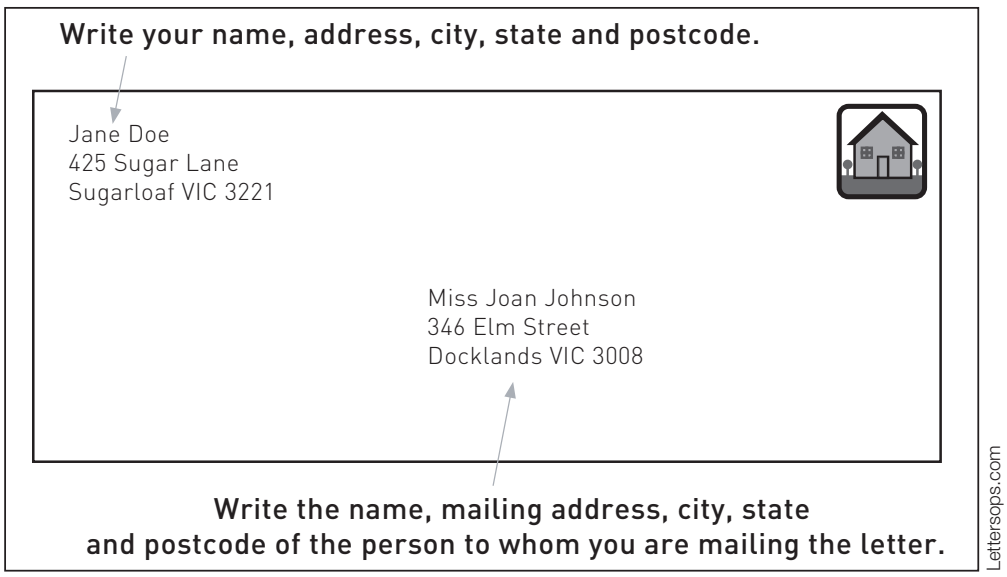


FIGURE 7.20 Conventions followed when addressing an envelope

With data visualisations appearing in so many different formats, there really is a range of conventions that would apply to each format.

Some common conventions that would improve the effectiveness of the visualisation would include:

- clearly title the visualisation and explain its purpose
- label axes if appropriate
- use a key or legend to identify different data items
- include the name of the author and the source of the data
- identify the units of measurement
- choose colours that match the information being discussed.

THINK ABOUT COMPUTING 7.10
Study a number of data visualisations you find effective. Try to identify the formats and conventions that have been used to make the solution stand out.

Software tools and functions

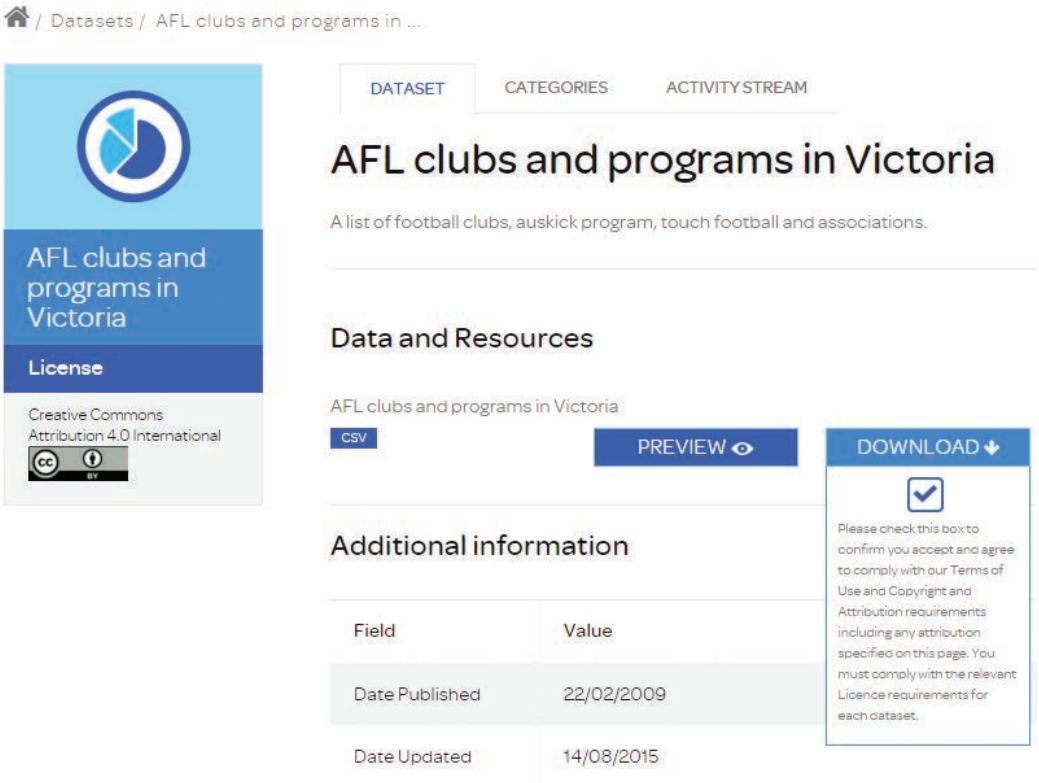
For this Outcome, you should follow two discrete steps to create your data visualisations. The first step involves extracting data from data sources. The second step involves presenting the data using a data visualisation.

Extracting data

Spreadsheet software

Many data sources provide data in a file format that is compatible with existing spreadsheet software applications.

The image below shows a data set available from the Victorian Government website Data.gov.vic.au, containing all of the details and locations of AFL teams and programs in Victoria. The data set is available in CSV file format that is compatible with spreadsheet software (Figure 7.21).



Copyright © The State of Victoria, Department of Environment, Land, Water & Planning 2015, Creative Commons Attribution 3.0 Australia

FIGURE 7.21 Data about AFL clubs and programs in Victoria

	A	B	C	D	E	F	G	H	I
1	Name	Address	Suburb	Postcode	State	Category	LGA	Region	
2	AFL Auskick - Aberfeldie	Aberfeldie Primar	Essendon	3040	VIC	AFL Auskick	Moonee Valley (	North and West Metropolitan Region	
3	AFL Auskick - Airport West	Hansen Reserve	Airport West	3042	VIC	AFL Auskick	Moonee Valley ;	North and West Metropolitan Region	
4	AFL Auskick - Ajax	Princes Park	Caulfield South	3162	VIC	AFL Auskick	Glen Eira	Southern Metro	
5	AFL Auskick - Alexandra	5665 Maroondah	Alexandra	3714	VIC	AFL Auskick	Murrindindi	Hume	
6	AFL Auskick - Alfredton	Alfredton Reserve	Alfredton	3350	VIC	AFL Auskick	Ballarat	Grampians	
7	AFL Auskick - Allansford	Allansford Recrea	Allansford	3277	VIC	AFL Auskick	Program		
8	AFL Auskick - Alphington/Fa	irfield	Northcote	3070	VIC	AFL Auskick	Yarra ; Darebin	North and West Metropolitan Region	
9	AFL Auskick - Altona	Altona Football C	Altona	3018	VIC	AFL Auskick	Hobsons Bay	North and West Metropolitan Region	
10	AFL Auskick - Altona Meac	Comden Reserve	Altona Meadows	3028	VIC	AFL Auskick	Wyndham ; Hob	North and West Metropolitan Region	
11	AFL Auskick - Amateurs	Queens Park Oval	Newtown	3220	VIC	AFL Auskick	Greater Geelong	Barwon S/W	
12	AFL Auskick - Anakie	2155 Ballan Roac	Anakie	3221	VIC	AFL Auskick	Surf Coast , Grea	Grampians, Barwon S/W	
13	AFL Auskick - Anglesea	Ellimatta Reserve	Anglesea	3230	VIC	AFL Auskick	Surf Coast	Barwon S/W	
14	AFL Auskick - Apollo Bay	Apollo Bay Recrea	Apollo Bay	3233	VIC	AFL Auskick	Colac-otway	Barwon S/W	
15	AFL Auskick - Apollo Parkways		Greensborough	3088	VIC	AFL Auskick	Nillumbik ; Bany	North and West Metropolitan Region	
16	AFL Auskick - Ararat	Richardson Oval	Ararat	3377	VIC	AFL Auskick	Ararat Rural	Grampians	
17	AFL Auskick - Ararat North	Blake St	Ararat	3377	VIC	AFL Auskick	Program		

FIGURE 7.22 Data set downloaded into spreadsheet software

Spreadsheet software also contains a feature know as a web query that allows data to be acquired from data sources. One limitation of a web query is that the data needs to be recognised by the spreadsheet software.

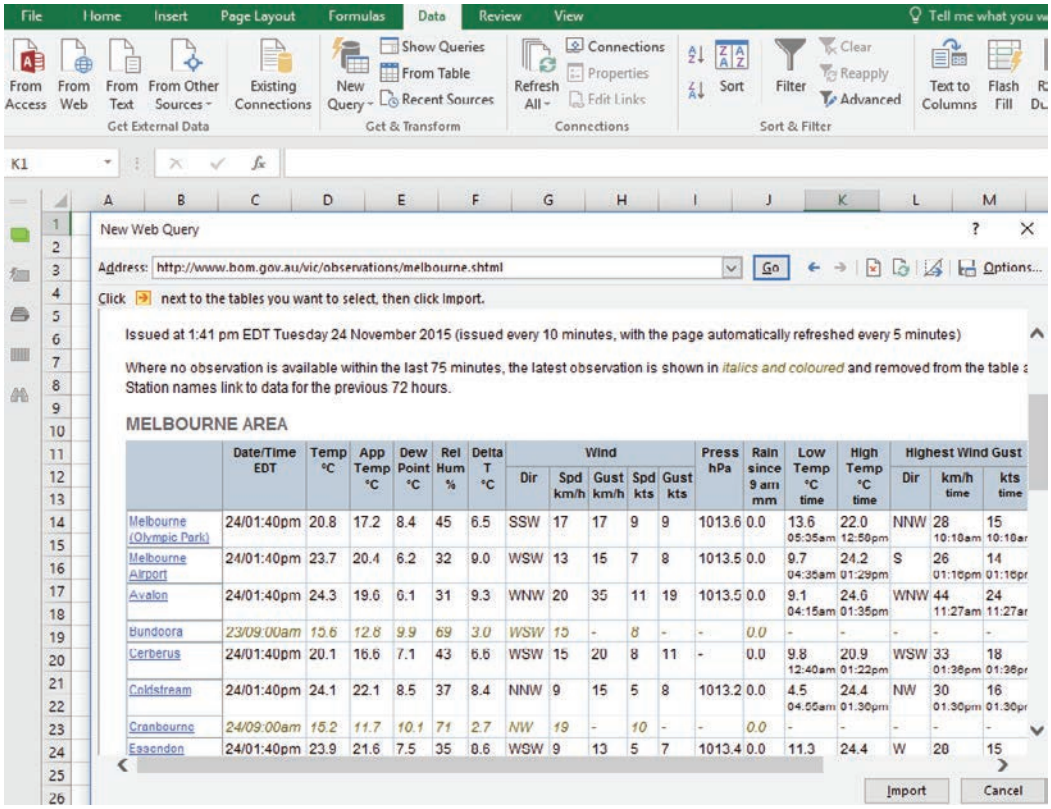


FIGURE 7.23 Example of a web query in spreadsheet software

	A	B	C	D	E	F	G	H	I	J	K	L	M
121		Date/Time	Temp	App	Dew	Rel	Delta	Wind					Press
122		EDT	°C	Temp	Point	Hum	T						hPa
123				°C	°C	%	°C						
124								Dir	Spd	Gust	Spd	Gust	
125									km/h	km/h	kts	kts	
126	Melbourne (Olympic Park)	24/01:50pm	20.6	17.5	8.9	47	6.2	SW	15	20	8	11	1014
127													
128	Melbourne Airport	24/01:50pm	23.6	18.9	5.2	30	9.3	SSW	19	22	10	12	1013
129													
130	Avalon	24/01:50pm	24.1	18.6	5.9	31	9.3	WNW	24	41	13	22	1013
131													
132	Bundoora	23/09:00am	15.6	12.8	9.9	69	3	WSW	15	-	8	-	-
133	Cerberus	24/01:50pm	20.2	15.9	6.8	41	6.8	WSW	19	22	10	12	-
134													
135	Coldstream	24/01:50pm	24.6	21	6.8	32	9.3	N	15	22	8	12	1013
136													
137	Cranbourne	24/09:00am	15.2	11.7	10.1	71	2.7	NW	19	-	10	-	-
138	Essendon Airport	24/01:50pm	23.9	19.6	7.1	34	8.8	SSW	19	20	10	11	1013
139													
140	Fawkner Beacon	24/01:50pm	-	-	-	-	-	SSE	19	20	10	11	-
141													
142	Ferny Creek	24/01:50pm	19.1	17.7	8.7	51	5.6	NW	6	7	3	4	-
143													
144	Frankston	24/01:50pm	17.2	15.9	9.3	59	4.1	WNW	6	6	3	3	-
145													
146	Geelong Racecourse	24/01:50pm	23.1	18.6	6.8	35	8.4	W	20	32	11	17	-
147													
148	Laverton	24/01:50pm	21.5	17.2	9.4	46	6.5	SE	22	28	12	15	1014
149													
150	Moorabbin Airport	24/01:50pm	20.6	16.6	9.2	48	6	SSW	20	28	11	15	1014
151													
152	Phillip Island	24/09:00am	15.7	11.4	10.6	71	2.7	WNW	24	-	13	-	-
153	Point Wilson	24/01:50pm	-	-	-	-	-	WSW	20	26	11	14	-

FIGURE 7.24 Data imported into spreadsheet software using a web query

Database software

Database software can also be used to acquire data from data sources. Using the external data function, data can be imported from a range of sources and file formats including HTML documents, Excel, text and XML files, as well as external databases.

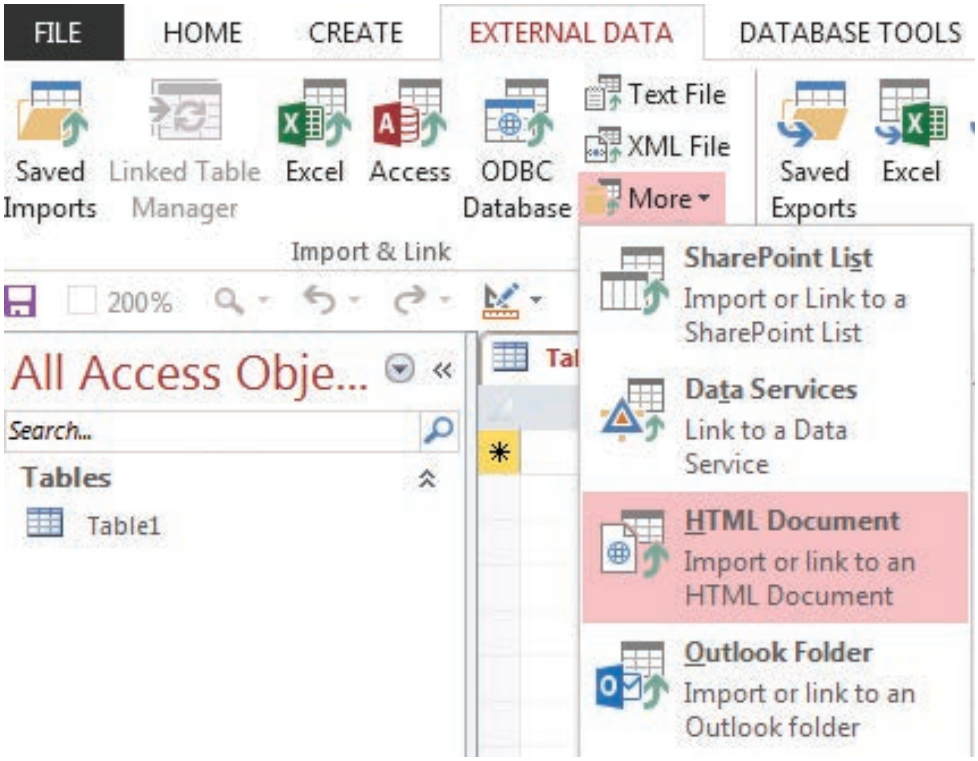


FIGURE 7.25 Options to import external data into database software

Programming languages

Some programming languages have the capability to acquire data from external sources. For example, Python can fetch data from websites. Applications can be created in Python that will fetch data from a website based on the page's HTML code.

Using the 'Inspect Element' command in a browser, you can view the HTML code of a webpage. The area that the data is located in the code can be identified and then imported into another application. Once data has been acquired, it can be imported into other applications, such as database or spreadsheet software.

Figure 7.26 shows a sample of Python code that will read all of the HTML code from Google.

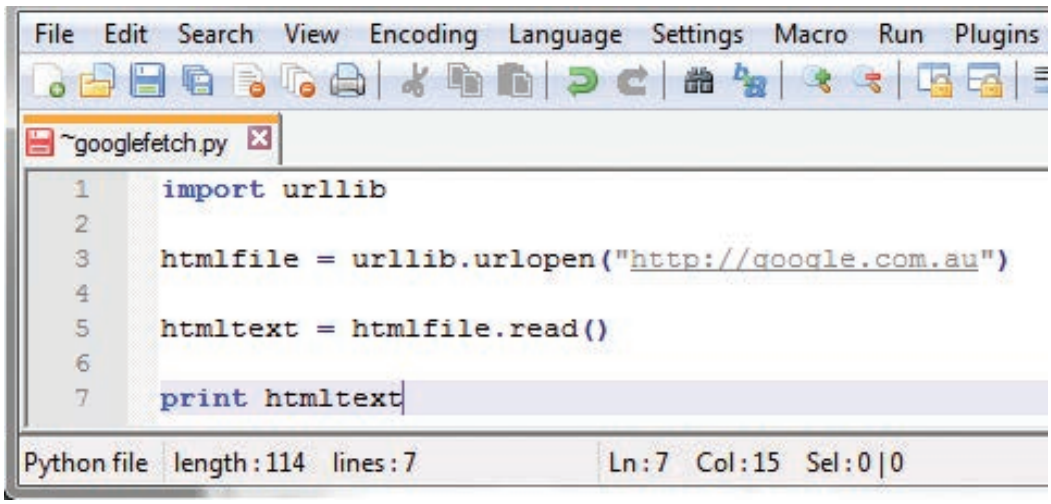


FIGURE 7.26 Sample Python code reading HTML code from Google

Displaying data

The following list shows tools and functions that can be used to display data.

Spreadsheet software

Many spreadsheet applications have a chart or graphing feature that helps manipulate the data into a visual form. Microsoft Excel has a range of different charts available.

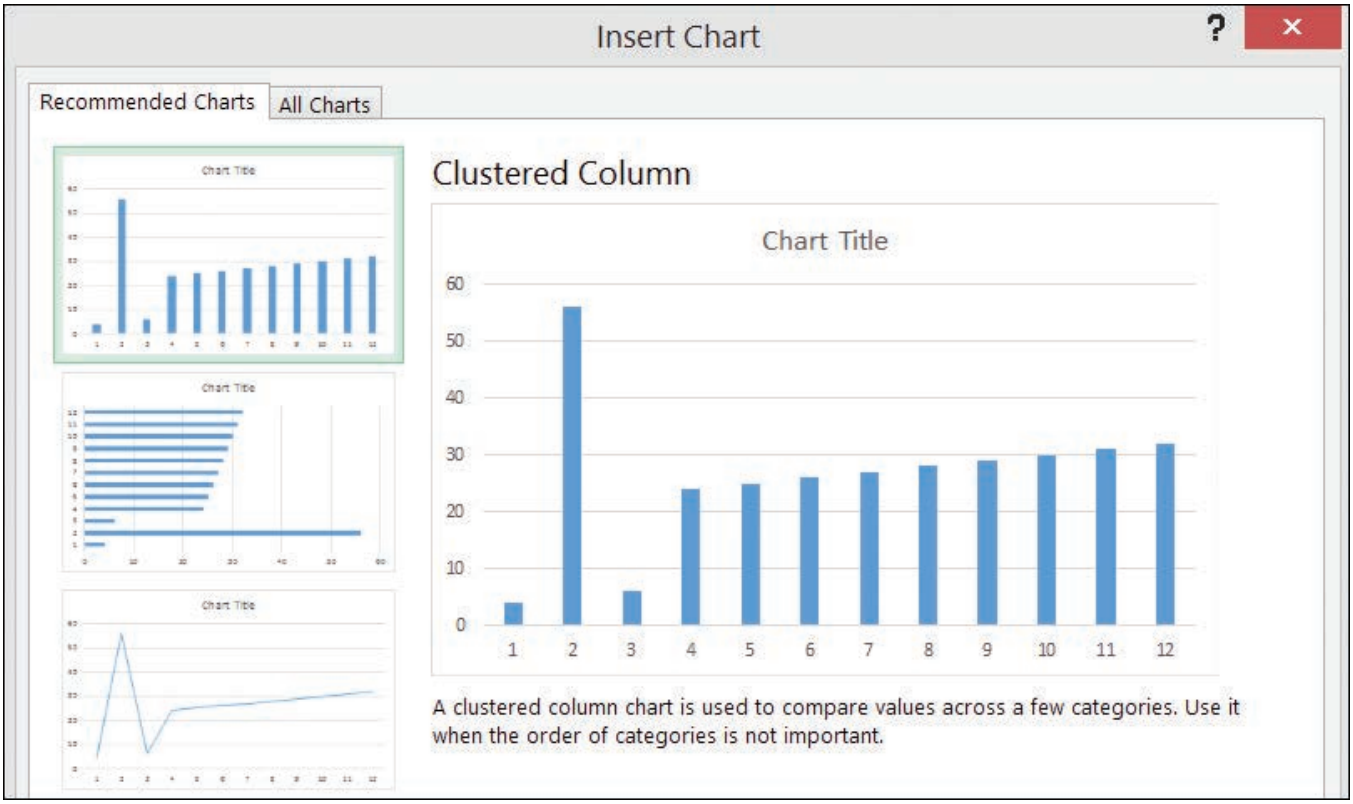


FIGURE 7.27 Spreadsheet chart function

Google

Google has a range of online cloud-based software tools that can manipulate data into a visual format. Google Sheets is spreadsheet software, Google Charts offer a range of live streaming of data from websites, and Google Motion has the functionality to develop motion charts that can be used to create time visualisations.

FIGURE 7.28 Motion chart created with Google Docs

Tableau Public

Tableau Public is free software that can allow anyone to connect to a spreadsheet or file to create interactive data visualisations for the web.

OpenHeatMap

OpenHeatMap allows users to create static or animated heat maps. Data can be saved into files including Google Docs or spreadsheets and then uploaded to the site to create the map. Heat maps represent values in a range of colours that indicate concentration, similar to a weather map.

Tableau Software

FIGURE 7.29 Tableau Public

THINK ABOUT COMPUTING 7.11

The tools mentioned are merely a sample of those available online. Research other online tools and functions available to create data visualisations.

File formats

Earlier in the chapter, files were discussed as a type of data structure. A data structure is a way to organise a collection of data items or a data set. Often the data set required to help create a data visualisation will be stored in a file. But just as there are many types and formats of data visualisations, there are many file types and formats that store the data sets themselves.

Below we will look at some common file formats used to store data used in visualisations.

XLM/XLMS file formats

A file with the **XLM** or **XLSX file** extension is an Excel Microsoft Office Open XML Format Spreadsheet file. Microsoft Office Excel is the primary software program used to open and edit XLM and XLSX files. Alternatively there are a range of free alternative spreadsheet programs that can use these files.

WMS file formats

A **WMS file** is a Web Map Service standard that involves retrieving map images over the internet from a webserver. WMS files are used in conjunction with GIS file formats to create geospatial visualisations.

Copyright © The State of Victoria, Department of Environment, Land, Water & Planning 2015, Creative Commons Attribution 3.0 Australia

FIGURE 7.30 Web Map Service (WMS) using geographic information system (GIS) file data to show industrial areas of urban development. The areas outlined with purple lines are flagged as 'industrial nodes'.

GIS file formats

A **GIS file** format contains geographical data. For example, it could contain the location of rivers and creeks in Victoria, stored as a series of longitude and latitude coordinates. For each river or creek, it could also contain attribute information, including the name and some history information. The GIS file will also contain data on how to display each creek and river in the visualisation.

CSV file format

A **comma-separated value (CSV) file** stores data in tabular format in a plain text file. The data is saved as individual fields using a comma to separate values. There are a number of advantages of saving data in a CSV format. First, a CSV is quite easy to import into (or export from) a range of software applications, including spreadsheet and database applications. Plain text files (such as notepad) also use very little storage space. Unlike some other file formats, they are saved as text, so are readable.

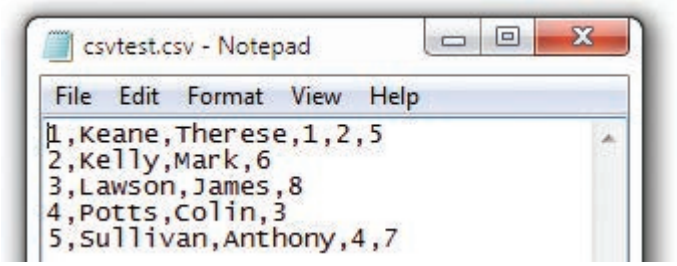


FIGURE 7.31 Comma separated value file format

API tools

An **application-programming interface (API) tool** is more a tool than a file format, but does allow the data required for a visualisation to be collected. An API tool contains the instructions and standards to allow live data to be collected from external websites and then be used in a data visualisation.

