Catchment Health Hackathon

Contents

Challenge Statement

Access to water is essential for human life, environmental sustainability, and society's overall prosperity. Most of the water used by industry or residents of towns and cities starts its journey in a catchment. A water catchment is an area of land that naturally collects rainwater or receives water from springs or groundwater (collectively termed “raw water”). This raw water eventually flows into rivers, streams, or underground aquifers and, in many cases, becomes the source for drinking water. The catchment area is known as a source drinking water supply catchment where this occurs.

Ideally, all drinking water catchments would consist of land undisturbed by human-related activity, whether agriculture, forestry, townships and dwellings, or recreational activity. Any human-related activity creates the risk of a range of contaminants, either biological or chemical, being present in the raw water. These contaminants need to be removed by water treatment processes to produce water that is safe to use as drinking water.

Understanding the contaminants in a water catchment and within the waterways, along with their concentrations and possible sources, is part of the risk management framework for producing safe drinking water. Managing this risk has traditionally been achieved by collecting water quality data from the waterways across catchment areas.

Collecting water quality data on a scale and regularity that is beneficial, from within a drinking water catchment is a challenging and expensive task. Processing the samples through traditional laboratories can take a long time. Water corporations are also faced with the challenge of having access to a wide variety of data sources that provide insight into the health of the catchment, but the challenge is how to best use the various bits of data to make judgements about catchment health, the source of contaminants, or the existence of meaningful data relationships that would help underpin improved risk management practice.

The topic for this Hackathon is the development of innovative ways to analyse data collected in drinking water catchments and being able to present this analysis in a concise, plain English and/or visual format to either the public, water quality professionals, or other stakeholders. Your solutions have the potential to significantly improve our understanding and management of water catchments. The concise problem statement for this challenge is:

“How can we care for the environment with data and analytics?”

Geography of the Hackathon Case Study

The case study for this Hackathon Upper Coliban River/Upper Campaspe River catchment areas, which are part of the Murray-Darlin Basin. Data is available for four locations, two in Five Mile Creek, one in Kangaroo Creek and the last on the Little Coliban River.

Water flows from the Northwestern slopes of Mt. Macedon, Five Mile Creek becomes one of the tributaries of the Campaspe River, entering the Campaspe at Carlsruhe. Kangaroo Creek, a seasonal tributary of the Coliban River, weaves and winds its way through both the Wombat State Forest and private property, emptying into the Malmsbury Reservoir. This water system forms the catchment for the water supplies of Kyneton, Castlemaine, and Bendigo.

