Acquiring and preparing data from online data repositories

**Years 7–8**

In this lesson, students develop skills to acquire, store and validate data from online data repositories. They learn how to navigate popular data sources and download data for use in spreadsheets. They also gain essential skills in cleaning and preparing data for analysis.

# Learning map and outcomes

By the end of this lesson, students will be able to:

* understand how to access online data repositories
* set specific goals for data acquired
* navigate data repositories and download data
* identify common issues in downloaded data
* clean and prepare data for analysis.

Learning hook
Introduce the scenario of planning a weekend getaway with friends. The location needs to have perfect weather, not too hot and not too cold.

Ask the following questions.

* How would you decide on the destination?
* Would you guess based on your existing knowledge, or would you use data to make an informed choice?

Explain that when planning a holiday, data plays a vital role. Weather-related data can help people make informed decisions about where to go at different times of the year, and can also help inform what clothing to pack. By analysing weather data for several locations, you can increase the probability of being sufficiently prepared for your trip. Data relating to weather can be found on websites such as <http://www.bom.gov.au/>.

Encourage student discussion with the following discussion points.

* Find out how students typically make decisions. Do they consider data or rely purely on instinct?
* Encourage students to share any personal experiences where data or information influenced and changed their choices.

# Learning input and construction

Acquiring and navigating data

Introduce several online data repositories. These repositories contain a range of data from various domains, and will help students with research, analysis and decision-making.

* **ABS (Australian Bureau of Statistics)**

The ABS provides a wide range of statistical data related to Australia, including demographics, economic indicators and social trends. <https://www.abs.gov.au/>

* **BOM (Bureau of Meteorology)**

The BOM offers weather and climate data, making it a valuable resource for climate researchers and anyone interested in weather-related data. <http://www.bom.gov.au/>

* **Kaggle**

Kaggle hosts datasets and data science projects. It offers a range of datasets contributed by data scientists and organisations worldwide. <https://www.kaggle.com/>

* **OurWorldInData**

OurWorldInData is a repository that focuses on global trends and statistics related to health, environment and socioeconomic indicators. <https://ourworldindata.org/>

Emphasise with students the importance of having a clear goal when collecting data. Using a clear goal helps students narrow down their search, ensuring that the data they obtain is relevant and covers what they are looking for.

An example of a clear data goal might be: ‘Obtaining rainfall records for a specific region to assess its suitability for agriculture.’

Use <http://www.bom.gov.au/climate/cdo/about/cdo-rainfall-feature.shtml> to look up rainfall data and use the website’s examples on how to filter through the data they are providing.



\*Source [**http://www.bom.gov.au/climate/cdo/about/cdo-rainfall-feature.shtml**](http://www.bom.gov.au/climate/cdo/about/cdo-rainfall-feature.shtml)

Now that we have a goal in mind for the sort of data we want to collect, it is important to understand how to navigate the selected repositories to locate and download relevant data.

The site we have linked to above shows a great example of how to navigate  <http://www.bom.gov.au/>

Walk students through the following data acquiring steps.

1. **Access the repository:** Show students how to access the repository's website or platform.
2. **Search for data:** Demonstrate how to use search features, keywords and filters to find datasets related to their specific goals (for example, searching for ‘rainfall records’ on BOM).
3. **Check dataset details:** Encourage students to click on a dataset to view details such as the dataset description, variables and data format.
4. **Download data:** Explain the process of downloading the selected dataset in a compatible format (CSV, Excel).

Data cleaning techniques

Walk students through the process of cleaning the provided sample dataset. This will help students understand how to read and sort data effectively, ensuring that what they have collected is relevant to the topic they are researching.

**Deleting unnecessary columns and rows**

* Explain the importance of identifying and removing columns and rows that contain irrelevant data that does not match what they are looking for.
* Demonstrate how to identify unnecessary data and safely delete it.

**Renaming columns and rows for clarity**

* Discuss the significance of clear and descriptive column and row names. Make sure to mention that having obvious columns and row names helps anyone reading our data understand what they are observing.
* Show students how to rename columns and rows to enhance clarity and readability.

**Adding units and labels for better understanding**

* Elaborate on the importance of units and labels to make data meaningful. For example, using temperature and/or labels when using dates in the dataset. Be sure to show how to add these units to the dataset.

## Using data validation to identify and handle issues

* Introduce data validation as a technique to identify and handle missing or bad data points. Data validation is a tool that helps you make sure the information entered in your spreadsheet is correct. It lets you create rules such as dropdown menus for choices and boundaries for numbers or dates to prevent errors.
* Show how to set up validation rules to flag or correct data issues (e.g. setting a range for valid values).
* Watch this brief tutorial (45 sec): How to Circle Invalid Data in Excel, which concisely demonstrates a method of data validation.

# Learning reflection

Allow students to share their experiences and insights from their hands-on practice.

# Curriculum alignment

Years 7–8

Digital Technologies

Acquire, store and validate data from a range of sources using software, including spreadsheets and databases ( [AC9TDI8P01](https://v9.australiancurriculum.edu.au/f-10-curriculum/learning-areas/digital-technologies/foundation-year_year-1_year-2_year-3_year-4_year-5_year-6_year-7_year-8_year-9_year-10/content-description?subject-identifier=TECTDIY78&content-description-code=AC9TDI8P01&detailed-content-descriptions=0&hide-ccp=0&hide-gc=0&side-by-side=1&strands-start-index=0&subjects-start-index=2&view=quick))