THINK ABOUT COMPUTING 7.12
There are many more file formats for storing data sets used in visualisations. Create a list of five other file formats you could use.

Evaluating data visualisations

Evaluation involves measuring how well the solution meets the information need and the needs of the client. Evaluation is used in two stages of the problem-solving methodology. First evaluation criteria are created in the design stage. Then, in the evaluation stage, two activities occur: determining an evaluation strategy and then completing an evaluation report.

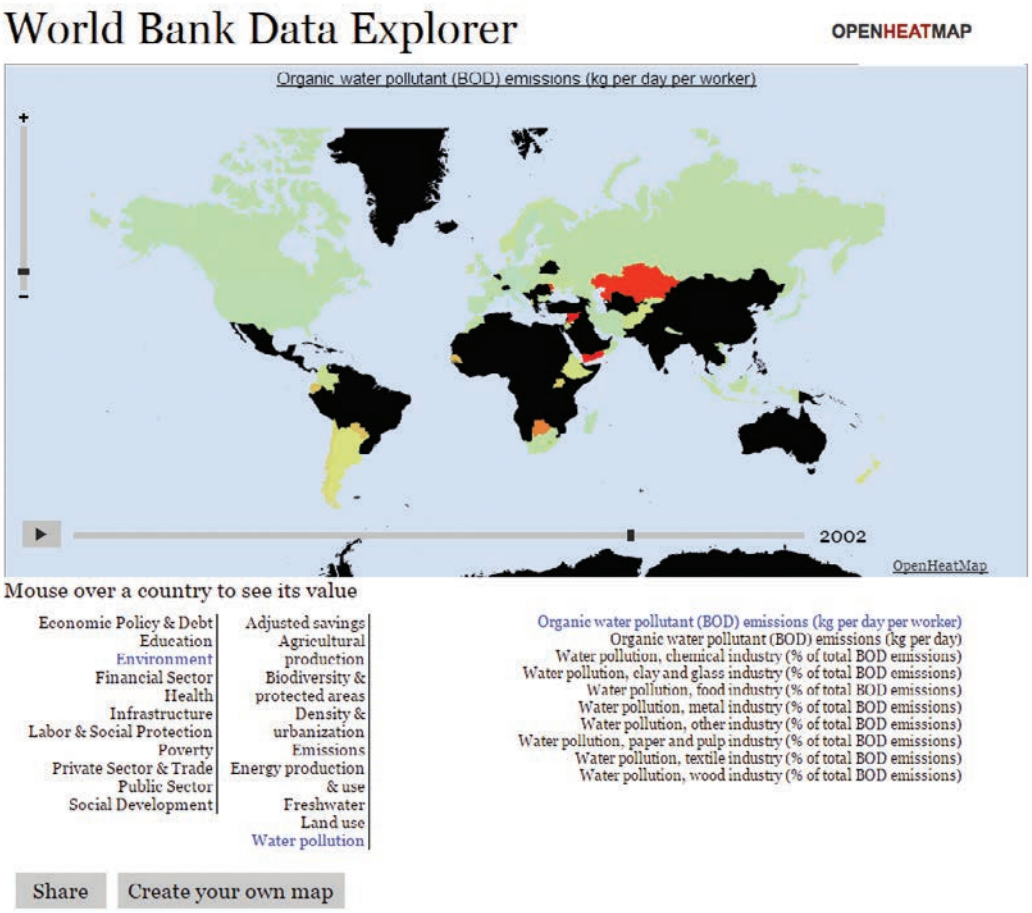
The evaluation stage is not completed until after the solution is developed and been used by actual users for a period of time, often three to six months after development. This time frame allows the users to learn and become familiar with the solution, so they can make clear judgements on whether the solution is meeting their information needs and any other requirements.

Evaluation criteria

Evaluation criteria are measures that will be used to judge whether the solution meets the information needs of the client or user.

Criteria should relate to the initial solution requirements because if all the requirements are contained within the solution, then the information need should be met. As the criteria are going to be used to judge the solution, they should be framed or written so that they are quantifiable or measurable.

OpenHeatMap: World Bank Data Explorer



© OpenStreetMap contributors

FIGURE 7.32 Dynamic heat map created with OpenHeatMap

Quantifiable or measurable criteria are measures whose satisfaction is easily determined. An example of a measurable criterion is 'the solution displays the visualisation within five seconds of loading the website'. It is relatively easy to then judge whether the solution can do this, or not.

Evaluation criteria should also cover both **effectiveness** and **efficiency**. Effectiveness criteria should relate to how well the solution works and if it provides the information needed. There are a number of characteristics related to effectiveness including accuracy, accessibility, attractiveness, communication of message, completeness, timeliness and useability. For a solution to be considered effective, it needs to have these characteristics. Based on the sample solution requirements, listed on pages 291–2, examples of evaluation criteria for effectiveness include:

- population sizes of towns are represented accurately
- visualisation contains appropriate contrast, space and balance
- data used is timely
- data for all towns in Victoria is included
- solution is easy to use.

Efficiency criteria relate to saving time, cost or effort when retrieving the information from the solution. Examples of evaluation criteria for efficiency include:

- town information can be accessed within three seconds
- zooming in and out can be completed without using the keyboard
- solution will reduce the amount of money spent on printing the data.

The evaluation criteria that have been created during the design stage are then available for the developers to refer to during the development stage to provide guidance for the project.

Evaluation strategy

The evaluation stage of the problem-solving methodology is not completed until after the solution is created and is being used by real users. As discussed, often a timeframe of three to six months after implementation is considered a reasonable time for the evaluation stage to begin. The first activity in the evaluation stage involves creating an evaluation strategy. An evaluation strategy involves deciding how each evaluation criteria can be measured. It includes creating a timeline for evaluation to take place, deciding on the data required to help judge each criterion, and looking at the way the data required will be collected and how the data can be used to evaluate each measure.

One method to display an evaluation strategy, using one criterion for effectiveness and one for efficiency from the example above, is to use a table format, as shown in Table 7.5.

TABLE 7.5 The evaluation strategy

Evaluation criterion	Data used must be less than 6 months old	Town information can be accessed within 3 seconds
Timeline	3 months after solution implementation	3 months after solution implementation
Data required	Date that the data used to create the visualisation	Time taken for a user to zoom in to a particular town and access the town information
Data collection method	Interview the staff involved in collecting the primary data about population to establish the date it was collected.	Observe a user use the solution and, using a stopwatch, time how long it takes to access the information of a particular town.
How the data can be used	Compare the date that the data was collected to the current date to determine how old the data is.	Compare the average time taken, over 10 attempts, to access the information to the benchmark (3 seconds).

A range of methods can be used to collect the data required for evaluation. We have previously discussed using interviews, survey and observations to collect data. In addition, checking download speeds, counting website hits, inspecting the solution output, reviewing error logs or timing how long it takes users to complete tasks are just some additional methods that enable evaluation criteria to be completed.

THINK ABOUT COMPUTING 7.13

Create a strategy table by selecting two other of the sample criteria listed on page 290. Complete the evaluation strategy for each.

Evaluation report

Finally, after the evaluation strategy has been completed, the evaluation report can be written. The evaluation report involves stating if the solution is meeting the information need and the needs of the user. To provide evidence for the final conclusion, each evaluation criterion must be assessed to identify if it has been achieved. If each of the evaluation criteria has been met, then the solution can be considered a success.

TABLE 7.6 Assessing each criterion to assist in preparing an evaluation report

Evaluation criterion	Data used must be less than 6 months old	Town information can be accessed within 3 seconds
Timeline	3 months after solution implementation	3 months after solution implementation
Data required	Date that the data used to create the visualisation	Time taken for a user to zoom in to a particular town and access the town information
Data collection method	Interview the staff involved in collecting the primary data about population to establish the date it was collected.	Observe a user use the solution and, using a stopwatch, time how long it takes to access the information of a particular town.
How the data can be used	Compare the date that the data was collected to the current date to determine how old the data is.	Compare the average time taken, over 10 attempts, to access the information to the benchmark (3 seconds).
Data	4 months	3.6 seconds
Achieved	Yes	No

CHAPTER SUMMARY

ESSENTIAL TERMS

accuracy correctness, without mistakes

application-programming interface (API) a tool that contains the instructions and standards to allow live data to be collected from external websites and then be used in a data visualisation

array a collection of data items generally of the same data type

authenticity in relation to data, data that is genuine, original and considered trustworthy

Boolean data type a logical data type; Boolean data can hold only two possible values, usually true or false

character a single letter, number or symbol that is usually stored in 2 bytes of memory; a data type

chart a method of displaying data as symbols

comma-separated value (CSV) file a file type that stores data in tabular format in a plain text file; the data is saved as individual fields using a comma to separate values

data types particular forms that an item of data can take, including numeric, character and Boolean

data visualisation the presentation of data in a pictorial or graphical format

floating point number a number with a fractional or decimal component

flow visualisations representing data that illustrates the flow pattern of an object or item

formats ways in which data and information can be presented

functional requirements those directly related to what the solution will do

geospatial visualisation overlay data related to a geographical area over mapping technology; Google Maps is an example of a geospatial visualisation

GIS file a file format that contains geographical data

hierarchy visualisations show the relationships and structure between data items

information need when particular information is required, yet is not being currently supplied; causes of an information need may include a current problem, an identified need or an opportunity

integer a number without a fractional or decimal component

map-based visualisations geographical data displayed in a visual format

matrix visualisations visualisations used to show the compensation of individual items in the sample size

network visualisations visualisations that show the relationship between different data items and data sets

non-functional requirements other requirements that the user or client would like the solution to have but they do not affect what the solution does

non-technical constraints limitations relating to areas other than hardware and software: social, legal and useability

record a collection of data items of different data types

relevance data that produces useful information required by the information need; if the data is not useful, it is not relevant

sample size the number of completed responses that you collect from interviews or surveys; the more responses completed the more likely the data being collected is accurate

solution constraints factors that may affect or restrict the requirements included in the solution

solution requirements what the client wants from the solution; can be broken down into functional and non-functional requirements

string data that consists of a string of characters

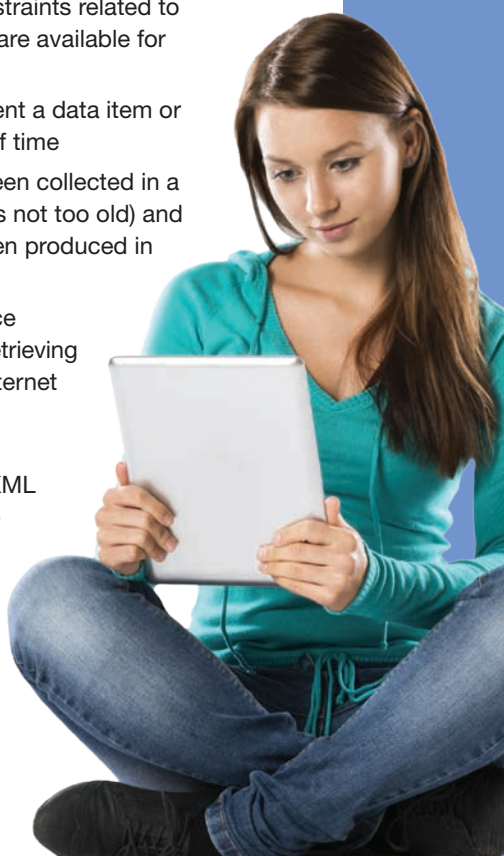
technical constraints constraints related to the hardware and software available for the project

time visualisations represent a data item or data set over a period of time

timeliness data that has been collected in a reasonable timeframe (is not too old) and information that has been produced in time to be useful

WMS file a web map service standard that involves retrieving map images over the internet from a webserver

XLM or XLSX file an Excel Microsoft Office Open XML Format Spreadsheet file



CHAPTER SUMMARY

IMPORTANT FACTS

1 Government departments collect and store vast amounts of data sets.
2 The Bureau of Meteorology is Australia's national weather, climate and water agency.
3 The Victorian government data directory contains a range of data sets from Victorian government departments.
4 The Australian Bureau of Statistics (ABS) is the statistical agency of the federal government.
5 Data.gov.au is similar to the Victorian government data directory in that it provides publically available data sets that have been collected by federal government departments.
6 A file data structure can hold numerous data items, arrays or records. A file is saved separately from the software program that utilises the file.
7 Infographics generally are used to communicate a particular message.
8 The first stage of the problem-solving methodology is the analysis stage. The analysis stage is considered with 'what' questions. The analysis stage is made up of three activities: determining solution requirements, solution constraints and the scope of the solution.
9 Economic factors include time and money.
10 Scope of solution helps to identify the boundaries or parameters of the solution and outlines what requirements will or will not be included.
11 The design stage is all about how the solution will be achieved.
12 The solution design involves planning how the solution will appear and function.
13 A layout diagram can be used to plan how the visualisation will appear.
14 Storyboards can be used to show how the data-visualisation animation might work.
15 An IPO chart is a design tool that helps to plan how a solution will function.
16 Flowcharts can be used to show the procedure that users need to complete to create a data visualisation.
17 Conventions are general rules that are followed when using a particular format.
18 Common conventions include clear title, axis labelled, key or legend used, name of the author and the source of the data, unit of measurements shown and matching colours.
19 There are a range of software tools and functions that can be used to create data visualisations.
20 Many spreadsheet applications have a chart or graphing feature that helps manipulate the data into a visual form.
21 Google has a range of online cloud-based software tools that can manipulate data into a visual format.
22 Tableau Public is free software that can allow anyone to connect to a spreadsheet or file to create interactive data visualisations for the web.
23 OpenHeatMap allows users to create static or animated heat maps.
24 A file is a type of data structure that can store and organise a data set. Files operate independently of the software application that uses the file, so after the application is closed, the file will still exist. There are many file types and formats.
25 A file with the XLM or XLSX file extension is an Excel Microsoft Office Open XML Format Spreadsheet file.
26 An application-programming interface (API) tool is more a tool than a file format but does allow the data required for a visualisation to be collected.
27 Evaluation involves measuring how well the solution meets the information need and the needs of the client.
28 In the design stage of the problem-solving methodology, evaluation criteria are created.
29 Evaluation criteria should also cover both effectiveness and efficiencies. Effectiveness criteria should relate to how well the solution works and if it provides the information needed.
30 Efficiency criteria relates to saving time, cost or effort in retrieving the information from the solution.
31 The evaluation stage of the problem-solving methodology is not completed until after the solution is created and is being used by real users.
32 An evaluation strategy is about deciding how each evaluation criteria can be measured.
33 Interviews, surveys, observations, checking download speeds, counting the number of website hits, inspecting the solution output, reviewing error logs and timing users are all methods used to collect evaluation data.
34 The evaluation report involves stating if the solution is meeting the information need and the needs of the user. To provide evidence for the final conclusion, each evaluation criterion must be assessed to identify if it has been achieved.

TEST YOUR KNOWLEDGE

INFORMATION NEEDS AND DATA VISUALISATIONS

Review quiz

1 Describe an information need.
2 List three reasons why an information need may occur.
3 Explain when an information need may result in a data visualisation being developed.

SOURCES OF AUTHENTIC DATA

4 Explain the difference between primary and secondary sources of data.
5 List three sources of primary data.
6 Explain when a survey would be better than an interview to collect data.
7 List three sources of secondary data.
8 Identify an issue related to using data from secondary sources.
9 Explain why government departments tend to be the source of numerous data sets.
10 Explain the role of the Bureau of Meteorology.
11 Identify how the Victorian government data directory differs from Data.gov.au.
12 Explain how the role of the Australian Bureau of Statistics differs from that of Data.gov.au.

INTEGRITY OF DATA

13 Define 'integrity of data'.
14 List four measures used to judge the integrity of data.
15 Outline a measure that can be used to improve the chances of data being accurate when collected.
16 Timeliness does not mean as quickly as possible. Describe the concept of timeliness.
17 List three attributes of authentic data.
18 For creating a visualisation, what would be considered relevant data?

DATA TYPES AND DATA STRUCTURES

19 Define 'data type'.
20 List four data types.
21 Explain the relationship between a character data type and a text (string) data type.
22 Explain how an integer differs from a floating point.
23 If collecting data for the question 'Are you 18 or over?', which data type would be the most efficient in storing the data? Give reasons for your answer.
24 Explain how an array is different from a record.
25 Give two reasons why files are an effective method of storing data sets.

TYPES AND PURPOSES OF DATA VISUALISATION

26 Define 'data visualisation'.
27 List three types of visualisations that can fit into the chart category.
28 Explain the concept of geospatial visualisations.
29 Discuss one advantage of geospatial visualisations.
30 Describe a situation when you would use a network visualisation.
31 Explain how timeline data differs from time series data.
32 List three examples of hierarchy visualisations.
33 Explain the purpose of a flow visualisation. Give an example to support your answer.
34 Discuss how a matrix visualisation is often laid out.

ANALYSIS STAGE

35 List the three activities involved in the analysis stage of the problem-solving methodology.
36 Define 'solution requirements'.
37 Explain the difference between functional requirements and non-functional requirements.
38 Define 'functional constraint' and 'non-functional constraint'.
39 Explain how constraints can affect the solution.
40 State the purpose of defining the scope of the solution.

DESIGN TOOLS

41 Explain the purpose of the design stage.
42 List the two activities involved in the design stage.
43 Outline the purpose of a design tool.
44 Explain the difference between appearance design tools and functionality design tools.
45 List two design tools for appearance.
46 Discuss the advantage of creating a number of layout diagrams representing the data in a different format.
47 Explain the purpose of an IPO chart.
48 Outline a situation when using a flowchart would be appropriate.

FORMATS AND CONVENTIONS

49 Define 'format'.
50 Define 'convention'.
51 List four conventions usually followed in data visualisations.

SOFTWARE TOOLS AND FUNCTIONS

52 List three software tools used to create data visualisations.
53 Explain the purpose of a motion chart.
54 Explain one similarity between Tableau Public and OpenHeatMap.

FILE FORMATS

55 Explain the relationship between files and software applications.
56 List three types of files commonly used to store data sets.
57 Explain the relationship between WMP files and GIS files.
58 Describe how a CSV file is structured.
59 Outline the purpose of an API tool.

EVALUATING VISUALISATIONS

60 Define 'evaluation'.
61 Identify when an evaluation takes place. Explain why an evaluation takes place at this point.
62 List the two activities involved in the evaluation stage.
63 Identify an activity completed in another stage of the problem-solving methodology that provides input into the evaluation stage.
64 Explain the purpose of the evaluation strategy.
65 List three methods that can be used to collect evaluation data.
66 Discuss the purpose of an evaluation report.
67 Explain the link between the evaluation criteria and whether a solution can be deemed a success.

APPLY YOUR KNOWLEDGE

Using the following data sets, create visualisations that will provide useful information.

1 Analyse the data to determine, in any three-month period (for example, Feb to Apr) which city would be the best to visit for warmer weather.

TABLE 7.7 Sydney versus Melbourne weather

City	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Melbourne	25.9	25.8	23.9	20.3	16.7	14.1	13.5	15.0	17.2	19.7	22.0	24.2
Sydney	25.9	25.8	24.7	22.4	19.4	16.9	16.3	17.8	20.0	22.1	23.6	25.2

2 **a** Calculate each state's proportion of Australia's population and display as a data visualisation. All states have twelve senators in the Australian Senate, while the Australian Capital Territory and the Northern Territory have two.

b Calculate the number of people in each jurisdiction for each senator. Create a data visualisation of the results.

TABLE 7.8 Australian population by state and territory

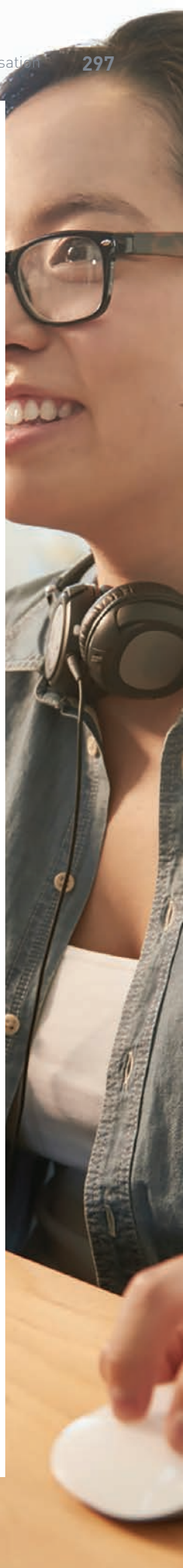
State	Population
New South Wales	7 272 800
Victoria	5 603 100
Queensland	4 516 361
Western Australia	2 296 411
South Australia	1 644 642
Tasmania	507 626
Australian Capital Territory	358 894
Northern Territory	229 675

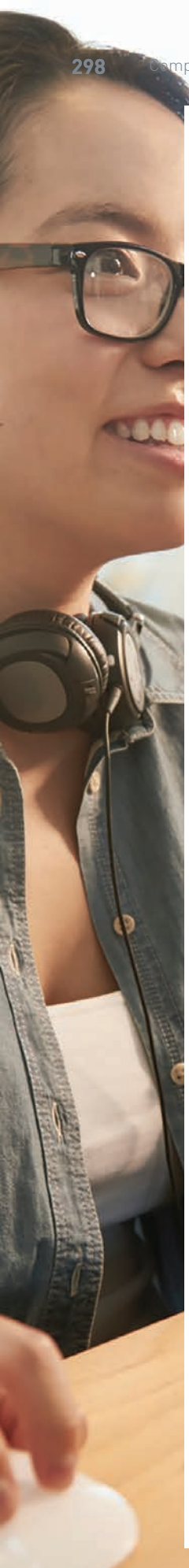
3 Create a visualisation showing the total number of students at the school and the proportion of boys to girls over time.

TABLE 7.9 Student enrolments at the Melbourne School

Year	1961	1971	1981	1991	2001	2011
Boys	81	450	600	750	750	900
Girls	0	0	0	0	150	400
Total	81	450	600	750	900	1300

These figures are estimates.





4 Create a visualisation to display the number of members of each AFL club from each state, assuming that all members live in the state where their team resides.

Collingwood Magpies	77 405	Port Adelaide Power	40 510
Hawthorn Hawks	62 393	Sydney Swans	35 612
Richmond Tigers	60 053	North Melbourne Kangaroos	34 511
West Coast Eagles	57 830	St Kilda Saints	32 562
Essendon Bombers	54 903	Melbourne Demons	32 847
Carlton Blues	49 224	Western Bulldogs	29 641
Adelaide Crows	45 000	Brisbane Lions	23 760
Fremantle Dockers	43 638	Greater Western Sydney Giants	12 631
Geelong Cats	41 935	Gold Coast Suns	12 350

PREPARING FOR UNIT 2 OUTCOME 2

Produce a data visualisation that meets the need of a user. This will involve using a complex data set and accessing software or online tools that will enable you to convert this data into a more meaningful format

Apply the problem-solving methodology and use appropriate software tools to create a data visualisation to meet user needs.

OUTCOME MILESTONES

1. Familiarise yourself with the design brief and analyse the information problem that exists.
2. Determine the type of data visualisation needed to solve the information problem.
3. Select appropriate sources of data and identify relevant data.
4. Determine the suitability of different data types and structures for creating visualisations.
5. Select types of visualisations that are appropriate to the data.
6. Select and apply appropriate tools to plan the design of the visualisations.
7. Apply software functions to locate and acquire data to input and manipulate.
8. Use appropriate software tools, and select and apply a range of suitable functions.

STEPS TO FOLLOW

Create the solution using the four stages of the PSM: Analysis, design, development and evaluation.