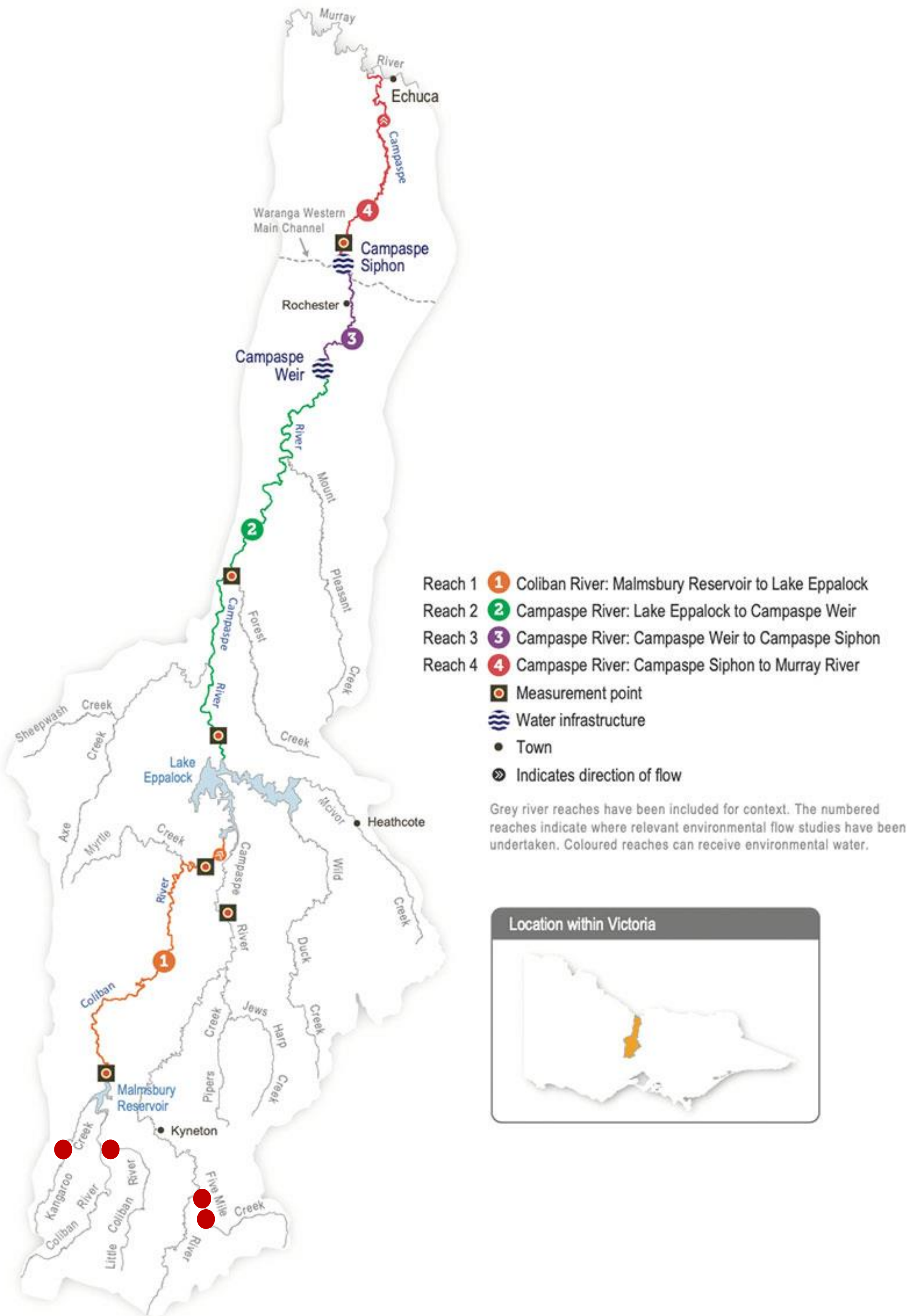

Figure 1: Campaspe River Catchment (Source: Victorian Environmental Water Holder, vewh.vic.gov.au) and approximate EcoDetection monitoring stations.

The Little Coliban River is a minor inland perennial river. The headwaters of the Little Coliban River rise on the northern slopes of the Great Dividing Range and descend to flow north into the Coliban River within the impounded Upper Coliban Reservoir. The Coliban River below the connected Upper Coliban Lauriston and Malmsbury Reservoir, flows north into Lake Eppalock, which also receives flow from the Campaspe River.

The two other locations relate to data collected upstream and downstream from the Woodend Recycled Water Plant in Five Mile Creek. The Five Mile Creek runs through the centre of the Woodend township and is significant open space for the community of Woodend. Five Mile Creek also drains into the Campaspe River just north of the study area.

EcoDetection Trial Data

The primary data for this Hackathon revolves around a trial project to collect automated samples from trial locations. These automated units, supplied by Eco Detection. Figure 2 provides a schematic of the units. These water quality monitoring units consist of various subsystems:

- Eco Sensor Device
- External batteries
- Additional electronic peripherals (auxiliary sensors)
- External antennas
- Solar panel sub-system

