



A coordinated and consistent approach to managing and enabling soil data

Final report for research project 2.2.005

Nathan Robinson, Jude Channon, David Medyckyj-Scott, Simon Clarendon, Kay Steel, Megan Wong, John Friend, Peter Dahlhaus, Xiaodi Huang

June 2024



CONTENTS

Executive Summary	5
Objectives	6
Results	6
Introduction	8
Background	11
Methodology	16
Results	32
Discussion	76
Conclusion	84
Recommendations	85
Acknowledgements	88
References	89
APPENDIX A: The soil research data management lifecycle	92
APPENDIX B: Soil CRC data management plan and guide	96
APPENDIX C: Data maturity survey - invitation	116
APPENDIX D: Data maturity ethics application	128
APPENDIX E: Data maturity survey PLIS	153
APPENDIX F: Focus group/interview agenda	156
APPENDIX G: Data repository survey	158
APPENDIX H: Data repository ethics application (amendment)	161
APPENDIX I: Data repository survey PLIS	170
APPENDIX J: Minimum metadata elements	174
APPENDIX K: Data repository survey responses	175



PROJECT PARTICIPANTS



Charles Sturt
University



Department of
Primary Industries



Federation
University



Manaaki Whenua
Landcare Research

UNIVERSITY of
TASMANIA 

DISCLAIMER

Any opinions expressed in this document are those of the authors. They do not purport to reflect the opinions or views of the Soil CRC or its partners, agents or employees.

The Soil CRC gives no warranty or assurance and makes no representation as to the accuracy or reliability of any information or advice contained in this document, or that it is suitable for any intended use. The Soil CRC, its partners, agents and employees, disclaim any and all liability for any errors or omissions or in respect of anything or the consequences of anything done or omitted to be done in reliance upon the whole or any part of this document.

PEER REVIEW STATEMENT

The Soil CRC recognises the value of knowledge exchange and the importance of objective peer review. It is committed to encouraging and supporting its research teams in this regard.

The author(s) confirm(s) that this document has been reviewed and approved by the project's steering committee and by its program leader. These reviewers evaluated its:

- originality
- methodology
- rigour
- compliance with ethical guidelines
- conclusions against results
- conformity with the principles of the [Australian Code for the Responsible Conduct of Research](#) (NHMRC 2018), and provided constructive feedback which was considered and addressed by the author(s).

KEYWORDS

Enabling, sharing, soil data, data management, FAIR.

CITATION

Robinson, N., Channon, J., Medyckyj-Scott, D., Clarendon, S., Steel, K., Wong, M., Friend, J., Dahlhaus, P., Huang, X. (2024). A coordinated and consistent approach to managing and enabling soil data. Project 2.2.005 – Final Report. Cooperative Research Centre for High Performance Soils (Soil CRC), Newcastle, NSW. 202p.

EXECUTIVE SUMMARY

The Enabling Soil Data project (a coordinated and consistent approach to managing and enabling soil data) is a Soil CRC-supported project focused on making it easier for all Soil CRC partners, including researchers, farmer groups, growers and their advisors, to contribute their existing data and new sensor data into automated and Findable, Accessible, Interoperable, Reusable (FAIR) systems. This will be achieved by providing the Soil CRC with reliable, accessible and secure data management processes that resolve data governance, licensing, standards and protocols within and across key data provision for current and future projects. This will ensure that relevant data are managed in terms of appropriate documentation, sharing, quality, analytics, storage, and other protocols. The project commenced in March 2021 as a collaboration between three universities and two government agencies.

Six keystone activities were proposed:

1. Define roles, responsibilities, and structure of key guiding groups in data management.
2. Identify high value use cases linked to current and future Soil CRC project milestones.
3. Propose and refine technical policies and procedures (including standards) and recommendations for making soil data generated through the Soil CRC projects discoverable and reusable.
4. Establish and implement data management practices and policies for Soil CRC programs.
5. Assess impact and success in review of use cases to achieve project outcomes of sharing data for reuse and benefits to users of this soil data.
6. Engage and educate stakeholders on soil data sharing and reuse for current and future Soil CRC projects.

Research engagement from the project focused on matters linked to supporting data management and sharing. The Soil CRC (including program leaders) were engaged on aspects of data governance and where the project could make a significant contribution to data management in the Soil CRC. Feedback and suggestions were used to reshape and prioritise the direction of the project. This high-level direction and guidance set the ensuing pathway for the project.

Policies and practices were reviewed, and an aspirational policy was developed for the Soil CRC. Such a policy has clearly defined roles and responsibilities, best practices defined as exemplars for others to follow, the securing of assets generated through the Soil CRC, community adherence to rules and regulations, and transparency in reporting and processes for managing data.

An approach deployed to support best practice was the formation of a Soil CRC Data Management Plan (DMP). Existing DMPs in literature and used in practice were reviewed to identify a fit-for-purpose DMP to satisfy Soil CRC requirements. This was then shared with various stakeholders in the Soil CRC including the Commercialisation and Intellectual Property Advisory Committee. Feedback was used to refine the DMP making it less onerous while retaining links to key steps in the data management lifecycle.

To best support those managing data in the Soil CRC, a data maturity survey was undertaken to understand where participants in soil research were in their data management practices and understanding of these constructs. This would then help focus efforts on training and development to uplift the Soil CRC in research data management (RDM). A mixed-method approach was deployed with a survey and qualitative engagement through focus group and interviews. The survey ran for three months (August to November in 2022) with a follow-up focus group and interviews in March and April 2023.

To support best practice in data sharing, the identification of suitable repositories to store and manage data for FAIR purposes was deemed a priority. A Data Repository survey was undertaken focusing on partner organisations in the Soil CRC including universities, farmer groups, government agencies and industry organisations. The principle aim of the survey was to understand the suitability of these repositories as long-term storage solutions for research data generated by the Soil CRC. A combination of long- and short-form survey responses were used here to gather this information.

To develop practices and support the sharing of data and the benefits of this for current and future projects, use cases were undertaken from the direction of program leaders. These use cases were engaged to identify their data management needs, issues and how the Soil CRC can support them in achieving their project objectives. Engagement also explored the utility of the DMP for their purposes with feedback used to hone the plan further to meet user needs. Use cases complemented the data maturity survey with findings and interpretations cross-referenced with experiences and sentiments expressed through use cases.

To support this uplift in RDM skills of the Soil CRC, a range of engagement events were undertaken. These included presentations to the HDR cohort of the Soil CRC, presentations of research findings and developments to the Soil CRC and national audiences including the National Soil Science Conference, webinars on DMPs and RDM, and the provision of educational material and resources via the Soil CRC website.

OBJECTIVES

Overall purpose: To deliver a cross-program, whole of Soil CRC, collaborative solution to data sharing and management that supports the secure and reliable storage, discovery and reuse of analysis-ready data, including for use in design-ready tools and systems to assist with Program 1 and 4 project-related data analytics, modelling, decision support systems and visualisation.

Long term objective: To make soil data generated through Soil CRC projects discoverable and reusable among the Soil CRC community, and to increase its value to stakeholders through linking to decision support tools, data visualisation and decision enabling dashboards, and grower ready applications for use in decision-making processes and research innovations.

RESULTS

For the Soil CRC, given there are many stakeholders and organisations involved, data governance and management can be difficult as many will need to adhere to their own institutional policies. The complexity of the Soil CRC ecosystem was described, and an aspirational data policy defined to set expectations from a governance and technical implementation perspective. The goals were presented in the proposed data governance

policy and remain outcomes the Soil CRC should aspire to. To fulfill this, clearly defined roles and accountabilities need to be enacted with respect to governance and stewardship.

A key finding of the project was the level of understanding and practices used in data management from participants in the Soil CRC. It was clear from survey results that many were unfamiliar with RDM terminology, most projects did not have a DMP, and respondents found it difficult to see the value of data management apart from risk of data loss. There were concerns raised over licensing, security and privacy of research data. The lack of familiarity on data management was echoed in many responses and the desire for formal training in RDM. This response has been the focus of training material resources now provided to the Soil CRC from this project.

It remains unclear how capable Soil CRC participants are with delivering metadata records for their institutions or their use of persistent identifiers (PIDs) and advocating for standards including PIDs that help enable FAIR data. There is potential use and adoption of controlled vocabularies and ontologies being made available as standards in the description of Australian soils (the Australian Soil Handbook Series) through the work of the Australian National Soil Information System (ANSIS) standards project. Likewise, there has been substantial progress made through the Visualising Australasia's Soils project supported by the Soil CRC.

To support best practice in RDM, there is a need to have resourced and secure data repositories to store and manage the data as an enduring asset. It was evident from the survey on data repositories that details on repositories is not always easy for researchers to find and their practices in managing or willingness to accept third-party data is varied. While most university repositories mint digital object identifiers (DOIs) and support application programming interfaces (APIs), metadata standards and associated practices in getting data into these systems are also varied. For farmer groups, government agencies, natural resource managers (NRM) and industry organisations, there is considerable variation in how institutions and organisations manage their data and the level of documentation and resourcing to manage this.

A set of training and development material has been collated and produced for the Soil CRC to assist researchers in making data generated through the Soil CRC FAIR. Examples include a Soil CRC DMP, the Data Asset Register, guidelines on metadata, FAIR and data standards, licensing and sharing data, data archiving, privacy and identifiers, to name a few. These education resources are all accessible to members online through the Soil CRC website.

Considerable effort has been exercised in helping those in need with RDM. Webinars, training and engagement sessions, presentations and use cases have all formed part of this package of engaging and educating on soil data sharing and reuse. Feedback confirmed these activities were valued and that the level of awareness on data management has increased as a result of the project.

From the project, it was clear there were opportunities for the Soil CRC to make a significant contribution to RDM and ensure the legacy of research and data generated across projects. A data steward role would make a valuable contribution here to achieving this outcome. Likewise, an online catalogue of data and information resources from the Soil CRC (Virtual Data and Information Catalogue) would ensure value in perpetuity for the Soil CRC beyond the existing 10-year investment cycle.

INTRODUCTION

Sound data management is an important part of the responsible conduct of research. Data that you reuse, collect or create during your research is an asset which needs to be managed throughout the research project and cared for into the future. This project responds to the critical need for a cross-program, whole of Soil CRC, project on data management to ensure secure and reliable storage, sharing, analysis and visualisation for soil-related data. Soil data that is both spatially and temporally explicit will help researchers transition to data use for foresight and allow multiple outcomes from data, particularly predictive and hence decision-making capability critical to enabling farmers to optimise their soil productivity. Through co-design and trial with Soil CRC projects in Programs 1, 2 and 4, project priorities included the development of guidelines, processes and policies that support discovery and reuse of research data and make it easier for researchers, farmer groups, growers and advisors to contribute soil data including sensor data streams into automated and FAIR systems. Research explored how near real-time in-field sensor and other soil data could be used with Soil CRC-related data analytics, modelling, decision support and visualisation dashboards.

The overall purpose of the project is:

To deliver a cross-program, whole of Soil CRC, collaborative solution to data sharing and management that supports the secure and reliable storage, discovery and reuse of analysis-ready data, including for use in design-ready tools and systems to assist with Program 1 and 4 project-related data analytics, modelling, decision support systems and visualisation.

Project aims include:

- Developing a fit-for-purpose approach to data management for Soil CRC projects that will enable soil data to be discoverable and reusable to underpin key milestones linked to predictive decision-support tools and applications.
- Defining solutions for wider sharing and reuse of soil data including sensor data streams across Soil CRC programs to advance access to near real-time and on-farm data for data analytics and future machine learning purposes.
- Demonstrating how sharing and reuse of soil data between researchers, farmers and advisors through use cases that link research outcomes and data with automated and FAIR systems benefits contributors and users on this data.

By fulfilling these aims, the project will make soil data generated through Soil CRC projects discoverable and reusable among the Soil CRC community, and increase its value to stakeholders through linking to decision-support tools, data visualisation and decision-enabling dashboards and grower-ready applications for use in decision-making processes and research innovations.

DATA AND THE SOIL CRC

The Soil CRC (Cooperative Research Centre for High Performance Soils) has 39 participant organisations in a collaboration that is focused on addressing lost opportunities experienced by farmers and regional communities due to infertile and underperforming soils. The collaborative efforts bring together a range of disciplines (including social science, economics,

biology, chemistry, agronomy and soil science) to provide farmers and land managers with the knowledge and tools they need to make decisions on extremely complex soil management issues (<https://soilcra.com.au/>). This knowledge and these tools are founded on high quality research data and information.

As stated by Dahlhaus et al. (2018), the Soil CRC 'is in a unique position to establish an Australian soil data portal ... and supply federated, harmonised and standardised soil data and modelled data to the "competitive space" to greatly improve Australia's agricultural profitability' (p 155). Furthermore, the authors state that 'access to these data would greatly benefit both the research and industry partners of the Soil CRC by ensuring that all research would be built on the best available and most current data sets available' (p 155). A recommendation from this scoping study was for a data stewardship and governance model to be developed to support a data federation and the sharing and exchange of data leading to new research outcomes (knowledge) and tools that benefit farmers in bridging the soil performance gap.

Data is a valuable business/organisational asset. As such it requires:

- Stakeholder care in terms of data ownership
- Regulatory compliance
- Proper management strategy
- Trust in that management strategy
- Adequate licencing of intellectual property.

Having these principles in place will result in:

- Sound data quality
- Improved efficiency (good trustworthy information facilitates good decisions)
- Minimised risk
- Greater transparency of processes and information
- Improved discovery
- Increasing adaptation to a changing digital world.

The vision is an advanced soil data ecosystem which enables frictionless data access and sharing across the researchers, institutions and partners participating in the Soil CRC to meet the CRC's mission efficiently and effectively.

THIS PROJECT

A collaborative work program between the project partners Federation University, Manaaki Whenua Landcare Research, Department of Primary Industries, Charles Sturt University and University of Tasmania that aimed to:

- Deliver a cross-program, whole of Soil CRC, collaborative solution to data sharing and management that supports the secure and reliable storage, discovery and reuse of analysis-ready data, including for use in design-ready tools and systems to assist with Program 1 and 4 project-related data analytics, modelling, decision support systems and visualisation.

