11ASM Full Paper

Wilson et al., - GBR Catchment Loads Monitoring Program: Digital Products and FAIR Data Principles

Great Barrier Reef Catchment Loads Monitoring Program: Digital Products and Commitment to the FAIR — Findability, Accessibility, Interoperability, and Reusability — Data Principles

Wilson, E. [1,2]*, Turner R.D.R.[1,2], Clech-Goods C.[2], Roberts C.[2], Villa A.[3], Kaminski H.[2], Welk K. [2], Huggins R.[2], Ferguson B.[2], Marsh A.[1,2], Orr D.N.[2], Neelamraju C. [1,2], Warne M.St.J.[1,2], and Mann R.M.[1,2,4]

University of Queensland, Brisbane, Queensland, 4000

**Corresponding and presenting author: Eloise.wilson@uq.edu.au*

1. Reef Catchments Science Partnership, School of the Environment, University of Queensland, Brisbane, Queensland, Australia

2. Water Quality & Investigations, Department of Environment, Science and Innovation, Brisbane, Queensland, Australia

3. Information and Digital Science Delivery, Department of Environment, Science and Innovation, Brisbane, Queensland, Australia

4. Sustainable Minerals Institute, University of Queensland, Brisbane, Queensland, Australia

Key Points

- the Great Barrier Reef Catchment Loads Monitoring Program (GBRCLMP) is Australia's largest water quality monitoring program,
- the launch of Tahbil, the Queensland Government's Water Quality Data Portal, provides a centralised repository for water quality data collected by the GBRCLMP, improving data accessibility and promoting reusability,
- automation is crucial for quality assurance processes, especially with near real-time data collection, to ensure data is published in a timely manner while maintaining data quality standards, and
- the high volume of data that the GBRCLMP generates stands out as the primary barrier to the adoption of the FAIR (Findable, Accessible, Interoperable and Reuseable).

Abstract

The Great Barrier Reef Catchment Loads Monitoring Program (GBRCLMP) is a geographically extensive water quality monitoring program that monitors water quality discharging into the Great Barrier Reef lagoon from 430,000 square kilometres of catchments along Queensland's east coast. The data produced by the GBRCLMP helps inform management decisions and investments that can ultimately reduce land-based runoff to the Great Barrier Reef lagoon. In 2022, an international review was completed to provide a benchmark assessment against similar international programs and allow for the strategic development of the Program moving forward. The international review panel commended the GBRCLMP on its progress towards providing FAIR — Findable, Accessible, Interoperable, and Reusable — data to the public and key stakeholders and supported its continued development. The Queensland Government's Water Quality Data Portal – Tahbil – was launched in May 2024 and provides access to water quality data collected by the GBRCLMP, the South East Queensland Loads Monitoring Program, and other third parties. This paper considers Tahbil and its adherence to the FAIR data principles, using the Australian Research Data Commons FAIR Data Self-Assessment Tool, and highlights the other digital products currently maintained by the GBRCLMP. Ensuring that data collected by the GBRCLMP is FAIR will facilitate efficient data dissemination, empower stakeholders to use it for decision-making, support academic and scientific research, enhance transparency and public trust, and provide educational value.

Keywords

Automation, FAIR data principles, pesticides, Shiny – Rstudio and water quality monitoring

Introduction

The Great Barrier Reef (GBR) receives run-off from six Natural Resource Management (NRM) regions along the east coast of Queensland, Australia. The 2017 Scientific Consensus Statement: Land use impacts on Great Barrier Reef water quality and ecosystem condition (State of Queensland, 2017) is a major independent review of land-based impacts on Reef water quality and ecosystem condition that underpins the Reef 2050

Water Quality Improvement Plan. The Scientific Consensus Statement attributed the decline of marine water quality to land-based runoff from adjacent catchments and deemed it a major cause of the current poor state of many coastal and marine ecosystems of the GBR. The Consensus Statement also highlighted the need to focus efforts on improving water quality to enhance the resilience of the coastal and marine ecosystems of the GBR (State of Queensland, 2017).

The main pollutants associated with agricultural run-off in the GBR catchments are nutrients, fine sediments, and pesticides, which pose risks at both the individual species level and the wider ecosystem level (State of Queensland 2017). Nutrient imbalances can lead to an increased abundance of plankton, macroalgae, and epiphytic algae, which may have negative consequences for seagrass and coral communities. Increased nutrient loads have been linked to an increase in epiphytic algal growth on seagrass leaves and reduced light availability for seagrass leaves (Wazniak et al., 2007). Excess nutrient availability may also increase the susceptibility of coral to disease (Morrow et al., 2012). Increased levels of fine sediments can cloud the water and decrease the amount of light that is able to reach coral and seagrass communities, potentially smothering them (Fabricius, 2005). Decreased light levels that are sustained for months may result in a decrease in the spatial distribution, contraction of depth range and death of plant communities (Chartrand et al., 2016; Collier et al., 2012). A study led by Warne et al., in 2023 monitored 86 pesticide active ingredients between 2015 and 2018 at 28 sites that discharged water into the GBR lagoon. Of these 86 pesticides, 22 were selected, and their combined risk (i.e., when they co-occur) was calculated. The study found that the Total Pesticide Risk for these 22 pesticide active ingredients ranged from less than 1% to 42% of aquatic species being affected. However, approximately 85% of the total pesticide risk estimates were greater than 1%, indicating they did not meet the Reef 2050 Water Quality Improvement Plan's pesticide target for waters entering the GBR, thereby posing a significant risk to freshwater ecosystems.