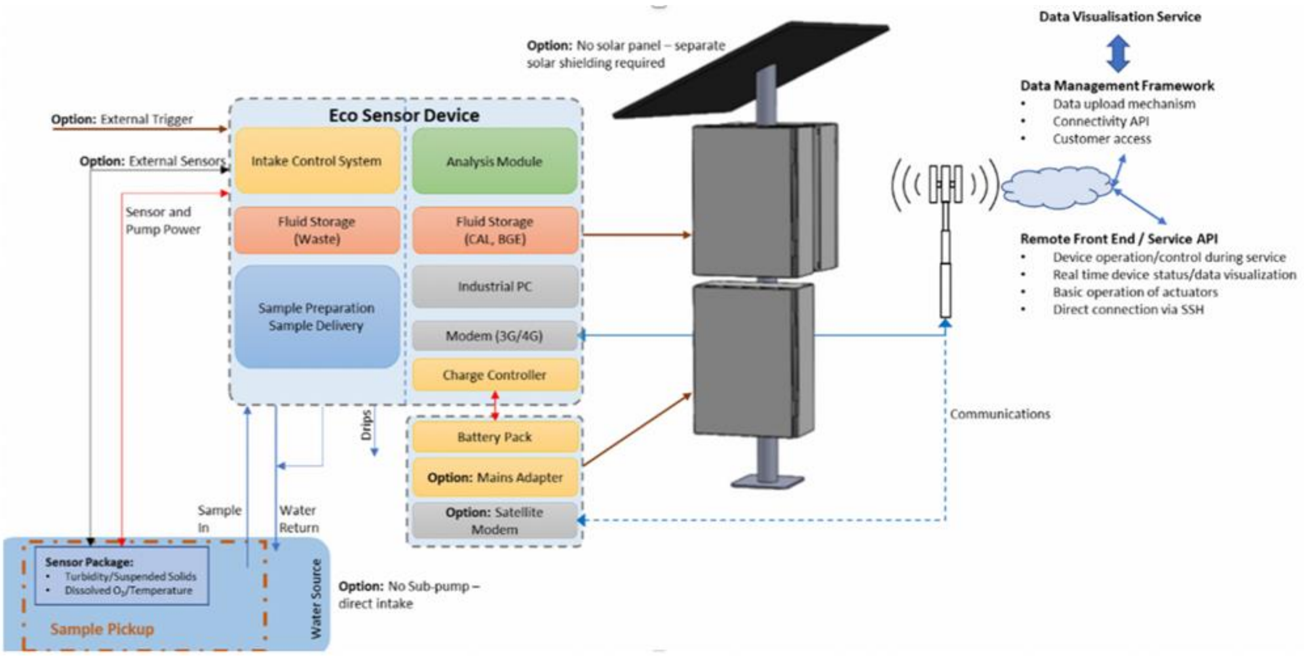

Figure 2: EcoDetection monitoring stations.

The system automatically collected 4–6 measurements per day for the trial duration of six months, with over 1,000 measurements being collected, to provide diurnal data for scientific analysis. The EcoDetection units measure ten analytes (Table 1).

Table 1: EcoDetect analytes.

Measurement	Unit
Chloride Concentration	ppb
Conductivity	µS/cm
Fluoride Concentration	ppb
Nephelometric Turbidity Units	NTU
Nitrate Concentration	ppb
Nitrite Concentration	ppb
Oxygen	mg/L
Phosphate Concentration	ppb
Sulphate Concentration	ppb
pH	-

Across the catchment areas of interest, Eco Detection have installed four units. The first two are located upstream and downstream of a wastewater treatment plant that releases treated wastewater into Five Mile Creek.

Other Data Sources

The main data source for solving this problem statement are the EcoDetect trial data for the four locations. Other data sources are available to contextualise and verify this data:

- Meteorology (temperature and rainfall)
- Stream flows
- Soil moisture profiles
- Laboratory samples collected by ALS on behalf of Coliban Water and Greater Western Water
- WaterWatch data

Catchment Health Background Information

A risk management framework to produce safe drinking water requires understanding the contaminants present in a drinking water catchment, their concentrations, and possible sources.

More broadly, what is being sought is the protection of what might be described as "catchment health". Whilst the protection of catchment health is not currently measurable against a standard or numerical value, it does encapsulate the following concepts:

- Drinking water supply catchments and waterways are complex entities. They are dynamic systems that require a comprehensive and adaptive approach to management.
- Our catchments are facing escalating risks. These risks, including the intensification of development, increased recreational demand, and extreme climatic events, are threatening the health and safety of our water supplies.

The Australian Drinking Water Guidelines (ADWG) assert that the most effective and efficient means of assuring drinking water quality and protecting public health is to adopt a preventive management approach that encompasses all steps in the water supply chain, from catchment to consumer.

Therefore, effective management of our water supplies requires governments, water utilities, land management agencies, landholders, and other custodians to collaborate and implement a range of measures across multiple barriers.

Notably, while the ADWG advocates that source waters should be protected to the maximum degree practicable, there appears to be a recent trend away from this and a greater reliance on water treatment solutions.

Challenge Statement

The core challenge for this hackathon is to find ways to best communicate the story that this data is telling the user, answering the core question: "How can we care for the environment with data and analytics?"

