

Work with Data and Files

WORK WITH DATA AND FILES

Digital Essentials

THE UNIVERSITY OF QUEENSLAND LIBRARY

The University of Queensland



Work with Data and Files Copyright © 2023 by The University of Queensland is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/), except where otherwise noted.

CONTENTS

Module overview	1
1. What is data?	2
2. Uses of data	9
3. Sources of data	13
4. Clean data	19
5. Data storage	22
6. File naming and formats	27
7. Analyse and display data	32
8. Cite data	39
9. Check your knowledge	41
10. Conclusion	43

MODULE OVERVIEW

Aims and objectives

This module will:

- explain what data is
- examine how data is used
- provide guidance and information on managing data and how this can impact and support your studies
- introduce tools to improve productivity
- examine how data is used
- explore ways to analyse data.

After completing this module, you will be able to:

- find, clean and use data
- identify various file types and distinguish their usage
- create storage and backup plans
- evaluate and select tools to analyse and display data.

Module sections

1. [What is data?](#)
2. [Uses of data](#)
3. [Sources of data](#)
4. [Clean data](#)
5. [Data storage](#)
6. [File naming and formats](#)
7. [Analyse and display data](#)
8. [Cite data](#)
9. [Check your knowledge](#)
10. [Conclusion](#)

[Download versions](#) are available on the homepage.

Duration: 

Approximately 30 minutes

Student partnership 

This module was developed with UQ student partners as part of a [Student-Staff Partnership](#) project.

Graduate attributes

Knowledge and skills you can gain from this module will contribute to your [Graduate Attributes](#):



Courageous thinkers

This module is part of [Digital Essentials](#), a series of online modules to help you quickly build your digital skills so you can succeed in study and work.

Return to [UQ Library](#).

1. WHAT IS DATA?

- [Data becomes information](#)
 - [Types of data](#)
 - [Metadata](#)
 - [Datasets](#)
 - [Statistics](#)
-

Data becomes information

[Data](#) is raw information or numeric files.

Data can take a variety of forms, including numbers, text or people's opinions. Data can be manipulated, analysed and interpreted to give it meaning. Once data has meaning, it becomes **information**. Data can be collected from measurement, observation, experience or experiment.

Note: *Data* is the plural of *datum*. However, *data* is commonly used as a singular collective noun. You will see both "These data are ..." or "This data is ..." in scholarly information. Check with your lecturer or school for the form you are expected to use.



If you observed ducklings around UQ Lakes, what kind of data could you collect?

- The number of ducklings
- Their size
- Where they go
- What they eat

If you continued to observe the ducklings over time you could gather more data. If you organise and analyse the data you gather, it could reveal information about patterns of behaviour and identify potential impacts. e.g. growth or decline in the number of ducks around the lakes.



Source: [UQ Ducks](#) by Flic French, [CC BY 2.0](#)

Types of data

- **Primary data** is data you have collected yourself.
- **Secondary data** is data collected by others, including public datasets.

Data can be described in many different ways. One way to describe data is to use the terms qualitative and quantitative.

Qualitative — data that records qualities e.g. descriptions, concepts and opinions. Qualitative research is often focused on how and why something occurs.

Quantitative — numerical or spatial data that records quantities, measurements or frequencies e.g. size, location or scores.

Qualitative data can be analysed through quantitative approaches, such as statistical analysis, by giving the data a numerical value or ranking. Qualitative data is generally harder to analyse than quantitative, as it is usually in a format that is difficult to analyse quickly with basic statistical techniques.

Check your knowledge



An interactive H5P element has been excluded from this version of the text. You can view it online here:

<https://uq.pressbooks.pub/digital-essentials-data-and-files/?p=70#h5p-15>

Metadata

Metadata is data that describes other data. It is useful for finding, sharing and evaluating datasets. The metadata may be automatically generated and contained within the data file, such as with images, or it may be manually created and exist external to the file.

“Metadata provides the descriptive, structural and administrative information necessary to effectively access and utilise digital information objects.”

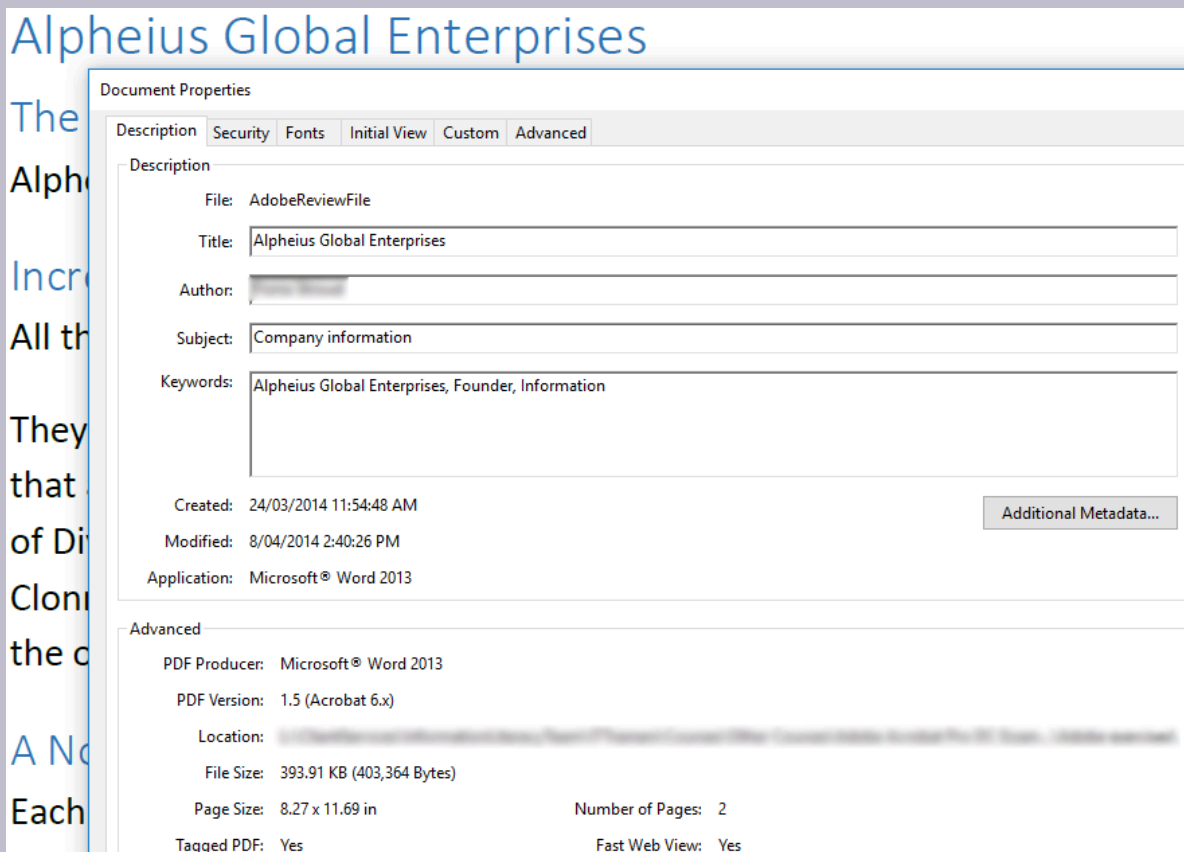
Source: Alemu, G., & Stevens, B. (2015). [An emergent theory of digital library metadata: enrich then filter](#) (1st edition). Chandos Publishing.