The Great Barrier Reef Catchment Loads Monitoring Program (GBRCLMP) is a geographically extensive water quality monitoring program that monitors water quality discharging into the Great Barrier Reef lagoon from 430, 000 square kilometres of catchments along Queensland's east coast. As of May 2024, the GBRCLMP monitors for sediment and nutrients at 61 sites, and for pesticides at 43 sites across the east coast of Queensland. Ambient (low flow) water quality samples are collected monthly during the dry season and up to weekly during the wet season, depending on the site. During high flow events in both the wet and dry seasons, sampling intensifies (e.g., daily or multiple times per day) to capture most land-based runoff. The objectives of the GBRCLMP include:

- monitoring of event and ambient concentrations of sediment, nutrients, and pesticides,
- calculating annual and daily loads (mass) of sediment (measured as total suspended solids) and nutrients (nitrogen and phosphorus total, particulate and dissolved fractions), and a Pesticide Risk Metric (PRM) based on the concentrations of 22 pesticides,
- providing sediment and nutrient loads data to calibrate and validate Source Catchment models for Reef 2050 Water Quality Improvement Plan targets assessment (The Commonwealth of Australia 2023)
- providing PRM data for use in the biennial Reef Water Quality Report Card,
- providing sediment and nutrient concentration data and PRM data for use in regional waterway condition report cards, and
- investigating methods to track long-term trends in water quality entering the GBR lagoon.

Since its inception, the GBRCLMP has expanded rapidly in response to an increasing range of stakeholder needs. The GBRCLMP has been successful in meeting these needs, and in 2022, an international review was completed to provide a benchmark assessment against international programs of a similar nature and allow for strategic development of the Program moving forward (Warne et al., 2022). The international review panel commended the GBRCLMP on the progress it has made towards providing FAIR — Findable, Accessible,

Interoperable, and Reusable — data to the public and to key stakeholders and supported its continued development. The FAIR Guiding Principles for scientific data management and stewardship were first described by Wilkinson et al., (2016) and comprise a "measurable set of principles ... that may act as a guideline for those wishing to enhance the reusability of their data holdings". Wilkinson et al., (2016) notes that these principles are applicable not only to conventional "data", but also to the pipelines, data workflows, metadata, or data architecture that enable data management and accessibility.

The continued implementation of the FAIR data principles within the GBRCLMP will facilitate efficient data dissemination enabling further analysis and interpretation by stakeholders and the wider public. Acknowledging the positive impact that the adoption of the FAIR data principles make within the GBRCLMP, the aims of this paper are to: a) review the current process of data collection, processing and management, b) assess the recently launched Tahbil – Water Quality Data Portal and its alignment with the FAIR data principles as per the Australian Research Data Commons FAIR Data Self-Assessment Tool (ARDC 2022), and c) highlight the innovative data visualisation tools that the GBRCLMP currently employ.

Data Management

Data Collection

Water quality monitoring is completed by the Water Quality & Investigations (WQI) team as part of the GBRCLMP, and by regional samplers who are trained in the techniques, methods and standards for sample collection, handling, quality assurance and quality control as prescribed by the GBRCLMP internal Standard Operating Procedures (SOPs) and outlined in the Monitoring and Sampling Manual: Environmental Protection (Water) Policy (DES 2018). Water quality samples are primarily collected by either, manual grab sampling or automatic sampling devices that chill and store the collected samples until they can be collected by a regional representative and delivered to Brisbane prior to analysis by the NATA (National Association of Testing Authorities) accredited laboratories of Queensland Health Forensic and Scientific Services, and the Queensland Government Chemistry Centre. The frequency of sediment and nutrient sampling is determined by flow events; intensive sampling (daily or every few hours) is completed in high-flow conditions, while ambient sampling (monthly) is completed in low or base-flow conditions. For pesticides, intensive sampling (daily or every few hours) is conducted during high flow events, with fortnightly or monthly sampling occurring during low flow and dry season conditions, respectively. The GBRCLMP collects more than 210, 000 discrete data points every year using manual grab sampling and automatic sampling devices. This includes approximately 43, 000 data points related to total suspended solids and nutrients, and 170, 000 for pesticides.

The WQI team also collects data using fixed in situ automated instrumentation across approximately 61 monitoring sites (as of May 2024) in GBR catchments to collect near real-time water quality data. Near real-time water quality instrumentation includes:

- YSI EXO multi-parameter Sondes deployed at approximately 19 sites (as of May 2024), measuring conductivity (µS/cm) and turbidity (NTU),
- Trios OPUS/NICO instruments installed at around 53 sites (as of May 2024), providing data on nitrate-N concentrations (mg/L), Total Suspended Solids Equivalents (mg/L), along with associated spectral values and quality metrics, and
- water level sensors, including vented/absolute and compact bubbler systems, acoustic doppler current profilers (ADCP) and computer vision stream gauging (CVSG) systems deployed at numerous sites to collect water velocity/discharge measurements to assist with load calculations.

Near real-time water quality data acquisition is hourly, with most equipment logging at 15-minute intervals. This results in a near constant data stream and the GBRCLMP is collecting over 40, 000 individual data points per day.

Discrete Data Handling

All manual and automatic grab samples are collected and stored with standardised metadata requirements. The metadata requirements vary depending on project or analyte requirements; however, the critical metadata elements that are collected for all water quality samples include:

- the project for which the sample was collected for,
- sample ID: a unique identifier that will link sampling information to the analytical result,
- site ID, and
- sample collection date and time.

Upon reception of the analysis results and laboratory reports, our team, along with a dedicated Quality Assurance/Quality Control officer, checks them to ensure their accuracy and validity. Each piece of data is then assigned a quality code according to the Quality Management Framework.