Analysis stage

1. Determine the solution requirements.
2. Determine the solution constraints.

Design stage

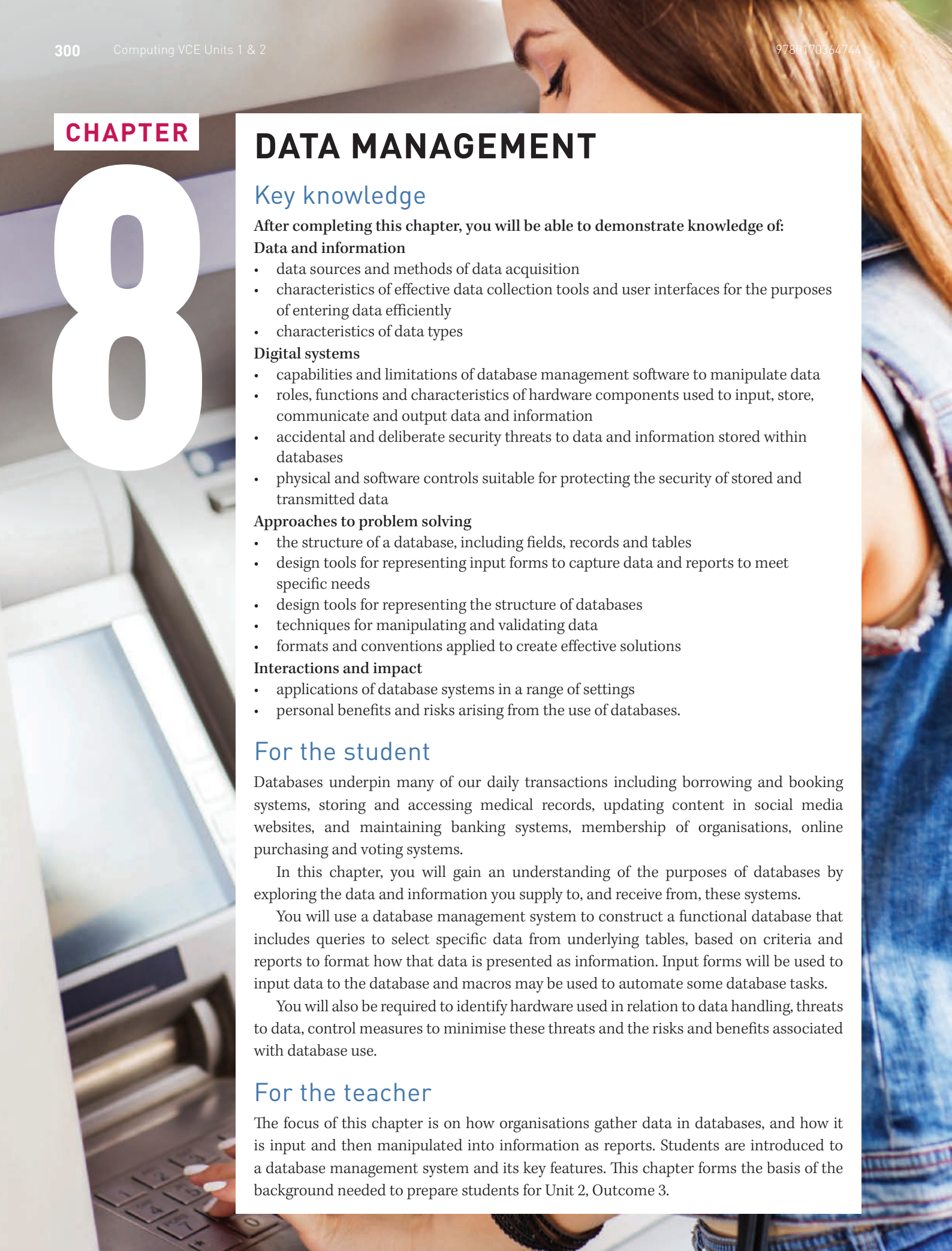
1. Plan the solution: Use design tools for appearance.
2. Plan the solution: Use design tools for functionality.
3. Establish evaluation criteria.
4. Extract relevant data from data sources.
5. Manipulate data for use.

Development stage

Create a data visualisation using appropriate software types and functions.

Evaluation stage

1. Create criteria to evaluate if the data visualisation meets users' needs.
2. Select appropriate techniques to determine if the data visualisation meets users' needs.



CHAPTER 8

DATA MANAGEMENT

Key knowledge

After completing this chapter, you will be able to demonstrate knowledge of:

Data and information

- data sources and methods of data acquisition
- characteristics of effective data collection tools and user interfaces for the purposes of entering data efficiently
- characteristics of data types

Digital systems

- capabilities and limitations of database management software to manipulate data
- roles, functions and characteristics of hardware components used to input, store, communicate and output data and information
- accidental and deliberate security threats to data and information stored within databases
- physical and software controls suitable for protecting the security of stored and transmitted data

Approaches to problem solving

- the structure of a database, including fields, records and tables
- design tools for representing input forms to capture data and reports to meet specific needs
- design tools for representing the structure of databases
- techniques for manipulating and validating data
- formats and conventions applied to create effective solutions

Interactions and impact

- applications of database systems in a range of settings
- personal benefits and risks arising from the use of databases.

For the student

Databases underpin many of our daily transactions including borrowing and booking systems, storing and accessing medical records, updating content in social media websites, and maintaining banking systems, membership of organisations, online purchasing and voting systems.

In this chapter, you will gain an understanding of the purposes of databases by exploring the data and information you supply to, and receive from, these systems.

You will use a database management system to construct a functional database that includes queries to select specific data from underlying tables, based on criteria and reports to format how that data is presented as information. Input forms will be used to input data to the database and macros may be used to automate some database tasks.

You will also be required to identify hardware used in relation to data handling, threats to data, control measures to minimise these threats and the risks and benefits associated with database use.

For the teacher

The focus of this chapter is on how organisations gather data in databases, and how it is input and then manipulated into information as reports. Students are introduced to a database management system and its key features. This chapter forms the basis of the background needed to prepare students for Unit 2, Outcome 3.

Applications of database systems

In the last few years, an unprecedented quantity of data has been generated as people flock to a range of online or networked systems. Apart from business uses, social media websites such as Facebook, Twitter and Tumblr, and even Google searches, accumulate vast amounts of data. When you borrow a book or DVD from a library, or book concert tickets online, data is generated and stored. When you visit your local doctor, your health records are updated so that your complete medical history can be viewed by other doctors in the future.

This data can be used for a variety of useful purposes but in order for that to happen, it needs to be stored somewhere. A **database management system (DBMS)** is software designed to store data and allow users to search and select from it in order to extract the information they need in order to make decisions.

You will find DBMSs used in a variety of contexts. Businesses use databases to record transactions such as bill payments or shopping purchases. Banks are particularly heavy users of databases because of the volume of financial transactions they must be able to process and store every second. Bank customers now also perform much of their own banking online, which is recorded in their bank's DBMS. Libraries catalogue books and store details of borrowers in their databases. Scientists and museums use DBMSs to catalogue their findings and specimens. When you join a club or other group, you will be asked to supply a number of personal details. These details will be collected and stored in a database. Many people purchase goods online through websites such as Amazon and eBay. To register for these types of transactions, you must be able to supply at least your email address, delivery address and payment details. In some cases, you may even be involved in online voting, which also requires a database to store and tally the votes for each option and report on the most favoured popular choices. Rating sellers on eBay, videos on YouTube or 'apps' for mobile devices are common types of voting systems you may encounter (Figure 8.1). Google and Wikipedia use powerful DBMSs to provide a list of relevant websites or topics when you type in a search term, and every item in your Facebook profile is also stored in a massive database. A list of names and phone numbers of friends to invite to your party would also be considered a database.

Well-known non-digital databases include Ancient Egyptian papyrus scrolls with lists of quantities of crops harvested for the Pharaoh and the Domesday book of William the Conqueror, which listed the property and value of every landholder in England, and was used to calculate taxes.

Well-known DBMSs include Microsoft Access, FileMaker, MySQL, PostgreSQL, Microsoft SQL Server, Oracle, SAP and IBM DB2.

THINK ABOUT COMPUTING 8.1

Think of some other examples of databases you may regularly encounter.

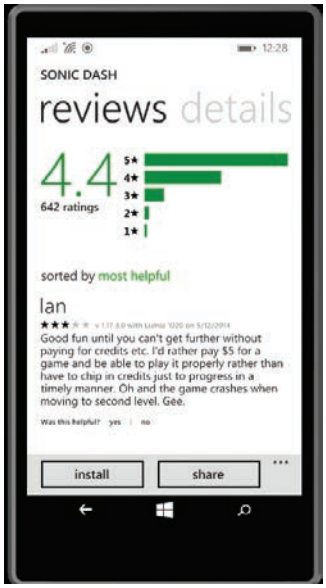


FIGURE 8.1 A rating from online voting for a popular game in the Windows Store

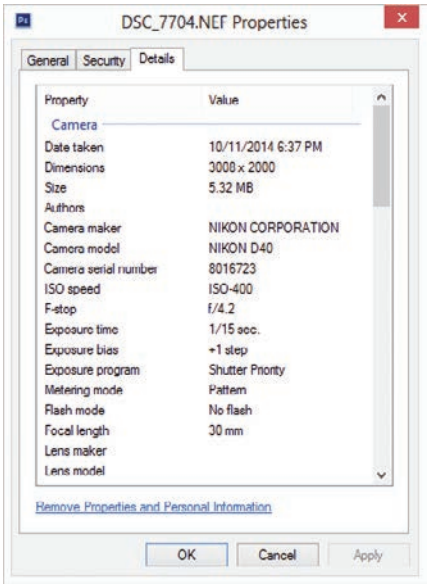


FIGURE 8.2 An example of metadata saved in a photo taken on a digital camera. Many images also save GPS coordinates to allow for geotagging.

THINK ABOUT COMPUTING 8.2
List the data that a bank might need in order to process a supermarket EFTPOS transaction.

In addition, all digital files now contain **metadata**, the primary purpose of which is to allow databases to locate files that match certain search criteria more easily. Figure 8.2 on the previous page provides an example of metadata for a photograph. DBMSs also contain a specific type of metadata that record the details of structural elements in a database file. These details are often referred to as a **data dictionary** and may include details such as the size and data type of each field and characteristics of tables and other objects. Essentially, it is data about data!

Database management systems (DBMSs)

For data handling, no other software type can compare with a database package. While each package may differ in terms of its look and features, several key capabilities are common to all DBMSs.

Capabilities and features

- **Data redundancy** is reduced by centralising different data sources into one database file. Data consistency is a flow-on effect of reducing redundancy.
- Sharing of data allows multiple users to access the same data simultaneously. For example, an accounts department might want to send an invoice to a customer, while the marketing department wants to send an invitation to a special product launch. Both departments will send their emails addressed the same way and including the same personal details because they are both taking the data from the one company database.
- Data integrity can be maintained through the use of validation (see below).
- Data security measures mean only authorised users can access certain data. Data is often **read only** to most users in order to restrict the number of people who can edit it.
- Multiple data elements can be updated in different tables when a transaction takes place.
- A common **structured query language (SQL)** is used in all packages to select data.
- Input forms can be created to make data entry easier for users.
- Reports can be created to display information extracted from database files. They can be saved and modified for later reuse. Reports can include graphic representations such as charts.
- Concurrency of data is maintained when two users try to edit the same record simultaneously.
- Data is independent of the database application, so properties of fields can be modified without altering the application or the database file itself.

Limitations

DBMSs also have a number of limitations that need to be considered if you are planning to implement a database.

- Equipment (software and hardware) can be costly, especially if regular upgrades are needed.
- Conversion from a computer file used for data storage, such as a spreadsheet, to a DBMS can be costly and time consuming. Fields and tables must be carefully constructed to ensure that no data is lost.
- Staff need training on how to use the new system. This can be costly.
- Specialised database administrators need to be employed.
- If a database is corrupted, (for example, as a result of a technical failure), it has a huge impact across the entire organisation.

Benefits and risks

A DBMS can be a very useful tool for a variety of reasons. By storing all of your relevant data in one place, the time and effort spent searching for information is greatly reduced, especially when compared with a manual search. This is especially useful for medical histories and banking details. Being able to find information more quickly could also save you money, especially in the form of wages for data-entry personnel.

When you need to ask a particular question of your data, such as, 'How many widget sales have been recorded in Dandenong during the previous three months?' or 'What is the distribution of Pobblebonk Frogs across Victoria?' you can answer the question in a timely manner with both relevance (only the information you require) and completeness (all the information you require). Using validation and calculation formulas can guarantee the accuracy of data and information, respectively.

An easy-to-use interface and selection tools can make your data more accessible and usable. Selected data can be formatted into well-presented and attractive reports that make the information clearer and more readable to a wider audience on a variety of different platforms. These all contribute to delivering the message that the information needs to convey to the end user.

Benefits of a database solution are classified in terms of efficiency (time, cost, effort) and effectiveness (completeness, readability, attractiveness, clarity, accuracy, accessibility, timeliness, communication of message, relevance and useability).

There are also several risks that need to be considered before and while using a DBMS.

As per the **garbage-in equals garbage-out (GIGO)** principle, the input must be validated to avoid inaccurate data being entered. You may also have difficulties with version control of data.

Privacy must be protected. How will you prevent unauthorised users accessing confidential data? Security must also be protected. How will you prevent hackers and other unauthorised users from stealing or damaging data?

Database systems can be complex, difficult and time-consuming to design. Training is required for all programmers and users, or data can be damaged. In addition to the costs of training and running the system, the costs of purchasing suitable hardware and software can be high.

Database structure

A DBMS is composed of several key parts. All the objects defined in the following sections can be contained within a single database **file**.

Fields, records and tables

This information is perfect as a sample entry for a database:

> Mrs Maria Johnson is 26 years old. She lives at 23 Example Street, Fairfield, 3078. Her mobile phone number is 04 123 456.

Please note that some personal information is redacted for privacy reasons.

The individual data elements highlighted below could be described as database **fields**:

> Mrs Maria Johnson is 26 years old. She lives at 23 Example Street, Fairfield, 3078. Her mobile phone number is 04 123 456. Her customer ID number at the local newsagent is 4820M.

A field is designed to hold a single type of raw data only. It cannot contain a formula. For example, a field may hold a name, a postcode or a telephone number. Specific data types used by fields are discussed below.

All the data that relates to one 'entity', in this case Maria Johnson, is called a **record**. Records are distinguished from one another by creating a field in which the data it contains has a unique value for each record. This is called a **primary key**. In the example above, the customer ID number would be the primary key to uniquely identify Maria's record.

A set of records is stored in a database **table**. When viewed, each column of the table represents a field while each row represents a record. See Figure 8.3.

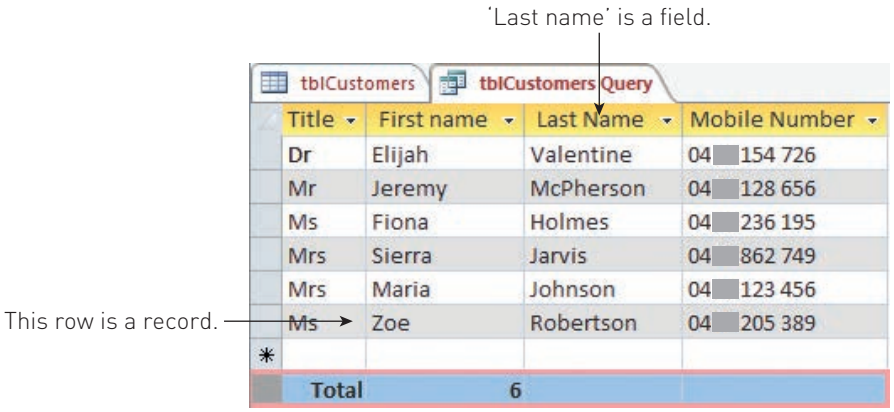


FIGURE 8.3 An example of DBMS table with a field (`Last Name`) and a record (Zoe Robertson) indicated.

Input forms

While a table holds the DBMS data, it is not necessarily a user-friendly object. To make acquisition of data easier and more accurate, **input forms** are used. Their structure guides users to understand what has to be entered and in what format. Forms use labels to explain the purpose of fields, hiding the complex field name from the user. Data entered via a form goes directly into the fields of the underlying table.

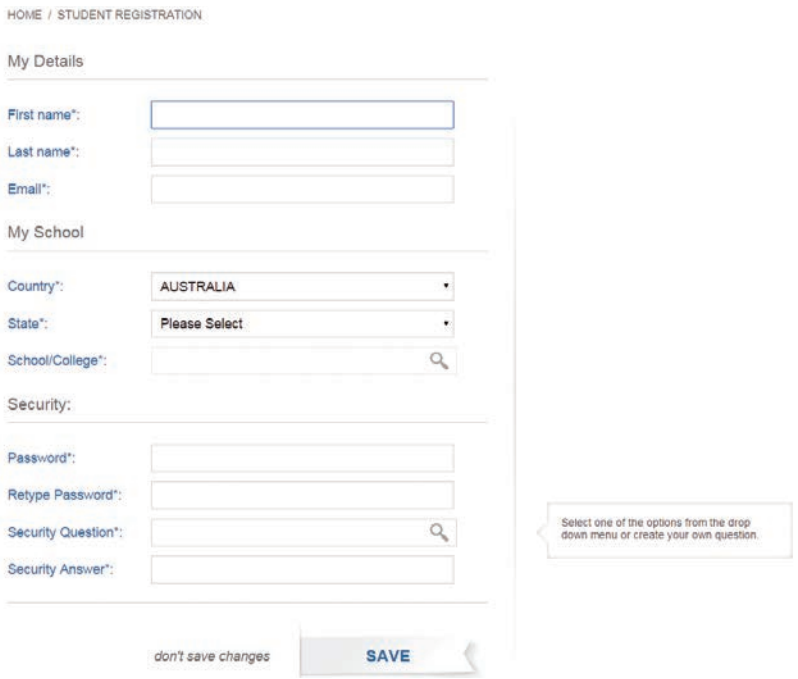


FIGURE 8.4 An example of an online input form

Queries

An important feature of a database is the ability to select a set of specific data based on a series of criteria. The database object that defines this selection is called a **query**. The query criteria are the result of a question we might have of the data. For example, 'How many of our customers in Fairfield are between the ages of 20 and 60?' (Figure 8.5) or 'Can I have a list of names and numbers for all our customers in Fairfield or Northcote?' (Figure 8.7).

Modern DBMSs use a structured query language or SQL as a way of standardising how data is managed in databases. Even DBMSs with GUI-based query creation, such as Microsoft Access, use SQL in the background (Figure 8.9).

Field:	cusCustomerID	cusCustomerTitle	cusCustomerFirstName	cusCustomerLastName	cusCustomerAge	cusCustomerStreet	cusCustomerSuburb	cusCustomerPostCode
Table:	tblCustomers	tblCustomers	tblCustomers	tblCustomers	tblCustomers	tblCustomers	tblCustomers	tblCustomers
Sort:								
Show:	☑	☑	☑	☑	☑	☐	☐	☐
Criteria:					Between 20 And 60		"Fairfield"	
or:								

FIGURE 8.5 The design of a query that selects all customers living in Fairfield between the ages of 60 and 60

qryFairfield20to60

Customer ID	Title	First Name	Last Name	Age (years)
6825Y	Ms	Zoe	Robertson	52
4820M	Mrs	Maria	Johnson	26
2026X	Mrs	Sierra	Jarvis	41

FIGURE 8.6 The results of a query of all customers living in Fairfield between the ages of 20 and 60

Field:	cusCustomerTitle	cusCustomerFirstNam	cusCustomerLastNam	cusCustomerSuburb	cusCustomerMobile
Table:	tblCustomers	tblCustomers	tblCustomers	tblCustomers	tblCustomers
Sort:					
Show:	☑	☑	☑	☐	☑
Criteria:				"Fairfield" Or "Northcote"	

FIGURE 8.7 The design of a query to list the names and mobile phone numbers of all customers living in Fairfield or Northcote

tblCustomers | tblCustomers Query

Title	First name	Last Name	Mobile Number
Dr	Elijah	Valentine	04 154 726
Mr	Jeremy	McPherson	04 128 656
Ms	Fiona	Holmes	04 236 195
Mrs	Sierra	Jarvis	04 862 749
Mrs	Maria	Johnson	04 123 456
Ms	Zoe	Robertson	04 205 389
*			
Total	6		

FIGURE 8.8 The results of a query to list the names and mobile phone numbers of all customers living in Fairfield or Northcote. The suburbs are not shown in the final query because they are not relevant to the needs of the user.

In Figures 8.5 and 8.7, each query has selected only the data we requested. A properly structured query helps to ensure any information we create is both relevant and complete.

You query a database of websites when you type words in the search box on the Google main page.

THINK ABOUT COMPUTING 8.3

Experiment with the advanced search tools in Google to see how to select a smaller set of results.

qryFairfield20to60

```
SELECT tblCustomers.cusCustomerID, tblCustomers.cusCustomerTitle, tblCustomers.cusCustomerFirstName, tblCustomers.cusCustomerLastName, tblCustomers.cusCustomerAge
FROM tblCustomers
WHERE (((tblCustomers.cusCustomerAge) Between 20 And 60) AND ((tblCustomers.cusCustomerSuburb)="Fairfield"));
```

FIGURE 8.9 The SQL for the query in Figure 8.5

Elements of effectiveness related to solutions like reports include completeness, readability, attractiveness, clarity, accuracy, accessibility, timeliness, communication of message, relevance and useability.

Reports

As useful as a query is, it is merely a simple list of selected records. To convert this list into real information, you must manipulate its formatting and layout, and add summary totals. A report is the information produced at the end of this process. This is what you communicate to the person for whom the information is created. It will be laid out as effectively as possible to make the information stand out and be easy to read and interpret.

Ralph's Roofing
60 Wellington St Kew 3101
Phone: 03 9884 1234

Accounts 30 days in arrears

Report created by MLS on 1-Aug-16

CustomerID	Last Name	Invoice No.	Size (m²)	Materials cost ($)	Labour cost ($)	Total ($)
2146	Aitken					
		B213	39	1045	2565	3 610
		B249	45	1450	2800	4 250
		B303	45	1850	2450	4 300
	Total money owed by customer: Aitken					12 160
1569	McCormack					
		B318	45	1110	2300	3 410
	Total money owed by customer: McCormack					3 410
2003	O'Donnell					
		B295	38	1870	2050	3 920
		B317	60	1795	3150	4 945
		B321	32	1670	2550	4 220
		X399	35	1495	2200	3 695
	Total money owed by customer: O'Donnell					16 780
	Total amount in arrears					32 350

FIGURE 8.10 An example of a database report. The information has been grouped and sorted, and summary totals calculated to make the report easier to understand.

Macros

In many cases, we may have to run a series of tasks whenever a database is opened or while it is in operation. For example, a data set may regularly need to be imported from another data source. The series of steps required to complete this process can be automated by a macro. The benefit of a macro is that it carries out the exact steps we ask of it when a single button or key combination is selected. This ensures the steps are followed accurately every time, and saves time and effort.

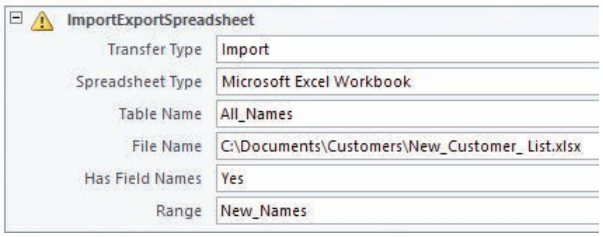


FIGURE 8.11 An example of the coding of a macro. This macro imports data from a spreadsheet file and appends it to a table called `All_Names`.

Characteristics of data types

The data held by a DBMS must be in a format suitable for storage. Typically, each field will be created to hold a specific type of data. This ensures data consistency in each field and assists with validation, because a field will refuse to accept data of the wrong type. For the purposes of this study, you must be familiar with text (string), numeric (integer and floating point), date, character and Boolean data types. Each of these is discussed in the section to follow Table 8.1.

Data types are also discussed in Chapters 6 and 7.

TABLE 8.1 The most common data types used in databases

Data type	Characteristics or uses	Examples
Text	• Alphanumeric up to 255 characters, searchable	• Name, address, postcode and phone number
Numeric	• Numbers only (see below for different forms)	• Any number to be used in a calculation
Currency	• Numbers, but in dollar amounts formatted with $ symbol and includes .00 at the end as a default	• Any number used to represent a financial value, usually only applied to a total, rather than values in a list
Date/Time	• A variation of numbers formatted to represent a date and/or time	• Any date can be used in calculations
Boolean	• Represents one of two states, usually True/False can be used in logical tests	• Also represented as Yes/No and On/Off

Text (string)

The majority of fields can be set as a text data type. Text may also be referred to as a string, particularly in connection with programming. This type of field holds a mix of characters (letters, numbers, special characters), also referred to as alphanumeric, to a limit of 255. Names and addresses would be considered to be text data. Also, postcodes and telephone numbers are normally formatted as text because they may contain spaces and are not intended to be used in the same way as a numeric value. It is also more efficient to store the values as text rather than as a large numeric value.

Numeric – integer, floating point

Another key type of data format used in databases is numeric. These fields will only allow numbers to be entered. They are often used when the value is to be used in a calculation of some kind. For example, the quantity of an item purchased might need to be multiplied by its price in order to calculate a total amount payable. This sort of calculation cannot be performed on a field formatted for text. Numeric fields can also be categorised into different variations. Integer is used for whole numbers, including negative numbers. Where decimal numbers are required, such as when dealing with financial transactions or percentages, then the floating point data type is used. Depending on the database package you use, the specific names of these numeric data types may vary, but their function will be the same.

Date

Strictly speaking, a date format is another variation of a numeric data type. The value is normally based on the number of days since the 'zero' day built into the operating system or DBMS. For example, the day '1' might be displayed as 01 January 1900, while '42673' would be displayed as 30 October 2016. Calculations can be performed on dates, which can be handy when comparing the difference between them. Dates can be formatted to show a combination of years, months, days, hours, minutes and seconds, depending on the needs of the user. In terms of time, they can also display 12- and 24-hour clocks.

Character

This is a text field that will only accept a single alphanumeric character. It is used where there are multiple options for a value, but they can be represented with a single character to make data entry easier and to save storage space. For example, a wooden box that comes in small, medium and large sizes might be entered as S, M and L, respectively.

Boolean

Boolean data type is used when the data to be entered falls into one of two categories. This is usually True/False, but it may also represent Yes/No or On/Off. It is sometimes shown as a checkbox or radio button on forms.