Examples of metadata

Documents

You can use metadata to determine who created a document and when it was created. Information available will include:

- when it was created
- when it was modified
- who owns the document
- who created the document.



Screenshot of metadata from a PDF document

Image metadata

Image metadata may show:

- when the image was taken
- what device was used to take the image
- if the device had GPS turned on, where that picture was taken.

In raw image files, the file properties may show where an image was taken. The image metadata provides the longitude and latitude that you can search using [Google Maps](#). The [Google Maps Help page](#) has more information on searching by latitude and longitude.

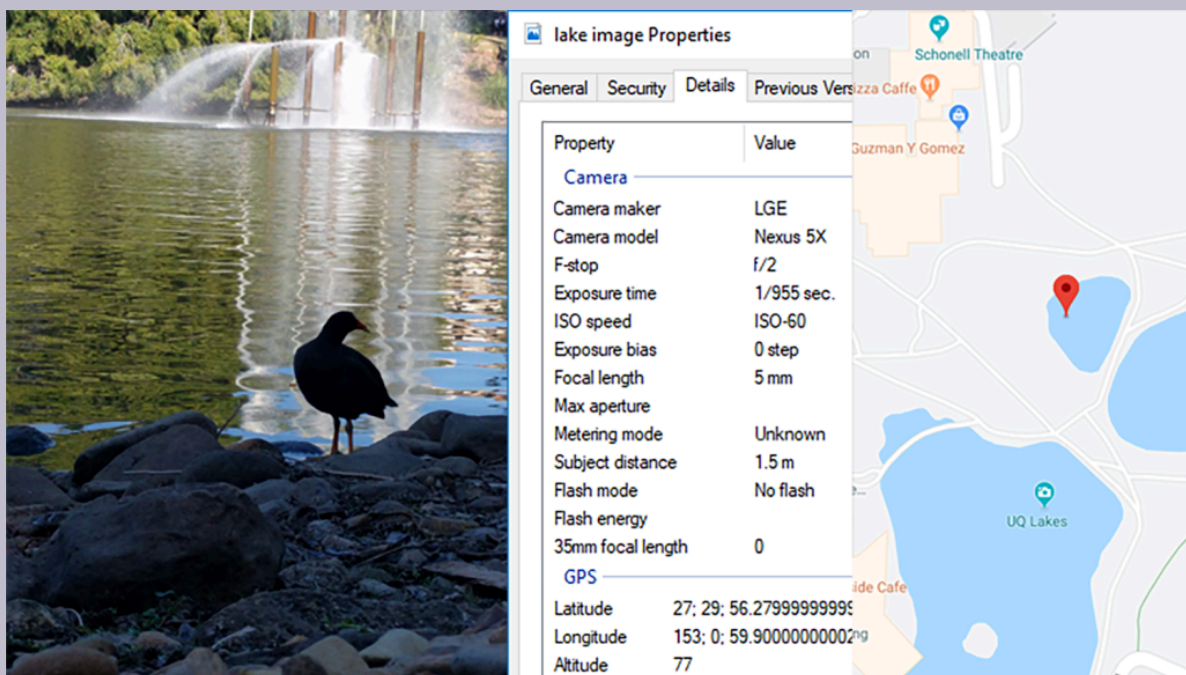


Image metadata may show the location and other data

Social media

In social media posts, the use of a hashtag is essentially acting as metadata for a post. Adding a hashtag (#UQ) to your tweet, Instagram or other posts creates a topic or tag for that post so that someone searching that platform for 'UQ' is more likely to find your post. The [Social Media](#) module has more information on hashtags.

Data discovery

Metadata is useful for data storage as it improves the discovery of files. It can help you avoid situations where you are not sure which file contains the information you are looking for. The quickest and easiest way to start using metadata is through a better file naming strategy.

Metadata and privacy

An individual's metadata can be used to learn a lot about that person's life and habits. When the telecommunications data retention scheme was first introduced in Australia, journalists were making their metadata available to the public to show the amount of information kept in the metadata. The ABC in 2015 published a summary of the findings on journalist Will Ockenden's metadata, [What reporter Will Ockenden's metadata reveals about his life](#).

Four-star General and former Director of the NSA & CIA discussed how powerful an individual's metadata is at [The John Hopkins Foreign Affairs Symposium \(YouTube, 1h14m\)](#) in 2014.



Google Maps has a feature called "Timeline" that will show you all the locations you have been to in the past. This only shows information if you use location services and Google Maps. [Try it out](#) and see how much Google knows about where you have been.

Datasets

Datasets are collections of data. Datasets can be in a wide variety of formats, including:

- Spreadsheets
- Text
- Transcripts from interviews and focus groups
- Videos
- Images
- Code.

The [File formats and naming conventions](#) section provides information on the different file formats that you may encounter and the software needed to open the files.

Statistics

Statistics are not the same as data. Statistics are:

- data that has already been processed and analysed
- an interpretation or summary of data
- used to make comparisons
- often presented in tables, charts, graphs or percentages.

Get more information on statistics:

- [Finding statistics guide](#) for journal articles, census data and websites related to statistics
- [Statistical language](#) from the Australian Bureau of Statistics (ABS) explains a range of statistical concepts and terms
- [Statistics for research students](#) has chapters on statistical concepts and the main types of statistical analyses.

2. USES OF DATA

- [How is data used?](#)
- [Ethics](#)
- [What kind of data is gathered on you?](#)

How is data used?

Data can be used to:

- reveal answers to questions
- make informed decisions
- tell a story.

At university, you may be asked to use data to:

- provide evidence to back up your conclusions
- explain complex information
- show trends or relationships
- understand behaviour or why things happen.

Interpreting data can help us to build our knowledge and to inform research. It can tell us what is happening around us, or around the world, and how it might affect us. We can use the data to plan and evaluate. Researchers use data to understand why something occurs and to find solutions to problems.



Find out how your [Spotify history data](#) could help predict what's going on with the economy. Researchers try to predict “economic sentiment” by identifying the emotional components of popular songs.

Ethics

If you are collecting data or using existing data, it is important to do so ethically. Unethical use of data would include behaviour such as:

- Breaching privacy
- Using fake data
- Plagiarising other people's data by not acknowledging the source
- Infringing copyright e.g. when scraping text and images from the web (Find out more about [web scraping in the next section](#))
- Not getting permission from your research subjects.