- Develop a fit-for-purpose approach to data management for Soil CRC projects that will enable soil data to be discoverable and reusable to underpin key milestones linked to predictive decision-support tools and applications.
- Define solutions for wider sharing and reuse of soil data including sensor data streams across Soil CRC programs to advance access to near real-time and on-farm data for data analytics and future machine learning purposes.
- Demonstrate how sharing and reuse of soil data between researchers, farmers and advisors through use cases that link research outcomes and data with automated and FAIR systems benefits contributors and users on this data.
- Make soil data generated through Soil CRC projects discoverable and reusable among the Soil CRC community, and to increase its value to stakeholders through linking to decision-support tools, data visualisation and decision-enabling dashboards and grower-ready applications for use in decision-making processes and research innovations.

BACKGROUND

As detailed in the introduction, this project responds to the critical need for a cross-program, whole of Soil CRC, project on data management to ensure secure and reliable storage, sharing, analysis and visualisation for soil-related data. To fulfill this need, six keystone activities were proposed:

1. Define roles, responsibilities, and structure of key guiding groups in data management.
2. Define high value use cases linked to current and future Soil CRC project milestones.
3. Propose and refine technical policies and procedures (including standards) and recommendations for making soil data generated through the Soil CRC projects discoverable and reusable.
4. Define and action data management practices and policies for Soil CRC programs.
5. Assess impact and success in review of use cases to achieve project outcomes of sharing data for reuse and benefits to users of this soil data.
6. Engage and educate on soil data sharing and reuse for current and future Soil CRC projects.

The relevant background literature to these activities is described below.

DATA GOVERNANCE AND ADVISORY GROUP

In broad terms, governance is about collective decision making. There are many varied and entangled 'levels of detail' concerning definitions of 'Data governance'. Put simply, Data governance is the exercise of authority and control over the management of data (DAMA International 2009). Governance is generally concerned with the decisions and who makes those decisions (rights) with respect to the data asset, while management involves the implementation of those decisions (Alhassan et al. 2016; Fu et al. 2011; Khatri and Brown 2010).

Data governance is central to the fabric of 'good data management'. The social architecture (the social aspects of infrastructure development and maintenance with a particular focus on governance, participation and the management of the agreements that together define how stakeholders collaborate to achieve collective goals (Box and Lemon 2015) along with technical infrastructure are the cornerstones of a data ecosystem that meets the needs of data providers and users such as those in the Soil CRC.

While data governance is critical in setting and making the decisions concerning data assets (steering), there is a need to support and encourage those responsible (rowing) for the research data collected and generated to achieve these 'collective goals' (Box and Lemon 2015). Ultimately, the purpose of data governance is to build trust in data – specifically, in how data is collected, analysed, shared, used and stored. This is where this project can make a valuable contribution to the Soil CRC in meeting its core objectives.

USE CASES TO INFORM THE DEVELOPMENT OF POLICIES AND PROCESSES

Use cases are an approach used mainly in information technology and system design to identify an optimal approach to manage a process that is fit for purpose. It is generally achieved through a sequence of steps or operations to fulfill a particular need and desired outcome. Generally, use cases have three key elements: an actor, a goal and a system or sequence of events to fulfill the goal. They are driven from a user (next or end) perspective to achieve outcomes sought by all parties, in this case, the user and the Soil CRC.

As detailed by Brush (2022), a use case should:

- Organise functional requirements
- Model the goals of the system and actor interactions
- Record paths—called scenarios—from beginning to end goals
- Describe the central flow of steps and alternate pathways
- Be multilevel, so that one use case can use the functionality of another one.

Use cases are also done for scenario planning and design purposes. For AgReFed (<https://www.agrefed.org.au/>) use cases were established with a FAIR data goal. Their approach was focused on identifying: data to be shared as FAIR data, how to make the data discoverable and accessible for research (either as open access or under specified conditions of licensing), technical solutions to make this possible, and the benefits to the user.

A use case ultimately provides the sequence of operations to deliver a meaningful outcome (goal). Use cases for the Soil CRC were focused on high-impact projects where data is recognised for its reuse and commercialisation potential and will enable multiple outcomes from the data (e.g. predictive and decision-making capabilities).

TECHNICAL AND INFORMATION (INCLUDING STANDARDS) RECOMMENDATIONS FOR MAKING SOIL DATA GENERATED THROUGH THE SOIL CRC PROJECTS DISCOVERABLE AND REUSABLE

Good data management and governance can improve the visibility and reuse potential of this data for multiple purposes such as research, development and extension in soils and agriculture. Understanding an organisation or individual's level of 'data maturity' enables the prioritisation of activities and policies that lead to improved data management and increased value of the asset. Importantly, for the data to be a valued asset it needs to be managed and curated using processes and practices that ensure data qualities are retained. To make generated soil data available and accessible for reuse purposes, making all Soil CRC data 'FAIR data' is the ideal solution to this problem. However, there are various factors that make this ambition difficult and problematic including privacy, commercial and intellectual property (IP) protection, understanding and use of data standards and vocabularies, to name a few.

Sharing of data via open access agreements such as Creative Commons (<https://creativecommons.org/>) has been increasing across the globe. Australia, unfortunately, lags other regions such as Europe in their publishing of research and data as open access (Montgomery et al. 2021). To share the data, it also needs to be somewhere that will be findable and enduring.

To support soil data access and sharing between the researchers, institutions and partners participating in the Soil CRC to meet the CRC's mission efficiently and effectively, a data repository is a seamless way of achieving this outcome. Many partners in the Soil CRC have

access to existing infrastructure such as data repositories supported through their institution. However, there are also partners who do not have such systems in place for enduring data sharing and access.

DEFINE AND ACTION DATA MANAGEMENT PRACTICES AND POLICIES

To support soil data access and sharing between the researchers, institutions and partners participating in the Soil CRC, understanding the current state of data management practices in research is fundamental. Data management is one of the key pillars of support for collaborative outcomes in research (von Davier et al. 2017). Data management in scientific research is critical from an institutional and individual researcher perspective as it increases time savings, mitigates risk of loss or uncertainty in data, and culminates in improved resource efficiencies over time (Whyte and Tedds 2011). Furthermore, it aids in preventing errors and improving quality of the analysis and provides protection of IP through proper licencing arrangements. By providing accessible data that allows others to validate and replicate findings, this can lead to powerful collaborative datasets that provide new insights that otherwise may have not been found in a single unshared dataset (Briney et al. 2020).

To embed effective data management as a worthwhile practice for an individual researcher, understanding the individual, project and institutional benefits are rarely defined. Therefore, understanding where individuals are at in their data management practices and understanding (data maturity) is critical to identifying opportunities for improvement and the setting of good practices and policies for data management in the Soil CRC. Understanding an institution's, or individual's, level of 'data maturity' enables the prioritisation of activities and policies that lead to improved data management and increased value of the data for primary and secondary use purposes.

Data that is reused, collected or created during research is an asset which needs to be managed throughout the research project and cared for into the future. By having a set of skills and knowledge of data management practices allowing work to be better organised and more efficient, you can:

- Find and understand your data when you need to use it
- Avoid unnecessary duplication of effort
- Ensure continuity of the research if project staffing changes
- Maintain the data underlying publications which allows for the validation of your findings
- Ensure your data can be readily shared which can facilitate research collaboration
- Meet data management obligations under the Australian Code for the Responsible Conduct of Research
- Avoid catastrophes such as the loss of data or violations of privacy.

A Data Management Plan (DMP) describes the activities that will be associated with the data other than the direct use of the data and may include data creation, storage, retention and disposal, documentation, long-term preservation and provisions for access, and data sharing and reuse.

ASSESS HOW USE CASES ACHIEVED PROJECT OUTCOMES OF SHARING DATA FOR REUSE AND BENEFITS TO USERS OF THIS SOIL DATA

As detailed earlier, use cases provide an opportunity to identify, optimise and support the sharing and access to data using a staged approach that is fit for purpose. The use cases are driven from a next- or end-user perspective to achieve outcomes sought by all parties including the Soil CRC.

To identify meaningful use cases for the Soil CRC community, a focus on use cases that can have functionality of another one is desired. The reuse proposition and benefits to users of this data also needs to be understood. To support this process and the identification of use cases, key areas of consideration include:

- The ownership of data within and beyond the research investment
- Managing the collected, acquired or produced data as an asset in projects
- Roles and implementation of data governance, policies and advocated practices
- Identifying next- and end-user beneficiaries of data reuse
- Data management and the use of a DMP to support the use case.

ENGAGEMENT TO INFORM AND EDUCATE ALL PROJECTS

Providing projects and the researchers of these teams with a fundamental skillset for data management is required for improved research and development outcomes. As the responsibility for making research data available and accessible generally falls to the researchers (Borghi and Van Gulick 2021), they need to be supported and educated to fulfill demands from institutions, funding agencies, and collaborative research groups such as the Soil CRC. Training and education that covers the topics pertinent to specific positions supporting the data management lifecycle are necessary. RDM courses should be integrated into the curriculum (Zhou et al. 2023) as it will provide researchers with an enduring skillset that is easily deployed for instant benefit but is also transferable between and for future jobs.

Studies including Rafiq and Ameen (2022), Borghi and Van Gulick (2021) and Bishop and Borden (2020) note the low rate of formal training on data management for researchers and consequently impacts or losses of data with a potential reuse proposition. Of concern in these studies is a general lack of knowledge of data management and associated skills and processes. Researchers' knowledge on data management has therefore largely been derived from their peers or research groups rather than from formal training (Bishop and Borden 2020). This inadvertently can lead to poor data management outcomes as learning through on-the-job work may be sub-optimal or even out of date with current best practice in data management.

One can argue that high quality, accurate, complete, consistent and reliable data is an outcome of best practice. Thus, applying accepted best practice data management methods facilitates optimal decision making or sound research outcomes. Research organisations wishing to provide value to their relationships with partners, stakeholders, contributors and the wider research community, recognise the value of good data. As such, inherently, it must also recognise the value of effective (or good) research data management (RDM). Krahe et al. (2019) state that RDM is an essential yet undervalued part of the research process that

involves the creation and stewardship of research materials to enable their use for as long as they retain value. By engaging and upskilling all participants in research projects on RDM, this will enable researchers to be better prepared to manage data they are dealing with to fulfill project commitments and unrealised benefits of the future (e.g. citation, positive reputation with funders, visibility of research and profile, and efficiencies in further research).

METHODOLOGY

DATA GOVERNANCE

Soil CRC Program (1 to 4) Leaders were engaged in May 2022 at the initiation of the project to obtain insights and suggestions on project directions, support for data governance and advisory roles, and where the project can make a significant contribution to data management in the Soil CRC. Five key questions were posed:

1. From the point of view of the Soil CRC, where do you think the ownership of that data lies within the Soil CRC and beyond?
2. What are your thoughts about an exit strategy for the CRC and data that has been collected or acquired or produced under projects?
3. What is the right fit in terms of data governance if we were to advocate something through this project to the Soil CRC?
4. Do you have any expectations of this project and for you, as a program leader, is there something we could do that would really make a big difference?
5. In terms of data management, we are starting to think about how projects/researchers manage data within the project and the use of a DMP, but beyond the project as well. Is that something that would be of interest?

DEVELOPMENT OF POLICIES AND PROCESSES OF SOIL DATA SHARING AND REUSE FOR THE SOIL CRC

A series of policies and processes were identified through the project to support the sharing of data and data management in the Soil CRC community. The approaches used and guiding principles of these are detailed below.

Aspirational Research Data Governance Policy for the Soil CRC

The goal of a data governance policy (DGP) is to facilitate discovery, management and reuse of research data thus making it easy for researchers, farmer groups, growers and advisors to contribute soil data including sensor data streams into automated and FAIR systems. Its purposes are to:

- Define the roles and responsibilities for different types of data creation and usage, cases and/or situations, and to establish clear lines of accountability and responsibility throughout the data life cycle
- Develop best practices for effective soil data management within the CRC including the use of data by secondary parties
- Protect the data against internal and external threats (e.g. security breach, breach of privacy and confidentiality, breach of usage rights)
- Ensure that those working in the CRC comply with applicable laws, regulations, and standards

- Ensure transparency regarding data management and data use by having processes associated with collecting, accessing, retrieving, sharing, managing and storing of data and tracking of data as it moves through the data life cycle.

The aspirations of a DGP should:

- Encompass a succinct DGP document which stays at a high level
- Be cognisant of the complex environment in which the CRC operates
- Be implementable
- Realise the financial and time constraints on what can be achieved in the project
- Be supported.

Data Management Plan

Existing DMPs were reviewed by project team members to identify key fields and information to capture in a DMP for the Soil CRC. Many resources were reviewed and assessed including those provided by ARDC. Key resources used in the development of a DMP for the Soil CRC were the Australian Antarctic Data Centre, Digital Curation Centre and Federation University Australia.

The DMP was drafted and reviewed by various stakeholders in the Soil CRC including the Commercialisation and Intellectual Property Advisory Committee. Feedback was used to refine the DMP to make it less onerous while retaining links to the key steps in the data management lifecycle (Appendix A).

Data Asset Register

Engagement from the project team with the Soil CRC realised the important recognition of data as an asset owned by the Soil CRC (excluding third party data used in projects). A request realised by the project team was to develop a Project Data Asset Register where projects (past, current and future) identify the key data products generated.

This register is based upon the “What data are you creating?” template of the DMP with additional fields to support the Soil CRC in the oversight of generated data assets.

Data Stewardship

Data stewardship is the oversight and management of data including data collection (acquisition, aggregation, usage, storage), deidentification and release (Rosenbaum 2010). A data steward may be a formalised role by an institution or organisation to assist in the management and guidance of good data management practices for individuals, projects or institutions.

A Soil Data Steward would be responsible for the quality and integrity, implementation and enforcement of data management within the CRC. The role could be to advise and guide CRC programs and research projects. Data owner responsibilities may be delegated to a data steward, but data owners are still the accountable party. An identified role description for a data steward was identified to fulfill this need from Wong et al. (2021) in AgReFed (<https://www.agrefed.org.au/>).

Soil Data Standards and FAIR Implementation

There are many metadata schema and resources that are used in the description of data. Common schemas include Data Catalog Vocabulary (DCAT), Dublin Core and the Data Documentation Initiative (DDI). Generally, metadata schemas are not something that researchers are aware of or need to select. Different organisations' data repositories and different specialist repositories (e.g. genomics) will use different schema/metadata profiles that are fit for users' purposes to describe data in a way that meets end user expectations.