DATA TYPES

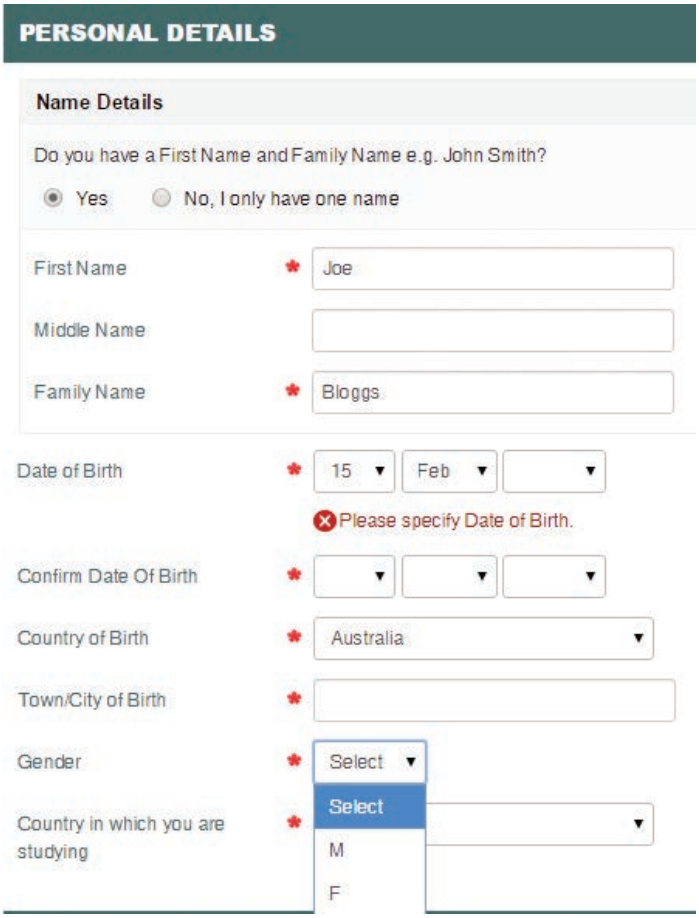
List two examples of each of the data types below. Note: You may also find it useful to refer back to Chapters 6 and 7.

1 Text (string)
2 **a** Numeric – integer
   b Numeric – floating point
3 Date/Time
4 Boolean

Data sources and methods of data acquisition

In this Outcome, students are expected to acquire their own data via a data collection tool that they design. The data collection tool does not have to be online, instead it will most likely be a form constructed from within their DBMS.

In Chapter 1, you covered various data sources and methods by which data can be acquired. These included various types of surveys and questionnaires involving a variety of open and closed questions. As has been discussed earlier in this chapter, to make use of any online system, such as banking, to join a group, to purchase products or to vote, you will be asked to register. As a minimum, you will need to supply basic contact details. Authentication may only require a valid email address, but banks would also ask for formal identification such as a driver's licence or birth certificate. To supply the requested data, you would complete an input form, similar to Figure 8.12 (characteristics of these forms are covered in more detail on page 315). When an online form is completed and submitted, the data is checked or validated, then written to a database. If a hardcopy form is used, then the details are either manually transcribed or scanned electronically into the database.



Department of Education and Training Creative Commons Attribution 3.0 Australia Licence

FIGURE 8.12 An example of an online application form. Note the mandatory fields marked with a red * and the use of dropdown boxes to assist with data validation.

Requirements for safeguarding personal information are discussed on page 332.

THINK ABOUT COMPUTING 8.4

List the personal details you need to supply in order to join a social media site like Facebook. Why might they need each of those specific elements of data? What items of data are optional?

Collection tools and user interfaces for data entry

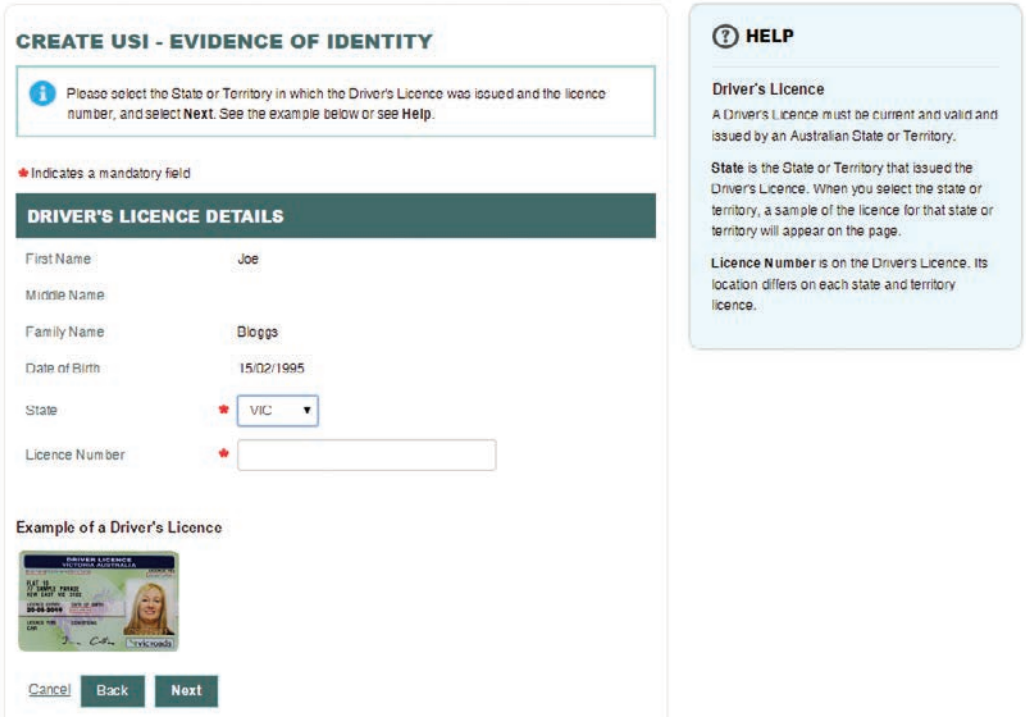
To ensure that data gathered via input forms is as accurate as possible, there are several techniques that make it easier for users to enter data. The form must guide users as to what types of data must be entered in different fields and what the accepted parameters or boundaries are for that data. The form itself needs to be arranged logically, usually to allow data entry starting from the top, and ensure that a user cannot complete the form without entering key data elements, like an email address and password.

Forms will involve several types of data validation to ensure the data entered is as accurate as possible. You will recall from previous chapters that validation involves checking that data exists in a field, is the correct type of data for that field and that it fits within the field's accepted value boundaries. **Existence checking** can be achieved by the use of mandatory entry fields. These are often marked with an asterisk (*) or might be in a different, eye-catching colour to the other fields. **Data type** validation can be achieved if you include labels showing examples of the expected values beside each field, and provide appropriate data-entry tools, such as numeric keypads and calendars, to encourage the correct data type to be entered. Labels can also indicate **correct ranges** of values that can be entered or by using dropdown lists with

prefilled options. The form might generate a suggested list of correct spellings if it detects the user has misspelled something important, such as part of an email address. These tools can also reduce errors caused by inconsistency.

An effective input form will also incorporate a number of other important design features to make the experience better for the user. It is important that forms reduce the amount of onscreen clutter by only displaying fields relevant to the user. In this way, forms can be adaptive; for example, when a user selects a certain option (such as delivery method for goods purchased), only fields relevant to that option (the type of delivery) are shown. Alternatively, as a user types in a credit card number, the type of card is determined and the options available change based on the card.

Forms must also display an appealing and attractive design overall.



Department of Education and Training Creative Commons Attribution 3.0 Australia Licence

FIGURE 8.13 Before continuing with a registration, a user must be validated by providing details of an accepted type of personal identification. Notice how the form tries to assist the user by providing an example of the form of identification to use and other helpful advice.

The role of a user experience (UX) designer is to create an online environment that encourages the user to interact with the system with ease by providing an interface that is both usable and accessible.

Checklist of common effective form elements

- Explain form fields clearly with adjacent labels
- Use indicators to show strength of passwords
- Use inline validation of fields
- Automatically suggest corrections for misspelled email addresses
- Provide a simple interface with large buttons for mobile phone users
- Adapt forms depending on options chosen (for example, delivery method, type of credit card)
- Convert text to a consistent format
- Vertically align labels and fields
- Provide context-sensitive help or information
- Provide clear steps if a user must navigate several input screens
- Use input tools that match the type of data to be entered (such as numeric keypads, sliders and tick boxes)
- Reduce form clutter so only what is relevant is on the screen
- Use an appealing overall design

The case study below is broken into several parts throughout this chapter. Its purpose is to show you a worked-through example of a database solution. Put yourself in the role of Saro. You, as the student, will be expected to come up with your own design brief and acquire your own data via a data-collection tool that you design. In this example, an online data-collection tool is used, but generally you are more likely to use a form you create within your DBMS.

CASE STUDY

KALUGANGA JUNIOR SWIMMING CLUB

The Kaluganga Junior Swimming Club is located in Doncaster. It provides swimming training and its members compete in regular local swimming meets. As this is a junior swimming club, members are 20 years old or younger, except for committee members, who are adults. The club records address, age and contact details for each member as well as their best times for several 50 m events. At present, each swimmer's record is stored in a separate word processing document that uses their name as the filename. After training sessions or after competition meets, the new times are compared with the existing times and updated if necessary on each swimmer's record. The club coach and other committee members often need to use the members' details to determine who the best swimmers are for certain events and who is up to date with membership fees.

There have been some problems with the current system. Sometimes the lists of swimmers are inaccurate because names are mistakenly put into the wrong age group or left off the list totally! The club's finanaces are not in order, as it has not accurately identified and followed up members who have not paid their annual fees. This has led to embarrassment for club officials when payment notices have been sent to the wrong club members.

As a result, Felix, the swimming club president, has asked his friend Saro to look at how to solve the problems with a digital-systems-based solution.

Saro has conducted an analysis of the problems and suggested a database solution that includes the ability to add and edit swimmer details and generate lists for the coaches and club treasurer. She has established the following requirements.

Functional requirements

- New swimmers must be able to be added.
- Existing swimmers can have their details edited.
- A list of all swimmers, grouped by age level and event can be generated.
- A list of all swimmers who owe club fees can be generated.
- The list for coaches must include all relevant performance details for each swimmer, including their best times.
- A macro will be used to open and print reports.

Non-functional requirements

- Interface for entering data must be simple to use.
- Results from reports must be accurate.

Constraints

The club wants to use its existing PC running a recent edition of an operating system and a popular office package that includes a DBMS.

Felix and the other committee members are confident with PC use, but unfamiliar with **database software**.

The solution needs to be in use before the next swim meet in a month's time.

Case study continues ▶

◀ *Case study continued*

Scope of solution

The solution will allow the club to:

- add, remove and edit records of swimmers
- generate lists of swimmers by age level and event within a certain swimming time and grouped by gender
- generate lists of swimmers with overdue payment of fees.

The solution will not:

- allow swimmers to edit their own details
- support online payments of fees.

This solution will benefit the club because:

- it will create accurate lists of swimmers and age groups for the coaches, allowing them to choose from the right pool of members for each competition event
- it will create accurate and up-to-date lists of members who are yet to pay their annual fees, which will assist in improving club revenue and reducing financial difficulties.

Database design tools

Database structure and appearance can be planned using a variety of common planning tools. Data dictionaries, data-structure designs, flowcharts, input–process–output (IPO) charts and query diagrams are useful for representing structure. **Layout diagrams** and mock-ups can also be used to plan the appearance of input forms and reports.

Several of these design tools will be further explained in the context of a DBMS in the following section.

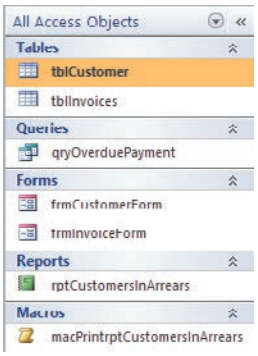


FIGURE 8.14
Examples of named objects within a Microsoft Access database file

Naming conventions

Every field and object within a database must be named properly so that it can easily be identified by a developer or a user. For example, you might prefix tables with `tbl` (such as `tblCustomer`, `tblProducts`) and fields within the Customer table with `cus` (such as `cusCustomerID`, `cusAddress`). Queries might be prefixed with `qry`, forms with `frm` and reports with `rpt`. Because users will interact with forms and reports often, ensure that the rest of the name is descriptive; for example, `frmInputCustomerDetails` or `rptOverdueInvoices` (Figure 8.14). From these examples, note that spaces and underscores are avoided in the object names. To aid understanding there is, however, a mix of upper- and lower-case characters in each name. In most databases a user would interact with a menu system with properly labelled buttons, rather than the individual objects in the main database screen.

Design tools to represent the structure of databases

The most common tool used to represent the basic structure of a database table is a **data dictionary**. This tells a database developer exactly how to setup the properties of each field in the tables. Remember that tables do not hold formulas, but can include validation. When planning fields, think about how the data might be used and how it might be best broken down. For example, the name 'Dr Andrea Smith' might be broken into three fields of `Title` (Dr), `FirstName` (Andrea) and `FamilyName` (Smith). Also look carefully at the type of data that the field will hold to determine the most appropriate data type to select.

TABLE 8.2 Example of a data dictionary to store invoices for a roofing contractor

Field	Data type	Field size	Input mask	Caption	Description	Validation rule	Validation text
invInvoiceNum	Text	4		Quote number	Quote number (primary key)		
invCustomerID	Text	4		Customer ID	Customer's identification code		
invSize	Number	Integer			Size of roof	Between 0 and 500	Normal building size should be between 0 and 500
invCostMaterials	Number	Integer		Cost of materials ($)	Cost of roofing materials used	Between 0 and 100 000	Cost of materials must be between 0 and 100 000
invCostLabour	Number	Integer		Cost of labour ($)	Cost of labour used	Between 0 and 20 000	Cost of labour must be between 0 and 20 000
invColour	Text	15		Roof colour	Colour of roof		
invJobDate	Date/ Time	dd-MM-YYYY		Date of job	Date the job was completed	Between #01/01/16# and #31/12/17#	Job completion must be in 2016 or 2017
invAccountPaid	Yes/No			Account paid?	Indicates whether or not the account has been paid		

When relational databases are designed, a data-structure diagram is used to show the relationships between each of the tables.

CASE STUDY

KALUGANGA JUNIOR SWIMMING CLUB – PART 2

Saro has taken the existing list of swimming club members and their details to determine how they will be represented in the database table. She has also added fields and included validation on the gender, postcode, date-of-birth and members-fees fields.

TABLE 8.3 Data dictionary for the Kaluganga Junior Swimming Club swimmers database `tblSwimmersDetails` table

Field	Data type	Field size	Input mask	Caption	Description	Validation rule	Validation text
swmID	Text	5		Membership number	Membership number		
swmFirstName	Text	20		First name			
swmFamilyName	Text	20		Family name			
swmGender	Text	6		Gender (M/F)	Male or Female	'Male' or 'Female'	Gender must be Male or Female
swmAddress	Text	100		Address			
swmSuburb	Text	20		Suburb			
swmPostcode	Text	4		Postcode		Between 3000 and 3999	Postcode must be between 3000 and 3999 inclusive

Case study continues ▶

◀ *Case study continued*

Field	Data type	Field size	Input mask	Caption	Description	Validation rule	Validation text
swmMobNumber	Text	14		Mobile number			
swmParent	Boolean	N/A		Parent contact?	Is the contact number that of a parent (True) or the swimmer (False)?		
swmDOB	Date	N/A	dd-MM-yyyy	Date of birth		<=Date()	DOB must be today or earlier
swmMemberFees	Numeric	3		Annual fee ($)	How much a member needs to pay each year	90	Annual fees are $90
swmFeesPaid	Numeric	3		Fees paid ($)	How much a member has paid this year		
swmFreestyleTime	Text	12		Freestyle time	Best recorded time for 50 m freestyle		
swmBreastStrokeTime	Text	12		Breaststroke time	Best recorded time for 50 m breaststroke		
swmBackstrokeTime	Text	12		Backstroke time	Best recorded time for 50 m backstroke		
swmButterflyTime	Text	12		Butterfly time	Best recorded time for 50 m butterfly		

Layout diagrams are also discussed in Chapter 5.

The other elements that need to be designed for the database include the input forms, queries to select data, reports to format the query results and macros to make database elements more efficient.

The key design tool used for input forms and reports is the layout diagram. It is essentially a sketch of what an input form or the reports of the solution will look like. It shows an interface developer the location of elements such as headings, labels and fields. This layout is based on appropriate use of formats and conventions. The layout diagram is annotated to show formatting details for all elements, including font type, size and style and options for selection lists. Any formulas to be added to a form or report will also be shown on a layout diagram. Like other design elements, layout diagrams are usually done by hand (see Figure 8.15), although pre-printed 'templates' can be used for some common elements or structures.

Balsamiq is a free handy online tool for creating layout diagrams. Your school can create a free account.

Design tools for representing input forms to capture data

When planning a layout diagram for a form, you need to ensure that it contains all the fields a user needs for their purpose. Elements of effective forms were explained earlier in this chapter. The fields should be laid out in a logical order, usually from top to bottom and in columns if there are too many fields for a single column. Headings can be used to arrange relevant fields into logical groups. Ensure that they have clear labels and make use of dropdown lists to assist with data entry. All formatting must also be clearly indicated. See Figures 8.15 and 8.16 for examples of annotated layout diagrams for input forms.

CASE STUDY

KALUGANGA JUNIOR SWIMMING CLUB – PART 3

Saro has designed a simple input form that Felix and the treasurer can use to manage (add, edit and delete) swimmers. She has incorporated a dropdown list for the gender field and a tick box for the field that notes if the mobile phone number field is a parent's number.

She has also designed a second form for the coaches that only shows key personal details and swimming event information.

Both forms use the same underlying table, but they only show data relevant to the needs of each user.

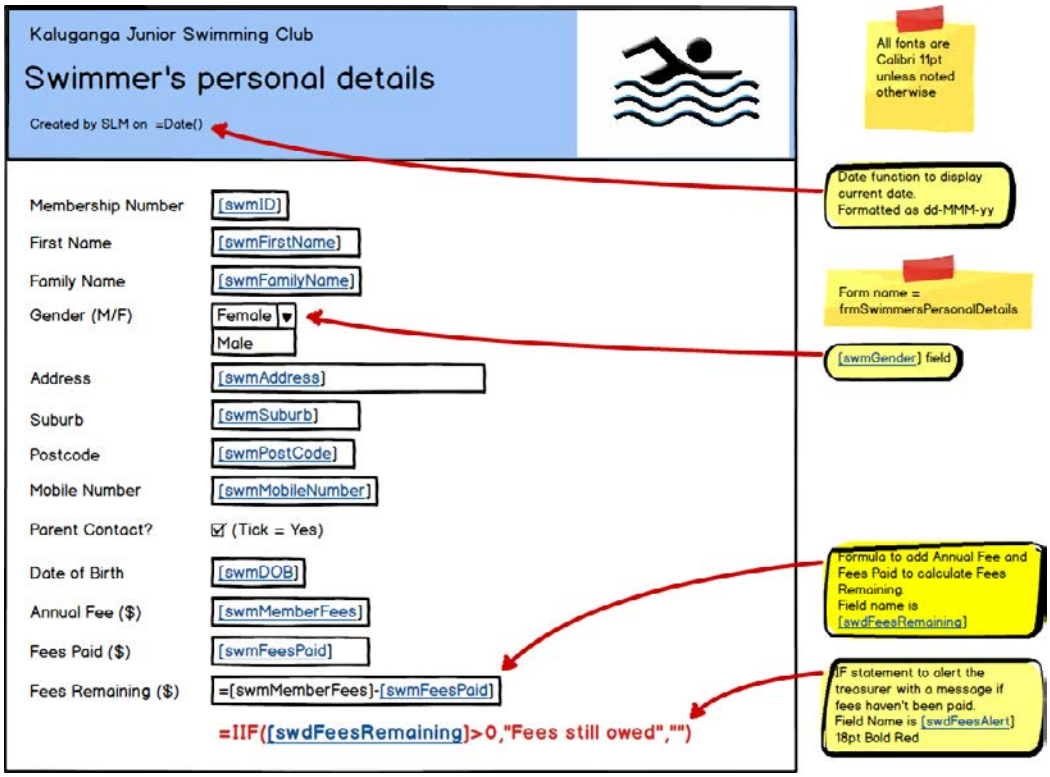


FIGURE 8.15 Saro's layout diagram for the main input form for the Kaluganga database `frmSwimmersPersonalDetails`. She has listed all the key personal details for each swimmer plus their membership fee payment details. The form has a heading section with the title, club logo and the current date. There are no swimming event times on this form, as it is used by Felix and the club treasurer.

Case study continues ▶

◀ *Case study continued*

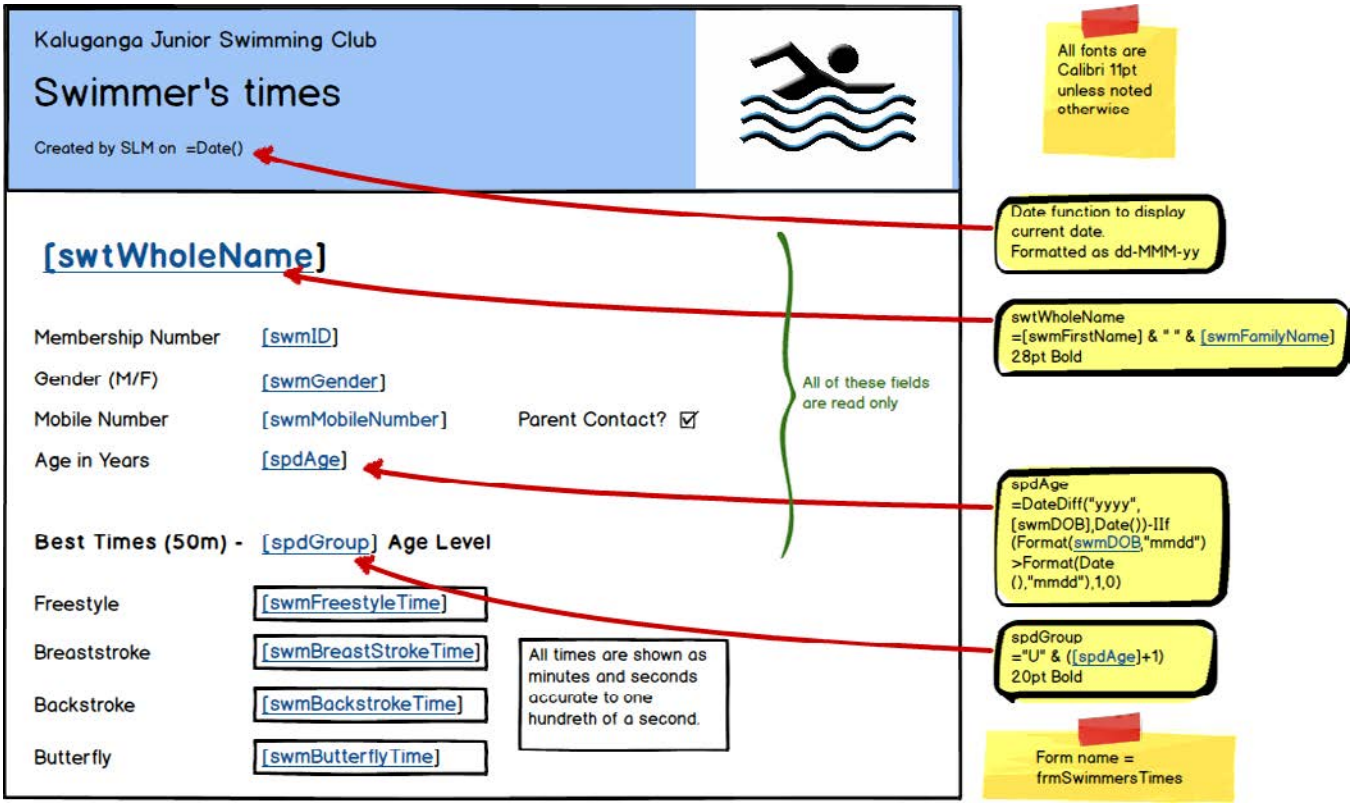


FIGURE 8.16 Saro's layout diagram for the input form coaches can use to update swimmers' times for events (`frmSwimmersTimes`). This form only has the personal details relevant to what the coaches need. They cannot edit any details except the times. This form also includes a calculation to show each swimmer's current age, in years, and their age group. It also contains a header with the same information as the other form.

Designing queries

Queries are used in databases to select only those records that meet the query criteria. They can also be used as a control structure to hold formulas used on forms and on other queries. The result of a query is a list of selected records organised by specified fields. This list is then used as the basis for a printed or electronic report.

The design of a query should specify the fields to be included and the tables to which they belong. The design will also include the criteria for the query and how the resulting records are to be ordered. The query design is usually hand drawn, although you may set up a basic table as a template. In your solution, you should use at least two criteria and both a primary and a secondary sort (sorting is explained in 'Setting up the queries' later in this chapter). Also, be aware that not all fields used for selecting data need to be shown in the final list.

If formulas are created in the query, you should ensure that they have the correct data type (for example, text or numeric) and a proper name applied.

TABLE 8.4 A sample query design for calling up records of overdue accounts. The records will be ordered by the last name of the customer and then the invoice number, since the leftmost sort designation (primary sort key) is assigned to `LastName`.