UQ has policies that guide ethical behaviour at UQ, including:

- [Responsible Research Management Framework Policy](#)
- [Student Integrity and Misconduct](#).

[Data at UQ](#) explains the responsibilities and policies that we must comply with at UQ.

Any research undertaken in Australia, including at UQ, must comply with the [Australian Code for the Responsible Conduct of Research](#).

What kind of data is gathered on you?

Governments, organisations and businesses plan services based on the data they gather from us e.g. planning transport services or how much bread to put on supermarket shelves. Netflix uses data analysis to make predictions about the type of shows you will enjoy, to keep you coming back for more.



What kind of cookie are you?

If you have never done an online quiz you might find that a strange question. If you have done online quizzes, you will be familiar with this type of “just for fun” quiz.

Think about how much data they can gather from our answers. Not just about cookies!

In May 2023, [Buzzfeed had over 94 million visitors](#) to its website. So potentially, BuzzFeed has gathered a lot of data from us.



Photo by [Whitney Wright](#) on [Unsplash](#)

Big data

[Big data](#) refers to extremely large, complex sets of data that are generated by connected devices. Devices that connect to the Internet, often referred to as the [Internet of Things \(IoT\)](#), have the potential to produce massive amounts of data. Governments, organisations and companies can [analyse and gain valuable insight using big data](#).

[Privacy vs pandemic: government tracking of mobile phones could be a potent weapon against COVID-19](#) – this article looks at the implications of governments using location history data from mobile phones to track the spread of infection.

Issues and challenges with big data:

- how to store the data
- getting software powerful enough to handle that much data
- ethical concerns.



Read about some of the ethical concerns surrounding big data:

- Can it [affect your life opportunities](#)?
- [Are you ready? Here is all the data Google and Facebook have on you](#)
- [Social cooling](#) discusses how the feeling of being watched, through our data, may change our behaviour.

Find out how to protect your privacy in our [Digital security](#) module.

Our data can be really useful for researchers and others, as long as it is collected and used ethically.



Bin chicken Google searches

Did **your Google search** contribute to the [boom of the “bin chicken”](#)?

```
trends.embed.renderExploreWidget("TIMESERIES", {
  "comparisonItem": [{
    "keyword": "bin chicken",
    "geo": "AU",
    "time": "2004-01-01 2023-06-21"
  }],
  "category": 0,
  "property": ""
}, {
  "exploreQuery": "date=all&geo=AU&q=bin%20chicken",
  "guestPath": "https://trends.google.com:443/trends/embed/"
});
```

Source: [Google Trends](#) and the [Ibis icon](#) is by [parkjisun](#) on the [Noun Project](#).

3. SOURCES OF DATA

- [What is the question you want to answer?](#)
- [Open data](#)
- [Dataset quality](#)
- [Collecting data](#)

What is the question you want to answer?

The question you want answered, or the problem you want to solve, will determine the type of data you need and analysis you need to do on that data. You can then use existing data, or gather the data yourself.

Who, what, when, where?



Think about:

- Who or what — is the subject of the data? e.g. a group of people, food, animals etc
- Where — is the location important? e.g. local or international, a particular suburb
- When — what is the appropriate time period to look at?

There may be an existing dataset that you can analyse to answer your question. Otherwise, you can use methods to gather your own data.

Open data

Open data is publicly available for reuse. Open data can be accessed from a range of sources. High quality data can be found on government websites and institutional repositories:

- [Research data](#) — this guide lists a range of high quality public research data sources
- [Text data](#) — this guide lists a variety of sources for open text data
- [Spatial data](#) — this guide lists Queensland, Australian and Global spatial data sources
- [UQ eSpace](#) — the repository for UQ research publications and research datasets

- [Google Dataset search](#) — allows searching across multiple repositories. Limits you can apply include download format and usage rights.

[GovHack](#) is an annual event, where people are invited to apply their creative skills to open government data.

? Using Census data

Do you want to use [Census data](#) from the Australian Bureau of Statistics (ABS) in your assignment or project?

[TableBuilder](#) is a free, online data tool, from the ABS, for creating tables, graphs and maps of Census data. [Basic tables \(YouTube, 4m45s\)](#) shows how to select data sets and create basic tables in TableBuilder:



One or more interactive elements has been excluded from this version of the text. You can view them online here: <https://uq.pressbooks.pub/digital-essentials-data-and-files/?p=100#oembed-1>

Dataset quality

Evaluate the quality of the dataset, just as you would [evaluate any information](#) you find, before you use it in your assignments or projects. Click the plus symbols to find out what you need to consider when assessing datasets:



An interactive H5P element has been excluded from this version of the text. You can view it online here:

<https://uq.pressbooks.pub/digital-essentials-data-and-files/?p=100#h5p-16>

Text version of Evaluate dataset quality:

Authority – Check:

- Who collected the data?
- Is it an educational institution, government or a reputable organisation?
- If an individual has produced the dataset, are they associated with a reputable organisation or are they well-known in their field?

Coverage – Check: –

- Were enough samples taken to be representative of the total population or group being researched?
- Is the time frame relevant or up-to-date?

Purpose – Check:

- Why was the data collected?
- Was there any bias in the collection methods?
- Who was the intended audience?

Accuracy – Check:

- Is the dataset complete?
- Are there responses missing or other errors?

Terms of use – Check:

Are the conditions for acceptable use of the data clearly stated and suitable for your needs?

The metadata or description of the data should include information to help you evaluate the dataset.



Check how the *Additional information* allows you to easily evaluate the quality of

this dataset — [Australia's threatened species, life history characteristics, and threatening processes](#).

Collecting data

You may need to collect your own data to answer your research question, if existing data is not available or suitable.

Storing the data

If you collect the data yourself, you will need to think about where to store it.

If you are collecting data for a small project or assignment, using local or online storage, such as Google Drive or OneDrive, would be suitable, as long as you do not have any identifiable, personal data. The [Data storage](#) section gives an overview of local and online storage options and how to back up your files.

If you are conducting research as part of your Winter/Summer Research project, Honours, Masters, or Higher Degree by Research degree:

1. [Manage research data](#) has information on how to manage, store and secure your project's data.
2. Discuss with your supervisor/coordinator the suitability of using the [UQ Research Data Manager](#) to manage your research data.

Sample size

Decide how many responses or observations you need to have a good sample size. Larger sample sizes are more likely to allow you to draw accurate conclusions than smaller samples. Get information on [sample design](#) from the Australian Bureau of Statistics.

Methods for collecting data

Observation

In observation, processes, activities or behaviours are observed. The subjects being observed may or may not be aware that they are under observation. A description of what occurs, or a checklist looking for a particular event, is used to record the observation.

Find out more about the [observational method](#).

Surveys or polls

A survey is a method of collecting data on behaviour, attitudes and opinions. Plan your survey questions carefully so the information you get from participants is useful to answer your research question.

A poll is a type of survey but usually quite short. Polls often have only one multiple choice question.