Participants answer this questions with the available data by analysing and visualising in a way to answer the question. There is no single story that this data can tell. Different stakeholders have varying perspectives on the data. The resulting data product can be as a written report or a digital application that visualises the data and answers the problem statement.

Participants are free to use any data tool they see fit, including spreadsheets, codes solutions in Python, R or Julia or any other language, or data visualisation tools such as Power BI or Tableau. The judging criteria do not include the complexity of the tool used.

The Hackathon consists of two deliverables:

1) Data product
    a. The final report, dashboard, spreadsheet, online application, or any other suitable medium that tells the data story.
2) Presentation
    a. Five-minute presentation at the Hackathon final day on 20 September 2024 at the Melbourne Water offices.

The recommended workflow consists of four stages (Figure 3).

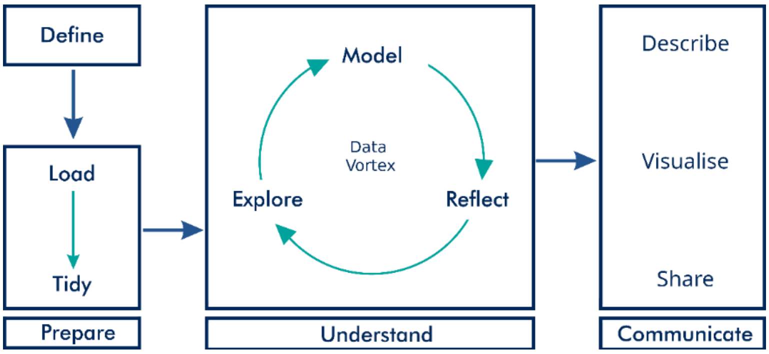

Figure 3: Data Analysis Workflow.[1]

Define the Problem

Changes in water quality affects a wide range of people and organisations, all who have an interest in the changes in water quality data. When communicating the data, insights, and water predictive methodology, it is important to consider the priorities and background of the stakeholder type you are communicating to.

[1] Peter Prevos, *Data Science for Water Utilities* (Routledge 2023).

The table below lists some of the stakeholders affected by changes in water quality in the Upper Coliban River and Upper Campaspe River catchments. Below is only some of the stakeholders affected by water quality, you may choose to develop your final data product for a Stakeholder not listed below if you wish.

Table 2: Catchment health data stakeholders.

Persona	Priorities	Background
Regulator	The Environment Protection Authority Victoria (EPA) seeks to: – Monitor and manage water quality. – Investigate any water quality pollution incidents. – Understand key drivers for and stressors on water quality.	*Scientific Literacy:* High degree of knowledge of the scientific components of water quality. *Data Literacy:* High degree of data literacy. *Knowledge of local geography:* Medium-Low knowledge of local geography as they may not be based in the region.
Water Corporations (e.g. Coliban Water)	The Environment Protection Authority Victoria (EPA) seeks to: – Monitor and manage water quality. – Support and facilitate local stakeholders such as landowners to implement measures improve water quality.	*Scientific Literacy:* High degree of knowledge of the scientific components of water quality. *Data Literacy:* High degree of data literacy. *Knowledge of local geography:* High knowledge of local geography
Landowners	Landowners seek to ensure catchment water quality remains high: – Understand the impact of their activities and by others and seasonality on water quality. – Maintaining the amenity of their property. – Extract water from the catchment.	*Scientific Literacy:* Low knowledge of the scientific components of water quality. *Data Literacy:* Low data literacy. *Knowledge of local geography:* High knowledge of their local area
Public	The public of the Upper Coliban River/Upper Campaspe River catchments want to: – Be informed of changes to water quality which will affect their health and activities (e.g., swimming in rivers).	*Scientific Literacy:* Low knowledge of the scientific components of water quality. *Data Literacy:* Low data literacy.

	– Be informed about catchment health out of a personal interest in environmental health. – Report any behaviours they may witness which may affect water quality.	*Knowledge of local geography:* Medium knowledge of their local area, may not be as interested in non-recreational waterways

Participants need to select a stakeholder as the target of their data product. When choosing a stakeholder not listed in this table, then participants need to define the persona of the end user of their analysis.