Using machine-actionable controlled vocabularies and ontologies to describe data, including within metadata, has become a high-value activity. It helps ensure that data can be transferred between systems across the internet and that meaning is not lost when it is transferred. Using FAIR vocabularies (findable, accessible, interoperable and reusable) is a key component of the FAIR data principles (**I2** – (Meta)data use vocabularies that follow FAIR principles).

Researchers may use term (concepts) "persistent identifiers" (PIDs) from any number of vocabularies and ontologies that are fit for their purpose. An example of an agricultural controlled vocabulary is [AGROVOC](#). These should be machine-actionable themselves (i.e. FAIR) (Cox et al. 2021). Terms can be searched using engines and repositories (e.g. [Research Vocabularies Australia](#), [Ontology Lookup Service](#) and [AgroPortal](#)).

In other metadata fields, it is useful for PIDs to be used for subjects (FOR Subject codes) and roles (e.g. ISO role codes). Examples of other forms of persistent and unique identifiers that are useful in describing research to make data FAIRER include ORCID IDs, Research Organisation Registry (ROR) IDs and International Generic Sample Number (IGSN) (Klump et al. 2021).

There should also be recognition of other standards and identifiers including global soil sample identifiers where a system such as the IGSN could be used for a globally persistent identifier for physical samples.

A Virtual Data and Information Catalogue for the Soil CRC

To support an advanced soil data ecosystem and with no provision of a dedicated Soil CRC data repository beyond the lifetime of the Soil CRC, discussions in the project team quickly turned to the concept of a Virtual Data and Information Catalogue (VDIC) for the Soil CRC. A VDIC would bring data and information metadata captured in institutional repositories into a catalogue with access and permission controls in place. Initial discussions began in 2021 prior to the project being formally funded and contracted. A follow up meeting between Federation University and Manaaki Whenua Landcare Research in December 2022 mapped the soil data ecosystem for the Soil CRC with further refinements after this event. The concept of a VDIC was further refined with the benefits, potential limitations and constraints in establishing a VDIC and the value proposition for the Soil CRC detailed.

Through the conceptualisation of a VDIC it was evident that ARDC would be a key stakeholder and potential collaborator towards enabling a VDIC. Discussions began in August 2022 and continued to December 2022 on the objectives, outcomes sought, and where support and assistance can be provided by ARDC to achieve the VDIC concept.

This led to further questions on the fitness of institutional repositories to support a VDIC such as:

- How many metadata records per year?
- How many different metadata sources?

- What are these different metadata sources?
- How many of these metadata records/datasets have a digital object identifier (DOI)?
- Would you be able to mandate DOI minting for all datasets?
- How prepared are the numerous metadata stores with being able to both identify their data and provide a metadata feed?
- Does a data management framework blueprint for multi-partner programs exist?
- How does the Soil CRC deal with third party data of data for those institutions that do not have a repository?
- Have there been any past experiences/contributions towards CRCs, that is, are we missing something that has already been done/used/implemented by a CRC?
- Have there been any research into the frequency and occurrence of “lost” or “dark” data collected through CRC investments including, implementation or deployment of sustainable data governance policies?

Significant engagement and progress were made with ARDC (<https://ardc.edu.au/>) on a virtual catalogue concept with recommendations on the use of institutional repositories and, more broadly, on data management. The last meeting (13 December 2022) proved to be extremely informative with web systems/app developers/PID roadmap experts joining to provide recommendations on potential solutions for the Soil CRC.

The meeting with ARDC confirmed the critical need to meet with the Soil CRC and program leaders and get endorsement for proposed practices to be adopted and implemented for the remainder of the Soil CRC. This was an absolute priority as directions for the project will stem from confirmed Soil CRC directives.

DATA MANAGEMENT PRACTICES AND POLICIES

The project undertook two surveys. The first established the current level of understanding and practices for data management, and the second assessed the institutional repositories of the partners in the Soil CRC and their potential use for long-term storage of research data generated through the Soil CRC.

Data Maturity Survey

A mixed-methods approach was deployed to understand where participants in soil research were in terms of their data maturity, data management practices, and understanding of these constructs. The data maturity survey (Appendix B) ran for three months (August to November 2022).

The survey questionnaire focused on the following key areas:

- The types of research data being created and collected
- The use of data management plans and responsibilities for data management
- Data storage and security practices
- Data sharing and access practices
- Data ownership and licensing practices
- Intellectual property assignment and rights

- Data management practices.

The survey was founded on key data management concepts and topics from a series of pre-existing tools, publications and studies on data maturity (e.g. Marchildon et al. 2018; Data Science for Social Good's [Data Maturity Framework](#)). These topics and associated questions were defined by the authors to cover a range of data management fields to establish the level of knowledge and practices in RDM for researchers in the Soil CRC. The survey was open to all participants in the Soil CRC and included 42 core questions (and sub-questions) linked to a set of overarching topics (Table 1). Questions included a combination of multiple choice, open- and close-ended questions.

The survey was evaluated with input from three researchers working in data management, FAIR data and social research at Federation University Australia. Feedback from these researchers helped refine the wording of questions and consolidated the survey objectives and sequence of survey questions. Input into perceptions on what constitutes “good data management” practices were obtained for relevant questions in the survey from domain experts in data technologies, data management and scientific research. The web-based survey platform Qualtrics (<https://www.qualtrics.com/>) was used as the survey tool with hard copies provided for those participants who preferred to complete the survey using a paper-based method. Assistance was offered for all survey participants, extending to survey completion over the phone if required.

Ethics for this research project was obtained through Federation University Australia via the Human Research and Ethics Committee (Approval number: 2022-143).

Survey questions, the rationale for their use and associated data management theme are listed below in Table 1.

Ethics

Soil CRC project team members from over 75 projects were invited to participate in this research and to complete an online survey.

The Plain Language Information Statement (PLIS) (Appendix C) outlined the purpose of the study, the researchers involved with it, what participation in the study involved, and the ethical procedures under which the study was conducted. The PLIS emphasised that participation in this research was voluntary and specified how the research data would be used and stored.

The survey comprised of 44 questions that sought insights about key aspects of data governance and management in alignment with principles and responsibilities of the Australian Code for the Responsible Conduct of Research. The survey generally took participants approximately 15 to 20 minutes to complete.

Survey Implementation

The survey was shared online through Soil CRC program leaders to project leaders for their distribution amongst project team members. At the annual Soil CRC Conference (23–25 August 2022), the survey was released and shared with participants able to access the survey through a QR code (in presentations at the event).

After analysing survey results, a focus group (and interviews for those that could not make the focus group) were conducted in March and April 2023. These events were semi-structured with a series of key questions defined to further explore and supplement survey findings.

A list of contact names and email addresses, supplied by the Soil CRC, was used to invite participants to respond to this research.

Use cases were defined in partnership with Soil CRC program leaders after the 2022 Soil CRC conference.

Focus Group

As part of assessing the data maturity of Soil CRC participants, survey participants were provided with an opportunity to participate in a follow-up focus group or interview. The aim of the focus group was to explore several themes arising from the data maturity survey as well as to provide participants with an opportunity to discuss issues they saw in their organisations' or institutions' data management processes and ideas for improving them.

The focus group was held on 22 March 2023, in the form of a Microsoft Teams meeting. It consisted of seven participants from a range of organisations, including universities, grower groups and industry partners. Participants included those creating and collecting hard science data, such as lab or geospatial data, as well as those collecting social science data and spanned all relevant programs of the Soil CRC.

Interviews

Participants undertaking the data maturity survey were asked to indicate whether they would like to participate in a follow-up interview or focus group. The aim of the interviews was similar to that of the focus group but with an emphasis of exploring themes arising from the data maturity survey from an individual researcher perspective. Like the focus group, the individual interviews provided participants with an opportunity to discuss issues they had within their own data management practices and the challenges they faced as researchers in embedding best practice RDM into their work.

Interviews were held between April and May 2023 in the form of Microsoft Teams meetings with each individual interview participant. There were five participants interviewed in total: two from universities, one from a grower group, one from a government organisation and one PhD student.

Table 1. Survey questions, rationale (why the question was asked) and related data management theme.

No.	Question	Why the question was asked	Core theme
Q6	If you are bringing existing data to your project/s, have you noted this in the Soil CRC intellectual property (IP) register?	How many participants are complying with Soil CRC guidelines regarding IP registers?	COMPLIANCE - SOIL CRC
Q19	Do you know who owns the research data resulting from your research?	Are participants aware of data ownership rules that apply to this research investment?	COMPLIANCE - SOIL CRC
Q20	If you answered “yes”, who owns this data? If your project involves different data, please focus your answer on each dataset.	Are participants aware of data ownership rules that apply to this research investment?	COMPLIANCE - SOIL CRC
Q30	Who owns the IP rights associated with the data arising from your research?	Are participants aware of IP registers and ownership rules?	COMPLIANCE - SOIL CRC
Q13	Is there a backup or duplicate copy of the data used or produced through your research?	Are data retention and maintenance practices “acceptable” or “good”?	DATA HANDLING PROCEDURES
Q15	How do you deal with different versions of your working data files?	Are data handling practices for datasets “acceptable” or “good”?	DATA HANDLING PROCEDURES
Q39	Are there factors relating to soil data and associated data that makes their management different to other types of environmental data?	Do participants think soil data is different to any other forms of data and its management?	DATA HANDLING PROCEDURES
Q34	Did the project funding include provision for data management activities, including data sharing and publication?	What proportion of participants note this “hidden” cost in their project budget?	DATA MANAGEMENT COSTS

No.	Question	Why the question was asked	Core theme
Q35	Can you estimate how much funding you have/will spend on activities related to management of data (e.g. storage, supporting sharing of data, creation of metadata records) for this project? If - Yes, - Please specify the amount \$	How have participants estimated this?	DATA MANAGEMENT COSTS
Q36	If you selected “yes” above, is this amount spent:	How have participants estimated this?	DATA MANAGEMENT COSTS
Q37	Do you have any comments regarding the time and cost, incentives or lack of, to managing project data?	How have participants estimated this?	DATA MANAGEMENT COSTS
Q24	Will the data you collect or create raise any data privacy issues?	Are participants noting privacy issues and does this align with different types or data, program, role, organisation?	DATA PRIVACY
Q24A	If “yes”, how will these be handled?	What things are participants with data privacy issues doing to manage this issue?	DATA PRIVACY
Q28	How long after the completion of your research project must the data be retained for?	Are participants aware of data retention rules?	DATA RETENTION
Q28A	How long after the completion of your research project would you prefer the data to be retained for?	Are participants aware of data retention rules and practices?	DATA RETENTION
Q29	Who do you think is responsible for the ongoing management (preservation) of the research data resulting from your project?	Are participants aware of data retention rules and practices?	DATA RETENTION
Q5	Are you bringing any existing data to your project/s?	Are projects reusing data in their research projects?	DATA REUSE

No.	Question	Why the question was asked	Core theme
Q33	Have you heard of the “FAIR” data principles?	What proportion of participants know about “FAIR” as a construct?	DATA REUSE
Q16	How have you published, shared, or otherwise disseminated the data resulting from this research – or how do you intend to do so?	Are data sharing practices “acceptable” or “good”?	DATA SHARING
Q17	If you selected “I choose not to publish, share or disseminate the data”, please specify your reasons for this.	Are there reasons participants do not want to share the data?	DATA SHARING
Q18	Who will have access to data resulting from your research?	Who do participants think should have access to the data	DATA SHARING
Q25	If applicable, how will you manage access restrictions and data security for the project/s duration?	Are participants using “acceptable” or “good” practices?	DATA SHARING / DATA SECURITY
Q26	Where will you store your dataset at the conclusion of your research project?	To identify where the data is stored at the completion of the research and if this is a good or poor practice.	DATA STORAGE
Q27	Please provide us with the name(s) or URL of the data management repository(ies) that you are planning to deposit the data in.	Are participants using data repositories and what are these?	DATA STORAGE
Q11	What storage locations do/did you use for the data from the research associated with your project/s?	Are storage locations being used by participants “acceptable” or “standard” or “good” practice?	DATA STORAGE/SECURITY
Q12	How do you ensure the security of your data?	Are the data security practices “acceptable” or “good”?	DATA STORAGE/SECURITY
Q14	If you answered “yes” or “partial backup” to the last question, what storage	Are these storage locations being used by participants “acceptable” or “standard” or “good” practice?	DATA STORAGE/SECURITY

No.	Question	Why the question was asked	Core theme
	locations do/did you use for the backup?		
Q9	How is the data generated through your project/s going to be used?	Why are you collecting this data?	DATA USE
Q7	Do your project/s have a data management plan (DMP)?	How many participants have a project with a DMP?	DMP
Q10	Who is primarily responsible for the day-to-day research data management during the active phase of the project/s?	Who is viewed as being responsible for data management in a project/s?	DMP
Q21	Thinking about the key data sets, what data licence(s) have you used/will you use when you publish, share, or disseminate the data from your research?	Do participants know which license to use for their dataset?	LICENSING
Q22	Who created – or assisted with creating – the metadata for the data that you generated, published, shared, or otherwise disseminated?	Who creates metadata records for datasets – if at all?	METADATA
Q23	What other supporting documentation do you keep about your data?	Are participants retaining/using other information to supplement metadata records for datasets?	METADATA
Q31	Have you read your institution's research data policies about data management, data ownership, and data governance?	What proportion of participants have read these, and are they complying with these policies?	POLICIES / PROCEDURES
Q32	If "yes", to what extent do you believe you adhere to the policies?	What proportion of participants have read these, and are they complying with these policies?	POLICIES / PROCEDURES

No.	Question	Why the question was asked	Core theme
Q1	Which of the following best describes your affiliation/organisation?	Are their behavioural, attitude or training requirement differences between these organisations? (We suspect so)	SAMPLE DETAILS
Q2	Which of the following best describes your role/s across Soil CRC projects?	Are their behavioural, attitude or training requirement differences between these roles? (We suspect so)	SAMPLE DETAILS
Q3	Which of the following programs aligns to your project/s?	Are their behavioural, attitude or training requirement differences between these roles? (We suspect so)	SAMPLE DETAILS
Q38	What role do you think the Soil CRC has in assisting you and your project/s with data management in relation to, for example, policies, guidance on data licences, technical assistance, provision of tools to help with dissemination of data, provision of a data repository for long-term archiving?	Do participants think the Soil CRC should be doing something/investing in better data management?	TRAINING
Q40	Have you received any training in research data management practices (e.g. creating metadata, file naming, file versioning, data storage)?	Are participants and those managing data trained?	TRAINING
Q41	If “yes”, what training have you received and who provided it?	What training have they received and are there differences (role/program/organisation) in who has been trained?	TRAINING
Q42	If “no”, what would you most value getting training on?	What training are participants requiring?	TRAINING

No.	Question	Why the question was asked	Core theme
Q4	Please describe the types of soil data and associated data you are collecting and/or using for the research associated with your project/s.	What types of data are being collected and or used as part of soil research projects?	TYPES OF DATA
Q8	What types/file format of data has or will be collected?	This will identify the different file formats being collected and managed during the research.	TYPES OF DATA
Q25A	Are you using domain accepted, standard vocabulary terms (e.g. from lists, glossaries or ontologies) to describe the data within your datasets (e.g. your data column headers)?	Are those involved in soils related research using standards and vocabularies?	VOCABULARIES / ONTOLOGIES
Q25B	If “yes”, which vocabularies did you use?	Are these actual vocabularies, or industry methods?	VOCABULARIES / ONTOLOGIES

Data Repositories

In the original ethics application (2022-143), approval was granted to conduct focus group meetings and/or interviews to explore data management practices and provide insights that would be incorporated into use cases defined in partnership with Soil CRC program leaders.