Information on planning a survey:

- [Setting up surveys](#) from the Australian Bureau of Statistics
- [Types of questions](#) to use and how to word them
- [How to create an effective survey in 15 simple tips](#)

You can conduct your surveys or polls face-to-face or you can use online tools.

Survey tools

Tools for conducting online surveys or polls

Tool	Free account available	Guides
Qualtrics	Available to UQ students and staff	Survey basic overview
Google Forms	Yes	How to use Google Forms
SurveyMonkey	Yes	SurveyMonkey Help Center
Crowd Signal	Yes	Crowd Signal Support Center

Interviews or focus groups

An interview typically involves asking either structured or unstructured questions of a single participant. Usually, the questions will be open-ended to allow for more in-depth insight on a topic than a survey can give.

A focus group involves a group of selected people (usually 6 to 12 individuals) participating in a group interview, guided by a moderator. It is a good way to get a social context on a topic.

In both techniques, you may need an audio or video record of the discussion, or have an observer record the details.

There are some factors to consider when [deciding whether to use focus groups or interviews](#).

Scraping

Scraping is a method of getting text and images from websites and social media. The practice can be problematic, depending upon the amount of reproduction and the intended use. Researchers may use automated methods to access any publicly available information on the web where they are specifically

engaged in legitimate research and study that makes use of the data and where they do not further publish the material. Get [more information or advice on copyright](#).

The [Find and use media](#) module provides information on how to comply with copyright.

Web APIs

Web APIs can be used to request data from a site using a URL. API stands for Application Programming Interface. Usually, some programming knowledge is needed to use APIs. [Web APIs for non-programmers](#) explains how web APIs work, gives tips on using them and lists some popular free APIs.

Tools for scraping

Tools to scrape text and images from websites and social media.

Tool	Freely available	Guides
Python	Yes	Downloading webpages with Python
Scrapy	Yes	Scrapy documentation
Reaper (for social media)	Yes	Help information is available on the Reaper site.



Interested in Instagram data? A [comparison of Instagram scraping tools](#) lists a range of tools.

Our Text mining and text analysis guide lists more methods for [gathering social media data](#). Text mining is a way of extracting text data from documents, for analysis.

4. CLEAN DATA

Data needs to be in a usable format to conduct an analysis. This often requires some preparation and “cleaning” of the data. Cleaning refers to the removal or modification of incorrect, incomplete or irrelevant data. Even open datasets may require some cleaning.

- [Structured or unstructured](#)
- [Preparing data](#)
- [Data formats](#)
- [Tools for cleaning data](#)

Structured or unstructured

Structured data is data that is well-organised. It follows a specific order, in a specific format. The data is categorised by different characteristics. Unstructured data has a range of information that is not organised into separate categories. It needs to be structured into a [machine-readable](#) format so that a computer can “read” the data.

Preparing data

You may have to sort or standardise the data. Dates, names and addresses are common data that require preparation.



Survey respondents may provide a range of answers to a question:

e.g. “Where do you live?”

- Brisbane, QLD, Australia
- Alice Springs, NT

- Hobart, Aus

You would need to sort the responses into separate fields and standard formats.

Example of standard data

City	State	Country
Brisbane	Queensland	Australia
Alice Springs	Northern Territory	Australia
Hobart	Tasmania	Australia

Data formats

Data that you analyse by computer must be in a format that software can open and read. The **delimited** format, where each variable is separated by a comma, tab or space, can be imported into software for analysis.

The ***.csv** (comma-separated value) extension is useful because it maintains **field separations** across software. If you have created a table in Microsoft Excel, and want to export it to another type of software, it is better to save the file as a csv.

Save an Excel file as a csv:

1. Click **File**.
2. Select **Save As**.
3. Then select the **CSV (Comma delimited) *.csv** option from the list.
4. Choose the location to save the file.
5. Click **Save**.

A csv file may not import correctly into other software if there are:

- blanks lines
- special (non-standard) characters e.g. # & @
- missing commas between fields.



Data in a “clean” csv file should appear like this example of female Oscar winners:

Index, Year, Age, Name, Movie

1, 2016, 26, Brie Larson, Room

2, 2017, 28, Emma Stone, La La Land

3, 2018, 60, Frances McDormand, Three Billboards Outside Ebbing Missouri

4, 2019, 45, Olivia Colman, The Favourite

5, 2020, 50, Renee Zellweger, Judy

Download this [example data file \(xlsx, 10KB\)](#) to see how it appears in a spreadsheet.

A data file may require special statistical software to open it.

Tools for cleaning data

You can manually sort data into the correct fields but if you have a lot of data it can be very time consuming. You can use formulas and functions in software to do the task much faster.

Data cleaning tools

Tools with functions to clean data

Tool	Freely available	Guides	Tutorials
OpenRefine	Yes	OpenRefine Documentation	Introduction to Google Refine (YouTube, 6m47s)
R	Yes	Essential R resources	Cleaning Bad Data in R (LinkedIn Learning, 1h 54m) UQ login required
Microsoft Excel	Available through Microsoft 365 (with your UQ Microsoft account)	Top 10 ways to clean your data	Clean data imported into Excel (LinkedIn Learning, 5m 13s) UQ login required

Our [Text mining and text analysis](#) guide has a lot of information about preparing text for analysis.

5. DATA STORAGE

- [What to consider about your data storage](#)
- [Local and online storage](#)
- [Storage capacity](#)
- [Backups](#)

What to consider about your data storage

Digital storage is a growing sustainability challenge for the world. As the amount of content creation keeps skyrocketing, all of this content needs to be stored somewhere. This demand for storage comes with questions and confusion about how to best store this data. The option you choose depends on your personal or professional wants and needs.

Examples of different options:

- computer hard drive
- USB drive
- CD
- external hard drive
- cloud.

The range of options means there is no ‘best’ pathway when it comes to where to store your files. Self-assessing what you require when it comes to data storage should always be the first step.

1. **How much space** do I need? You may not know the answer to this straight away. If you like to keep everything, you might be looking for the biggest solution you can get. However, you may work better with a smaller limit to better organise your files and only keep those that are necessary.
2. Do I want to **pay** or am I just after a **free option**?
3. Do I need a **simple solution** or one with **additional features and customisation**?
4. How concerned am I with **privacy**? Consider the sensitivity of the information and the accessibility of the storage platform.
5. Do I require **data creation or file sharing abilities** in addition to just storage?
6. Does it matter if I can’t **access files while there is no internet**? Access to cloud storage is impacted by internet outages and poor performance.
7. Am I working with research data as **part of my honours/master/PhD/higher degree**? If you are, talk to your supervisor as the [UQ Research Data Manager](#) may be the solution for you. Note: UQ staff should check [Where to store files and information](#) to ensure you are meeting UQ requirements.

Local and online storage

Local

UQ computer

The computers available for all to use at UQ allow you to plug in and use a USB drive or portable hard drive. UQ students also have access to a network hosted H:/ (Home) drive while using these computers. It has a capacity of 100Mb and can also be accessed online via the my.UQ dashboard.