Field	cusCustomer ID	cusLast Name	invInvoice Num	invSize	invCost Materials	invCost Labour	TotalCost:= [invCost Materials]+ [invCostLab our]	invJob Date	invAccount Paid
Sort		Ascending	Ascending						
Show	✓	✓	✓	✓	✓	✓		✓	
Criteria								<#01/08/ 16#-30	False

Query criteria

The query criteria can use <, >, =, <=, >=, and <> symbols when dealing with numbers or text. Alternatively, plain language such as 'between 0 and 20 000' or 'less than 500' can be used. The '*' and '?' symbols are also valid wildcards. Text used in query criteria needs to be placed inside single inverted commas. Other expressions such as 'is null' or 'is not null' can be used to show records where a field is either empty or contains some data. A query can designate a date by placing '#' marks at each end of the date. If more than one field has a query criterion assigned to it, the criteria are applied in an 'and' relationship. That is, both the criteria must be met before data is returned by the query.

CASE STUDY

KALUGANGA JUNIOR SWIMMING CLUB – PART 4

Saro has designed two queries for the Kaluganga Junior Swimming Club database. One of the queries will select all swimmers who have overdue fees. Because Felix wants to call the parents of these swimmers to discuss payment, the query will also only show those swimmers whose parent's mobile phone number is listed. The other query will select all under-17 swimmers with swimming times below a certain threshold in backstroke (40 seconds) to enable the coaches to select the fastest swimmers for that race.

TABLE 8.5 The design of the query to select swimmers with overdue fees (`qryOverdueFeesParentContact`). A formula called `swdFeesRemaining` has been created to calculate the value of unpaid fees. If it is greater than 0 (0 means all the fees have been paid) then the record is selected.

Field	swmFamily Name	swmFirst Name	swmID	swmMobile Number	swmMember Fees	swmFees Paid	swdFeesRemaining: [swmMemberFees]- [swmFeesPaid]	swm Parent
Sort	Ascending	Ascending						
Show	✓	✓	✓	✓	✓	✓	✓	
Criteria							>0	True

Case study continues ▶

◀ *Case study continued*

TABLE 8.6 The design of the query to list the fastest U17 backstroke swimmers (`qryU17BackstrokeBestTimes`). It selects swimmers who have a backstroke time of less than 40 seconds and who are also under 17. The long formula is intended to calculate the correct age in years as accurately as possible. All swimmers are under 17, so that field does not need to be shown when the query runs.

Field	swm Gender	swmBack StrokeTime	swmFamily Name	swmFirst Name	"U" & DateDiff("yyyy",[swmDOB], Date())-IIf(Format([swmDOB],"mmdd")> Format(Date(),"mmdd"),1,0)+1
Sort	Ascending	Ascending	Ascending		
Show	✓	✓	✓	✓	
Criteria		<"00:40.00"			"U17"

Design tools to help reports meet specific needs

As discussed earlier in this chapter, the result of a query is used to generate a report. A layout diagram is used to represent the structure of a report. Reports have common sets of formats and conventions. Other essential elements specific to reports include the organisation's details and date in the report header, grouping of information and summary statistics. The records that are returned from the query are placed in the detail section of the report.

Sections	Content/Layout	Formatting
Report header	Ralph's Roofing 60 Wellington Street, Kew 3101 Ph 03 9884 1234 Accounts 30 days in arrears Report created by ALB on =Date()	Garamond 24pt bold, blue TNR 10pt TNR 18pt bold, red TNR 10pt
Page header	CustomerID Last name Invoice number Size (sq m) Cost of materials ($) Cost of labour ($) Total cost ($)	TNR 10pt bold
cusLast Name header	[cusCustomerID] cusLastName	TNR 12pt bold, blue
Detail	invInvoiceNum invSiz invCostMaterials invCostLabour TotalCost	TNR 9pt
cusLastName footer	Total money owed by customer cusLastName =Sum([TotalCost])	TNR 9pt bold
Page footer	="Page" & [Page] & "of" & [Pages]	TNR 8pt
Report footer	Total in arrears =Sum([TotalCost])	TNR 10pt bold
Report name:	rptCustomersInArrears	**Source:** qryOverduePayment
Page orientation:	A4 portrait	**Grouping/sort:** Group by cusLastName, sort by invInvoiceNum

FIGURE 8.17 An example of a grouped report design. This is the report shown in Figure 8.10.

Note the report elements down the left-hand side. These correspond with the section headings used in the Design View of the report in Microsoft Access.

CASE STUDY

KALUGANGA JUNIOR SWIMMING CLUB – PART 5

Saro has designed two reports for the swimming club. Each report is based on one of the queries designed above. Saro shows the designs to Felix and the treasurer, who make suggestions for minor changes that will make the reports clearer and more readable.

Sections	Content/Layout	Formatting (All fonts Calibri 11pt unless specified)
Report Header	Kaluganga Junior Swimming Club Report of Overdue Fees with Parent Contact Available Report created by SLM on =Date()	20pt 20pt bold
Page Header	Family Name First Name Membership Number Mobile Number Annual Fee ($) Fees Paid ($) Remaining ($)	Bold
Detail	swmFamilyName swmFirstName swmID swmMobileNumber swmMemberFees swmFeesPaid swdFeesRemaining	
Page Footer	="Page " & [Page] & " of " & [Pages]	
Report Footer	Total number of overdue accounts is =Count([swmFamilyName]) Total amount overdue is =Sum([swdFeesRemaining])	Bold

Report Name: rptOverdueFeesParentContact **Source:** qryOverdueFeesParentContact
Page Orientation: A4 Portrait **Grouping/Sort:** Sorted by swmFamilyName, then swmFirstName

FIGURE 8.18 This is the design for Felix's 'Report for overdue fee with parent contact available' report. The report is to be sorted on the 'FamilyName' field, then the 'FirstName' field.

Sections	Content/Layout	Formatting (All fonts Calibri 11pt unless specified)
Report Header	Kaluganga Junior Swimming Club U17 Best Times for Backstroke Report created by SLM on =Date()	20pt 20pt bold
Page Header	Gender Backstroke Time Family Name First Name	Bold
swmGender Header	swmGender	Bold, 12pt
Detail	swmBackstrokeTime swmFamilyName swmFirstName	
swmGender Footer	="Total number of " & [swmGender] & " swimmers listed is" =Count([swmFamilyName]	
Page Footer	="Page " & [Page] & " of " & [Pages]	
Report Footer	Grand Total of swimmers available is =Count([swmFamilyName	Bold

Report Name: rptU17BestTimesForBackstroke **Source:** qryU17BestTimesForBackstroke
Page Orientation: A4 Portrait **Grouping/Sort:** Group by swmGender, then sort ascending by swmBackstroketime

FIGURE 8.19 This layout diagram shows the design of the report for the coaches, which selects the best under-17 backstroke swimmers. This report is grouped by gender, then sorted in ascending order by time.

Designing macros

Pseudocode, also known as structured English, is discussed in Chapter 6.

Macros can be planned using pseudocode or a flowchart. When using pseudocode to design a macro, be sure to use indenting and a logical order of steps, so that it is easy to see what it is intended to do and to make testing easier.

CASE STUDY

KALUGANGA JUNIOR SWIMMING CLUB – PART 6

Saro has decided to create two simple macros that print the two reports. The tasks that she wants to perform include opening the report, printing it to the default printer and then closing it. Her pseudocode design for one of these macros is shown in Figure 8.20. Saro will also create a macro that opens the swimmer's times form for editing. This macro may be used in conjunction with a button on a 'Dashboard'. The pseudocode for this macro is shown in Figure 8.22.

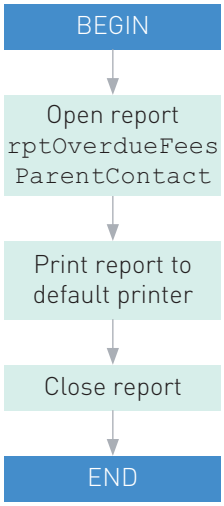


FIGURE 8.21 This flowchart represents the same steps as the macro design in Figure 8.20.

```
Begin
    Open Report 'rptOverdueFeesParentContact'
    Print report to default printer
Close report
End
```

FIGURE 8.20 Saro's pseudocode design for one of the macros that opens a report and prints it, then closes the report

```
Begin
    Open Form 'frmSwimmersTimes'
    Show on screen
End
```

FIGURE 8.22 Saro's pseudocode design for the macro that opens the swimmer's times form

Development of the DBMS

Once the designs have been finalised and agreed to by the client, the development of the database can take place. The DBMS file is first created and stored; then, each of the key objects inside the database is created, starting with the tables.

Creating tables

It is crucial that the developer constructs the tables properly as all other parts of the database depend on this. They must ensure that the correct names and data types, as well as validation rules, are assigned to each field.

Creating user input forms

Now that tables have been constructed, input forms can be created. Remember that an input form is designed to enable easy data entry. It can also be used to calculate formulas, provide selection lists and other data input control tools. Layout and various formats and conventions to follow were discussed earlier in this chapter.

Setting up the queries

The queries that will select the required subsets of data are constructed once the tables and reports have been finalised. Each database might involve multiple queries with different criteria.

Creating reports

Reports can be generated once the queries have been set up. Each report uses the data selected by its corresponding query. The report turns the data into information by presenting it in a more useful format for the end user.

Creating macros

Once the key objects in the database have been constructed, macros can be recorded to automate common tasks. Macros can even be placed on forms to create an interactive menu.

CASE STUDY

KALUGANGA JUNIOR SWIMMING CLUB – PART 7

Saro creates a database file called 'Kaluganga Swimmers Details'. She then constructs the 'tblSwimmersDetails' table, which will hold all of the swimming club data relevant to the swimmers (Figure 8.23). Once happy with the table, Saro constructs the two forms – for individual swimmers' personal details (see Figures 8.25 and 8.26) and for individual swimmers' best times (see Figures 8.30 and 8.31). These are followed by the two queries – regarding parent contact for swimmers with overdue fees (Figures 8.32 and 8.33) and for swimmers' best times for backstroke (see Figures 8.35 and 8.36) – and then the two reports – the overdue fees parent contact report (see Figures 8.39 and 8.40) and the best times for backstroke report (see Figures 8.41 and 8.42). Finally, Saro creates the macros to make it easier and faster to generate the reports (see Figures 8.43 and 8.44). At this stage, Saro is satisfied that the software side of the solution is complete.

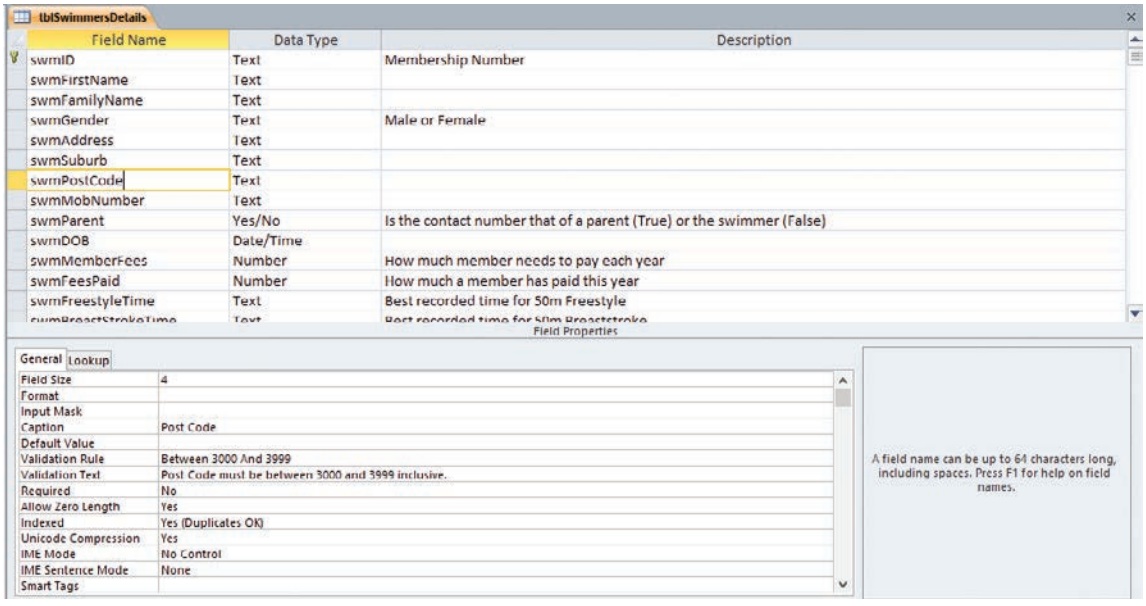


FIGURE 8.23 The `tblSwimmersDetails` table in Design View in Microsoft Access. Note how validation has been set on the postcode field.

Case study continues ▶

◀ Case study continued

tblSwimmersDetails

Membership Number	First Name	Family Name	Gender (M/F)	Address	Suburb	Post Code	Mobile Number	Parent Co
L8785	Sharon	Baird	Female	21 Crescent	Doncaster East	3109	04 311 241	☑
E5687	Audrey	Barry	Female	3/2 Avenue	Doncaster East	3109	04 154 500	☑
U8645	Paula	Bond	Female	2 Avenue	Doncaster	3108	04 842 762	☐
F3480	Clare	Bray	Female	10 Street	Doncaster	3108	04 188 714	☐
T6305	Erica	Burnett	Female	14	Templestowe	3106	04 042 614	☐
U6941	Eric	Chang	Male	39 Street	Doncaster	3108	04 433 530	☐
H9523	Rajah	Chowdhry	Male	71 Street	Templestowe	3106	04 233 176	☑
G0141	Benjamin	Church	Male	5/18 Court	Doncaster East	3109	04 555 039	☐
R5010	Callum	England	Male	93 Street	Doncaster East	3109	04 486 504	☐
K8148	Robin	Evans	Female	87 Court	Templestowe	3106	04 864 486	☑
J6691	Justine	Evans	Female	98 Lane	Doncaster	3108	04 052 620	☐
B1463	Carolyn	Fleming	Female	56 Court	Templestowe	3106	04 577 123	☐
Z2664	Jesse	Gardner	Male	1/5 Avenue	Doncaster	3108	04 091 612	☐
V2944	Nina	Haynes	Female	1 Place	Templestowe	3106	04 339 144	☐
D1544	Cheryl	Hooper	Female	23 Crescent	Doncaster East	3109	04 400 150	☑
R8102	Phillip	Hunt	Male	482 Way	Templestowe	3106	04 882 884	☑
U6018	Warren	Luna	Male	63 Close	Doncaster East	3109	04 036 414	☑
C7170	Isabelle	Lynch	Female	63 Street	Doncaster East	3109	04 544 492	☑
J1252	Jacqueline	Mercer	Female	3/2 Court	Doncaster	3108	04 974 204	☑
M8892	Helen	Nieves	Female	124 Avenue	Doncaster East	3109	04 444 906	☑
E5099	Kamal	Reeves	Male	75 Court	Doncaster East	3109	04 552 016	☑
M3927	Lois	Schroeder	Female	58 Street	Templestowe	3106	04 024 862	☑
P4895	Tarik	Shannon	Male	156 Place	Doncaster	3108	04 361 471	☑
P7262	Kirk	Shelton	Male	3 Court	Doncaster East	3109	04 272 692	☑

FIGURE 8.24 Part of the swimmers' details table with data entered.

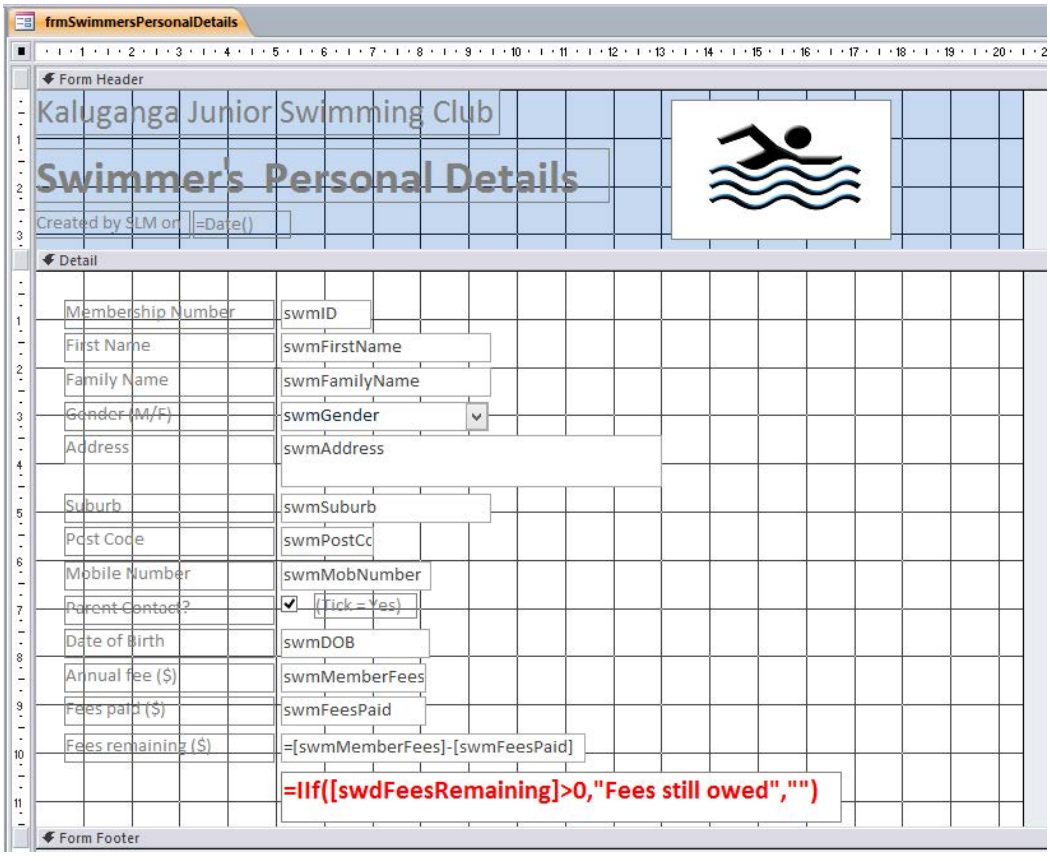


FIGURE 8.25 The `frmSwimmersPersonalDetails` form in Design View in Access.

Case study continues ▶

◀ Case study continued

FIGURE 8.26 The swimmer's personal details form as it appears for data entry or editing. Note that because the 'fees remaining' field is greater than 0, the 'Fees still owed' message is displayed to alert the club treasurer.

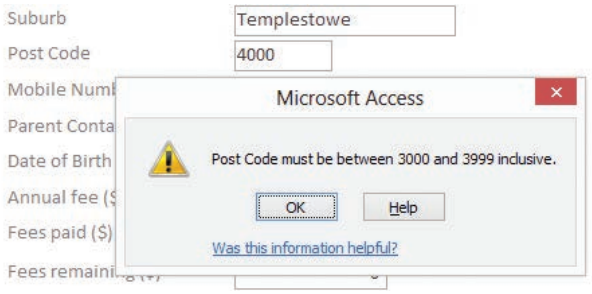


FIGURE 8.27 This is the message that appears when a range checking validation rule is activated caused by an incorrect value being entered in the postcode field on the form.

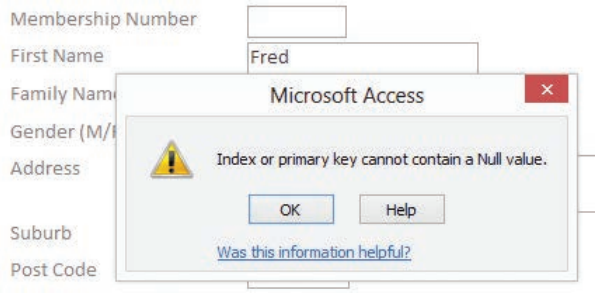


FIGURE 8.28 This is an alert for an existence check on the membership number field. The membership number is the primary key so a value must exist for each record.

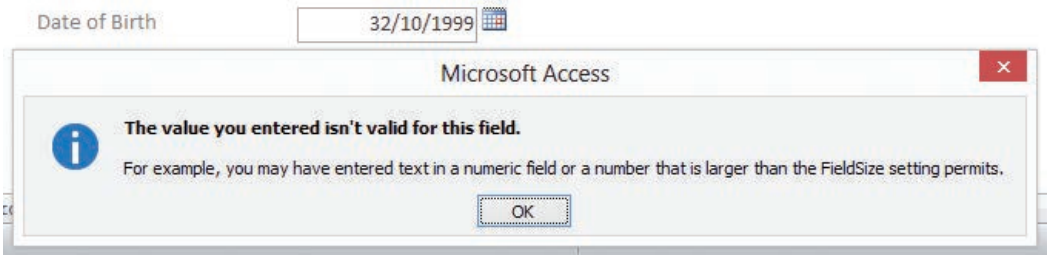


FIGURE 8.29 A data-type check alert occurs when an invalid date is entered in the date-of-birth field. A similar message will also appear when a non-numeric value is entered into the fees-paid field.

Case study continues ▶

◀ Case study continued

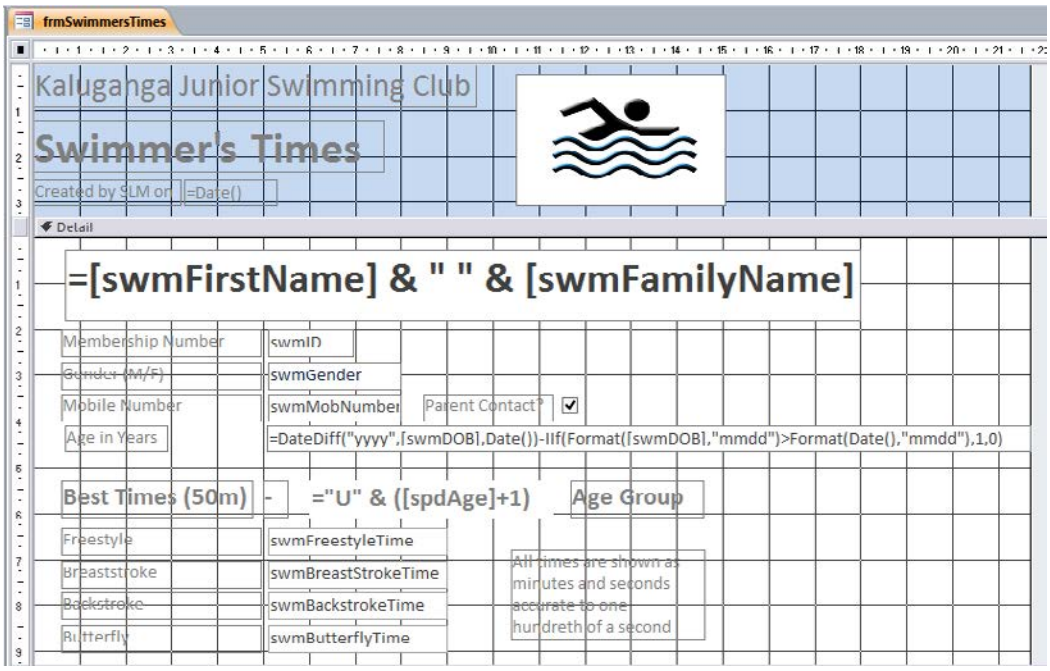


FIGURE 8.30 The `frmSwimmersTimes` form in Design View in Access. The fields have been expanded so the various formulas can be seen.