A complementary survey that involved the collection of information on institutional data repositories from storage administrators was approved (Appendix D). Data from this repositories survey provided further insights to support the research findings about understanding the suitability of these repositories as long-term storage solutions for research data generated by the Soil CRC.

Data Maturity Focus Group and/or Interview

As specified in the ethics application for this project, a focus group and/or interview was intended to explore how successful the project had been in supporting Soil CRC projects and individuals in better data management practices. The focus group and interviews (Appendix E) enabled the collection of data with deeper insight into themes (e.g. compliance, financial costs and perceived benefits of data management) presented as selected questions in the data maturity survey.

The focus group and/or interviews (interviews were offered to interested participants unable to attend the focus group) posed questions that captured insights about themes including data management costs, data retention and reuse, data storage and repositories, data management plans, perceived benefits and costs of better data management, policies and procedure adherence, and training and education requirements. This offered new insights and understandings about data management practices enabling this research project to submit a report to the Soil CRC proposing recommendations to fulfill project aims and objectives to improve management and sharing of data across the Soil CRC.

In the data maturity survey, 16 participants expressed an interest in being involved in a focus group or interview for this research. Those who indicated their interest in further participation in this study include representation from all organisation sectors and varying roles as defined in the data maturity survey ensuring cross representation of the data maturity survey results.

A covering email and formal letter of invitation, the PLIS and consent form, and details on the focus group (date, time, online meeting location, questions) were provided to all invitees.

Data Repositories Survey

The data repositories survey comprised a set of 28 questions (Appendix D) to assess these institutional repositories and their application for long-term storage of this research data. The project partners as identified in the ethics application were invited to complete the survey. The survey content focuses on themes including: the use of permanent data identifiers, storage and allowance for third party data lodging, metadata, archival and disposal processes, application programming interfaces (APIs), support for those sharing data, and repository specifications.

A shortened version of the survey (rapid survey) was used to capture responses from organisations at the Soil CRC conference and via phone where appropriate. These responses are collated in the results with responses anonymised.

Consistent with the recruitment methods (outlined in the original ethics application), university participants were informed and invited to further participate in the research through the

completion of a brief survey. Survey responses were collated and analysed by the authors of this report.

The remaining Soil CRC members (farmer groups, industry partners and state government agencies) were also invited to participate in this research.

MAKING SOIL DATA THROUGH THE SOIL CRC DISCOVERABLE AND REUSABLE

Training and Education Materials

The data maturity survey asked two key questions:

1. Have you received any training in RDM practices (e.g., creating metadata, file naming, file versioning, data storage)?
2. If you answered “no”, what would you most value getting training on?

Survey responses to these questions, and the alignment with other questions and interpretation of these response by the project team, suggest that training and education are key priorities to improve data management practices in the Soil CRC.

Survey respondents were asked if they had received any training in RDM practices such as storing data or creating metadata. For those who did receive RDM training, it was variable in quality and breadth depending on the institution.

One participant noted that they had received “*standard university data management policy training which wasn't terribly helpful for these kinds of issues*”, whilst conversely another noted that they had received very specific training on “*...how to make your data FAIR, version control, data management plan, etc. QCIF [Queensland Cyber Infrastructure Foundation] provided for the training*”.

Those participants who had not received training were asked to identify the areas of data management they would obtain most value from. There was considerable interest for training across all listed categories, with DMP, data standards, metadata and FAIR data being the areas of training of most interest.

As noted by some participants, there was considerable interest in further training and support in data management:

A data management workshop/training course would be an excellent opportunity to better ensure data management is done correctly. This should involve working with data already created or about to be created, and sufficient time set aside.

This survey has made me realise how little I know – formally – about data management. More training in this area would be very valuable.

In response to these findings, the project collated a range of resources including pre-existing materials and resources (where applicable) for access by the Soil CRC.

RDM Frequently Asked Questions

From a variety of events conducted by the project (e.g. webinars, presentations, use cases), questions posed by researchers and project leaders were collated and organised into a series of frequently asked questions on RDM.

USE CASES FOR SOIL CRC PROJECTS

Nominated use cases (from Soil CRC program leaders) proceeded with a focus on the implementation and revision (based on feedback) of a Soil CRC DMP. These use cases were initiated through online meetings with project leaders (Table 2) where responses and feedback from project leaders helped inform and guide revision of the DMP.

Table 2. Use cases for the project.

Project	Title	Project leader
2.1.006	Matching soil performance indicators to farming systems	Nathan Robinson
2.1.008	Measuring soil microbes	Mick Rose
4.3.006	Optimising soil constraint management through computer-based learning and modelling	Chloe Lai
2_S.011	In-paddock variability of plant available water	Peter Weir

An additional use case focused on various data management challenges and issues for the Soil CRC. This use case involving the University of Tasmania (UTAS) was undertaken on “an exit strategy for managing and transferring data and information”. As UTAS could no longer continue as part of this project due to other extenuating circumstances, an interview revealed valuable insights for the Soil CRC (and those managing projects) as the Soil CRC passes into the last third of its lifespan and how to manage data through this process.

ENGAGEMENT WITH THE SOIL CRC COMMUNITY

Webinars

Two webinars hosted by the Soil CRC were undertaken by the project. Both webinars were recorded and are now accessible through the Soil CRC members login area.

National Soil Science Conference

Results of the data maturity survey were presented at the 2023 Australian Soil Science Conference in Darwin (June 2023).

Soil CRC Conference Presentation on RDM

For the 2022 Soil CRC conference in Adelaide (25 August 2022), a presentation on data management and governance was delivered. There were over 100 attendees where messages focused on valuing data as an asset, use and knowledge of DMPs, why data management matters, and sharing results of the survey on data maturity for the Soil CRC.

Workshop Presentations on Data Management to High Degree by Research (HDR) Students

To coincide with the 2022 and 2023 Soil CRC conference, presentations to the HDRs in attendance were provided on RDM and the latest tools and developments to assist HDRs with their research journey.

Engagement with Soil CRC and Program Leaders

Soil CRC program leaders were engaged at the beginning of the project to obtain insights and suggestions on project directions and where the project can make a significant contribution to data management in the Soil CRC. Feedback and suggestions were then used to reshape and prioritise the direction of the project.

Program leaders were also engaged to share the data maturity survey out through projects in their program and the identification of priority use cases for the project.

RESULTS

PRODUCTS FOR THE SOIL CRC

SOIL CRC Data Management Plan

A DMP was developed based on the data management lifecycle (Appendix A) and is now accessible to Soil CRC members online via the Soil CRC website (<https://soilcra.com.au/members-login/>). This is also presented as a separate guide to the DMP (Appendix B). A glossary of associated terms is accessible via the Research Data Management Area of the online portal (Figure 1).

SOIL CRC

Subscribe | Soil Data Portal | My Account | Search

About | Research | Education & Training | News & Events | Documents | Publications | Contact

RESEARCH DATA MANAGEMENT

Research Data Management

Science, and in particular soil science, is evolving. At the heart of science is data, but too many times data gets lost or is never seen again once the lights get switched off or the filing cabinets are emptied into the skip. We need to improve the visibility of a staggering amount of information that is currently inaccessible.

As researchers, we need to ensure a legacy for soil research data. Dr Nathan Robinson from Federation University is leading a Soil CRC project to facilitate good practice research data management for data generated by Soil CRC projects.

This page contains useful information for Soil CRC researchers and PhD students. Dr Robinson's research team will continue to add new content here as it becomes available, including useful documents, frequently asked questions and training videos and webinars.

What's in it for me?

We know that research data management can sometimes feel like just another administrative task, but we would like to propose that it's so much more than that. It can actually work for you! Here's how:

TIME POOR?
Being more organised makes it easier for you to find, understand, relay, share and use data

IS FUNDING LESS?
Avoids costs of duplication of research and facilitates improved time and resource management

RESPONSIVE TO FUTURE NEEDS
Being able to access, use and reuse data can help you respond to changes in legal or other funder requirements quickly and competitively, if needed.

KNOW WHERE YOUR DATA GOES!
Ensures that your data stays secure and in safe hands where you control access to it and use of it through proper attribution and licencing arrangements

CAN'T FIND YOUR DATA?
Operationalising your DMP reduces the risk of data loss and makes sure your data is backed up and safely stored at the start, during and after your project has ended.

NEW FINDINGS!
Good data management built into the design and ongoing life of the project facilitates robust data quality thereby allowing for more effective decision making

SO.....WHY DOES RESEARCH DATA MANAGEMENT MATTER?

Figure 1. Available online research data management resources from the Soil CRC portal.

Data Asset Register

Soil CRC research project final report: A coordinated and consistent approach to managing and enabling soil data | Page 32 of 199

The foundations of the Data Asset Register are based on the DMP and have been ratified by the Soil CRC as part of the package of information shared on 28 December 2023: Final Report Packages – new templates available. The packages (for projects and PhDs) both include the Data Asset Register form. The fields used in this asset register are listed below (Table 3).

Table 3. Data Asset Register example – now linked to final project reporting for the Soil CRC.

Name of dataset	<i>National Agricultural Land Sandy Soil Map</i>
Who is the person to contact about this data and what is their email address?	<i>Nathan Robinson (n.robinson@federation.edu.au)</i>
Description of the data you collected.	<i>This is an aggregated map of sandy soils from existing state and territory government mapping with review and refinement to suit project requirements.</i>
What has the data been used for?	<i>The map has been used in reporting for the Soil CRC and in presentations by project team members on their aligned research.</i>
Where is the data set stored?	<i>The data is stored at the Federation University under the CeRDI SharePoint folder for the project.</i>
What format is this dataset in?	<i>ArcGIS shapefile</i>
Volume - How large is the dataset?	<i>250MB</i>
How can the Soil CRC access this data set?	<i>The data will be loaded onto the Fed Uni FigShare repository. Access controls can then be administered from there under Soil CRC direction.</i>
How can the data set be shared?	<i>Currently the data can be shared via email/DropBox/OneDrive. This will be replaced by FigShare soon.</i>
Does the dataset contain any personal or sensitive data?	<i>No</i>
Does the dataset contain any third party data? If so, are you licenced to use and share that?	<i>The data used in the creation of the map will be acknowledged in the metadata record. All is licensed under Creative Commons.</i>
Is there any additional information (samples, field notes, etc.) that need to be stored with this dataset?	<i>There is a milestone report (Milestone 7: Detailed plan for mapping and grouping of sands) which has background context to the mapping methodology used and metadata.</i>
Are there any risks/impacts identified with this dataset	<i>No</i>

Training and Education Materials

Whilst it is not the intention of this project to create the “super data manager” in scientific research, it is important that all researchers have a basic set of skills in RDM to improve how they manage their data to support findability, reuse, collaboration, publication, and general integrity of research outcomes. A set of training and education resources (Table 4) were created or collated for the Soil CRC community to access and use in their data management activities.

Table 4. Training and educational materials supplied to the Soil CRC.

Training / education need	Response or activity	Implemented (✓ or ✗)
Data management plan (DMP)	DMP developed and revised from Soil CRC Research and Adoption Committee (RAC) and use case feedback. This is accessible via the Soil CRC members area. A guide to the DMP has been also created to support users of the DMP. This is accessible via the Soil CRC members area. A glossary of Data Management Terminology has been produced and is accessible via the Soil CRC members area.	✓
Metadata	Links are provided on guides and information provided by ARDC on metadata. For institutional metadata requirements, best to consult with data management librarians or support staff. https://ardc.edu.au/resource/metadata/ https://ardc.edu.au/resource/storing-metadata/	✓
Data repositories	Insights from institutional repositories have been collated into an information sheet. Links to the repositories and key contacts are also included. A data repository decision-tree to assist and guide users in repository selection for data has been produced as a fact sheet. A glossary of the repositories and links to further details and information are included.	✓
How to make your data FAIR (Findable, Accessible, Interoperable, Reusable)	Resources from ARDC on FAIR have been collated and shared. Links to other online resources also provided for those interested in FAIR data principles.	✓
Data standards (including vocabularies)	Summary information is provided on soil data standards and vocabularies that can be accessed and used.	✓
Data security	The National Health and Medical Research Council (NHMRC) have a guide on the management of data including safety, security and confidentiality and that responsibilities align to institutional policies.	✓

Training / education need	Response or activity	Implemented (✓ or ✗)
Data licenses and sharing	Resources from ARDC and online (via web) have been collated and shared. Links to other online resources are provided for those interested in data licensing. Links have been provided to Soil CRC guides on project management (e.g. IP management and training) accessible through the Soil CRC members login. A license selection decision tree has been created and shared with the Soil CRC to support those managing data in making these decisions.	✓
Data privacy	Web links to useful information on data privacy and dealing with sensitive data, for those interested in de-identifying data (from ARDC) including the Five Safes framework. For further information, refer to institutional policies and procedures as they will differ in compliance of legislative requirements for states and territories. https://www.youtube.com/watch?v=wsKNDiB_WW8 https://ardc.edu.au/resource/sensitive-data/ https://ardc.edu.au/resource/identifiable-data/ https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/	✓
Data archiving	Web links are provided to a video on why to archive data, what to store and where. A data repository decision tree has been created to guide users in the archiving of data to a repository. https://vimeo.com/293572872 https://www.deakin.edu.au/library/research/manage-data/store/what-data-do-i-need-to-keep-and-for-how-long https://libguides.anu.edu.au/c.php?g=881167&p=6340584	✓
Data identifiers	A link below is also provided to support researchers in assigning identifiers to data, software or research including: ORCiD, DOIs, RAiD, PURL, etc. https://ardc.edu.au/resource/citation-and-identifiers/	✓

Training / education need	Response or activity	Implemented (✓ or ✗)
Final naming conventions	File naming conventions and data versioning guides. https://authors.library.caltech.edu/records/mmnpf-cez11 https://researchtoolkit.library.curtin.edu.au/data/research-data-management/file-management/ https://www.ands.org.au/working-with-data/data-management/data-versioning	✓
Data requirements - publishers	A set of slides are provided with details on various journals research data policies and data availability statements. Authors are encouraged to seek guidance from their institution/organisation or funder if they have queries.	✓
Data maturity – self assessment	A Research Data Management scorecard was developed for use by individuals and/or projects to assess their data maturity. This included 11 equations with self-assessment ranked from low to advanced.	✓
Top 10 tips for data management	A set of 10 helpful suggestions are provided to help users to manage data better.	✓

Research Data Management Frequently Asked Questions

A set of 13 questions were collated with responses to assist those in managing research data.