Prepare the Data

All data is provided to participants upon registration. Chapter 0 describes the available data in detail. The data is provided as provided by the various sources and needs to be transformed into a format suitable for analysis and visualisation.

Participants don't have to use all data sets and are encouraged to use other data sets not provided in this pack that could be relevant to the story that you endeavour to tell, such as geospatial data.

Understanding the Data

The core process of analysis is an iterative process of exploring the data, model or visualise the outcomes and reflect on the results from the perspective of your chosen stakeholder.

The analysis should focus on three criteria:

- Usefulness: Are the insights actionable so that catchment health can be improved?
- Soundness: Is the analysis logical and statistically sound?

Communicate the Results

The resulting data product needs to answer the problem statement from the perspective of the chosen stakeholder. The final product will provide visualisations and a short explanation on how to interpret the results.

The final data product must clearly communicate the intention of the analysis so that the needs of the stakeholder are met.

Participants are encouraged to use best-practice in data visualisation. Some sources useful sources on this topic are:

- The Data Visualisation Catalogue (datavizcatalogue.com)
- How to Choose the Right Data Visualization | Atlassian
- Tufte's Principles (thedoublethink.com)

Hackathon Rules of Engagement

Registration

Registration for the hackathon open on 29 July at the VicWater conference. All participants will receive a copy of this challenge statement and the related data upon registration.

We endeavour to include all participants who register. All applications are assessed based on representation from all IWN members. We encourage diverse teams that don't only consist of data analysis experts. Consider recruiting from a broad section of the water utility, such as external communication experts, catchment subject health subject matter experts and data analysts.

Process

An online pre-event-session will be held for all participants to introduce the problem. During this session, teams can ask questions about the challenge, data, and deliverables. A channel on the VicWater Teams site will be available to ask questions during the analysis process.

Hackathon Day Program

The Hackathon will culminate with a face-to-face workshop held at the Melbourne Water Head office in Melbourne on 20 of September 2024.

08:00 – Doors open

08:30 – Sign-in opens

09:30 – Opening / briefing (Acknowledgement of Country, accessibility conversation at the welcome)

10:00 – Commence Hack

13:00 – Hacking Lunch

14:30 – Work on presentation (judge arrival)

15:00 – Presentations

16:30 – Judges deliberation

16:45 – Ceremony and award

17:30 – Drinks reception

Judging Criteria

The final data product and presentation is judged on three criteria:

1) Useful/Actionable
    - Does the data product have clear value for the target stakeholder audience?
    - Is there a clear result, message and/or action from the data product?
2) Soundness

- Has the data product been developed in a way that is technically and logically rigorous?
- Is there clear documentation for all assumptions and calculations made in the development of the data product?

3) Aesthetics
    - Is the data product easy to understand and visually appealing?
    - Does the data product encourage users to use it frequently?

Data Sources

All raw data is available in the `data` folder in the challenge pack. Table 3 list all available data files. The spreadsheet with monitoring stations lists all measurement locations available in this challenge.

The following sections provide a data dictionary for each of the files.

Table 3: Register of raw data files.

Filename	Source	Description
monitoring_stations.xlsx	IWN	List of monitoring stations
export-ecod72.Coliban Water.Kangaroo Creek.csv	EcoDetection	EcoDetect data Kangaroo Creek
export-ecod72.Coliban Water.Little Coliban River.csv	EcoDetection	EcoDetect data Little Coliban River
export-ecod72.Greater Western Water.Five Mile Creek - Woodend RWP Site 1.csv	EcoDetection	EcoDetect data Five Mile Creek 1
export-ecod72.Greater Western Water.Five Mile Creek - Woodend RWP Site 2.csv	EcoDetection	EcoDetect data Five Mile Creek 2
cw_catchment_sampling.csv	Coliban	Laboratory testing results
IDCJAC0009_088037_1800_Data.csv	BoM	Rainfall Lauriston Reservoir (88037)
IDCJAC0009_088061_1800_Data.csv	BoM	Rainfall Woodend (88061)
IDCJAC0011_088051_1800_Data.csv	BoM	Min temp Redesdale (88051)
IDCJAC0010_088051_1800_Data.csv	BoM	Max temp Redesdale (88051)
406266.csv	DEECA	Flows Five Mile Creek
406280.csv	DEECA	Flows Little Coliban River
406281.csv	DEECA	Flows Kangaroo Creek

EcoDetection

The core Hackathon data involves four monitoring stations, two in the Coliban River catchment and two in Five Mile Creek.