FIGURE 8.31 The swimmer's times form to be used by the coaches for updating details for best times in each swimming event. Personal details cannot be changed on this form, only the times.

qryOverdueFeesParentContact

```
SELECT tblSwimmersDetails.swmFamilyName, tblSwimmersDetails.swmFirstName, tblSwimmersDetails.swmID, tblSwimmersDetails.swmMobNumber, tblSwimmersDetails.swmMemberFees,
tblSwimmersDetails.swmFeesPaid, [swmMemberFees]-[swmFeesPaid] AS swdFeesRemaining
FROM tblSwimmersDetails
WHERE ((([swmMemberFees]-[swmFeesPaid])>0) AND ((tblSwimmersDetails.swmParent)=True))
ORDER BY tblSwimmersDetails.swmFamilyName, tblSwimmersDetails.swmFirstName;
```

FIGURE 8.32 The `qryOverudueFeesParentContact` query, displayed as an SQL statement

Case study continues ▶

◀ Case study continued

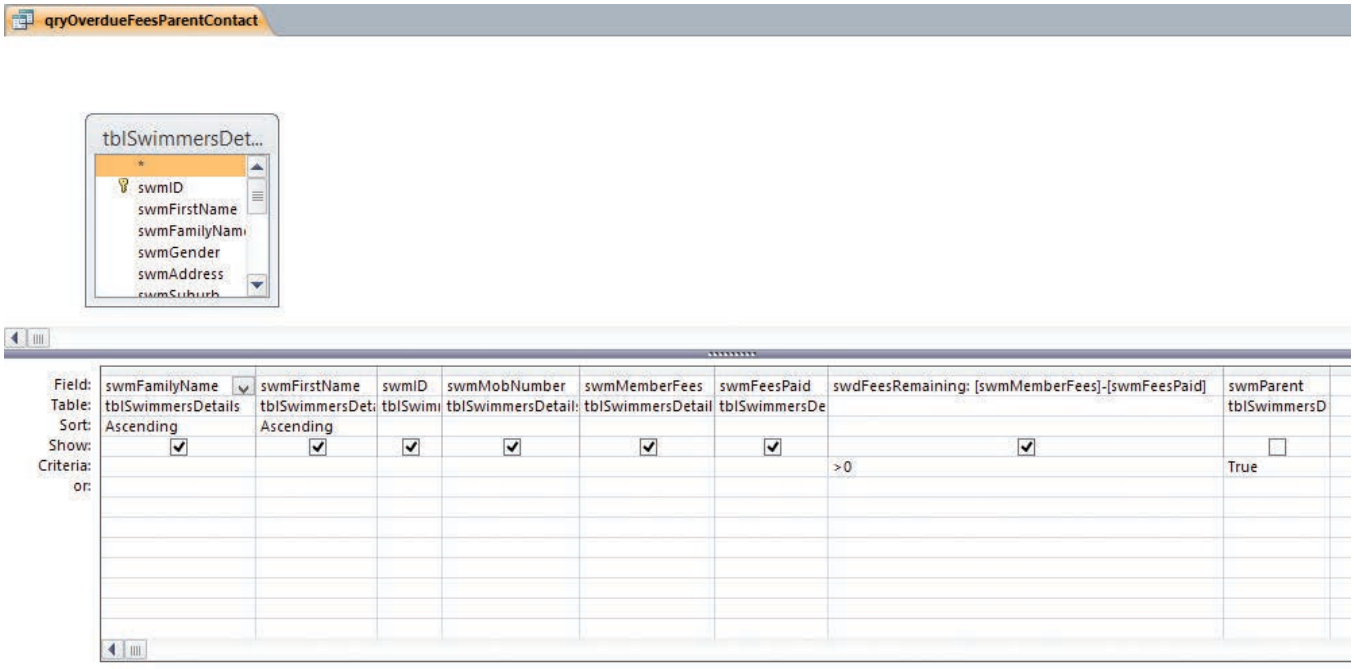


FIGURE 8.33 The Design View in Access for the `qryOverdueFeesParentContact` query. The widths have been changed on some of the fields so that the formula calculating the fees remaining to be paid can be seen.

qryOverdueFeesParentContact

Family Name	First Name	Membership Number	Mobile Number	Annual fee ($)	Fees paid ($)	Fees Remaining ($)
Hunt	Phillip	R8102	0412 882 884	90	53	37
Lynch	Isabelle	C7170	0450 544 492	90	8	82
Mercer	Jacqueline	J1252	0421 974 204	90	36	54
Shannon	Tarik	P4895	0461 361 471	90	64	26
Shelton	Kirk	P7262	0434 272 692	90	5	85
*				90	0	
Total		5				284

FIGURE 8.34 When run, the overdue fees query selects only those swimmers with outstanding membership fees and whose contact number belongs to their parents. Only fields relevant to the treasurer's needs are used so that only the required data is displayed.

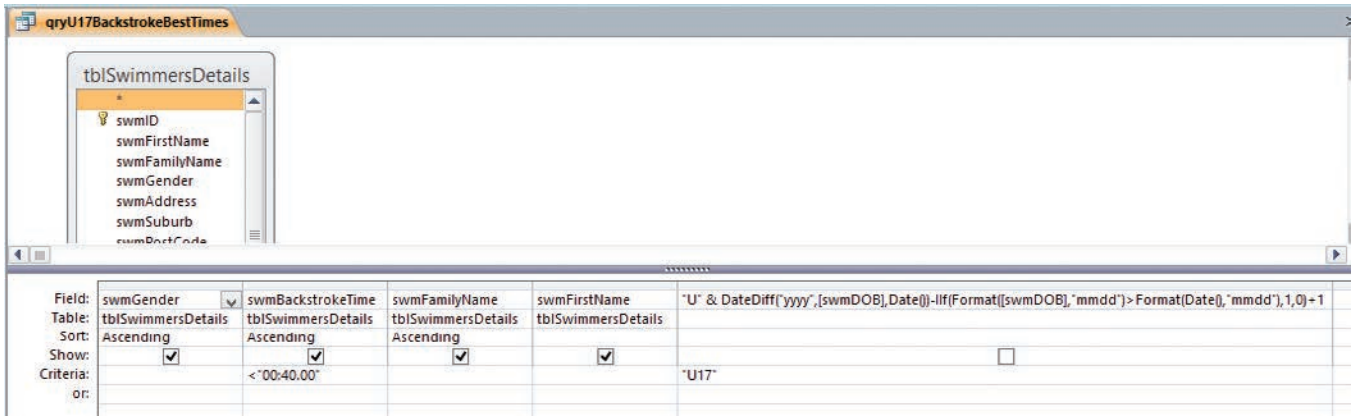


FIGURE 8.35 The Access Design View of the `qryU17BackstrokeBestTimes` query. It selects swimmers with a backstroke time of less than 40 seconds and whose age group is under 17. The age-group field does not need to be shown when the query is run because all records listed will have the same age group, which makes this data redundant.

Case study continues ▶

◀ Case study continued

qryU17BackstrokeBestTimes

```
SELECT tblSwimmersDetails.swmGender, tblSwimmersDetails.swmBackstrokeTime, tblSwimmersDetails.swmFamilyName, tblSwimmersDetails.swmFirstName
FROM tblSwimmersDetails
WHERE (((tblSwimmersDetails.swmBackstrokeTime)<"00:40.00") AND (("U" & DateDiff("yyyy",[swmDOB],Date())-IIf(Format([swmDOB],"mmdd")>Format(Date(),"mmdd"),1,0)+1)="U17"))
ORDER BY tblSwimmersDetails.swmGender, tblSwimmersDetails.swmBackstrokeTime, tblSwimmersDetails.swmFamilyName;
```

FIGURE 8.36 The under-17 backstroke swimmers query in SQL view.

qryU17BackstrokeBestTimes

Gender (M/F)	Backstroke Time	Family Name	First Name
Female	00:34.35	Evans	Robin
Female	00:35.11	Nieves	Helen
Female	00:38.09	Baird	Sharon
Female	00:38.16	Mercer	Jacqueline
Male	00:32.61	Shelton	Keith
Male	00:33.18	Shannon	Tarik
Male	00:35.28	Gardner	Jesse
Male	00:36.33	Shepherd	Brenden
Male	00:36.58	Church	Benjamin
Total		9	

FIGURE 8.37 The query selects the best under-17 backstroke swimmers and is sorted by gender, then by time with the fastest at the top (ascending). It only selects the data the coaches need.

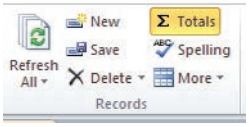


FIGURE 8.38 In Access, simple totals like sums and counts can be shown in queries by choosing the totals option.

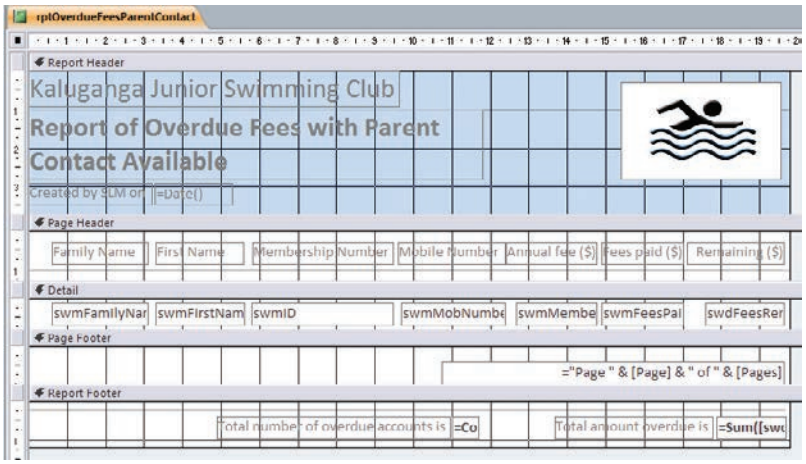


FIGURE 8.39 This is the design of the `rptOverdueFeesParentContact` report in Access. A Count and a Sum function are both used at the end of the report to provide summary statistics. These increase readability and communication of the message.

Report of Overdue Fees with Parent Contact

Kaluganga Junior Swimming Club

Report of Overdue Fees with Parent Contact Available

Created by SLM on 12-Mar-15

Family Name	First Name	Membership Number	Mobile Number	Annual fee ($)	Fees paid ($)	Remaining ($)
Hunt	Phillip	R8102	0412 882 884	90	53	37
Lynch	Isabelle	C7170	0450 544 492	90	8	82
Mercer	Jacqueline	J1252	0421 974 204	90	36	54
Shannon	Tarik	P4895	0461 361 471	90	64	26
Shelton	Kirk	P7262	0434 272 692	90	5	85

Total number of overdue accounts is 5 Total amount overdue is $284

Page 1 of 1

FIGURE 8.40 The final version of the overdue fees report for the club treasurer. It includes a summary total of the number of records listed as well as the total amount owed to the club. It is designed to be printed on an A4-size portrait page.

Case study continues ▶

◀ Case study continued

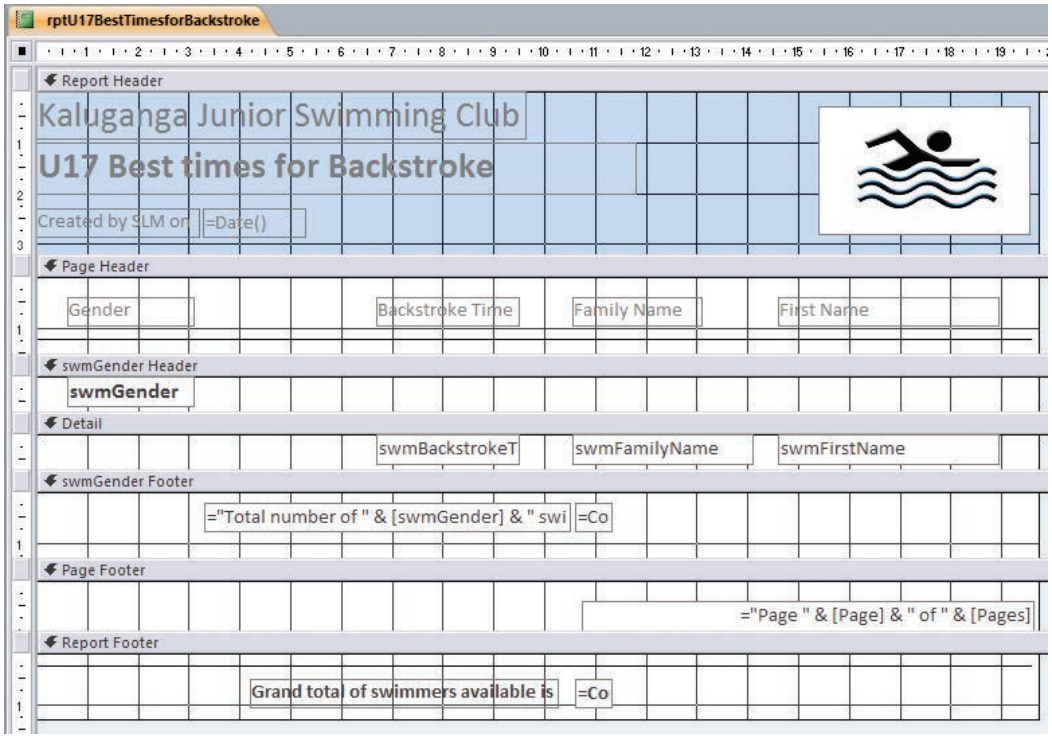


FIGURE 8.41 The `rptU17BestTimesforBackstroke` report in Access Design View. Note the header and footer sections for the gender field because it is being used to group data.

Kaluganga Junior Swimming Club

U17 Best times for Backstroke

Created by SLM on 12-Mar-15

Gender	Backstroke Time	Family Name	First Name
Female			
	00:34.35	Evans	Robin
	00:35.11	Nieves	Helen
	00:38.09	Baird	Sharon
	00:38.16	Mercer	Jacqueline
Total number of Female swimmers listed is	4		
Male			
	00:32.61	Shelton	Keith
	00:33.18	Shannon	Tarik
	00:35.28	Gardner	Jesse
	00:36.33	Shepherd	Brenden
	00:36.58	Church	Benjamin
Total number of Male swimmers listed is	5		
Grand total of swimmers available is	**9**		

FIGURE 8.42 This is the grouped list of the best under-17 backstroke swimmers sorted by time for the coaches. It includes a subtotal for each gender group as well as an overall total of swimmers who fit the selection criteria. This report is designed to be printed on an A4 portrait page.

Case study continues ▶

◀ *Case study continued*

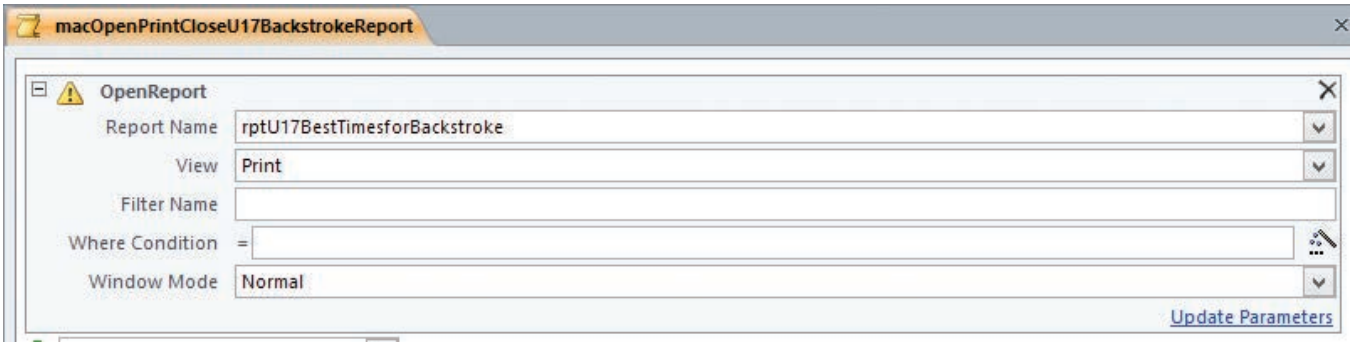


FIGURE 8.43 The Design View for the macro that opens the under-17 backstroke swimmers report, prints it, then closes it.

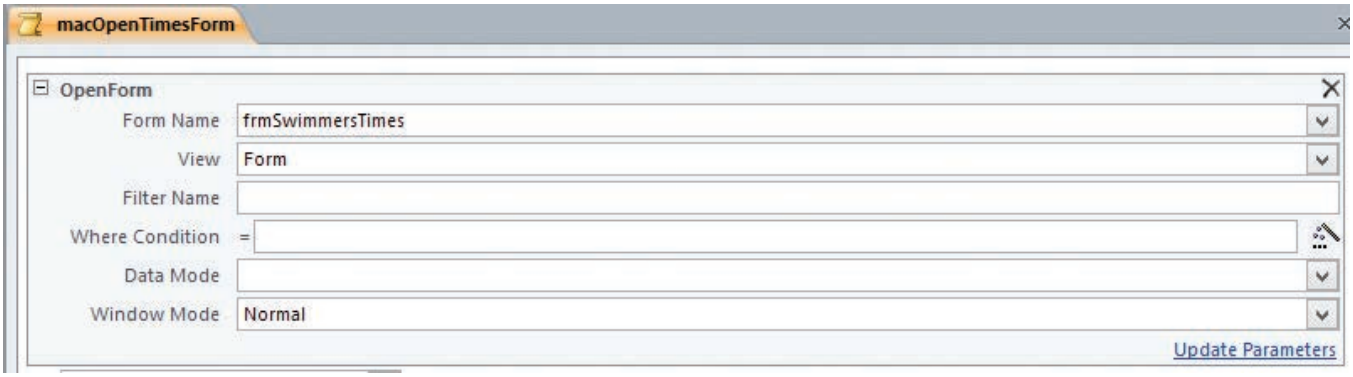


FIGURE 8.44 This is the Design View for the macro that opens the Swimmer's Details form when run.

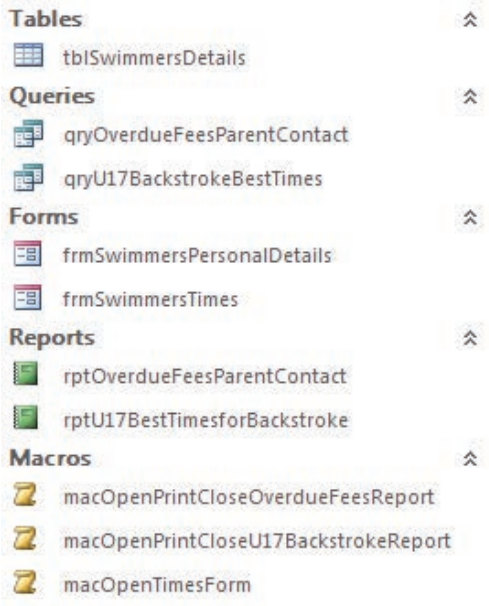


FIGURE 8.45 This is a listing of all the database objects Saro has created and their names.

Roles, functions and characteristics of hardware components

In Chapter 5, you were introduced to hardware used for input, storage, communication and output. In this section, we explore common hardware devices that perform these actions in the context of databases. You may recall that each piece of equipment needs to be identified by its role, functions and its characteristics. It is important that you are able to advise a client, or confirm for yourself if you are creating the database for your own needs, of suitable hardware that they may need to operate the database solution you create for them.

Input devices

The keyboard is a frequently used hardware device for typing data directly into fields. This may be users filling in a form on a webpage or staff transcribing data from a printed form completed by hand. A mouse is also used to interact with dropdown lists or as a convenient way to activate tick boxes or similar controls. But these devices are inefficient when a database requires a high volume of data to be entered rapidly from a non-electronic source. For example, imagine how much longer shopping would take if a person at the checkout had to manually enter the barcode of every product purchased!

To improve the speed of data entry, a range of scanners and reading devices exist. These include optical scanners and readers, barcode readers, RFID readers, magnetic stripe card readers, MICR readers and various electronic sensors. Often these input devices are grouped into handheld data-collection devices so that a user can input data at the point where the item or person is located. They will often communicate wirelessly with the database. Many waiters may use such a device when they take your order in a restaurant. You may also see them in use in large stores and supermarkets when managers walk along beside the shelves noting which items need to be restocked. In some cases, you may even have been asked to sign for a parcel on such a device.

TABLE 8.7 Input devices

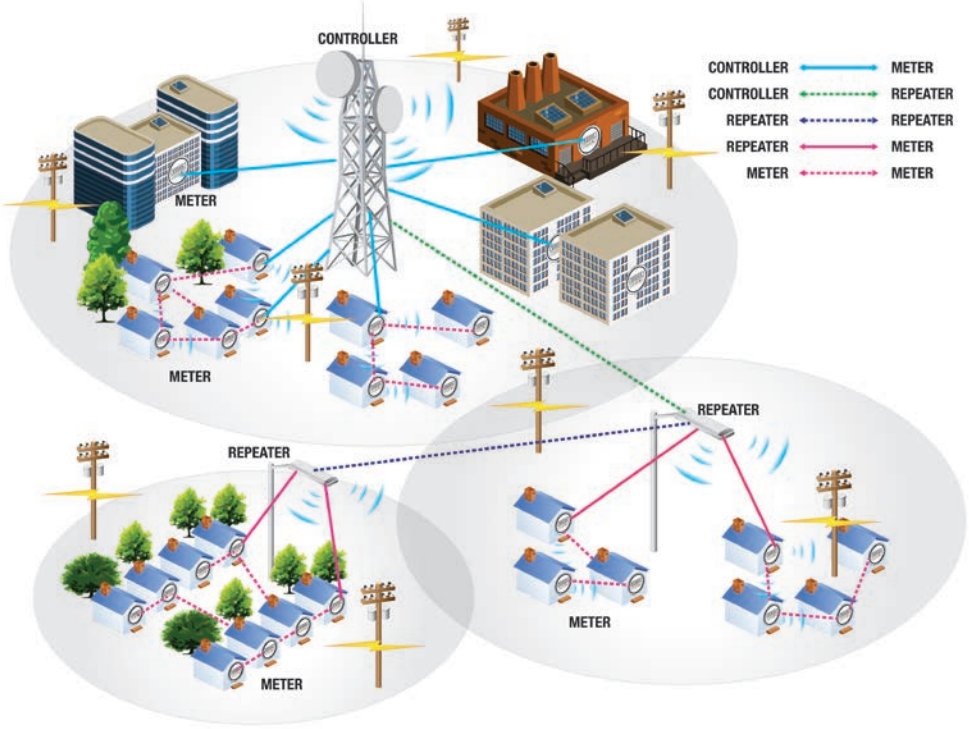
Device	Role	Function	Characteristics
Keyboard	Manual input of data	Users press keys to input data	From 101–5 keys Assignable special function keys Can handle rapid input of typed data
Mouse		Users press buttons to select data to input from premade lists	2–3 buttons Slow data speed
Optical scanner	Automatic data acquisition and input via optical reader or radio signal.	Source documents or devices either pass in front of, or nearby, the reader.	Can handle high volume of data input Source documents are brought to reading device Immediate input High level of accuracy Often wireless
Optical reader			
Barcode reader			
RFID reader			
Magnetic stripe card reader			
MICR reader			
Sensor			
Data-collection device	Manual or electronic input of data	Device can go to the location of data being input and use one of several input devices to send the data to the main database wirelessly.	Combination of different acquisition and input devices Immediate input Wireless Portable (hand-held)

Shutterstock.com/Ruslan Kuzmenkov

FIGURE 8.46 A barcode reader

Shutterstock.com/Voznikevich Konstantin

FIGURE 8.47 A data collection device that reads RFIDs and can communicate via wi-fi, Bluetooth and mobile phone network.



courtesy of EDX Wireless

FIGURE 8.48 This diagram represents a mesh network used to collect electricity usage data via Smart Meters. AMI stands for Advanced Metering Infrastructure.

Storage devices

You will remember that storage is the permanent saving of data for later retrieval. Storage hardware that is suitable for databases is the same as for any electronic files. These include internal or external hard disk and solid state drives, flash memory drives such as SD cards and USB flash drives. A more permanent backup copy of a database may be saved onto optical media like DVD or magnetic tape. Some databases may be stored in the cloud, but this is usually a massive data store with many hard drives connected using **RAID (redundant array of inexpensive disks)** to minimise data loss if there is some kind of threat, whether accidental or deliberate.

Getty Images/Erik Isakson

FIGURE 8.49 Inside a typical cloud-storage data centre

Communication devices

Databases can be transferred across and accessed over almost any networking device. You will recall from Chapter 2 that common networking hardware consists of network cards, wireless access points, routers, switches and cables. Data to be input might be transferred across any number of these devices if it is being entered remotely across the internet or an organisation's own network.

An important consideration is to ensure that any data is protected from interception and viewing by unauthorised users by using encryption.

Encryption is further discussed on page 334.

Output devices

Output is data that has been manipulated into a useful and meaningful form. It is also referred to as information. Information generated from a database is in the form of a report. The two most relevant output devices for information contained in a database would be via a screen (softcopy) or a printer (hardcopy). Information displayed on a screen is referred to as a softcopy because it is temporary and only exists electronically. The layout of the information is designed to suit the particular display device. For example, the format for Google search results will look different on a mobile device when compared with a large desktop screen.

In contrast, printed material is more permanent and exists physically, which is why it is called hardcopy. We might commonly also refer to hardcopy as a printout. The orientation of the hardcopy depends on the media being used and the layout of the information. Where possible, reports should be printed in a portrait orientation as this is generally easier to read.

Device	Role	Function	Characteristics
Screen (softcopy)	Displays information so it can be read and interpreted	Information is displayed on an electronic device	Temporary Only electronic
Print (hardcopy)		Information is printed onto physical media like paper.	More permanent Exists physically Can use a range of media

CASE STUDY

KALUGANGA JUNIOR SWIMMING CLUB – PART 8

Saro has investigated the hardware components that she thinks will be required to successfully allow input, output, communication and storage of the swim club's database.

For input she suggests that a mouse and keyboard will be adequate for entering swimmers' details and for updating financial information. The coaches will use a tablet with data collection facility and touchscreen so that times for events can be entered immediately. The tablets will communicate wirelessly with the database. The database itself will be stored on a hard drive on the club's server, but will also have copies saved in real time to a cloud-based storage service. Felix and the treasurer will be able to view swimmer details onscreen, although they can also print the report of overdue fees if required. The coaches can also view the list of swimmers by event and time on the tablets, but they are also able to receive a printed version of the report for when it is impractical for them to use the tablets.

Accidental and deliberate security threats

Once our database is constructed and in use, it is important that it be safeguarded. If the data is altered or damaged by someone who is not authorised to do so, it can easily be compromised. This will lead to inaccurate information being produced for users. Most of the key types of threats to data were explained in Chapter 2. The table below summarises these threats.

TABLE 8.8 Types of threats

Category of threat	Types of threat	Explanation
Accidental	Data altered by accident	An inexperienced or careless user may change or delete data. This creates inaccuracies.
	Stored files are lost	Database objects or data are stored on a portable device, like USB flash memory, and this device is misplaced.
	Not deleting data from an old storage device	A storage device that may have stored database records may be replaced by a newer device, but if the files are not removed then they could be recovered and accessed by whoever the storage device is passed on to.
Deliberate	Hackers	Some people may target a database in order to alter or delete its contents. A common method of attacking a database is by using SQL Injection.
	Malware	These types of malicious software may be designed to automatically attack a specific database, like your email contact list, or may damage database files as a by-product of targeting any files on storage devices
Event based	Power related	In these event-based threats, the results can be the same. Data may be lost or corrupted or the entire file might be destroyed.
	Crash or failure of a storage device	
	Remote connection fails	If a transaction, say from a web-based form, is interrupted, it may result in an incomplete record being saved.