1. What is data management?

Data management is the practical processes used to generate, collect, collate, organise, store, maintain, share, archive and destroy data. It also includes how to use or reuse data, publish data and access data.

2. Are data management and data governance the same thing?

No. Data governance refers to a collection of processes, roles, policies, standards, and metrics that ensure the effective and efficient use of data. Data management is more the operationalising of data governance.

3. What are my responsibilities with regards to managing my research data?

Researchers, together with their institutions, are responsible for the management and future potential of their research data. Researchers are also responsible for the quality and integrity of their research data. The NHMRC offer guidelines for researchers in the Australian Code for the Responsible Conduct of Research (<https://www.arc.gov.au/about-arc/program-policies/research-integrity/australian-code-responsible-conduct-research-2018>).

4. What is a data management plan (DMP) and why should one be used?

A DMP is a plan that clearly outlines actions that will be taken throughout the data life cycle. DMP requirements may be defined differently by different funders, programs, or organisations so to see specific requirements the organisation or funder contract needs to be consulted. DMPs are becoming a standard requirement of funded projects but can also help to design research projects so that high quality data is collected/generated and helps meet the research outcome. There is a DMP template on the Soil CRC portal.

5. What is a data licence, and which one should I choose?

A data licence is a legal arrangement from the owner or repository of the data specifying what users can do with that data. There are different forms of data licences that offer various levels of protections for data. Most commonly, copyright and creative commons licences are used. More about data licencing and decision making can be found at: <https://howtofair.dk/how-to-fair/data-licences/>. Institutions may also have a preferred licence they wish for researchers to use. If you plan to use third party data or associated data, review the data license to determine any usage constraints (or whether you have the right to use the data at all). This is particularly important if you plan to merge the data with other data or derive new data from the third party data, as some licenses may impose restrictions.

6. What is metadata?

Metadata refers to documentation or information associated with a data set. It may be embedded in the data itself or exist independently. Metadata typically describes details such as the ownership, purpose, methods, organisation, usage conditions and technical specifications of the data.

7. What is a digital object identifier (DOI) and how do I get one?

A DOI is an electronic “tag” created to uniquely identify objects or data. Often in institutions, a librarian or data steward may be able to assist in creating one. Alternatively, a Google search on creating DOIs will offer many options to create a DOI (e.g. DataCite or ResearchGate).

8. What is a persistent identifier (PID) and how do I get one?

A PID is a unique and long-lasting reference that allows for continued access to a digital object. Examples of persistent identifier systems include DOIs, handles, and Archival Resources Keys (ARKs). PIDs support interoperability and the reliable citation of digital content. Institution librarians may be able to help in generating a PID. Alternatively, websites such as ORCID and ARDC may also be of use in helping to generate a PID.

9. What is an ORCID ID and why do I need one?

An ORCID ID is an open digital identifier, like a reference number, that is unique to each researcher. Any researcher can apply for an ORCID ID. They are free and offer a great way for others to be able to view a listing of a researcher's body of work in one place.

10. Do I need to keep or share all of my data?

No, but it really depends on the contractual requirements. Some contracts are very clear on which data must be preserved and shared. Others less so. As a rule, it would be wise to keep and/or share data that cannot be replicated, that have been used for publications, or that have long term value. If users plan on sharing data, they should adopt open file formats and data standards, where they exist.

11. I collect sensitive data, does that need to be shared?

This is not a question that has a simple yes/no answer. It really depends on the type of sensitive data collected. There are very strict guidelines that determine whether sensitive data can be shared, and different types of sensitive data are governed by different legal and ethical guidelines. To get a better understanding in relation to the legal and ethical considerations for the data type, access further information here: <https://ardc.edu.au/resource-hub/working-with-sensitive-data/>.

12. I would like to publish a paper, and the journal is asking for the data. Can/should I share it?

It can be argued that good science should be backed up by good transparency of data, but only you can decide whether you want to share your data or not with a journal publisher. However, when it comes to publishing your research, you will probably want to reference your data. It may be a requirement to deposit the data in an open data repository. If you are sharing your data, make sure you assign a Digital Object Identifier (DOI) to your data so you can cite it in publications and reports. Most journals now provide standards for how to cite datasets (and most data repositories automatically generate a citation and DOI when a dataset is published, which includes the data authors, the repository where the data are archived).

13. What do I do with my data once my project is finished?

At the end of your project, you should store your data in a data repository or archive. This means the data is available to be used by others. The data should be stored even if it cannot be made openly available to others e.g. for legal reasons. Your institution can offer guidance on repository selection, deposition procedures, and any additional requirements. Ensure either you or your project leader, informs the Soil CRC Secretariat about any relevant dataset deposited in a data repository.

DATA MANAGEMENT AND THE SOIL CRC

The Soil CRC is complex with many stakeholders and organisations involved. Some of these will have their own data governance and data management practice in place. A challenge for

the Soil CRC and this project is to identify a solution for working in this ecosystem. There are various factors and considerations with respect to research data including governance and ownership, policies, processes, and standards.

Similar to other data ecosystems, such as in health data (Figure 2), the Soil CRC ecosystem will be complex. By defining this ecosystem and the roles and responsibilities for soil data governance and soil data management, the practices and policies required for this ecosystem to operate will become apparent.

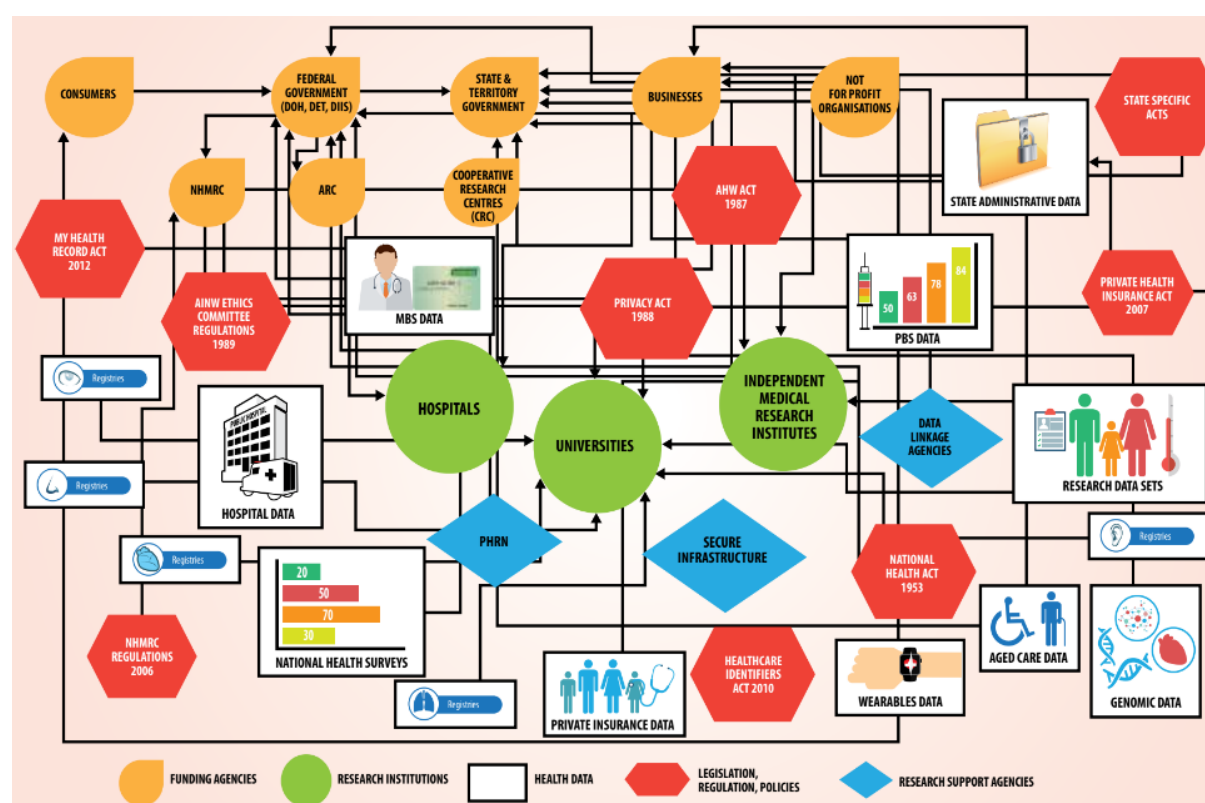


Figure 2. The Australian Health Data Ecosystem (from Srinivasan et al. 2018).

Soil Data Governance in the CRC: Goals and Principles

Benefits-based goals

- Maximise the fair use and re-use of the soil data collected by the CRC.
- Minimise or eliminate re-collection or reworking of data.
- Enable evidence based research.
- Improve traceability and repeatability of research.
- Optimise investment in data collection and researcher effectiveness.
- Increase trust in the data.
- Ensure transparency regarding data management and data use with stakeholders.
- Improve the soil data management maturity of the CRC.

- Break down data and people silos.
- Understand data use and reuse, and how these benefit meeting the goals of the CRC.
- Facilitate new research and innovation.

Risk-based goals

- Improve soil data security particularly with respect to privacy concerns.
- Improve data quality and accountability for data quality.
- Ensuring soil data collected/held by the CRC is used properly.
- Comply with regulatory requirements.
- Promote accountability.
- Standardise processes, practices and behaviours across the CRC.
- Provide effective oversight regarding data gathering, management including curation and data use.
- Protect institutional and individual rights and interests.

Principles

- Data is recognised and valued as a critical asset of the CRC.
- Data should be findable, accessible, interoperable and reusable (FAIR) to ensure it is soundly managed.
- Data should be usable and reusable through a shared understanding of what the data represents and the conditions of access and use communicated clearly.
- Data should be trustworthy and of known quality.
- The roles, responsibilities and accountabilities of those participating in the data life cycle are defined and understood.
- Data should be protected from loss, and unauthorised use and disclosure.
- Data governance and associated activities will be ongoing throughout the life of the CRC.

SOIL CRC ASPIRATIONAL DATA POLICY

An aspirational data policy for the Soil CRC should set reasonable expectations and can be delivered (implementable) from a technology and administrative perspective. Ideally the policy defines a vision for what can be delivered and enables stewards, stakeholders and administrators to see the policy through to fruition.

There are many core actions (define, implement and monitor) and areas of data governance that could be considered (roles and responsibilities, policies, processes and procedures, standards, strategy, technologies, guidelines, requirements) in an aspirational policy (Alhassan et al. 2016). There are also the key elements of what constitutes good data

governance as the people, policies, process and products (AIHW 2022). Considered in the context of the outcomes the Soil CRC should aspire to, these goals include:

- Developing a culture of “applied best practices” for effective soil data management within the CRC, including the use of data by secondary parties
- Defining the key roles and responsibilities to establish clear lines of accountability and responsibility throughout the data life cycle leading to FAIR data
- Protecting the data against internal and external threats (e.g. security breach, breach of privacy and confidentiality, breach of usage rights)
- Ensuring transparency regarding data management and data use by having processes associated with collecting, accessing, retrieving, sharing, managing and storing of data and tracking of data as it moves through the data life cycle – the DMP
- Ensuring that those working in the CRC comply with applicable laws, regulations, and standards.

The goals were presented in the proposed DGP and remain outcomes the Soil CRC should aspire to.

Technical and information policies detailed for AgReFed (MacLeod et al. 2020) should be adhered to where possible, including the information, computational and engineering and technology policies.

Membership policy is accessible to all partners and participants in the Soil CRC. Guiding principles of membership (active participation, independence and autonomy, responsible conduct of research) are advocated under cooperative principles (Wong et al. 2021).

Strategic and business decisions will be administered by the Soil CRC (accountable) with advice from participants in the Soil CRC as requested. Likewise funding and finance will be administered by the Soil CRC with advocacy from project partners and participants as required.

As the Soil CRC is the owner of data created through research projects, they will guide projects in the management of data through an assessment and alignment of institutional policies and procedures with goals of the Soil CRC (FAIR data). This may be facilitated through a Soil Data Steward. A **Custodian** (institution or project leader/data manager) will act on behalf of the Soil CRC in fulfillment of responsibilities defined in the Management of Data and Information in Research: A guide supporting the Australian Code for the Responsible Conduct of Research (NHMRC and ARC 2019).

A proposed set of roles and responsibilities (Table 6) is provided as a “Responsibility Assignment Matrix” (RAM), where the four key responsibilities (responsible, accountable, consulted, and informed) are considered for all defined roles. The order in which these occur in the proposed administration of data management in the Soil CRC should be considered. Note that the roles of Custodian and Data Creators have been assigned to the project leader or data manager (institution dependent). Data specialists are project or institutional dependent and are aligned with projects and under the guidance of a project leader or data manager.