USB drive

The advantages of using a USB drive are that it is simple to use and offers various storage sizes at a reasonable price. Many stores often have discount bins of cheap USB drives, however, it is often worth getting a good quality drive to ensure compatibility and reliability.

One of the largest disadvantages of a USB drive is the possibility of misplacing or damaging it and losing your files. You should store your files in more than one location so that a lost USB drive with your assignment on it is not the end of the world.

Online

Cloud storage

There are a variety of cloud storage systems. The solutions listed below are free with no trial period and may be a good place to start. There are other providers that may offer more but have trial periods, charges or other concerns.

OneDrive

[OneDrive](#) is a service offered by Microsoft. It can be freely used on most modern devices. OneDrive is available to UQ students and staff via your UQ [Microsoft 365](#) account. It is also available via a free, personal (non-UQ) Microsoft account. Much like Google Drive, OneDrive has great features:

- Easy link sharing
- Personal accounts have 15GB of free storage
- Useable on most devices
- Can be used with Office365 apps to edit online.

In regards to privacy, OneDrive has the best history among the big free services in Cloud storage. There have been no reports of large breaches in its history. However, there is still the risk.

Google Drive

[Google Drive](#) offers a lot of useful features for students:

- Shareable links to your files or folders that allow different levels of access to one or more people
- 15GB of free storage
- Easy to access on most devices
- Google apps may be a free replacement for Microsoft Office, depending on your customisation and compatibility needs.

It does have disadvantages, including privacy concerns. Google services are also not available everywhere in the world.

Apple iCloud

[iCloud](#) is the storage service offered by Apple that comes built into all Apple devices by default. You can use iCloud to store files, photos, videos, contacts, etc. It is free for anyone to get a basic iCloud account, and one of the benefits for Apple device owners is that it is quite easy and seamless to use iCloud for all storage. In the past iCloud has had data breaches. However, since then Apple has implemented [two-factor authentication](#).

Dropbox

[Dropbox](#) is a cloud storage solution that offers many of the same features as others with the exception of being able to create files and working collaboratively with others. It offers 2GB of storage with a free account but offers more storage for users who subscribe.

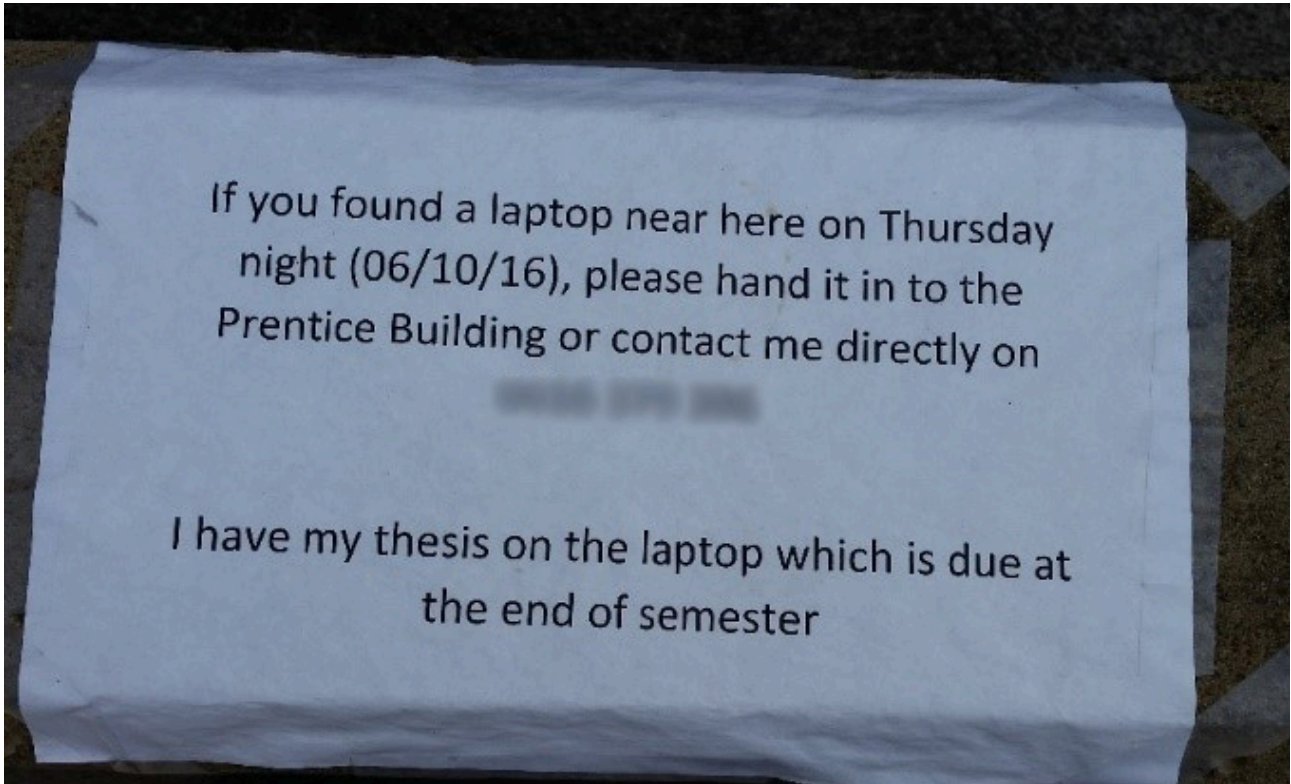
Storage capacity

To give you an idea of what the different storage solutions capacity actually means, the following table outlines how many documents each can hold. For this example, we used a 2000 word document with a few images and basic formatting in Microsoft Word 2016, which is 816KB in size.

Service	Storage size	Number of documents
H:/ (Home) drive	100MB	120 documents
USB drive	4GB	4900 documents
Google Drive	15GB	18,000 documents

Backups

You know that backing up your files is really important, but do you really do it? Taking the steps to actually back up your files and data doesn't need to be hard. You should develop a manageable backup plan that works for your needs so that you don't end up having to post a sign at University, like this one:



Example of someone who needed a backup plan

Backup plans

The basics of a backup plan are to keep copies of files in multiple locations. In practice, this means saving your file then copying that file to a different location, for example, USB drive, external hard drive or cloud storage, not just to a different folder on the same computer.

Not every file you possess requires an active and elaborate backup plan, as they are easily recreated or not important. Some applications or tools automatically backup. For instance, in the past, photos of family and friends required a safe and secure storage solution as they were irreplaceable memories. Nowadays most photos are taken with a smartphone.



Check your device

Apple, Microsoft and Google save a copy of your photos to your account that can be accessed from any computer or device. If you lose or replace your phone your photos are not lost. Check your device and account settings to ensure this is the case.

Automating backups

Backing up files to a USB stick, external hard drive and then to a cloud storage account after each update or change is a hassle and time-consuming process.