The analysis results and associated sample and results metadata are stored in HYDSTRA, a corporate database, and are the data point of truth. Metadata and data are imported daily into the WQI Microsoft® Azure Data Lake (data point of record). Microsoft® Azure Data Factory and Databricks, both Extraction, Transformation and Loading (ETL) tools, are used to process and made data easily accessible in relevant format to downstream applications such as operational dashboards in Microsoft® Power BI or visualisation tools developed using Shiny – Rstudio, a free and open source R package for developing web applications (Chang et al., 2024).

Near Real-time Data Handling

The geographic extent of the GBRCLMP, coupled with the sheer quantity of data generated by near real-time instrumentation, results in a high volume of data that is stored and not verified in a timely manner. It is not feasible to manually quality check and verify the 40, 000 individual data points that are collected each day; therefore, automation is necessary. The team employs an automation process to quality check data points as they are collected in near real-time. This process is analyte and instrument specific and checks for several quality issues such as outliers and spikes. This process uses rolling statistical analysis but also identifies values that fall outside the range specified by the manufacturer's recommendations and internally derived limit values. These form the basis of a script that flags exceeded values as anomalies. This script runs the raw data, analysing it using predefined rules specific to each probe setup and then assigns a defined quality code to each value. Once complete, the curated data is stored in the WQI Microsoft® Data Lake alongside the raw data and made available to all downstream applications. This curated data is displayed in several digital products that assist in data dissemination to the public.

Data Governance and Information Technology (IT) Standards

Products developed by the GBRCLMP follow the Queensland Government Data Governance policies ensuring data security and integrity. Development by the GBRCLMP follows the latest IT standards and best practices. The code developed by the team is housed on the Queensland Environmental Science Data (QESD) GitHub organisation, a tool for version control and collaboration on software development projects. It facilitates automatic version control and peer review of any proposed changes to code.

This approach includes a requirement that no data is ever deleted, and that a complete comprehensive history of each data point is preserved from its collection in the field to its visualisation on a web-based application.

Data Sharing with Tahbil

Tahbil - Water Quality Data Portal is a web-based Water Quality Data Portal (State of Queensland 2024; Figure 1) and was launched in May 2024. It provides access to water quality data collected by the GBRCLMP since 2005, the South East Queensland Loads Monitoring Program and other third parties (e.g. universities and Natural Resource Management groups). The data currently on Tahbil includes:

- concentration data for water quality parameters such as sediment, nutrients, and pesticides,
- calculated annual and daily Total Suspended Solids (TSS) and nutrients loads, annual TSS and nutrients yields, and,
- daily and wet season PRM.

Data downloaded from Tahbil includes metadata specific to the data sourced. For concentration data, metadata such as collection method, sampling depth, and analysis method are provided. Loads data includes metadata such as calculation method and the type of flow data used (e.g., monitored or modelled). PRM metadata includes the number of pesticides used in the calculation the first flush date. The provision of metadata allows the user to decide if the data is suitable for their purpose. It is only possible to provide this wealth of metadata because the team has put in place a system to manage its sample, result and calculated results metadata and data from the field to the calculation of its output (load and PRM calculation).

Since 2021, a seamless data management system has been developed. It starts in the field with an application designed and developed by the team that simplifies water quality data collection and enable samplers to gather detailed metadata while collecting physical samples in the field. This reduces the risk of errors during data transcription, enhancing the integrity of the collected data from start to finish. The metadata and results of the samples are available to the team and can be shared with users. This is facilitated by the full integration of separate components of the system, ensuring a hands-free approach and guaranteeing high data quality as a result of appropriate process management, quality control including corrective actions, and transparency.

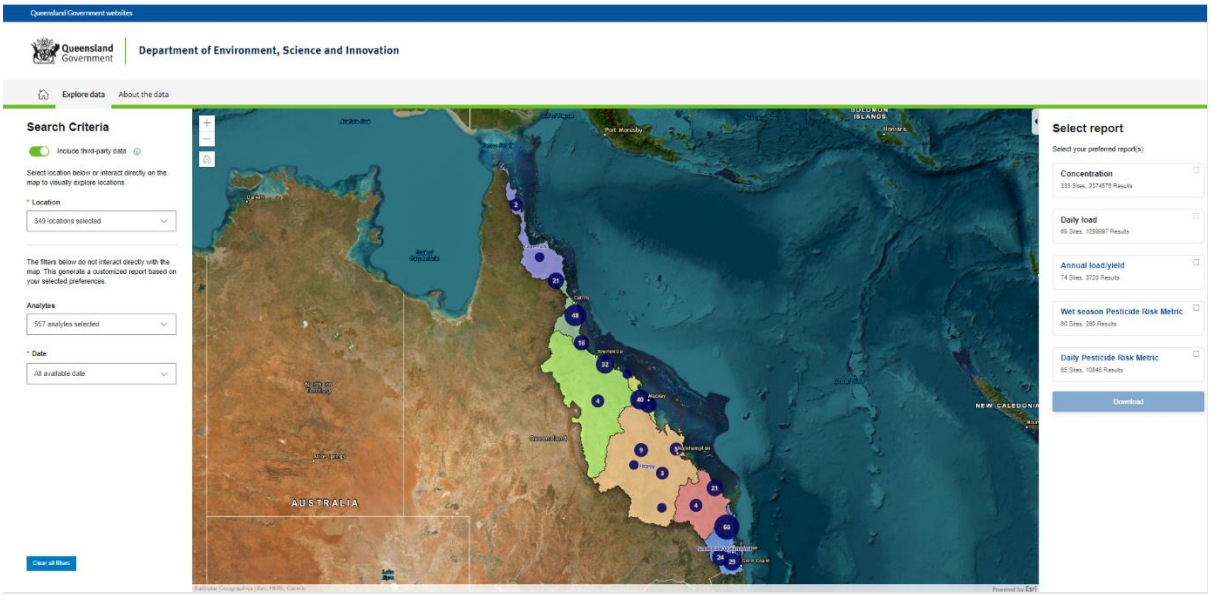


Figure 1. Tahbil – Water Quality Data Portal user interface (Water Quality & Investigations 2024)