From interviews with program leaders and the Soil CRC, it was evident that the appetite for a comprehensive data governance and management framework was tempered by a need to keep support staff and overheads low in the Soil CRC. What was also consistent in the responses was the need for solutions that may include an ongoing service to support the Soil CRC and projects in managing research data. Given the Soil CRC has reached its half-life,

the opportunity to undertake and implement a Soil CRC data management strategy appears limited.

Consequently, we recommend that an aspirational policy be adopted by the Soil CRC. In this document, the phrase “soil data” encompasses not only soil data but any associated research data. The policy itself applies to all soil data collected, supplied, derived and published within the Soil CRC program. More formally:

- Raw scientific data directly arising from a measurement, observation and/or experiment
- Derived data which has been subject to some form of standard or automated data procedure such as modelling or analysis
- Published data (i.e. data which underpins a publication and from which scientific conclusions have been derived or data which is provided as a product to third parties)
- Raw, derived or published data supplied to CRC funded researchers by third parties and CRC Partners for the purpose of use in CRC funded research.

Out of scope of this policy are:

- Software or code associated with the capture, extraction, transformation or loading of data into tools, or with the analysis and presentation of the data
- Administrative data required for the operation of the CRC
- Physical collections of items such as soil samples and cores
- The governance of any IT systems associated with the collection, management or sharing of soil data and associated data.

Roles and accountabilities

Table 5 is a list of proposed soil data governance roles. A more detailed version of this is presented using a ‘Responsibility assignment matrix’ e.g. **RACI**, where the four key responsibilities most typically used are: *responsible*, *accountable*, *consulted*, and *informed* are considered (Table 6). The order that these occur in the proposed governance structure also needs to be considered.

Table 5. Soil data governance roles and responsibilities.

Soil Data Governance Role	Soil Data Governance Responsibility	Comments
Soil Data Creators	Data Creators are researchers and technical staff who create (and collect) the original research soil data during the course of their involvement in the CRC (e.g. through participation in projects).	
Soil Data Owners	Data Owners are responsible for ensuring effective local protocols are in place to guide the appropriate use of their soil data. Access to, and use of, data will generally be administered by the appropriate Data Owner. Data Owners (or delegated Data Stewards) are also responsible for ensuring that all legal, regulatory, and policy requirements are met in relation to the specific soil data. Data Owners are responsible for ensuring that data conforms to Soil CRC data management policies, processes and standards. This includes maintaining data quality, documentation, and adherence to best practices in data governance within the CRC framework.	
Soil Data Custodian	Any individual or organisational unit such as a CRC Program is the Custodian of the soil data and any new data or information derived from the data. Custodians are responsible for maintaining and curating soil data in accordance with Soil CRC soil data policies, processes and standards. The Custodian ensures the integrity, security, and proper utilisation of the soil data throughout its lifecycle.	

Soil Data Governance Role	Soil Data Governance Responsibility	Comments
Soil Data Stewards	Responsible for the quality, integrity, implementation and enforcement of data management within their CRC program or CRC research project. Data owner responsibilities may be delegated to data stewards, but data owners are still the accountable party.	Need to decide if Stewards operate at project or program level. Custodian could be project and stewards at program level. But not stewards assume some ongoing involvement (funded) in a program.
Data Specialists	Data Specialists are technical experts who assist in the ongoing management of the soil data or information assets (e.g. associated tools, developers, database managers, IT). This includes overseeing associated tools, collaborating with developers, managing databases, and handling IT infrastructure relevant to soil data management. Their expertise and contributions are integral to ensuring the effective governance, security, and usability of soil data throughout its lifecycle within the CRC framework.	
Soil Data Users	Data users are any researchers, contractors, consultants or end users who access, amend, delete, extract, and analyse soil data while working on projects in the CRC. Data users are not generally involved in the governance process, however they have a responsibility to adhere to data governance policies, guidelines, and standards established by the CRC to ensure the responsible and ethical use of soil data during their project activities.	Can be direct and secondary users.
CRC Data Governance Steering Committee	The Data Governance Steering Committee is responsible for the overall management of the Soil CRC Data Governance. It is the governing body responsible for the strategic guidance of the soil data governance, prioritisation for the data governance projects and initiatives, approval of organisation-wide data policies and standards, as well as enabling ongoing support, understanding and awareness of the data governance program.	

Soil Data Governance Role	Soil Data Governance Responsibility	Comments
CRC Data Governance Lead	The Data Governance Leader is responsible for the overall implementation and management of the soil data management and governance.	Who is this? It's the key role. It's the person who does the heavy lifting of data governance.
CRC Executive Sponsor	A senior person within the Soil CRC who is charged with coordinating data governance activities and programs. Serves as the conduit between the data governance lead, the Steering Committee and the most senior stakeholders and the CRC Board and is authorised to make decisions and take actions.	Could report to the Exec instead of the Board. Could be Program leader (e.g. Richard Doyle) but may require higher seniority within the Soil CRC (board or Exec).
Soil Data Stakeholder	An individual or group that could affect, or be affected by, data governance decisions, processes, policies, standards, etc.	Would encompass data creators, owners, users, data suppliers (grower association). This could replace Data Users (to be discussed). Want to give the Partners and Associates some say in the data governance.

Table 6. Responsibility Assignment Matrix soil data roles and responsibilities.

Project Tasks	Data steward	Soil CRC administration	Program leaders	Project leaders	Data managers	Repository administrators	Stakeholders
Initiation							
Data policy administration	R	A	C	I	I	I	I
Data steward administration		A	R	C			
Data management advocacy	I	R	A				
Data management							
Training and education	R	A	C	I	I	I	I
DMP implementation	R	C	C	A	I	I	I
Data Asset Register							
Administration	R	A	C	I	I		
Population	R	C	I	A	I	I	I
Repositories							
Assessment	R	A	C	I	I	C	I
Liaison and support	R	C	I			A	
FAIR data							
Assessment	R	A	C	I	I	I	I
Liaison and support	R	C	I			A	

Responsible

Accountable

Consulted

Informed

Data Stewardship – A Soil Data Steward

1. Purpose

The purpose of the Soil Data Steward is to ensure that the projects, programs and individuals of the Soil CRC community are supported in adherence of data policies and standards. The Soil Data Steward may also act as a point of contact for external stakeholders wishing to engage with the collective assets of the Soil CRC.

2. Scope

The Soil Data Steward is responsible for the alignment of the collective data assets of the Soil CRC (i.e. those made available through research projects).

The Soil Data Steward reports to the Soil CRC IP Officer.

Specific responsibilities include:

- Supporting CRC projects and programs by providing advice, guidance and technical support such as advising on DMP implementation and collection of data asset metadata (such as working with new projects on developing DMPs to help navigate their data management journey), assisting with data asset registration, assisting with repository assessments, and delivering training and education to all interested parties within the Soil CRC.
- Assessing and maintain the Data Assets Register for the Soil CRC including FAIR compliance (see the AgReFed FAIR Data Policy and AgReFed Trusted Repository Policy).
- Liaising and collaborating with institutional repositories to provide relevant and up-to-date information and guidance for data assets to be shared through these systems.
- Providing outputs as guided by the Soil CRC IP officer and CEO.

3. Outputs

The Soil Data Steward will provide outputs as directed by the Soil CRC including maintaining a record of data assets that are approved for sharing through approved repositories, and tracking and reporting on progress, success and lessons learned through quarterly reports (e.g., updates on data assets realised, risks, challenges and barriers to participation, suggestions and opportunities).

4. Role filling

The Soil Data Steward may be an existing member of the Soil CRC community or belong to a third party tasked with the role.

The role will be filled by expressions of interest from candidates meeting selection criteria. The appointee will need to be supported through the terms of employment and processes of their employer.

5. Review

This Role Description will be reviewed on an annual basis and may be reviewed, updated or adapted at any time by the Soil CRC.

Soil Data Standards and FAIR Implementation

It remains unclear how capable Soil CRC participants are in filling out common fields in metadata records for their institution (e.g. Appendix G), or the use of PIDs in metadata content

where helpful, or to advocate for standards including PIDs that help enable FAIR data, research sharing and efficiencies across the CRC and broader research community. Given the results of the data maturity survey, we suspect that the likelihood of this occurring is low and when done the quality of the records may be questionable.

In other metadata fields, it is useful for PIDs to be used for subjects (FOR Subject codes) and roles (e.g. ISO role codes). Other examples of other forms of persistent and unique identifiers useful in describing research to help make data 'more FAIR' include ORCID IDs, ROR IDs (Research Organisation Registry) codes and physical sample codes (ISGN) (Klump et al. 2021).

The Visualising Australasia's Soils project (Dahlhaus et al. 2021) has used a bespoke set of soil data standards and vocabularies. These vocabularies have been refined and developed in alignment with Australian National Soil Information System (ANSIS), however there are also specific vocabularies developed to accommodate the diversity of soil observations and measurements submitted and shared by project partners in the Soil CRC.

Recently there have been several controlled vocabularies and ontologies being made available as representations of well-known standards used in the description of Australian soils (the Australian Soil Handbook Series) through the work of the ANSIS standards project. These are being made discoverable through Research Vocabularies Australia by searching the subject 'Soil Science' or through the ANSIS webpage (<https://ansis.net/standards/>). Terms (concepts) in these vocabularies and ontologies each have PIDs, notations, description, and links to related terms. Where relevant to describe researchers' data, researchers are encouraged to use these term PIDs to help describe their data. They can be included in the subject or keyword field of metadata records or within provenance descriptions in metadata or the description fields. Furthermore, they could be used within dataset/data products themselves.

Data Maturity Survey

The Soil CRC needs an overarching approach to data governance and management (i.e., data maturity) to improve these practices in projects that collect and manage soil-related data. Data management and governance can improve the visibility and reuse potential of soil-related research data for multiple purposes such as research, development and extension that can support farmers in their pursuit of optimised soil productivity.

To improve data management practices and policies for Soil CRC projects and programs, a benchmark was required to identify where the Soil CRC partners are with respect to its current RDM practices. The survey questions were based on similar surveys conducted elsewhere as well as best data management practices.

The survey was launched as part of a presentation and shared with HDR students at the 2022 Soil CRC conference. Eighty-seven participants opened the survey, with 66 responses included in the analysis. Of these included responses, 45 were fully completed with the remaining 21 partially completed to varying degrees (13–80% completion), from individuals working in a university (64%), government agency (13%), grower group (15%) or NRM agency, industry partner, CRC staff (8%). These groups reflect the major industry breakdown of the 39 participants in the Soil CRC. Researchers (30%) or PhD candidates (38%) were the major participants in the survey. Survey participants aligned equally across each of the four investment programs (the fifth being the PhD student program) with more than 15 participants from each. The breakdown of organisation and roles for survey participants is provided in Table 7.

Table 7. Survey participants organisation/affiliation and role in Soil CRC project/s.

Organisation	University	Government agency	Grower group	Industry partner	NRM agency	CRC staff
Role						
CRC Program Leader	3					
Project leader	4	2	1	1		
Researcher	10	6	3	2		
Post-doctoral researcher	1					
Technical support					1	
Administrative assistant	1					1
PhD candidate	23					
PhD supervisor [#]	1					
Extension			4			
Unknown			1			

[#] Four participants identified a secondary or tertiary role as a PhD supervisor. Other secondary roles included researcher and technical support.

The results are presented sequentially (question by question) with brief question insights discussed (as applicable).

For the context of this report, over two thirds (70%) of participants were from a university, likely due to the large contingent of participants in the Soil CRC being university partners. Roles were spread evenly with a tendency for more people having a role in a PhD program (candidate or supervisor). The remaining role descriptions were otherwise evenly spread.

Q3. Which of the following programs aligns to your project/s?

		Program 1	Program 2	Program 3	Program 4	PhD Program
Q3	Program 1	17	4	5	6	5
Q3	Program 2	4	24	4	5	9
Q3	Program 3	5	4	16	10	4
Q3	Program 4	6	5	10	25	5
Q3	PhD program	6	11	7	10	23

There was a spread of participants across all programs of the Soil CRC including the PhD program. Many participants had roles across more than one program (e.g. the PhD students where they generally also aligned themselves with one of the four major investment programs).

Q4. What types of soil data and associated data you are collecting and/or using for the research associated with your project/s

	Program 1	Program 2	Program 3	Program 4	PhD Program
Chemical	5	12	12	17	8
Physical	1	7	7	14	3
Biological	0	4	2	4	1
Profile	1	4	0	0	2
Water	1	10	2	5	4
Geophysical	0	1	0	1	0
Imagery	0	3	1	2	4
Plant	1	2	3	6	0
Climate	0	3	0	1	1
Terrain	0	2	0	0	1
Landuse/management	0	1	0	1	0
Trial/treatments	0	3	1	6	0
Sensor	0	2	0	0	0
Economics	2	0	0	0	1
Social	7	1	2	2	4

The data that participants were collecting or using were predominantly soil chemistry, physical and soil/water data. This question was initially asked to help identify what data architecture would be required to house and manage the different types of data, however this appears difficult due to the diversity of data.

Q5. Are you bringing any existing data to your project/s?

	Program 1	Program 2	Program 3	Program 4	PhD Program
Yes	8	9	6	7	6
No	7	12	9	15	16

37% of respondents were bringing existing data to their project.

Q6. If you are bringing existing data to your project/s, have you noted this in the Soil CRC intellectual property register?

Four respondents noted they were bringing data to their project in the IP register, eight had not, and seven people chose not to answer the question despite saying “yes” to bringing existing data. This highlights the need for IP management and the recognition of data as an asset.

Q7. Do your project/s have a data management plan?

Option	No.
Yes	18
No	17
No but I know what a DMP plan is	18
I don't know	5

31% (n = 18) of the survey respondents had used a DMP in their project while the remaining respondents (68%; n = 39) did not have a DMP or did not know if they had one for their project. This suggests a training or education issue, or possibly lack of communication on data management and responsibilities.

Q8. What types/file format of data has or will be collected?