Cloud storage solutions often have a desktop application you can download, that can back up files locally and on the cloud. There are also a number of options for USB stick and external hard drive syncing. These types of solutions allow you to set up schedules and rules so that your folders on your computer back up to your desired device. One option is [Apple Time Machine](#).



While automated backups are time-saving, they require setting up and may not be as effective or reassuring as manually backing up files yourself.

6. FILE NAMING AND FORMATS

- [File naming](#)
- [File formats and saving](#)

File naming

 Do you know the best way to name a file?



An interactive H5P element has been excluded from this version of the text. You can view it online here:

<https://uq.pressbooks.pub/digital-essentials-data-and-files/?p=110#h5p-18>

Naming files on your computer in a logical and understandable way is a good practice to get into. How you name and organise your files is completely personal, but you can save yourself a lot of time by getting into the habit of file naming using **descriptive** names that give **context**.



[Project structure by Danielle Navarro \(YouTube, 8m30s\)](#) – Part 1 of a playlist on project structure gives a good overview of what to consider when naming your files:



One or more interactive elements has been excluded from this version of the text. You can view them online here: [https://uq.pressbooks.pub/digital-essentials-data-and-](https://uq.pressbooks.pub/digital-essentials-data-and-files/?p=110#oembed-1)

[files/?p=110#oembed-1](https://uq.pressbooks.pub/digital-essentials-data-and-files/?p=110#oembed-1)

There is also a [slide version of Project structure](#).

Key points for file naming

You may be required to name files in a specific format for assignments, work or groups. For instance, in a group, you might need to add your initials so that it is easy to determine who created or edited that file.

The most important point is to be **consistent!** Other recommendations:

- Use an **underscore character** (`_`) instead of a space. Spaces can cause issues, including when uploading files.
- Use letters, numbers, hyphens and underscores only. Avoid special characters.
- Use a **hyphen** to separate **different words** that are part of the same “chunk”.
- Use an **underscore** to separate **different “chunks”**.
- Add the date at the start of the filename in the format of **YYYY-MM-DD**. For example, 2018-09-14_LAWS1700_reading-1.pdf. This file was saved or created on the 14th of September 2018.

If you are making edits on a document and wish to keep a record of the drafts, you should name the different drafts as ‘versions’ by appending the filename with “v1, v2, etc”. This ensures that if a file corrupts or need to a previous version you can access this easily.

Examples of bad file names

- Essay1.docx
- essay2.docx
- Doc1 – Copy (1).docx
- presso.ppt
- pic for assignment.jpg

Examples of good file names

- 2018-09-14_LAWS1700_reading-1.pdf
- LAWS1700_assignment-1_v1.docx
- Essay1_picture_BB.jpg
- EDUC1130_assignment_2.xlsx

File names can get long if you are constantly noting which course they are associated with. You can use folders to organise files by courses.



Check your files

Go to your files and see if you can easily identify what each file is without opening it. If you can't, consider renaming those files. Additionally, try showing your files to a friend and see if they can guess what the file may contain.

File formats and saving

Most software saves files in a default file format. However, quite often you have the power to save in a different file format. Why would you want to do this? You may need to upload something in a required format or may need to open the file in another piece of software to work on it. For instance, Microsoft Word 2003 won't open .docx files. Despite there being several updates to Word since 2003, you can still come across some computers that only have 2003 version of Office installed.

A lot of file types started out as proprietary and therefore only certain applications would support that file type. As technology has developed those restrictions have reduced, allowing a lot more choice in which application can be used. Most word processing software can open any text file type that you come across.

The main difference between file types is how or if they compress the data. Compression is the act of trying to reduce the size of a file. There are two types of compression – [lossy and lossless](#).



File extensions



An interactive H5P element has been excluded from this version of the text. You can view it online here:

<https://uq.pressbooks.pub/digital-essentials-data-and-files/?p=110#h5p-19>

There are a lot of file types. [FileInfo](#) is a website that keeps a record and definition of almost every file type.

Common file types

Text files

File extension	File type name	Used by
.doc	97-2003 Word Document	Microsoft Word
.docx	Word Document	2006 and newer Microsoft Word
.odt	Open Office Document	Apache Open Office Writer
.txt	Plain Text File	All word processing applications
.pages	Pages Document	Apple Pages

Image files

File extension	File type name	Used for
.jpg	Joint Photographic Experts Group	Lossy compression of digital images. JPEG compression can be adjusted towards smaller file size or better image quality
.gif	Graphics Interchange Format	Simpler digital images that do not require large colour palletes, such as logos. GIF supports 256 colours. It is mostly used for creating animated digital images.
.png	Portable Network Graphics	Lossless compression of digital images. PNG is a raster graphics file format, which means it does not lose quality, or cause pixelation, when the image size is increased or decreased.

Audio files

File extension	File type name	Used for
.mp3	This is the third audio format of the Moving Picture Experts Group-1 (MPEG-1) standard (formally MPEG-1 Audio Layer III & MPEG-2 Audio Layer III)	Lossy data compression of digital audio. MP3 compression reduces components of a sound file that are beyond the hearing capabilities of humans.
.wav	Waveform Audio File Format	Uncompressed digital audio. WAV files can be compressed but are commonly used for uncompressed audio to store maximum audio quality.
.wma	Windows Media Audio	Lossless data compression of digital audio. WMA compresses digital audio files without loss of quality.
.flac	Free Lossless Audio Codec	Lossless data compression of digital audio. FLAC is an open format that can compress to 50-70% of the original file size.

Video files

File extension	File type name	Used for
.mp4	MPEG-4	Storing digital video and audio. It can also store subtitles and still images, as well as stream content from online.
.avi	Audio Video Interleave	Storing digital video and audio. Created by Microsoft, it can also store subtitles and still images, as well as stream content from online.

7. ANALYSE AND DISPLAY DATA

Data analysis is the process of examining data to find answers to questions. There are many ways that you can analyse and display data. This module will give an overview of some common analysis methods and tools. The data can be read through and manually organised by you, or processed using software.

- [Get an overview of your data](#)
- [Descriptive data analysis](#)
- [Inferential data analysis](#)
- [Data visualisation](#)
- [Misrepresenting data](#)
- [Tools for analysing and visualising data](#)

Get an overview of your data

1. Look through, or listen to, all the data you have gathered to get an overall idea of what you have gathered.
2. Transcribe any audio content so it is ready to be organised. You can [record directly in Word for web or upload an audio file to get a transcription](#).
3. Think about how you can reveal patterns or meaning in the data.
4. You may be able to sort the qualitative data in a way that it can be measured.



Read the blog – [How we helped our reporters learn to love spreadsheets](#).

You can access the [Times data training program in Google Drive](#), including training information, datasets and tip sheets to learn skills for Google Sheets and using data to tell a story.

Descriptive data analysis

Descriptive analysis provides a summary of the data that has been gathered. For **quantitative data** this involves presenting the data visually in tables and charts, and measuring values such as the mean.