The Australian Research Data Commons (ARDC) FAIR Data Self-Assessment Tool found Tahbil to score 73% across all FAIR criteria (Appendix A; ARDC 2022). The *Findability* of Tahbil was deemed moderate (47%). To enhance its *Findability*, future improvements could include assigning a globally unique and citable identifier, such as a DOI (Digital Object Identifier), and providing formal machine-readable metadata (Appendix A; ARDC 2022).

The *Accessibility* of Tahbil was rated high (80%), with the data being publicly accessible and the metadata record available even if the data is no longer accessible. Currently, Tahbil's data is downloadable from an online location; enhancing *Accessibility* could involve providing a standard web service Application Programming Interface (API). However, the current download method is effective and user-friendly, allowing individuals and stakeholders with all levels of computer and technology proficiency to access both online and downloadable data.

Tahbil's *Interoperability* was found to be moderate (50%). Currently, data available for download is formatted as structured, machine-readable comma-separated values (CSV) files. While providing download files in a JSON format may enhance *Interoperability*, CSV files are accessible to users of all levels of computer and technology proficiency, with software readily available on most machines. Further enhancements for increased *Interoperability* may include using standardised open and universal global identifiers that link to explanations. Additionally, providing links to other machine-readable metadata will boost overall *Interoperability*.

The *Reusability* of Tahbil is rated very high (100%). It is licensed under Creative Commons Attribution 4.0 International to promote open access while upholding copyright standards, and provenance information has been fully recorded in a machine-readable format to facilitate data reuse.

With the launch of Tahbil, the Queensland Government's Water Quality Data Portal, the public can now openly access sediment and nutrient load and concentration data, which were previously only accessible through direct data requests. This directly addresses the need for improved accessibility to data, highlighted in the international review (Warne et al., 2022) regarding sediment and nutrient data availability. Similarly, pesticide concentrations and PRM calculations, which were previously view-only, are now openly accessible via Tahbil. These both represent significant steps towards achieving FAIR data principles by the GBRCLMP and the wider WQI team at the Queensland Government Department of Environment, Science, and Innovation.

Innovative Data Visualisation Tools

Pesticide Reporting Portal (PRP)

The GBRCLMP routinely collects and analyses water samples to monitor pesticide concentrations, which can be assessed against guideline values to identify exceedances. These exceedances occur when the concentration values exceed the pesticide guideline values for the protection of specific environmental values. This information was previously reported on, and distributed within, relevant Government agencies via email and pdf; however, the web-based public Pesticide Reporting Portal (PRP) has since superseded this method of data communication. The PRP displays laboratory analysed pesticide data, remotely captured water level data and a graphical representation of the different guideline values available, including irrigation water guidelines, drinking water guidelines and aquatic ecosystem protection guidelines using Shiny - Rstudio (Figure 2).

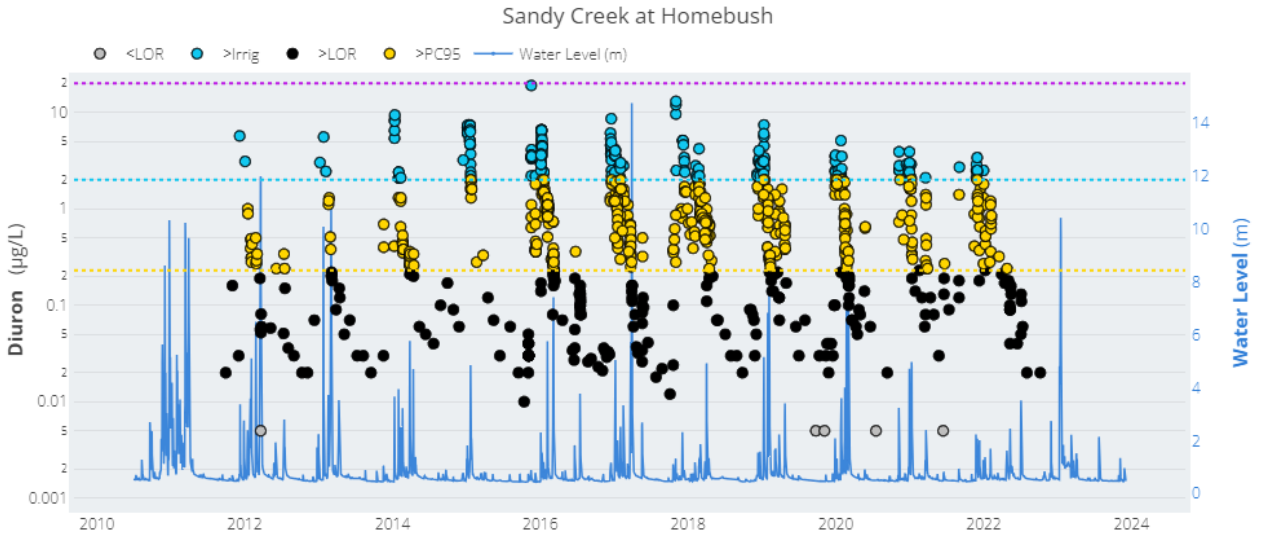


Figure 2. Graphical data from the Pesticide Reporting Portal: yellow dots indicate concentrations exceeding aquatic ecosystem protection guidelines, blue dots indicate those exceeding irrigation guidelines, and black dots represent concentrations above laboratory reporting limits but below any guideline threshold. Coloured lines correspond to the guideline values for aquatic ecosystem protection (yellow), irrigation (blue), and drinking water (purple) (Water Quality & Investigations, 2020).