Data type	No.
Numerical or statistical (e.g. CSV, MAT, XLS, SPSS, R, etc.)	50
Text (e.g. TXT, DOC, PDF, RTF, HTML, XML)	37
Geospatial	21
Instrument specific	12
Models	9
3D imaging	2
Images	26
Video	4
Other – please specify	3

The high number of respondents to images appeared to align with the high number of reports and papers.

Q9. How is the data generated through your project/s going to be used?

Use	No.
Unsure	3
VAS	2
Extension/communication	7
Reports/papers	17
Thesis	7
Toolkit/application	2
Engagement	2
Model development	8
Modelling	7
Visualisation purposes	5
Reuse/sharing	5
Sensor	3

The use of data was varied, with publications (e.g. reports or papers) being the most common.

Q10. Who is primarily responsible for the day-to-day research data management during the active phase of the project/s?

Day-to-day data management comes down to the individual. 100% (n = 20) of PhD candidates recognised this as their responsibility, with project leaders (n = 6) and researchers (n = 7) also recognising it was their responsibility. Survey respondents in other roles viewed the project leader as being responsible. Some participants did not know who was responsible for day-to-day data management (n = 6), with most of these participants being researchers (n = 5).

Q11. What storage locations do/did you use for the data from the research associated with your project/s?

Storage Location	No.
My work computer	43
Home computer	10
External hard drive/Portable storage/Flash drive (USB)	25
Storage provided by my department or school	6
Storage provided by my institution or organisation	22
Cloud-based data storage e.g. Dropbox, MS OneDrive, Google Drive	34
Physical hardcopy (retained in boxes, cabinets, etc.)	15

Most participants used their work computer, cloud-based storage or external storage device. There was quite a varied response, with some of the storage solutions not enduring or safe.

Q12. How do you ensure the security of your data?

Security method	No.
Data is protected via an authentication process	21
Password protected	34
Data is de-identified (i.e., only anonymous data stored)	9
Data is encrypted	2
Data is not secured	14

Data security options were varied with most adhering to secure practices. There were 14 participants who noted their data was not secure. This raises the question as to whether participants understand what data security is and what should or should not be secured.

Q13. Is there a backup or duplicate copy of the data used or produced through your research?

	No.
Yes	39
No	3
Partial back up of data	7
Unsure	8

Backing up or duplicating data is important for researchers to ensure their valuable data is not lost.

Q14. If you answered “yes” or “partial backup” to the last question, what storage locations do/did you use for the backup?

Storage location	No.
Home computer	1
External hard drive/Portable storage/Flash drive (USB)	12
Storage provided by my department or school	4
Storage provided by my institution or organisation	5
Cloud-based data storage e.g. Dropbox, MS OneDrive, Google Drive	16
Physical hardcopy (retained in boxes, cabinets, etc.)	4
Shared with multiple members	1

Backing up data may occur through existing data storage and sharing locations such as cloud-based storage, which is not an ideal for active data or backups. This raises the question about possible constraints around programs (backup processes) that organisations are forced to use.

Q15. How do you deal with different versions of your working data files?

Option	No.
I save multiple files and update the files with different names/naming convention	37
This is done automatically (e.g. Dropbox/SharePoint)	5
I use a versioning control program	5
Version management is done by the database I am using	6
I am not using versioning control	7
I only use one dataset version	4

Most respondents were using a form of version control.

Q16. How have you published, shared, or otherwise disseminated the data resulting from this research – or how do you intend to do so?

Option	No.
I have not released any data and don't intend to	17
Personal website or social media	2
Via email	7
Data server/service/web portal established by the project	4
Data server/service/web portal established by a CRC project e.g. Visualising Australasia's Soils	8
Institutional data repository	5
Disciplinary data repository	2
Dataset published in data journal	10
I choose not to publish, share or disseminate the data	1

30% of participants indicated that they were not sharing data, however it is not known how much of that data is shareable. Publishing data in a journal article is a more traditional way of data storage. The best current option is the disciplinary data repository. This raises the questions as to whether a disciplinary data repository is within scope of the Soil CRC, whether these data disciplinary repositories are Australia specific, and whether data needs to be uploaded to Research Data Alliance (RDA) or ARDC (as these are both catalogues, not repositories).

Q18. Who will have access to data resulting from your research?

Option	No.
Researchers in other Soil CRC Projects	29
Soil CRC partners and stakeholders	25
Other researchers, outside the Soil CRC	12
Industry organisations	11
The general public	13

Participants who completed this part of the survey indicated a willingness to share data within the greater Soil CRC. There are questions about the sharing of data outside the Soil CRC, if this was agreed to (in contracts) or discussed with the Soil CRC given the Soil CRC own data generated in funded projects.

Q19. Do you know who owns the research data resulting from your research?

Option	No.
Yes	18
No	6
Unsure	29

This question links back to the IP issues mentioned above and represents an education issue.

Q21. Thinking about the key data sets, what data licence(s) have you used/will you use when you publish, share, or disseminate the data from your research?

Option	No.
I have/plan to have data licences established for the data	4
I didn't licence the data	2
I don't plan on licensing the data	13
I haven't decided which licences(s) to use but have access to information to help us make that decision	5
I, are unsure which licence(s) to use	10
I am unsure what a data license is	9

It is interesting to note the high number of responses about not planning to license the data, perhaps due to commercial reasons. This requires further investigation and is a training issue. It is also uncertain whether those who responded that they did not license their data are referring to, for example, social data, that will be shared or not.

Q22. Who created – or assisted with creating – the metadata for the data that you generated, published, shared, or otherwise disseminated?

Option	No.
I haven't created any metadata	15
I have yet to create any metadata	5
Myself	16
Researcher involved in the project	12
Administrative assistant	1
Data specialist from my institution	3
An individual from another CRC project	0
PhD candidate	3
Post-doctoral researcher	2

From these responses it is clear that metadata is being created, but earlier responses show that there is no understanding of what metadata is or what to do with it.

Q23. What other supporting documentation do you keep about your data?

Option	No.
Description of methodology/protocol (e.g. method and materials)	35
Data dictionary/coding manual	6
Sources for third party data used in the project (source, terms of use, access restrictions)	10
Instrument metadata (manufacture, model, parameters, calibration)	8
Additional metadata (e.g. technical, preservation, copyright, access restrictions)	4
I do not record any additional data documentation	6

The high number of responses for methods/materials indicates that most respondents are publishing their research.

Q24. Will the data you collect or create raise any data privacy issues?

Option	No.
Yes	15
No	28

Two thirds of respondents indicated that the data they collect or obtain have no privacy issues.

Q25a. Are you using domain accepted, standard vocabulary terms (e.g. from lists, glossaries or ontologies) to describe the data within your datasets (e.g. your data column headers)?

Option	No.
No, I don't	17
Yes I do, but descriptions of these terms are not available online	2
Yes I do, they are available online (e.g. as webpages or pdf)	5
Yes I do, they are available as semantic-web resource (e.g. in CropOntology, AGROVOC or AgroPortal)	2
Unsure what data standards and vocabularies are	15

Overall, most respondents indicated that they do not use standards or semantic-web technologies, with only a few using some standard vocabularies.

Q26. Where will you store your dataset at the CONCLUSION of your research project?

Option	No.
Home computer	10
External hard drive/Portable storage/Flash drive (USB)	16
Storage provided by my department or school	6
Storage server, provided by my institution	22
Cloud-based data storage e.g. Dropbox, MS OneDrive, Google Drive	19
Physical hardcopy retained (in boxes, cabinets, etc.)	4
A repository nominated by the Soil CRC	9

Responses were varied, with a potential correlation with Q11 responses.

Q28. How long after the completion of your research project MUST the data be retained for?

Option	No.
1 year	2
5 years	5
10 years	0
Indefinitely	6
I do not know	21

The high proportion of responses of "I don't know" indicates a training and education need about data retention.

Q28a. How long after the completion of your research project would you PREFER the data to be retained for?

Option	No.
1 year	1
5 years	3
10 years	0
Indefinitely	22
I have no preference	16

This question confirms that respondents are uncertain about data retention. A training response is required, including the retention differences between types of data (e.g. social, ethical, economic, soil observations/measurements).

Q29. Who do you think is responsible for the ongoing management (preservation) of the research data resulting from your project?

Option	No.
Myself	7
Project leader	3
Researcher	0
Administrative assistant	0
Technical support	1
PhD candidate	2
Post-doctoral researcher	0
The institution you work for/conduct research for	13
The Soil CRC	4
I don't know	9

These responses suggest a lack of knowledge, hence “institution” or “I don’t know” responses. They also indicate a potential lack of understanding about data value, and whether there is a duty of care versus responsibility issue.

Q31. Have you read your institution’s research data policies about data management, data ownership, and data governance?

Option	No.
Yes	21
No	10
I am unsure whether there are any policies	11

Half of the participants responded “yes” and the other half “no” or “unsure”. This strongly suggests an education and training requirement from institutions.

Q32. If you answered “yes”, to what extent do you believe you adhere to the policies?

Option	No.
Fully	4
Mostly	13
Partially	4

Respondents believed they adhered to the policies of their institution.

Q33. Have you heard of the FAIR data principles?

Option	No.
Yes	25
No	18

Just over half of respondents had heard of the FAIR principles.

Q34. Did the project funding include provision for data management activities, including data sharing and publication?

Option	No.
Yes	5
No	14
Unsure	24

Five out of 43 participants had funding included. The rest did not or did not know. The “I don’t know” respondents were predominantly university students who may not be expected to know, as it is an assumed part of their project. The 14 participants responding with a “no” raises the question of who pays for data management.

Q35. Can you estimate how much funding you have/will spend on activities related to management of data (e.g. storage, supporting sharing of data, creation of metadata records) for this project?

Option	No.
Yes	5
No, we are not really managing the project data	8
No, I/we do not know	18
No, we have no idea, but it has probably been quite a lot, e.g. of people time	11

There were few participants who accounted for the costs of data management, with the majority unsure of the costs of data management. This must be a major issue in research projects effectively masking the true cost of the research.

Q40. Have you received any training in research data management practices (e.g., creating metadata, file naming, file versioning, data storage)?

Option	No.
Yes	9
No	31
Unsure	2

The majority of participants indicated that they had not received RDM training. This response is a true reflection of the whole survey and sums up where the problem lies with data management.

Q42. If you answered “no”, what would you most value getting training on?

Option	No.
Data management plan	26
Metadata	16
Data repositories	12
How to make your data FAIR (Findable, Accessible, Interoperable, Reusable)	16
Data standards (including vocabularies)	17
Data security	12
Data licenses and sharing	13
Data privacy	8
Data archiving	12
Data identifiers	3

These training requests align with many of the other questions in the survey.

Key Findings

- There is a lack of understanding or familiarity with the terminology of RDM among participants.
- Most Soil CRC projects had not completed a DMP.
- The major barriers to good data management were identified as lack of time and data management not being factored into the cost of the project, mirroring similar surveys and RDM literature (e.g. Bloemers and Montesanti, 2020).
- Most participants had not received any formal training in RDM practices.
- Researcher respondents found it difficult to see the value of data management apart from mitigating the risk of data loss, which was perceived as minimal, in managing their research data well.

- Concerns were raised over licencing, security and privacy of research data collected.

Focus Group

As part of assessing the data maturity of Soil CRC participants, participants were provided with an opportunity to participate in a follow-up focus group or interview. The aim of the focus group was to explore several themes arising from the data maturity survey as well as to provide participants with an opportunity to discuss issues they saw in their organisations' or institutions' data management processes and ideas for improving them.

The focus group was held on 22 March 2023 in the form of a Microsoft Teams meeting. It consisted of seven participants from a range of organisations, including universities, grower groups and industry partners. Participants included those creating and collecting hard science data, such as lab or geospatial data, as well as those collecting social science data, and spanned all relevant programs of the Soil CRC.

Key Findings

- There is a large variance in how institutions and organisations manage their data (i.e. diverse groups have different methods and often these methods are not streamlined or properly organised).
- Managing soil data effectively is a common challenge faced by researchers.
- Where and how to store soil data is an additional challenge researchers face.
- Whilst processes for managing and storing data are documented in some institutions, how these processes are managed in an ongoing capacity remain unclear (e.g. what happens to a data management plan or a data set after it is submitted).
- There is overwhelming support for the Soil CRC project to enable better data management of data arising from the Soil CRC.

Interviews

Participants undertaking the data maturity survey were asked to indicate whether they would like to participate in follow-up interviews. The aim of the interviews was similar to those of the focus group but with a focus of exploring themes arising from the data maturity survey from an individual researcher perspective. Like the focus group, the individual interviews provided participants with an opportunity to discuss issues they had within their own data management practices and the challenges they faced as researchers in embedding best practice RDM into their work.

Interviews were held between April 2023 and May 2023 in the form of Microsoft Teams meeting with each individual interview participant. There were five participants interviewed in total: two from universities, one from a grower group, one from a government organisation and one PhD student.

Key Findings:

- There is often no funding for data management or the funding that is provided is rolled into a single category of research and it is unclear how much (time and money) can be spent on data management.

- Any time taken to do data management properly is seen as time taken away from doing the actual research.
- Researchers sharing and accessing data internally is less of an issue, but accessing and sharing data externally is a challenge.
- Researchers have concerns about sharing raw data without contextual information
- Metadata records are not extensive or ideal.
- Privacy and security of data remain a concern.
- There is a recognition of evolving contractual obligations to include good data management in projects.

A Virtual Data and Information Catalogue for the Soil CRC

As the Soil CRC is focused not only on the science, but also how to translate this into real-world outputs that are actionable and transferable to the land managers of Australia, it is imperative that a catalogue is not only a system (portal) where research data can be discovered, but also where information resources (e.g. reports, journal publications, tools and associated decision support systems) can be accessed and located through the one shopfront. This will underpin the Soil CRC objectives to maximise investment impact by ensuring farmers have access to the latest findings and discoveries online in an easily understood and transferable form to increase their soil performance.

The proposed approach/operation for a Soil CRC virtual catalogue was also discussed including “appropriateness” of different repositories to store and facilitate access and findability of these datasets, reports and publications. This led to an initial whiteboard session to map the ecosystem (Figure 3) and, later, a refined ecosystem model (Figure 4).

What became apparent in these discussions was the complexity of relationships and interactions of the systems and the abundance of repositories already in existence for members of the Soil CRC (Figure 5). Much like the Australian Health Data Ecosystem (Figure 2), there are a variety of dependencies and relationships that require cultivation and investment to build and maintain these linkages.