For **qualitative data** the data is interpreted through grouping into categories or coding. Coding is the process of organising and sorting your data.

You can use one method or a mix of methods to do your analysis. Methods include:

- **Classification** — sorting the data into different kinds of things
- **Statistical analysis** — for qualitative data this could be counting different elements, such as keywords and phrases, to find out how often something is mentioned or has occurred
- **Thematic analysis** — looking for recurring themes in the data.

Thematic analysis of qualitative data

Looking for patterns or themes in your data can help to answer your original question or may lead to new questions. **Descriptive coding** is a useful technique for thematic analysis. Coding can be done by hand or by using software, such as NVivo.

Descriptive coding by hand



[What does coding look like? \(YouTube, 4m42s\)](#)



One or more interactive elements has been excluded from this version of the text. You can view them online here: <https://uq.pressbooks.pub/digital-essentials-data-and-files/?p=106#oembed-1>

Statistical analysis of quantitative data

Statistical analysis methods can range from simple to complex. Simple methods include:

- **Range** — the difference between the highest and lowest value
- **Minimum** — the smallest value
- **Maximum** — the largest value
- **Frequency** — the number of times a certain value appears
- **Mean** — the total of the values divided by the number of values
- **Median** — the middle value of any data after they are put in order
- **Mode** — the most frequently occurring value.

Learn about:

- [calculating percentiles](#)

- [confidence intervals](#).

You can use data analysis software to perform the statistical analysis.



How to use [features in Microsoft Excel to analyse data](#), including sort, filter, charts and pivot tables. It includes step-by-step tutorials on the different functions.

Inferential data analysis

Inferential analysis involves using the information from the data to make judgements about a topic or issue. For example, the results from a small group could be used to infer something about a larger group, or the results could be used to predict what might happen. Inferential analysis of qualitative data usually requires more advanced statistical methods



[Choosing which statistical test to use \(YouTube, 9m32s\)](#):



One or more interactive elements has been excluded from this version of the text. You can view them online here: <https://uq.pressbooks.pub/digital-essentials-data-and-files/?p=106#oembed-2>

[Which Stats Test](#) (UQ login required) is a question tool to help you narrow down the type of statistical test to use.

Data visualisation

Tables, graphs, maps and charts are used to summarise and display data. Once you have done your analysis you need to think about the best way to present the data.

Examples of data visualisations

The [Tudor Networks](#) visualisation brings together 123,850 letters connecting 20,424 people from the [United Kingdom's State Papers archive](#), dating from the accession of Henry VIII to the death of Elizabeth I (1509-1603).



[Charting culture \(YouTube, 5m 36s\)](#) is a visualisation of cultural mobility tracking the births and deaths of notable individuals like David, King of Israel, and Leonardo da Vinci, from 600 BC to 2014:



One or more interactive elements has been excluded from this version of the text. You can view them online here: <https://uq.pressbooks.pub/digital-essentials-data-and-files/?p=106#oembed-3>

The following visualisation shows the [average years of schooling per country from 1950 to 2017](https://ourworldindata.org/grapher/mean-years-of-schooling-long-run?time=1975):
<https://ourworldindata.org/grapher/mean-years-of-schooling-long-run?time=1975>

Choosing a chart

- [Chart Chooser](#) is a tool to help you find the right chart to display your data. After you choose the type of chart, you can download an Excel or PowerPoint template and then insert your own data.
- [Chartopedia](#) has a guide for choosing the right chart type.
- [Picking the right chart for your data \(LinkedIn Learning, 1h19m\)](#) Main topics include getting to the key idea you're trying to communicate; finding the right standard chart for your data type; and brainstorming and experimenting to come up with alternatives to the standards.

Displaying data

- Our [Data visualisation guide](#) has information on techniques, tips and tools to display data.
- The [Communication Learning In Practice for Scientists \(CLIPS\)](#) website has guidance on communicating results with scientific graphs.
- The [Data visualisation catalogue](#) shows different information visualisation types.
- Look at the [Periodic table of visualisation methods](#) for ways to display your data.

Misrepresenting data

Graphs can make data easier to understand but they can also be used to **misrepresent data**. Check graphs carefully. The graph creator can manipulate the design to inaccurately reinforce their own agenda.



Read [5 ways writers use graphs to mislead you](#), including:

- omitting baselines to make one group look better than another
- manipulating the y-axis to blow out the scale
- only including certain parts of the data
- choosing a type of chart that does not fit the data
- using colours, that alter long-held conventions or associations.

Tools for analysing and visualising data

There are many tools available for analysing and visualising data. You may want to use a tool that you have some experience with already, like Excel or Google Sheets, or you may want to try using software that is specifically for data analysis.

Spreadsheet tools

These tools can be used to analyse and visualise data

Tool	Useful features and limitations	UQ Library training	Online training guides and courses
Microsoft Excel	Microsoft Excel is the standard spreadsheet tool commonly used in workplaces. It is recommended if you need a full range of specialized functions for data analysis and organisation tasks.	Yes	Excel Essential Training (Microsoft 365) (LinkedIn Learning, 2h29m)
Google Sheets	Google Sheets is a free, web-based tool for creating spreadsheets and requires a Google Account to use. To access Sheets offline you need to use the Google Chrome browser and install Google Docs offline Chrome extension .	No	Google Sheets Essential Training (LinkedIn Learning, 1h52m)
Calc (LibreOffice)	Calc is a free spreadsheet tool you can download to your computer. Calc does allow you to share spreadsheets so multiple users can add data, however, real-time collaboration is not supported. File formats are compatible with Excel.	No	Learn LibreOffice Calc 6 in Under 30 Minutes (YouTube, 22m3s)

Data and text analysis software

These tools can be used to analyse and visualise data.

Tool	Useful features or limitations	UQ Library training	Online training guides and courses
Matlab	Matlab is downloadable software you can use for data analysis, modelling and visualisation. UQ students can download MATLAB to their personal computers.	No	Learning MATLAB (LinkedIn Learning, 1h13m) Course: MATLAB 2018 Essential Training (LinkedIn Learning, 3h15m)
Python	Python is a programming language that can be used for data analysis and visualisation.	Yes	LinkedIn Learning pathway - Python Pathway: LinkedIn Learning pathway - R, Resources - used in our different R courses.
R and RStudio	R is a language and environment for statistical computing and graphics. RStudio is an integrated development environment (IDE) to make R easier to use.	Yes	
SAS	SAS's Education Analytical Suite is a data analytics tool available to be downloaded by UQ students.	No	Getting Started with SAS Studio Video: Introduction to SPSS for data analysis (YouTube, 16m17s) . Course: SPSS Statistics Essential Training (LinkedIn Learning, 5h49m)
SPSS	SPSS Statistics is a software package used for statistical analysis and is available for UQ students to via Zenworks on Library computers and in HASS/HABS Student Computer Labs.	No	
KNIME Analytics Platform	Use KNIME Analytics Platform to create data workflows, perform different analysis steps and display the results as models or interactive views for interpretation. KNIME analytics platform is available as a free open source desktop application under the GNU General Public Licence .	No	Resources from KNIME to get you started. Data Science Foundations (LinkedIn Learning, 4h3m) - sections on KNIME.