Pesticide Risk Metric (PRM) Dashboard

The Pesticide Risk Metric currently includes 22 pesticides which address the majority of the toxicity of pesticides known to occur in GBR waterways (Spilsbury, Warne & Backhaus 2020; Warne et al., 2023). The combined toxicity risk of these 22 pesticides is estimated by the PRM calculator (Water Quality & Investigations 2023). Pesticide data used in the PRM calculation are obtained from the analyses of water samples collected manually or by automatic sampling equipment. The PRM is expressed as the percentage of species that are potentially affected, or conversely, protected by the combined toxicity of the 22 pesticides over a standardised wet season (Warne et al, 2020; Figure 3).

The PRM Dashboard, a public web-based portal, allows the user to explore the data in the following ways:

- Time series plots displaying the combined toxicity of the 22 pesticides for each site and year; where multiple water samples are collected in a day the risk score represents the average risk for the day.
- A summary table of average wet season PRM values for the standardized 182-day period following the first flush event (i.e., the wet season).
- The relative contribution of each pesticide to the overall toxicity risk across the wet season.

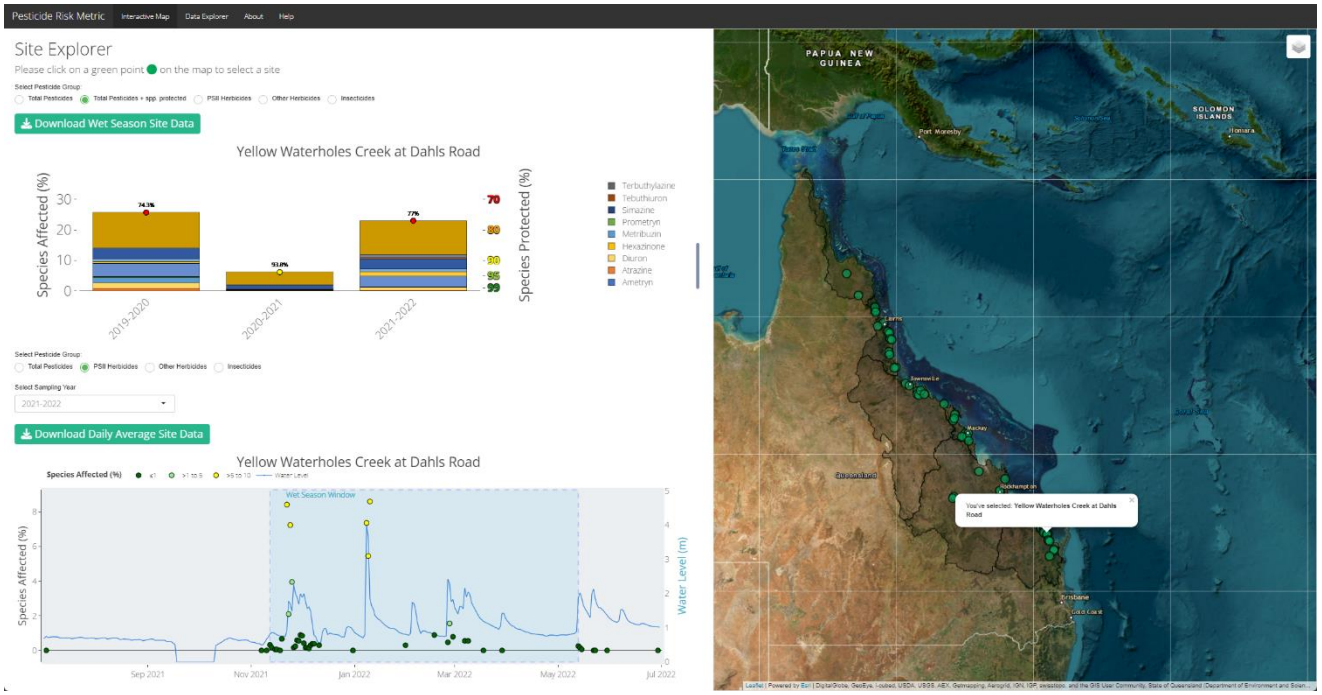


Figure 3. PRM Dashboard interactive map user interface including, top left: a graphical representation of total pesticide contributions and species affected and protected for various temporal periods. Bottom left: a daily percentage of species affected by exposure to all 22 pesticides. The shaded area indicates the 182-day risk window for the calculation of the PRM in the 2020-2021 monitoring year. Right: Map feature enabling user exploration of sampling regions and interaction with sites (Water Quality & Investigations, 2023a).

Current Challenges

The findings from the international review have highlighted additional steps that can be taken to further support the adoption of the FAIR data practices within the GBRCLMP. Overall, the high volume of data that the GBRCLMP generates stands out as the primary barrier to their adoption, which includes the following specific challenges:

- ensuring the GBRCLMP's data architecture supports storing, querying, analysing, and accessing data to meet diverse user needs,
- the diversity of metadata associated with each data source,
- the automation of quality assurance and quality control processes, that meet or exceed the same manual quality assurance checks,
- the ability to adapt to the future needs of the program,
- publishing near real-time data in a timely manner while meeting quality standards,
- staying up-to-date with ever-evolving technological advances, and
- ensuring that historical data remain incorporated and usable in current data processing streams.