SQL injection is a way of attacking databases via their input forms. Malicious SQL statements are inserted in data-entry fields in order to target and damage underlying data tables or steal their contents. Proper validation on any input box can prevent this form of attack on data.

CASE STUDY

KALUGANGA JUNIOR SWIMMING CLUB – PART 9

Saro has considered the likely threats the swim club database may face. She has assessed that the overall threat level from hackers is fairly low, but there is a larger possibility of data loss occurring by accident, as Felix is less familiar with how to operate a database package. She also thinks it is important to consider the possibility of malware infecting the server. The other key threats are the possibility of a blackout or the failure of the storage device, which could lead to lost data.

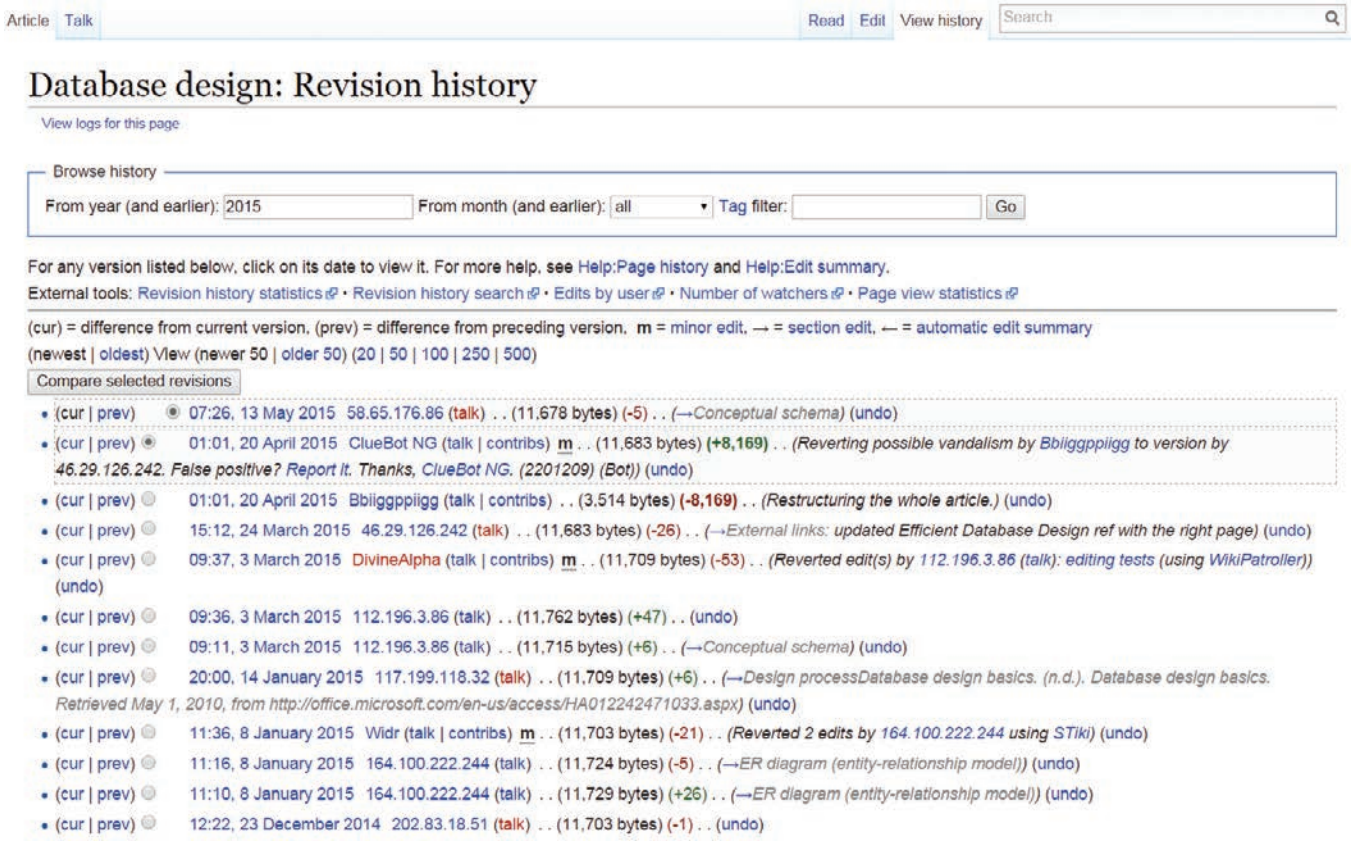
Physical and software controls for protecting security

All data in a database is valuable – we need to protect the integrity of the data so it can produce the most accurate information. A database may hold confidential data that must only be accessed by authorised users. Most government databases fall into this category, but this is crucial for businesses too as they store confidential information such as business contacts or customer details.

In Chapter 1, you were introduced to several physical and software-based controls that can protect data both when stored and transmitted. These measures help to reduce or eliminate the possible damage caused by the threats discussed above. These measures include backing up, electrical protection, usernames and passwords, using system security software and encryption.

Backing up

Databases are generally 'live', but they need the ability to have data 'rolled back' to an earlier time in case of data corruption or malicious activity. Full copies of databases can be made onto any of the storage devices listed above. This means the entire database can be restored to a previous point in time if necessary. Alternatively, many databases support version control, meaning they actually store changes to each data element in any record so that, if necessary, only that particular field is restored. A wiki site like Wikipedia is a good example of an implementation of this type of database 'protection'.



https://en.wikipedia.org/w/index.php?title=Database_design&action=history

FIGURE 8.50 Wikipedia stores backup copies of articles and changes made to them. This allows editorial staff to roll back to a previous version of an article without having to rewrite it.

Electrical protection

A common way of protecting against power events is to use a well regulated power supply that can intercept power surges or spikes. But a database may need to remain accessible even under adverse power conditions. For this reason, backup power supplies and batteries are often used. An Uninterruptable Power Supply or UPS can regulate electricity coming into a computer and supply backup electricity for a short time if a blackout occurs.

Usernames and passwords

Usernames and passwords can be used to restrict access to the entire database, so that only authorised users can open it. Alternatively, usernames and passwords allow certain users to edit records, while others can only read, but not alter them. These types of databases will also make a record of who changes what data as an auditing feature.

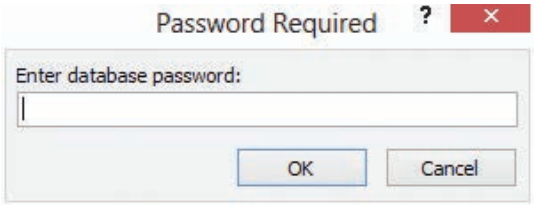


FIGURE 8.51 A password can be used in Access to prevent unauthorised users from opening a database file.

Systems security software

Malware is discussed in Chapter 3.

It is important to use a variety of systems security software to prevent malware or other malicious coding from running on your system and either damaging or destroying the contents of a database.

Encryption (storage and transmitted)

Encryption is used to prevent interception and theft. It involves coding data into a form that only authorised users can read and decrypt. The encryption and decryption processes both require a 'key'.

If data is entered via a website, then it will use SSL or TLS over an HTTPS connection. When SSL or TLS protocols are used over HTTPS (on port 443), an encrypted and secure 'pipeline' is created between the user and the database so that if the data being transmitted is intercepted, it cannot be read without the key. You will often notice that online application forms, registration forms and order forms use this type of encryption, as do internet transactions involving credit card details.

When the plain text is encrypted, it is called 'ciphertext'.

When stored, databases can also be encrypted. The contents can be encrypted inside fields. This is known as '**hashing**' and is often used to protect passwords. Hashing passwords is a way of taking a variable-length password and creating a cryptic, fixed-length password from it. You do this by generating and using a **salt value**. A salt value is a random value that you use to generate the hashed password. If the password field is viewed by a user, they cannot decrypt the hashed password.

The entire database file can also be encrypted so that if stolen, its contents cannot be opened and used by another person.

Encryption involves coding data into a form that only authorised users can read and decrypt. The encryption and decryption process both require a 'key'. When the plain text is encrypted, it is called 'ciphertext'. When SSL or TLS protocols are used over HTTPS (on port 443), an encrypted and secure 'pipeline' is created between the user and the database so that if the data being transmitted is intercepted, it cannot be read without the key. This system is commonly used for internet transactions involving registration or order forms, and for credit card details.

CASE STUDY

KALUGANGA JUNIOR SWIMMING CLUB – PART 10

For the Kaluganga database, Saro has limited access to personal and financial details to authorised users by placing a password and username on the files. The coaches can view basic details about each swimmer but can only make changes to their times for events. Any changes to data are logged. Anti-malware software will be installed on the tablets, PC and server while an encrypted connection will be used to transmit data between the server and tablets. The cloud storage service also acts as a backup and it keeps multiple versions of the database. A portable external hard drive is also used as a secondary backup of the database each night.

CHAPTER SUMMARY

ESSENTIAL TERMS

data dictionary a set of information used to design the structure of a database table; it includes field names, data types and size as well as validation rules

data redundancy where the same piece of data, a customer's address for example, is stored in two or more places in a database. This can waste storage space if the data is merely the same data stored elsewhere, but can also be useful if it is deliberately saved as a form of backup

database software designed to store data and allow users to search and select from it in order to extract information

database management system (DBMS) software designed to store data and allow users to search and select from it in order to extract the information they need

hashing a way of taking a variable-length password and creating a cryptic, fixed-length password from it

input form a form that allows user to enter data into a database via a more user friendly interface than a table

layout diagram a hand-drawn sketch that shows the elements to be included on an input form; The diagram would indicate the placement of fields and labels, the fonts to be used, and any graphics or other elements to be included

metadata data about data; in a DBMS context it consists of details about the size and type of each field within a database file

primary key a field attached to each record of a relational database; this is a unique identifier for each record, such as a customer ID number or phone number

query to select specific data based on a series of criteria in order to answer questions and make links between data; the criteria are the results of questions about the data; for example, 'How many of our customers are female?' or 'Can I have a list of names and mobile numbers for all customers aged 20 to 40, male, who live in Fairfield or Northcote?'

RAID (redundant array of inexpensive disks) banks of hard drives that are configured in such a way that if one or two fail the data they contain isn't lost. Instead the remaining disks can rebuild the missing data in their own storage areas.

salt value a random value that you use to generate the hashed password

SQL (structured query language) a language used by many database packages to provide a consistent way of selecting required data from a table

IMPORTANT FACTS

1. The problem-solving methodology (PSM) involves four stages: Analysis, design, development and evaluation.
2. DBMSs are used in a range of contexts including banking, membership, online purchasing and in voting systems.
3. Reasons for using a DBMS for data handling include reduction of data redundancy, sharing of data, maintenance of data integrity, increased data security, easy updating of data, the ability to select data (with a query), use of forms for data input, creation of structured reports, concurrency of data and data independence.
4. Limitations of a DBMS include expensive equipment, potentially costly conversion from other file types, staff training, the need for specialist database administrators and the impact across an organisation if data is corrupted.
5. Benefits of using a DBMS include improvements in efficiency and effectiveness.
6. Risks associated with using a DBMS include inaccurate data leading to inaccurate information being produced, unauthorised users accessing data (privacy and security); they are time consuming to design properly, the possibility of accidental damage to data by untrained users, cost of purchasing suitable equipment.
7. A field contains the same type of data for a series of records. The same field for a series of records will contain the same type of data.
8. A database record is a set of data about one entity; for example, a person, event or object.
9. A database table stores sets of records.
10. A primary key uniquely identifies each record in a database table.
11. Records that meet specified criteria can be selected from the total number of records by performing a query. Each criterion can be a number, a piece of text or an expression.
12. Reports format and present the data selected by a query as usable information.

CHAPTER SUMMARY

13 Common data types used in a DBMS include text, numeric, date, character and Boolean.
14 To gather data effectively, key characteristics of forms include validation, a logical arrangement of fields, uncluttered layout, and adaptive to what users select.
15 Use a consistent naming convention for DBMS objects to make it clear what they are (for tables and queries) and what they are part of (for fields). Examples are tblCustomers for a table, or qryAllProductsSoldToday. Fields should begin with letters that indicate the object they come from, such as cusFamilyName.
16 Common design tools used with databases include data dictionaries, data structure diagrams, data structure charts, flowcharts, input–process–output charts and layout diagrams.
17 A query design includes fields and criteria in order to select only the data required.
18 Reports are designed with a more structured style of layout diagram.
19 Macros can be designed with either pseudocode (structured English) or a flowchart.
20 Input–process–output (IPO) charts are used to identify inputs, outputs, and the processing steps required to transform the inputs into the outputs.
21 A list tool can be used to direct the entry of data where there is a limited number of allowed values. A list box on a form will display a predetermined set of values from which the user must select.
22 A range of devices is used to input, store, communicate and output data and information.
23 Input devices include a keyboard, mouse, optical scanner, optical reader, barcode reader, RFID reader, magnetic stripe card reader, MICR reader, sensor and data collection device.
24 Common storage devices include hard and solid state drives, flash memory devices, optical media and magnetic media. Data can also be stored in the cloud and for added security RAID can be used.
25 Communication devices include network interface cards, wireless access points, routers, switches and cables.
26 Types of output include screen (softcopy) and print (hardcopy).
27 Threats to databases can be classified as accidental, deliberate or event based.
28 Accidental threats include data that is altered by accident, stored files that are lost, and data not being deleted from old storage devices.
29 Deliberate threats to data include hackers and malware.
30 Hackers may wish to damage or steal data in a database.
31 Event-based threats to databases include the crash or failure of a storage device, failure of remote connection and power loss or surge.
32 Organisations use a range of software and physical controls to protect their data. These controls include backing up, providing electrical protection, using usernames and passwords, installing system security software and using encryption.

TEST YOUR KNOWLEDGE

Review quiz

1 Suggest a specific application/use for a database in each of the following contexts.
   - **a** Banking
   - **b** Membership of a club or organisation
   - **c** Online purchasing
   - **d** Voting systems

2 In a database context, what is metadata?

3 List three capabilities of a DBMS.

4 List three limitations of a DBMS.

5 What are three benefits of using a DBMS?

6 What are three risks associated with using a DBMS?

7 Define each of the following DBMS terms.
   - **a** Field
   - **b** Table
   - **c** Primary key
   - **d** Input form
   - **e** Query
   - **f** SQL
   - **g** Report
   - **h** Macro
   - **i** File

8 There are different data types that can be used in a DBMS. In the table below, write the most appropriate data type and an example of data beside each field.

Field	Data type	Example
Company name		
Australian postcode		
Mobile phone number		
Height in cm		
Preferred pizza size: S, M or L		
Would you like to be sent our newsletter?		
Date of dispatch		

9 Explain how the three validation checks could be used on a field that records people's weight in kilograms.

10 What are two characteristics of an effective input form?

11 What does the term 'user experience' mean?

12 Explain how the file-naming convention for database objects works.

13 How does a data dictionary assist in the design of a database?

14 How does the purpose of a layout diagram differ from a data dictionary?

15 Recall the purpose of query criteria.

16 Provide reasons why it is important to sort information in a report.

17 How do summary totals in reports assist with effectiveness of a solution?

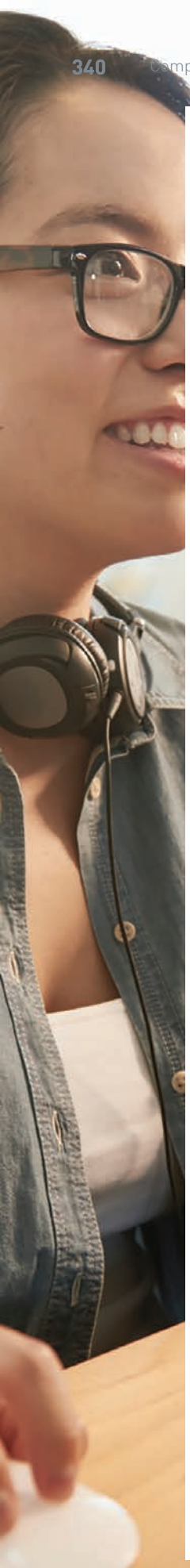
18 Recall the design tools that are commonly used to plan a macro.

19 State the role of input devices.

20 Contrast data input by a keyboard to that of a barcode reader.

21 What is an advantage of the use of a wireless data collection device in a supermarket?

22 List three differences between internal solid state drives and removable flash drives.
23 How can RAID help to protect stored data?
24 Why is communication of data critical for many databases?
25 What is the difference between a softcopy and a hardcopy? Provide an example of where each may be more appropriate.
26 How does a deliberate threat to data differ from an accidental threat?
27 Describe what is understood as an event-based threat to data.
28 How can backing up help to protect the security of data?
29 What types of electrical protection can be provided by a UPS?
30 Why is a combination of username and password used to protect a database, rather than an individual username by itself?
31 Recall the definition of malware.
32 How does encryption secure a data connection between a user inputting data and the database itself?



APPLY YOUR KNOWLEDGE

DATABASE ACTIVITY

Catherine works as a junior programmer for a small software company. She currently earns $2500 a month and pays $300 in rent. She has many other monthly expenses. Sometimes she has overspent and run out of money before her next pay period. This has led to embarrassment when she has been unable to buy food for herself or her cat at the local shop. In order to manage her budget better, Catherine has asked you to create a database in which she can store her bank transactions. At the end of each month she will use this database to analyse her spending and determine if she is in credit or debit. For the purposes of this task, she is working with the month of November 2016.

The two reports Catherine would like are as follows.

a A report that shows all the debits from her cheque account for the month of November 2016. It should be grouped by transaction category and sorted by transaction date within each category. Be sure to include summary statistics that show the percentage and subtotals for each group as well as an overall total so Catherine can see what money remained at the end of the month.

b A report that lists all expenses from all of Catherine's accounts for the 'Car/Travel' category for the month of November 2016. The report should be sorted by transaction date (from earliest to most recent) and then by amount (smallest to largest amount).

1 Table 8.9 shows the data set that Catherine has provided.

2 Looking carefully at the data, create a data dictionary to plan the database table. You may rearrange the data into different fields as you see fit. Also remember to include validation and to choose appropriate data types and formats.

3 Design a suitable data input form that will enable Catherine to add, review and edit her bank records. Be sure to make use of labels and dropdown lists to assist in more accurate data entry.

4 Create designs for the two queries that will be used to create the reports outlined above.

5 Design the two reports outlined above.

6 Design macros that will open, print and close each of the reports when they are run.

7 Using a DBMS package, create Catherine's bank transactions database. This will involve constructing the table, the input form, the two queries, two reports and the two macros you designed above. Remember to include the data validation and appropriate formulas and functions.

8 Create annotated screenshots of the various database objects you have created to show their key features and how you have met Catherine's needs.

9 In a paragraph, list the roles, functions and characteristics of suitable hardware components that Catherine could use to input, store, communicate and output the data and information in her database.

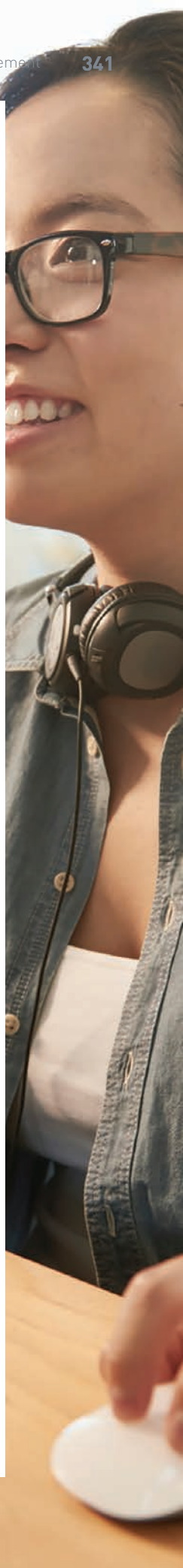
10 Identify one each of an accidental, deliberate and an event-based security threat to the data and information stored in Catherine's database.

11 Identify and explain suitable physical and software controls that will protect the data when it is stored in Catherine's database and when it is transmitted (both physically and electronically).

12 Evaluate the potential success of the DBMS you have created for Catherine by comparing relevant benefits and risks from its use.

TABLE 8.9 The table of data Catherine has supplied you with for her bank transaction database.

Transaction ID	Account (cheque or savings)	Date	Details	Category	Transaction type (debit or credit?)	Amount ($)
21	Cheque	01-Oct-16	Monthly salary	Salary	C	2500.00
30	Cheque	15-Oct-16	Rent	Miscellaneous	D	300.00
5	Cheque	15-Oct-16	Food shopping	Food/groceries	D	15.20
10	Cheque	01-Nov-16	Petrol for car	Car/Travel	D	34.56
14	Cheque	01-Nov-16	Movie at cinema	Entertainment	D	25.00
15	Cheque	01-Nov-16	Popcorn and ice-cream at cinema	Entertainment	D	15.00
6	Cheque	01-Nov-16	Food shopping	Food/Groceries	D	8.54
24	Cheque	01-Nov-16	Monthly salary	Salary	C	2500.00
22	Cheque	03-Nov-16	Mobile recharge	Miscellaneous	D	20.00
9	Cheque	05-Nov-16	Dinner with Nathan	Entertainment	D	125.87
7	Cheque	07-Nov-16	Food shopping	Food/groceries	D	185.21
34	Savings	08-Nov-16	Refund from tax dept	Salary	C	350.28
32	Savings	12-Nov-16	Birthday present and card for Mum	Miscellaneous	D	125.99
31	Cheque	15-Nov-16	Rent	Miscellaneous	D	300.00
11	Cheque	15-Nov-16	Petrol for car	Car/travel	D	56.28
36	Savings	15-Nov-16	Puncture repair	Car/travel	D	35.00
33	Cheque	15-Nov-16	MSO concert	Entertainment	D	85.00
2	Cheque	15-Nov-16	Car insurance	Car/travel	D	551.67
4	Cheque	17-Nov-16	Red shirt	Clothing	D	65.00
17	Cheque	17-Nov-16	Black pants	Clothing	D	35.50
1	Cheque	17-Nov-16	New shoes	Clothing	D	125.00
25	Cheque	18-Nov-16	Car service	Car/Travel	D	245.89
23	Cheque	18-Nov-16	Cat food	Miscellaneous	D	35.99
8	Cheque	20-Nov-16	Food shopping	Food/Groceries	D	57.66
16	Cheque	21-Nov-16	Placemats for table	Miscellaneous	D	30.00
12	Cheque	22-Nov-16	Petrol for car	Car/travel	D	72.45
29	Savings	22-Nov-16	Repair bike	Miscellaneous	D	65.80
28	Cheque	23-Nov-16	Movie and pizza with Bill	Entertainment	D	85.00
19	Cheque	25-Nov-16	Sleeping bag	Miscellaneous	D	156.00
20	Cheque	29-Nov-16	Sunscreen 50+	Miscellaneous	D	15.99
18	Cheque	29-Nov-16	Sun hat	Clothing	D	24.99
13	Cheque	01-Dec-16	Petrol for car	Car/travel	D	22.60
35	Cheque	01-Dec-16	Monthly salary	Salary	C	2500.00
3	Cheque	03-Dec-16	Food shopping	Food/groceries	D	34.55



AUTOMATIC NUMBERPLATE RECOGNITION

New South Wales launched an automatic numberplate recognition (ANPR) system in 2009. It can scan and check up to six numberplates a second from passing and oncoming cars. The system is typically mounted on police cars (see Figure 8.52). There is a terminal inside the car that displays information about cars that are scanned, in real time (see Figure 8.53) and alerts police if that car ought to be pulled over for further investigation. The numberplate details and the date, time and location of the scan for every car the ANPR system detects are transmitted immediately and stored in a police controlled database. This data is stored for a minimum of 5 years. Privacy advocates have raised concerns about the quantity of data being collected and the purposes for which it may be used, other than to help with highway patrols.

1. Identify one each of an input, storage, communication and output device that might be used in the ANPR system.
2. Suggest two reasons as to why privacy might be a concern with this system.
3. List some ways the data could be misused.
4. Suggest two threats that the ANPR database might face.
5. Explain a suitable control measure to protect the ANPR data once it is stored in the database and while it is transmitted to and from a police car.

Eastwood Highway Patrol, https://creativecommons.org/licenses/by/2.0/

FIGURE 8.52 Roof mounted lightbar and ANPR cameras on a NSW patrol car

Newspix/Tracee Lea

FIGURE 8.53 The ANPR terminal inside a police car

PREPARING FOR UNIT 2 OUTCOME 3

Apply the problem-solving methodology to create a solution using database management software, and explain the personal benefits and risks of interacting with a database

Devise a need or opportunity for a database solution, collect relevant data and create that solution using database management software (DBMS). The solution must make use of a table, input forms, queries and reports. Where appropriate, apply mathematical calculations to the data and create macros to automate repetitive tasks. Identify and explain suitable hardware that can be used with the database, as well as risks to data and information, and ways to minimise these risks.
You will apply all four stages of the PSM: Analysis, design, development and evaluation.