The data is made available through an online data portal on ecodetection.eagle.io, which requires a username and password. This platform provides basic time series data visualisation for each location.

The data from each analyser is exported to a CSV file with the location description in file name. The relevant data files are provided in the challenge pack.

The first three rows of the files contain metadata.

1. Numeric ID for each field
2. Column names
3. Units

The column names of the second row are Timestamp, followed by /-separated columns, e.g.: `ecod72/Coliban Water/Kangaroo Creek/Diagnostics Data/Background Electrolyte`, which contains five identities:

- Identifier (ecod72)
- Client name (Coliban Water)
- Location name (Kangaroo Creek)
- Measurement category, either:
  - Analytes Concentration
  - Diagnostics Data
  - External Sensor
- Measurement name (various)

Laboratory Testing

Coliban Water undertakes regular sampling of these locations, which are analysed in a NATA-accredited laboratory. Laboratory data is available for each of the trial locations:

- SITE2: Little Coliban River
- SITE17: Kangaroo Creek

Meteorology

The Australian Bureau of Meteorology provides access to meteorological data through their website. Data is freely available in CSV format for rainfall and maximum or minimum temperature.

The nearest weather stations for the monitoring sites with relevant data for this hackathon are Lauriston Reservoir (80037), Woodend (88061) and Redesdale (88051). Table 4 lists the nearest rainfall and temperature monitoring station for each of the four trial locations.

Table 4: Nearest meteorology stations.

Location	Rainfall	Temperature
Kangaroo Creek	88037	88051
Little Coliban River	88037	88051
Five Mile Creek 1	88061	88051
Five Mile Creek 2	88061	88051

The filenames indicate the product code and station number. The CSV files for meteorology contain the following fields:

- Product code
  - IDCJAC0009: Daily rainfall
  - IDCJAC0010: Maximum daily temperature
  - IDCJAC0011: Minimum daily temperature
- Product code
- Bureau of Meteorology station number

- Year
- Month
- Day
- Measurement, either:
  - Rainfall amount (millimetres)
  - Minimum temperature (Degree C)
  - Maximum temperature (Degree C)
- Period of measurement
  - Period over which rainfall was measured (days)
  - Days of accumulation of maximum temperature
  - Days of accumulation of minimum temperature
- Quality: Data quality indicator

Streamflows

The Water Management Information System (WMIS) managed by the Department of Energy, Environment and Climate Action (DEECA) monitors hundreds of locations in Victoria for both surface and groundwater. All four monitoring sites are in the Campaspe Basin (406) within the Murray-Darling Basin.

The CSV files contain a timestamp, the measured water level in the creek (m) or river and the associated discharged in ML/day. The Five Mile Creek monitoring station also contains rainfall data. Each measurement has a Quality Code, which is explained in the first rows of the file. The documentation folder contains detailed explanatory notes for the WIMS data (`ExplanotaryNotes.pdf`).

Table 5: Streamflow monitoring stations.

Trial Location	WIMS Stream flow
Kangaroo Creek	406281
Little Coliban River	406280
Five Mile Creek 1	406266
Five Mile Creek 2	406266

Soil Moisture

The Bureau of Meteorology publishes hydrological data in both a simple map grid and in the NetCDF format. The NetCDF format provides a three-dimensional matrix of spatial data over time. This means that the data can be analysed spatially as well as over time for point locations.

Agriculture Australia maintains a fleet of soil moisture monitoring stations around Victoria. The nearest location to the Coliban Water sites is Banyon.

The Australian Water Outlook website (Australian Water Outlook). This website provides historical water balance data as well as forecasts and projections. Gridded climate input data and hydrologic model outputs (as NetCDF4) can be directly downloaded at https://dx.doi.org/10.25914/6130680dc5a51.

WaterWatch Victoria

Waterwatch Victoria is a successful community engagement program connecting local communities with river health and sustainable water issues and management since 1993. Through the Waterwatch Program, citizen scientists are supported and encouraged to become actively involved in local waterway monitoring and on-ground activities.

The WaterWatch folder contains a series of data sets and background information generated by this program.