The quantity of data is the primary challenge in the adoption of the FAIR data practices. Diverse data capture methods necessitate multiple pipelines to funnel data into an integrated repository. Manually collected water samples undergo processing in a NATA accredited laboratory, and therefore, the process may take months

before data appears in the platform. Conversely, near real-time data sensors offer near-instantaneous data capture but may inundate the system with a significant volume of data requiring quality control checks. One of the challenges the GBRCLMP needs to address is how to make information publicly available in a timely manner, whilst ensuring this data is quality checked and "verified" information. While access to near real-time data presents opportunities for stakeholders, it must also be recognised that the use of unverified data may introduce uncertainty and misinformation to end-users, potentially impacting the GBRCLMP's reputation. Addressing this challenge may entail providing well-labelled "provisional" data for transparency and timeliness (e.g., marking the data as 'unverified' on public facing products), subsequently replaced by "verified" information post validation.

This presents a snapshot of the most important data challenges currently encountered by the GBRCLMP and emphasises the inevitability of change and the necessity for the team to review the long-term strategy to effectively manage change and navigate rapidly evolving technologies. It is important to acknowledge that the cost involved in maintaining such systems and solving these on-going challenges are not minimal and should be viewed as integral components of the funding structure for monitoring programs.

Future Directions

The 2022 international review (Warne et al., 2022) provided recommendations for the future direction of the GBRCLMP. It was acknowledged that the GBRCLMP has made significant contributions to pesticide research, and it is thought that a realignment of the GBRCLMP's focus could yield similar research outcomes for sediments and nutrients. As previously mentioned, the launch of Tahbil, the Queensland Government's Water Quality Data Portal, has facilitated direct access to sediment and nutrient load (both daily and annual) and concentration data. With an efficient metadata and data management system, the GBRCLMP can now deliver and maintain web visualisation applications and tools in a streamlined manner (i.e., a Sediment and Nutrient Loads Dashboard and a near real-time Nitrate Dashboard) that will allow the public to visualise the data in a contextual and meaningful way.

The GBRCLMP has grown since its inception, expanding both in scope and in workload. Further to this, since June 2023, the GBRCLMP has established a collaborative model for delivery of the GBRCLMP through a Consortium of organisations. The Consortium currently includes WQI in the Queensland Department of Environment, Science and Innovation, the Reef Catchments Science Partnership in the University of Queensland, and TropWater in James Cook University. In mid-June 2024, the Consortium attended a strategic direction workshop that went some way toward formulating a strategic plan. This plan will guide the delivery of the GBRCLMP for the next five to ten years.

Conclusion

The international review completed in 2022 estimated GBRCLMP to be the largest targeted water quality monitoring program in Australia. Additionally, the review panel was very impressed by the 'esprit de corps' amongst a motivated, talented, and thoughtful Program team; the outward and problem-solving focus of the GBRCLMP team; and the quality of the design, construction, and maintenance of the monitoring stations. This review also found that the current digital products maintained by the GBRCLMP demonstrate significant achievements with regards to data FAIRness, while also outlining areas needing improvement to better align with these principles. It is the GBRCLMP's intention that all generated data will be publicly accessible and downloadable, with the launch of Tahbil instrumental in facilitating this. The next steps in the move towards data FAIRness include managing the high volume of generated data, further investment in the automation of quality assurance processes, publishing near real-time data while maintaining quality standards, and keeping pace with evolving technology. These steps will be incorporated into the GBRCLMP's strategic plan, guiding its delivery for the next five to ten years.

Acknowledgements

We would like to thank the authors, caretakers and collaborators of the digital products mentioned in this paper. While many of these individuals are listed as authors, we recognise that there have been numerous people over the years who have contributed to the development of these products. To all those who have played a part, we express our gratitude for your contributions towards the achievements outlined in this paper.

Furthermore, we extend our gratitude to the wider Water Quality & Investigations team at DESI for their invaluable contributions in field technical support, operations, and stakeholder engagement. We acknowledge their commitment and collaborative efforts, which have been instrumental in the continued success of the Great Barrier Reef Catchment Loads Monitoring Program.

References

Australian Research Data Commons (ARDC) (2022). FAIR Data Self-Assessment Tool [Online tool]. Retrieved from https://ardc.edu.au/resource/fair-data-self-assessment-tool/

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A. & Borges, B. (2024). shiny: Web Application Framework for R. R package version 1.8.1.9001, https://github.com/rstudio/shiny, https://shiny.posit.co/

Chartrand, K.M., Bryant, C.V., Carter, A.B., Ralph, P.J., Rasheed, M.A., 2016. Light thresholds to prevent dredging impacts on the Great Barrier Reef seagrass, *Zostera muelleri* ssp. *capricorni*. *Frontiers in Marine Science*, 3. doi.org/10.3389/fmars.2016.00106.

Collier, C.J., Waycott, M., Ospina, A.G., 2012. Responses of four Indo-West Pacific seagrass species to shading. *Marine Pollution Bulletin*, 65(4–9), 342–354.

DES (2018). Monitoring and Sampling Manual: Environmental Protection (Water) Policy. Brisbane: Department of Environment and Science Government, Brisbane, Australia. Retrieved from: https://environment.des.qld.gov.au/__data/assets/pdf_file/0031/89914/monitoring-sampling-manual-2018.pdf

Fabricius, K. E. (2005). Effects of terrestrial runoff on the ecology of corals and coral reefs: review and synthesis. *Marine Pollution Bulletin*, 50(2), 125-146.