Geographic information systems (GIS) tools

GIS tools are used to capture, analyse and present spatial or geographical data

Tool	Useful features and limitations	UQ Library training	Online training guides and courses
Carto	Carto is a cloud-based GIS platform built on open source software. Students can use Carto for free via the GitHub Student Developer Pack .	No	Carto Academy Carto tutorials
ESRI ArcGIS	ESRI ArcGIS is a downloadable tool for building and analysing geographical information. Students may use ArcGIS within specific courses. View the UQ Map Gallery from ArcGIS .	No	ArcGIS Pro Essential Training (LinkedIn Learning, 3h17m) .
Google Earth	Google Earth is a free mapping tool available online, via your desktop, or on your mobile device.	No	Become a Google mapping expert
QGIS	QGIS is a free, open-source desktop application for creating, editing and analysing geospatial information.	Yes	Course: Learning QGIS (LinkedIn Learning, 2h57m) Guide: QGIS Training manual

Our [Geographic information systems \(GIS\)](#) guide lists other tools and software to use with spatial or geographical data.

You can find more tools for analysing, visualising and presenting data in our [Choose the right tool](#) module.

8. CITE DATA

You should cite any datasets you use in your reference list, just as you would books or journal articles.

Example of a data citation in APA 7th style

Rhodes, J. (2014). *Spatial map of resistance estimates for the koala (*Phascolarctos cinereus*) in South East Queensland*. [Dataset]. The University of Queensland. <https://doi.org/10.14264/uql.2014.6>

Digital object identifier (DOI)

A DOI is a persistent identifier that takes you to the current URL of a resource.



Do you know how to find a resource using the DOI?

1. Go to doi.org
2. Enter the doi of the dataset from the previous example — **10.14264/uql.2014.6**
3. Click **Submit**.

It will open the dataset record in UQ eSpace.

Referencing styles

These [Library referencing style guides](#) include information on how to cite a dataset in that particular style:

- [APA 7th](#)
- [Vancouver.](#)

If the style you use does not include specific information on how to cite datasets, cite the dataset elements in a similar manner to other formats in your style.

Standard elements

The standard elements to include in a data citation are:

- Author
- Year/Date
- Title
- Publisher
- DOI or URL.

The **Elements of a data citation** section in [How to cite datasets and link to publications](#) explains more about citing data.

The [Data citation](#) section, in our guide to managing research data, has more information on the importance of data citation.

9. CHECK YOUR KNOWLEDGE

All the answers to the quiz questions can be found in this module.

Your response data will **not** be gathered. [Take a screenshot](#) of **Your result** at the end of the quiz if you are required to show you have completed it.

? Work with Data and Files quiz

There are 9 questions to answer. After you answer a question, click the arrow to move to the next question.



An interactive H5P element has been excluded from this version of the text. You can view it online here:

<https://uq.pressbooks.pub/digital-essentials-data-and-files/?p=52#h5p-17>

If you are required to get a certificate with your name, take this version of the [Work with data and files quiz \(requires a UQ login\)](#).

Note for Teaching staff: You can [embed the module](#) and the [H5P quiz](#) in your courses.

Module summary

① What is data?

- Data can take a variety of forms.
- It must be analysed and interpreted to give it meaning.

② Uses of data

- Data is used to build knowledge and inform research.
- It must be used ethically.

③ Sources of data

- You can use existing data or collect your own.
- The quality of datasets should be evaluated.

④ Clean data

- Data may require cleaning to be usable.
- Data must be in a format that analysis software can open and read.

⑤ Data storage

- There are many different types and sizes of digital storage solutions.
- Asking yourself some simple questions can be helpful in figuring out your storage needs.
- It is important to think through how you will back up and to create a backup plan.

⑥ File naming and formats

- Using file names that can give context is a valuable practice as it can save time and frustration.
- Being able to recognise the different types of files, means you can then easily understand and make use of them.

⑦ Analyse and display data

- There are many ways that data can be analysed and visualised.
- Tools and software can help with this process.

⑧ Cite data

- Any data used in your assignments or projects should be cited.
- Different referencing styles have different formats for citing datasets.

10. CONCLUSION



You have completed the **Work with Data and Files** module.

Tell us what you think

Use our [Digital Learning feedback form](#) to give **anonymous** feedback on this module. You can provide your email if you would like us to reply to you.

Digital Essentials modules

Build your digital skills with Digital Essentials. Select modules from the 6 themes that match your interests and will help you succeed in study and work.



New to UQ?

- [Getting started at the UQ Library](#)

- [Use UQ systems](#), includes:
 - [Book rooms](#)
 - [Printing at UQ.](#)



Assignment essentials

- [Find and use media](#)
- [Information essentials](#)
- [Write, cite and submit](#)
- [Types of assignments](#)



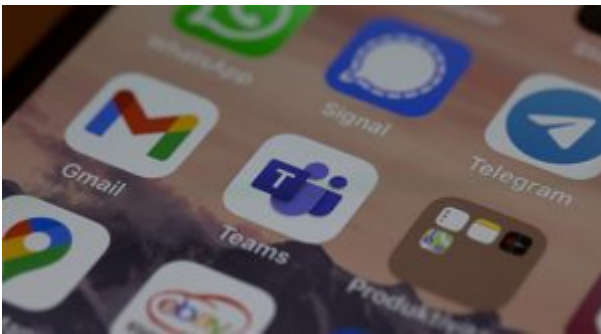
Digital creation

- [Accessibility](#)
- [Choose the right tool](#)
- [Intellectual Property](#)



Digital security and safety

- [Digital security](#)
- [Internet essentials](#)
- [Social media](#)



Professional identity and skills

- [Communicate and collaborate](#)
- [eProfessionalism](#)



Work with information

- [Artificial Intelligence](#)

- [Work with data and files](#)
- [Document your research data](#)

Teaching staff – use the modules in courses

Teaching staff can embed or link the modules in courses to help build your students' digital literacy.

- There are interactive elements throughout each module and a short H5P quiz at the end.
- The modules are also available for your students to download in EPUB, PDF and HTML format to make them more accessible.

Learn how to [add the Pressbook module](#) to your Learn.UQ (Blackboard) course.

Assess student learning

Most modules have a final short quiz created in H5P. You can download the H5P quiz from the module and embed it in your course if you would like to check your students' completions or to allow for the results to be transferred to the Grade Centre in your course.

The H5P quiz content will not record any completion data unless you download and add the H5P quiz directly to your course. Students can screenshot the quiz if they are required to show completion.

Learn how to [add the H5P content](#) to your course.