OUTCOME MILESTONES

1. Identify a need for a data-handling task that requires the use of a DBMS as its solution.
2. Prepare an analysis of the need, which includes an outline of the organisation, a list of functional and non-functional requirements of the solution, a list of constraints, and the scope of the solution.
3. Identify input data and information required to fulfil the need.
4. Design the solution.
5. Use relevant functions of DBMS software to develop the solution.
6. Use relevant software functions to demonstrate features of the software.
7. Identify relevant hardware components to input, store, communicate and output data and information.
8. Identify threats to data and information stored within databases.
9. Identify physical and software controls to protect data when stored and transmitted.
10. Identify risks and benefits associated with using database solutions.

STEPS TO FOLLOW

The analysis, design, development and evaluation stages of the problem-solving methodology will be used to create the solution, although you will not be required to create any documentation nor conduct any testing.

1. You must identify a user or organisation where a need for a DBMS exists. The user can be yourself and the database need may be as simple as recording your music collection or sports club statistics.
2. Analyse and define the need. In the analysis clearly identify:
   - **a** the solution requirements (needs of the users and problem to be solved, functional and non-functional requirements)
   - **b** the solution constraints
   - **c** the scope of the solution (solution boundaries and what it will and will not do).

The problem to be solved and the needs of the users are best written as simple statements; for example, 'Sometimes the lists of swimmers are inaccurate because they are mistakenly put into the wrong age group or left off the list totally! In addition, some financial problems have occurred for the club as they have not accurately followed up which members have not paid their annual fees. This also leads to embarrassment for club officials when payment notices are sent to the wrong club members'.

3 Design the solution. The following are required.
   - **a** Use of a data dictionary to represent the table plan
   - **b** Layout diagrams to plan the forms
   - **c** Query plans to design the queries
   - **d** Layout diagrams for the reports to be generated
   - **e** Pseudocode or flowcharts for any macros used

4 Develop the solution. For a DBMS-based solution, ensure that you have created all the objects required and named them correctly.

5 Validate the data and input it.

6 Generate the required information from the DBMS solution (the reports).

7 Evaluate the solution by commenting on the following.
   - **a** Appropriate hardware for input, storage, communication and output
   - **b** Accidental and deliberate threats relevant to the database
   - **c** Relevant physical and software controls to protect the stored or transmitted data
   - **d** Identification of benefits or risks for the user arising from the use of this database

DOCUMENTS REQUIRED FOR ASSESSMENT

1 Documentation produced during the analysis stage, including the solution requirements, constraints and scope of the solution

2 A representation of the design of each part of the solution

3 Printouts of the solution and information

4 Annotated copies of the solution and information indicating features you have used and conventions you have followed and how it meets the needs of the users

5 Documentation to evaluate the hardware, threats, security controls, risks and benefits associated with the database

ASSESSMENT

A set of assessment criteria will be prepared and distributed by your teacher before you start the task. Details on the actual criteria for this Outcome and the breakdown of how each criterion will be marked are available at http://computing1and2.nelsonnet.com.au.

INDEX

802.11 standard 103–4
accessibility (design principles for graphic solutions) 52–3, 184–7, 207, 236
accidental threats (network security) 116–17, 332
accuracy of data 9, 158, 268
affordance
  mobile devices 176
  user interface 236
agriculture and information systems 147–8
algorithms 231, 245
alignment (design principles for graphic solutions) 53, 188–9
alpha testing 248
American Psychological Association (APA) 13–14
analysis (problem-solving methodology) xii
animations (graphic solutions) 46–7
annotated diagrams (graphic solutions) 66, 67–8, 198–9
appearance (design principles for graphic solutions) 53–6, 187–94, 198–204
  alignment 53, 188–9
  annotated diagrams and mock ups 198–9
  balance 56, 193–4
  contrast 189–92
  conventions and formats 201–4
  flowchart 182
  functionality 182–7
  layout design 198–9
  repetition 189
  space 192–3
  test data 200
application-programming interface (API) tool 289
apps 105
arrays 220, 250–2, 270
arrows (graphic solutions) 59–60
asymmetric digital subscriber line (ADSL) 92
audio file 203
Australian Bureau of Statistics (ABS) 266
Australian Privacy Principles (APPs) 22–4
authenticity of primary data 269
automatic numberplate recognition 342

backups 20, 333
balance (design principles for graphic solutions) 56, 193–4
bandwidth 109
battery life (mobile devices) 161
beta testing 248
bias (information) 9–11
  through graphic representations 11
  through sorting 10
binary search 255
bluetooth 113
Boolean data type 270, 308
bottom-driven loops 243
brainstorming 170
broadband routers 91–2
bulletin boards 153
Bureau of Meteorology 265–6
buttons (design principles for graphic solutions) 205

card skimming 149, 150
cellular radio 114–15
central processor unit (CPU) 220, 223
channel 109
character (data type) 269, 308
charts
  as data visualisations 273
  graphic solutions 43–5
choices (information architecture) 179–80
classification (information architecture) 178
client-server networks 84–5
closed questions 7, 8
cloud computing 21, 159–60
cloud storage 100, 159–60
code identification 243
code modules 240–1
collaboration 168
colours (graphic solutions) 62–5, 190–2
communication of message (graphic solutions) 69
communications, facilitating network 125
communications channel 109–11
communications devices 81, 90–6, 331
  broadband routers 91–2
  network interface cards 93–4
  NBN devices 92–3
  routers 90–1
  switches 90
  wireless access point 94
  wireless adaptors 93–4
  wireless extender 95–6
communication hardware 221
communications satellite 115–16
communications software 96–8
compiler 223
completeness testing (graphic solutions) 68
complex instruction set computing (CICS) 220
component testing 248
compression 227
consent forms 15–16
contingency plans (project management) 168, 169
contrast (design principles for graphic solutions) 53–4, 62–5, 189–92
conventions (design principles for graphic solutions) 56–65, 201–3, 281–2

file naming 201–2
images 202–3
index page 202
style guides 203
video and audio file 203
Copyright Act 1968 (Cth) 89–90, 141
costing (project management) 169
counted loops 243
CSV file formats 289

data
accuracy 9, 158, 268
acquisition 308–9
correctness 159
dummy 206
entry 309–12
gathering 5–9
and information 4–5
integrity 151, 158–9, 267–9
manipulation (design tools) 196–7
primary 5, 155–7
processing data to create solutions 70–3
qualitative 4, 155–7
quality of 9–11
quantitative 3–4, 155–7
reasonableness 159
redundancy 302
sample size 10
secondary 5, 155, 157–8
security practices 123–4
sharing (networks) 125
sources 5–9, 264–7, 308–9
structures 269–71
test (problem-solving) 200
transmission cost 161
transmission speed 161
data analysis 2–25, 263–92
Australian privacy principles 22–4
design tools 66–8, 279–80
ethical dilemmas 24–5
file formats 288–9
formats and conventions 56–66, 281–2
gathering data 5–9
graphic solutions 42–51
graphic solutions, design principles 51–6
physical security controls 20–2
and privacy 19
problem solving 31–42
process data to create solutions 70–3
referencing primary sources 11–14
and research 3–4
seeking permission 15–18
software tools and functions 282–7
sources 5–9, 264–7
types of tests 68–70
validation 70
data collection
quality 9–11
seeking permission 15–18
techniques and methods 5–7
timing 10
tools 310–12
data dictionaries 229–31, 302, 312–13
data management 300–35
accidental and deliberate security threats 116–17, 332
communication devices 331
hardware components 329–30
physical and software security 333–5
data structure diagrams 231
data structures 269–77
data types 225–6, 269–77, 309
array 270
Boolean 270, 308
characteristics 269, 307–8
common types 307
file 271
floating point (or decimal) 270, 307
integer 270, 307
record 270
string (or text) 270, 307
data validation (programming concepts) 241–2
data visualisations 263–4
charts 273
common types 272
definition 271
evaluating 289–92
flow 276
hierarchy 275–6
maps 274
matrix 276–7
networks 274
time series 275
types and purposes 271–7
data.gov.au 267
database design tools 312–20
and database structure 312–14
designing macros 320
designing queries 316–17
input forms to capture data 315–16
layout diagrams 312
naming conventions 312
reports for specific needs 318–19
database management systems (DBMSs) 301, 302–3, 307
benefits and risk 303
capabilities and features 302
development of 320–8
limitations 302
database software 284–5, 311
database structure 303–6
fields, records and tables 303–4
input forms 304
macros 306

queries 305
reports 306
database systems 301–2
date (data format) 308
debugger 223
debugging (programming concepts) 244–8
desk check 246–8
logical errors 245
runtime errors 246
syntax errors 245
dedicated servers 84
default action 239
default values 239
deliberate threats (network security) 117
design
problem-solving methodology xii
website 171–6
design principles for graphic solutions 52–6, 181–94
accessibility 52–3, 184–7, 207
alignment 188–9
appearance 53–6, 187–94
balance 193–4
contrast 189–92
flowchart 182
functionality 52–3, 182–7
orientation 194–5
proportion 194
repetition 189
space 192–3
specific considerations 194–5
useability 52, 182–4, 207
design tools (graphic solutions) 66–8, 195–204
appearance 198–204
functionality 195–8
design tools (problem-solving methodology) 66–8, 279–81
designing solutions and output 33–4
designing websites 171–6
mobile devices 172–6
principles 181–94
specific considerations 194–5
tools 195–205
designs, network physical 121–2
desk check 246–8
developing websites 205–7
development (problem-solving methodology) xii
diagrams
graphic solutions 45–6
network 121–2
differential backup 20
digital information systems 146
digital signal 81
disclosure (information architecture) 177–8
domain name server (DNS) 98
downlink 116
dummy data 206

Earth-based station 115–16
e-commerce 148–9
editor 223, 224
electrical protection 334
electronic funds transfer (EFT) 125
electronic health records (EHR) 151
electronic validation (graphic solutions) 68, 70
email 99
encryption 20, 334–5
entertainment and information systems 140–4
Ethernet 101–2
ethical dilemmas
data analysis 24–5
information systems 142–4
programming 255
ethical responsibilities (networks) 122–6
evaluation (problem-solving methodology) xii
existence checking 309

Facebook 146
fibre-optic cable 111
file data structure 271
file formats (data analysis) 288–9
file naming 201–1
file transfer protocol (FTP) 99
files and records (programming concepts) 252–3
finance and information systems 148–51
firewalls 21–2, 120
first-in first-out (FIFO) 252
first-in last-out (FILO) 252
floating point (or decimal) (data type) 270
flow visualisations 276
flowcharts (graphic solutions) 45, 196–7, 280–1
focus groups 7
font selection (website design) 174, 189
footnotes 12
formal testing 248
formats
data 2
design principles for graphic solutions 56–65, 281–2
forums 146, 153
full backup 20
functional requirements (problem-solving methodology) 277
functionality (design principles for graphic solutions) 52–3, 196–8
functionality testing (graphic solutions) 69

games consoles 108–9
Gantt charts 170–1
garbage-in equals garbage-out (GIGO) 303
geospatial visualisations 274
GIS file formats 289
global positioning system (GPS) 106–7
Google 286–7

graphic solutions
annotated diagrams 66, 67–8, 198–9
appearance 198–204
audio file 203
contrast 53–4, 62–5
definition 43
design principles for 52–6
design tools 66–8, 195–204
flowcharts 45, 196–7
images 46, 200–3
layout designs 198–9
and processing data 70–3
purpose of 42–52
site map development 197
sources of data and legend 60–1
storyboards 197–8
test data 200
testing 68, 206–7
types of 43–51
types of tests 68–70
validation 68, 70, 206
video file 203
graphic user interface (GUI) 224
controls and structures 253–4
graphics processor unit (GPU) 220
groupware 171
growth (information architecture) 179

hacker 117
hand-held data-collection devices 106
hard disk drive (HHD) 219, 220
hardware
communication 221
components for data management 329–30
processing 220
and programming 219–21
sharing (networks) 125
storage 220–1
Harvard referencing style 158
hashing 334
health and information systems 151–2
Health Records Act 2001 (Vic) 151–2
hierarchies (graphic solutions) 46
hierarchy visualisations 275–6
histograms 157
home networks 83–4
HTML 82
HTTPS 334
Hungarian notation 230
hyperlinks 205
hypertext transfer protocol (http) 98
hypertext transfer protocol secured (https)–9 98
hypothesis 3

identity theft 149–50
image editing 205
images (graphic solutions) 46, 200–3
incremental backup 20
index page 202
infographics (graphic solutions) 47–51
information 4–5
quality of 9–11
information architecture 177–80
choices 179–80
classification 178
disclosure 177–8
growth 179
navigation 178–9
information needs 263–4
information systems 140–61
in action 140–52
acquiring data, methods and techniques 155–8
agriculture 147–8
data integrity 158–9
digital 146
definition 140
and entertainment 140–4
expressing opinions 153–4
and finance 148–51
forums and bulletin boards 153
and health 151–2
mobile devices 160–1
primary data collection methods 155–8
in programming 219
rating systems 153–4
secondary sources 157–8
social media 153
storing shared file 159–60
and sport 145–7
web design 160–1
informed consent 15
input devices 219, 329
input forms 304, 315–16, 320
input–process–output (IPO) charts 66–7, 196, 231–2, 280
integer (data type) 270, 307
Integrated Development Environment (IDE) 224
integration testing 248
integrity of data 151, 158–9, 267–9
interface mock-up 234
interfaces, user 309–12
interfaces, creating effective user 235–9
accessibility 236
affordance 236
consistency 239
legibility 238
structure 237
tolerance 239–40
useability 235–6
visibility 237–8

internal documentation (programming concepts) 242
internet peer-to-peer networks (P2P) 86
internet service software 98
internet services 98–100
  cloud storage 100
  email 99
  file transfer protocol (FTP) 99
  voice over Internet Protocol (VoIP) 100
  web browsers 98–9
interviews 7, 11–12, 155
intranets 82–3
IP addresses 103

Kaluganga Junior Swimming Club (case study) 311–12, 313–14, 315–16, 317–18, 319–20, 321–8, 331, 335
key concepts xiv
keylogger 118

languages, compiled and interpreted (programming concepts) 240
last-in first-out (LIFO) 252
layout designs 199–200
layout diagram
  database design tools 312
  problem-solving methodology 279–80
leecher 88
legal responsibilities (networks) 122–3
legibility
  user interface 238
  website design 174
linear search 255
lines (graphic solutions) 59–60
linker 223
local area networks (LAN) 82, 85, 90–1
logic (programming concepts) 244
logic bombs 118
logical errors 245
loops (programming concepts) 242–4

macros 32, 306, 320, 321
malware 117
malware protection 22
manipulation (graphic solutions) 205–6
manual validation (graphic solutions) 70
map-based visualisations 274
maps (graphic solutions) 47
matrix visualisations 276–7
Mbps 90
media data types 226
meta tags 205
metadata 302
microwaves 115
microwave station 115
mobile devices
  accessibility 173–4
  affordance 176
  connected to networks 104–9
  consistency 176
  definition 104
  games consoles 108–9
  hand-held data-collection devices 106
  and information systems 160–1
  interface design 172–3
  legibility 174
  navigation systems 105–6
  network-attached storage device 109
  read tap asymmetry 173
  smartphones 105
  tablets 104–5
  tolerance 174
  touch zone 173
  useability 173
  visibility 175
  wearable technology 107–8
mobile phone 114
mock-up 234
modular programming 240–1

National Broadband Network (NBN) 111–12, 168
  devices 92–3
navigation (information architecture) 178–9
navigation systems 105–6
near field communication (NFC) 113–14
network administrator 85
network analysis tools 97–8
network communications standards 101–4
  802.11 standard 103–4
  Ethernet 101
  sending and receiving devices 104
  TCP/IP 102–3
network diagrams 121–2
network interface cards (NIC) 93–4
network operating system 96–7
network protocol 101
network security 116–21
  accidental threats 116–17
  deliberate threats 117–18
  event-based threats 118
  firewalls 120
  measures 118–21
  passwords 21, 119–20, 334
  security threats 116–18
  uninterruptible power supplies 120–1
  usernames 21, 119–20, 334
  wireless security 121
network visualisations 274
network-attached storage device (NAS) 109, 220, 221
network(s) 81–125
  architecture 82, 84–90

benefits 124–5
client-server networks 84–5
communications channel 109–11
communications devices 81, 90–6
communications software 96–8
definition 81
ethical responsibilities 122–6
home networks 83–4
internet peer-to-peer networks (P2P) 86
internet services 98–100
intranets 82–3
legal responsibilities 122–6
local area networks (LAN) 82, 85
and mobile devices 104–9
National Broadband Network (NBN) 111–12
network communications standards 101–4
peer-to-peer networks (P2P) 85–6
personal security strategies 124
physical designs 121–2
risks 124, 126
responsibilities of users 124
security 116–18, 332
security measures 118–21
security practices 123–4
social 126
types of 82–4
users, responsibilities of 124
virtual private networks (VPN) 86–90
wide area networks (WAN) 84
wireless transmission media 112–16
networking software 86
node 82
non-functional requirements (problem-solving methodology) 277
non-technical constraints (problem-solving methodology) 278
numeric (data format) 270, 307

object description 232
object naming 230
observations 9, 12, 155
one-dimensional (1D) array 251
open-ended questions 8
OpenHeatMap 287
operating system (OS) 222
orientation (website design) 194–5
output devices 219, 331

P2P downloads: torrents 87–9
participation information statements 16–18
passwords 21, 119–20, 334
peer-to-peer networks (P2P) 85–6, 141
peers 88
permission, seeking 15–18
personal digital assistants (PDAs) 104
personal information, use and disclosure of 24
phishing 118, 150
physical security controls (data) 20–2
physical transmission media 109, 110–11
platform 224
ports 221
presentation testing (graphic solutions) 68–9
primary data 5, 155–7
primary data sources 264
primary key 304
primary storage 220
privacy 19
Australian principles 22–4
Privacy Act 1988 (Cth) 22–3
Privacy and Data Protection Act 2014 (Cth) 22
problem analysis 32–3
problem solving
approaches to 31–42, 168
design principles 181–94
design tools 195–204
design websites 171–6
designing solutions and output 33–42
developing websites 205–7
four stages of 31
functions of a spreadsheet 31–2
information architecture 177–80
project management techniques 168–71
specific design consideration 194–5
team solutions 168
problem-solving methodology (PSM) xii–xiii, 31, 196, 205–7, 219, 263
analysis 277–9
design tools 66–8, 279–81
evaluation criteria 289–91
evaluation report 291–2
evaluation strategy 291
evaluating data visualisations 289–92
formats and conventions 56–66, 281–2
graphic solutions 42–51
graphic solutions, design principles 51–6
presentation 207
problem solving 31–42
process data to create solutions 70–3
scope of solution 278–9
software development 228–35
solution constraints 278
solution requirements 277
types of tests 68–70
validation 70
process testing (programming concepts) 248–9
processing hardware 220
programming 218–55
creating effective user interfaces 235–9
developing software 228–39
hardware 219–21
information systems 219
modular 240–1

operating system 222
and scripting languages 222–3
software 222
software development tools 223–5
storage structures 225–8
programming, fundamental concepts 240–55
arrays 250–2
compiled and interpreted languages 240
data validation 241–2
debugging 244–8
files and records 252–3
GUI controls and structures 253–4
internal documentation 242
logic 244
loops 242–4
modular 240–1
process testing 248–9
queues 252
and scripting languages 222–3
searching 255
stacks 252
programming languages 285–6
project management
definition 168
techniques 168–71
tools 170
project table 170
proportion (website design) 194
protocol 87, 101
prototypes 223
pseudocode 232–4
rules of 233–4
public key encryption 233

qualitative data 4, 156–7
quality control (project management) 169
quantitative data 3–4, 156–7
queries 305, 316–18, 321
query criteria 317
questionnaires 5, 7, 12
questions
closed 7
graphic solutions 70
open-ended 8
queues (programming concepts) 252

RAID (redundant array of inexpensive disks) 330
random-access memory (RAM) 220
random files 253
ratings systems 153–4
read tap asymmetry (mobile devices) 173
receiving device 81
record (data type) 270
reduced instruction set computing (RICS) 220
redundant array of inexpensive disks (RAID) 330
referencing
examples 12–14
primary sources 11–14
relevance of data 9
relevance of primary data 269
relevance testing (graphic solutions) 69
reliability 11
reliability testing (graphic solutions) 68
repetition (design principles for graphic solutions) 53
reports (data management) 306, 318–19, 321
research
types of 3–4
understanding 3–4
resolution 227
retrieval 206
routers 90–1
runtime errors 246

salt value 334
sample size 10
screen resolution (mobile devices) 161
scripting languages and programming 222–3
searching (programming concepts) 255
secondary data 5, 155, 157–8
secondary data sources 264–5
secondary storage 220–1
security (networks) 116–18
breaches 126
measures 118–21
personal strategies 124
physical and software 333–5
practices 123–4
system software 334
seeds 88
sending and receiving devices 104
sending device 81
sequential files 252
shapes (graphic solutions) 58
skeuomorphism 176
smartphones 105
social media 153–4
social networks 126–7
software 222
database 284–5, 311
spreadsheet 286–7
software development 228–35
data dictionaries 229–31
data structure diagrams 231
functional requirements 228
input–process–output (IPO) charts 231–2
interface mock-up 234–40
non-functional requirements 228–9
object description 232
pseudocode 232–4

PSM stage: analysis 228–9
PSM stage: design 229–35
tools 223–5
software tools and functions (data analysis) 282–7
database software 284–5
extracting data 282–6
Google 286
OpenHeatMap 287
programming languages 285–6
spreadsheet software 286–7
Tableau Public 287
solid state drives (SSD) 220
solution constraints (problem-solving methodology) 278
solution requirements (problem-solving methodology) 277
source code 223
sources (data analysis) 5–9
sources of data and legend (graphic solutions) 60–1
sound editing 205
space (design principles for graphic solutions) 54–5, 190–2
sport and information systems 145–6
spreadsheet functions 31–2
spreadsheet software 286–7
storage devices (data management) 330
structure query language (SQL) 302
spywear 117
stacks (programming concepts) 252
stakeholders 5, 140, 141–2, 146–7, 148, 150–1, 152
storage 205, 220–1
storage hardware 220–1
storage structures 225–8
data types 225–6
definition 225
storing shared file 159–60
storyboards 197–8, 280
streaming 141
string (or text) (data type) 270, 307
Structure English 232
surveys 7, 156
switches 90
syntax errors 245
system security software 334
system testing 248

Tableau Public 287
tablets 104–5
tagging 205
TCP/IP (transmission control protocol/internet protocol) 102–3
team solutions (problem solving) 168
technical constraints (problem-solving methodology) 278
test data 200, 248
testing
graphic solutions 68, 206–7
process 248–9
tables 249
text (data type) 270, 307
text editing 205
text styles (graphic solutions) 57
time visualisations 275
timelines (graphic solutions) 47
titles (graphic solutions) 56–7
tolerance
user interface 238–9
website design 174
top-driven loops 243
torrents 87–9
transmission media 109
physical 109, 110–11
wireless 109, 112–16
trojans 118
twisted-pair cables 110
twisted-pair wire 110
typography 174

uncounted loops 243
unencrypted data 20
uninterruptible power supplies 120–1
universal serial bus (USB) 221
uplink 116
user interphase (UI) 177
user interfaces, creating effective 235–9
accessibility 236
affordance 236
consistency 239
legibility 238
structure 237
tolerance 239–40
useability 235–6
visibility 237–8
user interfaces for data entry 309–12
useability (design principles for graphic solutions) 52, 182–4, 207, 235–6
usernames 21, 119–20, 334

validation (graphic solutions) 68, 70, 206
validation testing 248
variables 11, 43
version control 171
vested interest 10
Victorian government data directory 265–6
video file 203
virtual private networks (VPN) 86–90
virus 117
visibility
mobile devices 175
user interface 237–8
visualisations *see* data visualisations
voice over Internet Protocol (VoIP) 100

wearable technology 107–8
web browsers 98–9
web design and information systems 160–1
web server 82
web-authorising software 196
webpage storyboards 197–8
websites, designing 171–6
  mobile devices 172–6
  principles 181–94
  proportion 194
  specific considerations 194–5
  tools 195–205
websites, developing 205–7
  manipulation 205–6
  testing 206–7
  validation 206
white space 54
wide area networks (WAN) 84, 90
wi-fi 103
wi-fi communication 112–13
wi-fi protected access (WPA or WPA2) 121
wireless access point 94
wireless adaptors 93–4
wireless broadband routers 92
wireless extender 95–6
wireless network technologies 95–6
wireless security 121
wireless transmission media 109, 112–16
WMS file formats 288
wordclouds 156
World Wide Web Consortium (WC3) 173
worms 117

XLM/ELMS file formats 288

NELSON CENGAGE Learning

Register online at:
www.nelsonnet.com.au
to access digital resources to support improved learning & teaching*

Overview

- Full and in-depth coverage of the VCAA VCE Computing Study Design 2016–2019
- Experienced authors and trusted resources that have been part of the market for more than 20 years
- A problem-solving methodology is applied to create digital solutions in a range of contexts that meet specific need
- Provides full preparation for the Outcome assessment tasks and end-of-year exam

ISBN: 978-0170364744

*Please note that complimentary access to NelsonNet and the NelsonNetBook is only available to teachers who use the accompanying student textbook as a core educational resource in their classroom. Contact your sales representative for information about access codes and conditions.

NELSON CENGAGE Learning
For learning solutions, visit **cengage.com.au**