Morrow, K.M., Ritson-Williams, R., Ross, C., Liles, M.R. & Paul, V.J. (2012). Macroalgal extracts induce bacterial assemblage shifts and sublethal tissue stress in Caribbean corals. *PLoS ONE*, 7(9), e44859.

Schulz, R. (2004). Field Studies on Exposure, Effects, and Risk Mitigation of Aquatic Nonpoint-Source Insecticide Pollution: A Review. *J. Environ. Qual*., 33: 419-448.

Spilsbury, Francis D., Michael St J. Warne, and Thomas Backhaus. "Risk assessment of pesticide mixtures in Australian rivers discharging to the Great Barrier Reef." *Environmental Science & Technology* 54.22 (2020): 14361-14371.

State of Queensland (2024). Tahbil - Water Quality Data Portal. Department of Environment, Science and Innovation, Queensland Government, Brisbane, Australia. Retrieved from https://apps.des.qld.gov.au/water-data-portal/

State of Queensland (2017). 2017 Scientific Consensus Statement: Land Use Impacts on Great Barrier Reef Water Quality and Ecosystem Condition. Department of Environment and Science, Queensland

Government, Brisbane, Australia. Retrieved from https://www.reefplan.qld.gov.au/__data/assets/pdf_file/0029/45992/2017-scientific-consensus-statement-summary.pdf

State of Queensland (2018). Reef 2050 Water Quality Improvement Plan 2017 – 2022. Department of Environment and Science, Queensland Government, Brisbane, Australia. Retrieved from https://www.reefplan.qld.gov.au/__data/assets/pdf_file/0017/46115/reef-2050-water-quality-improvement-plan-2017-22.pdf

The Commonwealth of Australia (2023). Reef 2050 Long-Term Sustainability Plan. Retrieved from https://www.dcceew.gov.au/sites/default/files/documents/reef-2050-long-term-sustainability-plan-2021-2025.pdf.

Warne MSt.J., Grayson R., Landers M. (2022). International Review of the Great Barrier Reef Catchment Loads Monitoring Program. Retrieved from https://www.publications.qld.gov.au/dataset/international-review-of-the-great-barrier-reef-catchment-loads-monitoring-program

Warne MSt.J., Neelamraju, C., Strauss, J., Smith, R.A., Turner, R.D.R. & Mann, R.M. (2020). Development of a method for estimating the toxicity of pesticide mixtures and a Pesticide Risk Baseline for the Reef 2050 Water Quality Improvement Plan. Brisbane, Australia: Department of Environment and Science, Queensland Government. Retrieved from: https://www.publications.qld.gov.au/ckan-publications-attachments-prod/resources/c65858f9-d7ba-4aef-aa4f-e148f950220f/pesticide-risk-baseline-project-report.pdf?ETag=a9665f53d62acabcddcc9fbe38e025b5

Warne, M. S. J., Neelamraju, C., Strauss, J., Turner, R. D. R., Smith, R. A., & Mann, R. M. (2023). Estimating the aquatic risk from exposure to up to twenty-two pesticide active ingredients in waterways discharging to the Great Barrier Reef. *Science of the Total Environment*, 892, 164632.

Water Quality & Investigations (2020). Catchment Loads Monitoring Program Pesticide Reporting Portal (Updated March 2024), Department of Environment and Science, Brisbane, Australia. Retrieved from https://pesticidereportingportal.des.qld.gov.au/

Water Quality & Investigations (2023). Pesticide Risk Metric Calculator. Department of Environment and Science, Brisbane, Australia. Retrieved from: https://prmcalculator.des.qld.gov.au/

Water Quality & Investigations (2023a). Pesticide Risk Metric Dashboard. Department of Environment and Science, Brisbane, Australia. Retrieved from: https://prmdashboard.des.qld.gov.au/

Water Quality & Investigations (2024). Tahbil - Water Quality Data Portal - Analyses, Quality and Methods for Calculating Pollutant Loads/Yields and the Pesticides Risk Metric. Department of Environment, Science and Innovation, Brisbane, Australia. Retrieved from: https://www.qld.gov.au/__data/assets/pdf_file/0024/414348/Data-Water-Portal_Methods.pdf

Wazniak, C.E., Hall, M.R., Carruthers, T.J.B., Sturgis, B., Dennison, W.C. & Orth, R.J. (2007). Linking water quality to living resources in a mid Atlantic lagoon system, USA. *Ecological Applications*, 17, S64–S78.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1-9.

Appendix A

Findable

The data has sufficiently rich metadata and a unique and persistent identifier to be easily discovered by others. This includes assigning a persistent identifier (like a DOI or Handle), having rich metadata to describe the data and making sure it is findable through disciplinary local or international discovery portals.

Does the dataset have any identifiers assigned? What is this?

- ○ Globally unique, citable, and persistent (e.g. DOI, PURL, ARK or Handle)
- ◉ Web Address (URL)
- ○ Local Identifier
- ○ No Identifier

Is the dataset identifier included in all metadata records/files describing the data?

- ○ Yes
- ◉ No

How is the data described with metadata?

- ○ Comprehensively using a formal machine-readable metadata schema
- ◉ Comprehensively, but in a text-based, non-standard format
- ○ Brief title and description
- ○ The data is not described

What type of repository or registry is the metadata record in?

- ○ Data is in one place but discoverable through several registries
- ○ Generalist public repository
- ○ Domain-specific repository
- ◉ Local institutional repository
- ○ The data is not described in any repository

47%

Accessible

The data is retrievable by humans and machines through a standardised communication protocol, with authentication and authorisation where necessary. The data does not necessarily have to be open. Data can be sensitive due to privacy concerns, national security or commercial interests. When it's not able to be open, there should be clarity and transparency around the conditions governing access and reuse.

How accessible is the data? What is this?

- ◉ Publicly accessible
- ○ Fully accessible to persons who meet explicitly stated conditions, e.g. ethics approval for sensitive data
- ○ A de-identified / modified subset of the data is publicly accessible
- ○ Embargoed access after a specified date
- ○ Unspecified conditional access e.g. contact the data custodian for access
- ○ Access to metadata only
- ○ No access to data or metadata

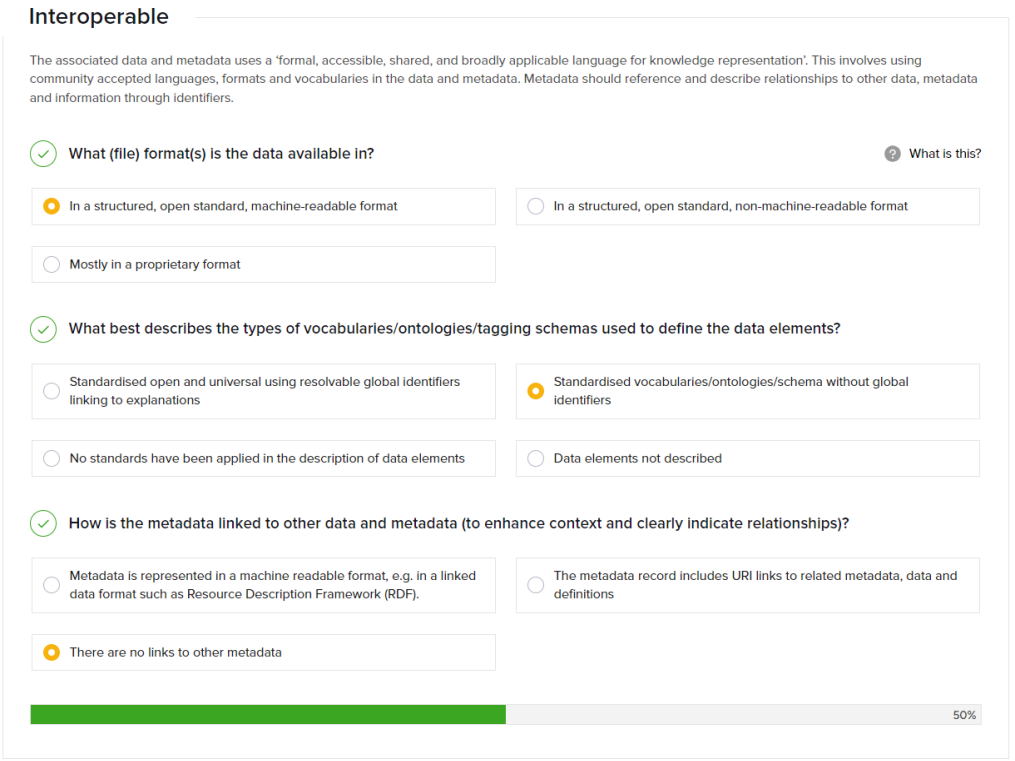
Is the data available online without requiring specialised protocols or tools once access has been approved?

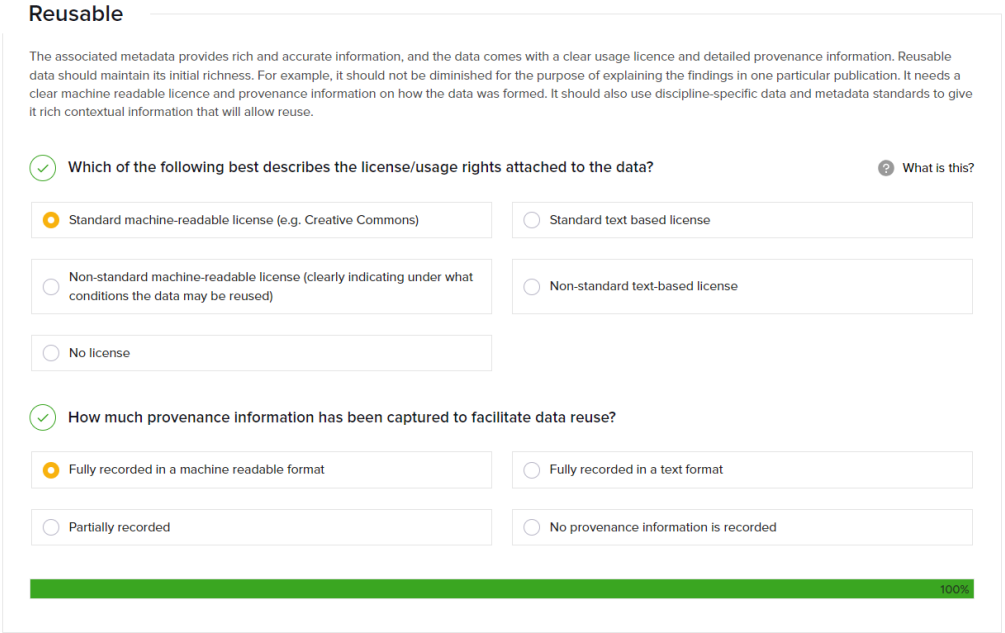
- ○ Standard web service API (e.g. OGC)
- ○ Non-standard web service (e.g. OpenAPI/Swagger/informal API)
- ◉ File download from online location
- ○ By individual arrangement
- ○ No access to data

Will the metadata record be available even if the data is no longer available?

- ◉ Yes
- ○ No
- ○ Unsure

80%





A 1 Results from the Australian Data Research Commons FAIR Self Assessment Tool for Tahbil Water Quality Data Portal