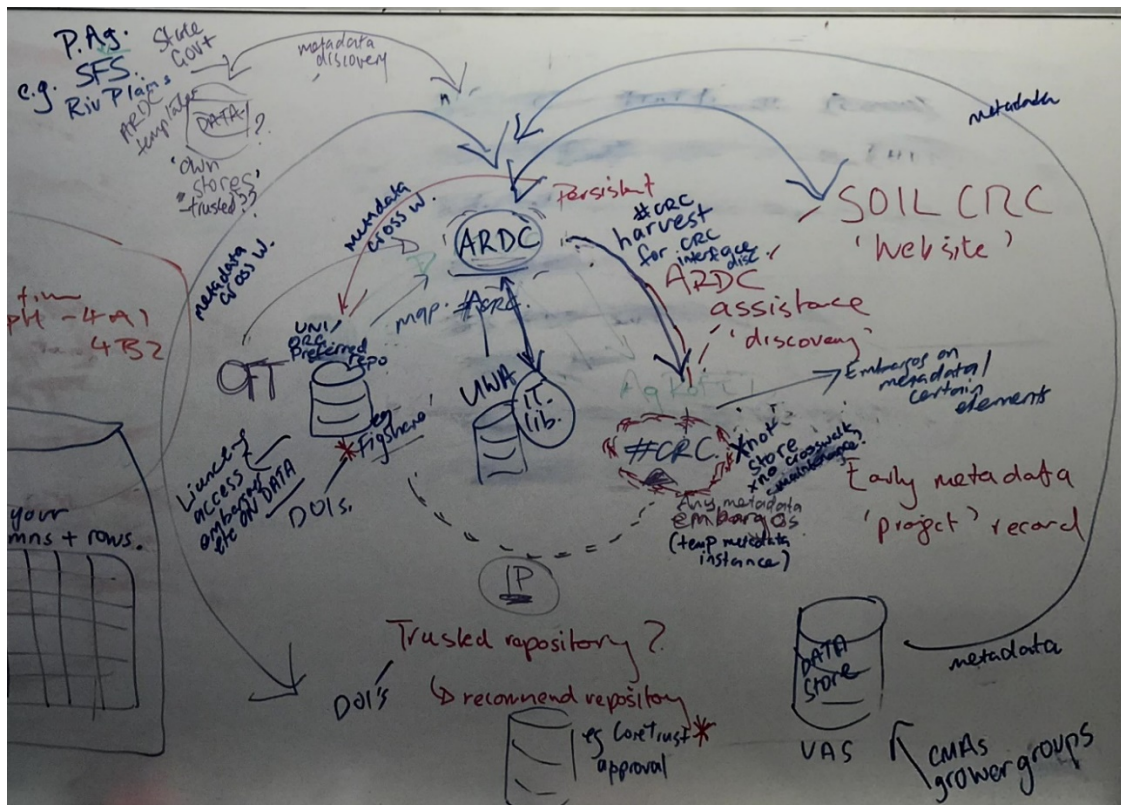


Figure 3. Initial project workings on creation of a soil data ecosystem for the Soil CRC.

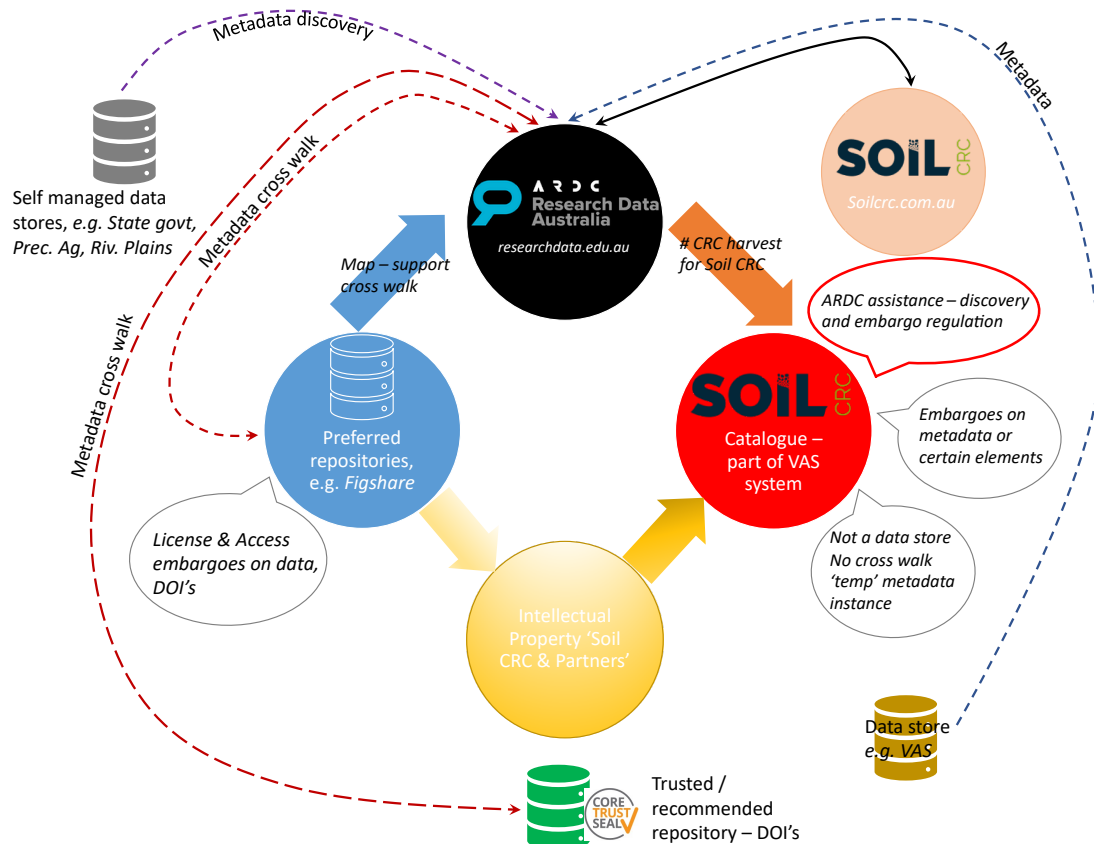


Figure 4. Refined soil data ecosystem model for the Soil CRC.

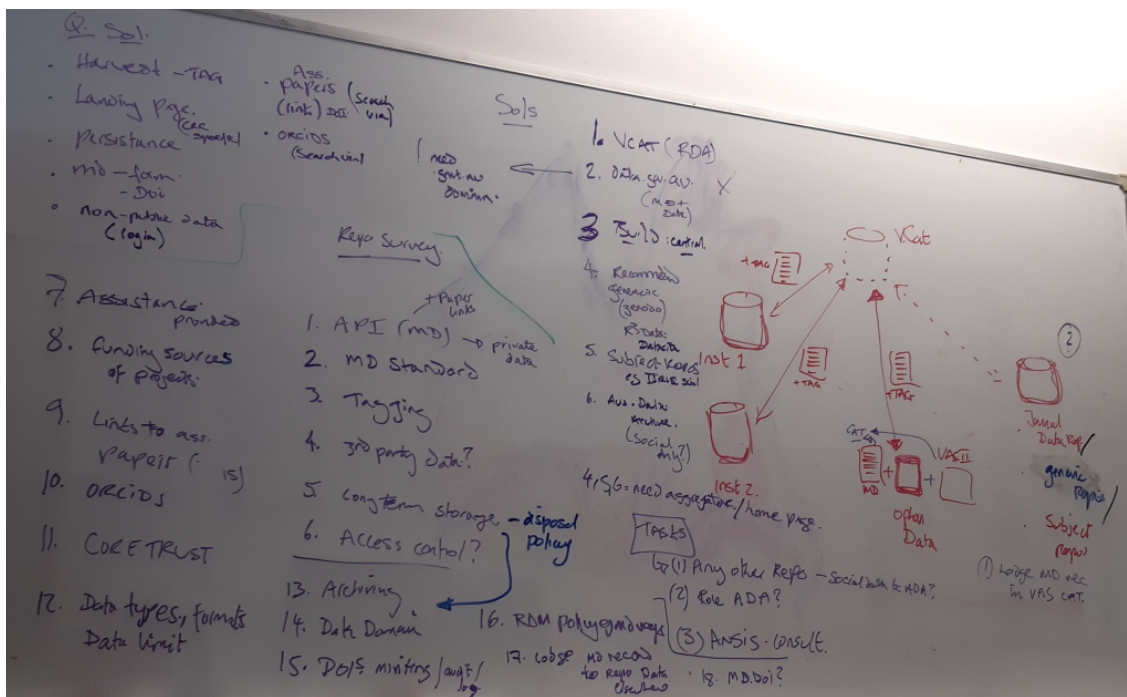


Figure 5. Institutional repository questions (left half) and the proposed virtual catalogue and repositories system (right half).

The project team discussions focused on a VDIC solution that could be tested and evaluated internally. The priority was to harvest metadata from institutional repositories for data records but also publications (e.g. journal papers and reports) that would be in these repositories. Data.gov.au (founded on Making Australian Government Data Available: MAGDA) is an example where MAGDA as an open architecture catalogue can pull data from many different sources into one easily searchable catalogue. In parallel, it was also suggested that a Comprehensive Knowledge Archive Network (CKAN) instance could be established to test the concept.

With the concept of a VDIC as an “aggregator”, there were questions on what instructions should be provided from the Soil CRC to projects on storing and sharing data created. Suggestions such as Zenodo, R3 or DataCite were discussed along with the Australian Data Archive or even the ISRIC Soil Data Hub as potential enduring solutions for soil data. All options would need development time and a Soil CRC landing page with search capabilities. ANSIS was also noted as a potential system that should be consulted.

The options of a CKAN instance and a MAGDA instance were identified as potential options from a proof-of-concept perspective.

It was recognised that the VDIC was dependent upon institutional repositories having connectors to these architectures and that the repositories were able to share metadata records with distinguishable tags with a VDIC for Soil CRC records.

The benefits of a VDIC are many with the principal advantages being:

- Low upfront costs and ongoing overheads as the catalogue benefits from existing systems and repositories are already in existence

- Data and information is made FAIR rapidly: while there may be embargo periods for data or some information products, this can be exposed early to the Soil CRC community which are bound to roles and responsibilities detailed in the Soil CRC
- Records can be exposed to other catalogues and systems such as Research Data Australia that enable the public to know data and information exists, and from here access can be arranged by negotiation or direct access which will support further R&D benefits
- The Soil CRC community can benefit by reuse of data and information within the lifetime of the Soil CRC, with new collaborations and research outcomes being achieved through the knowledge that this data and information exists
- On cessation of the Soil CRC, records will still be maintained and accessible through institutional repositories. Many records would then default to published and would be accessible via RDA as an enduring legacy of the Soil CRC. The concept of a virtual collection accessible for the Soil CRC was also advocated through RDA at this point
- Development of connections through metadata harvesters for the various repositories would benefit the development of other Virtual Catalogues beyond the CRC, therefore ensuring a legacy contribution of the Soil CRC to data ecosystems across Australia.

The limitations or potential constraints to establishing a VDIC include:

- The upfront costs and development of metadata harvesters: while still minor in contrast to repository establishment and upkeep, there is still financial resources required to formalise the soil data ecosystem
- Stability and changes to institutional repositories: as there are various changes occurring to data repositories across institutes, these changes and updates may result in data harvesters having to be maintained and updated accordingly
- Ongoing maintenance and costs with administration: while these would be relatively minor, to ensure the catalogue is functioning and responding to user requirements and needs, there would be support administration required
- There is duplication with other catalogues such as RDA which already harvests records from approved repositories across Australia, however, a VDIC would enable direct access and findability of such data and information in a controlled environment (e.g. member login)
- A functioning catalogue needs records to be of benefit to users, requiring projects and the Soil CRC to ensure that data and information assets are recognised and recorded promptly to be viewed in the system
- As the Soil CRC has a limited lifespan and the project and concept of a VDIC was supported at five-and-a-half years into a 10-year lifespan, the potential value and utility of a system is diminished in comparison to establishment at the initiation of the Soil CRC; however, there still exists benefits that could be realised even at this stage of the Soil CRC within and beyond the CRC.

Data and Publications – What is in, or not in, Existing Repositories

For a VDIC to have value there is a need to have records in institutional repositories for search and discovery purposes. To test this concept, a range of tests and evaluations were undertaken to identify what metadata records exist in established repositories. This focused on universities and used RDA as the tool to see what Soil CRC records were in existence or had been published.

In parallel, for the Sandy Soil Project (3.3.003), a map of sandy soils for Australia's agricultural lands was published online through Federation University FigShare repository (https://federation.figshare.com/articles/dataset/Sandy_soil_map_of_Australia_s_agricultural_lands/24515101). Through the search and filtering of RDA, it was evident that this was the first dataset from the Soil CRC published online. There were no other datasets identified as being of Soil CRC origin through this portal.

The Soil CRC shared a list of publications for the FY 2022–23 including journal papers, conference publications, books, and book chapters. From this list a focus was placed on journal papers to identify if there were metadata records (or associated data) in institutional repositories. Of 12 journal papers, one conference paper and three other publications assessed, 12 had a metadata record in the relevant institutional repository with no pre-print versions in existence. No associated datasets were identified in these repositories.

VDIC Prototype Testing and Evaluation

A brief evaluation of the VDIC concept was undertaken using the sandy soil map dataset at Federation FigShare in the test through a MAGDA instance (Figure 6). This was implemented using a Virtual Machine by Manaaki Whenua Landcare Research (<https://github.com/magda-io/magda-brown-bag/tree/main>). Documentation for MAGDA was found to be difficult to interpret and therefore implement.



Figure 6. MAGDA instance as a prototype for a Soil CRC VDIC.

Unfortunately, there was no pre-existing connection to FigShare (known as a connector), and resourcing was required to create a connector. Consequently, the proof of concept ceased in December 2023. Other connectors in existence (e.g. Google, Facebook, CKAN, AAF, VANguard (WSFed), ESRI Portal and Oktawhich) could be explored and assessed in the future.

Data Repositories

Participation in this research was voluntary and contact undertaken in a variety of ways including in person, by email and by telephone to ascertain participants' willingness to participate in the survey. Participants consisted of eight universities, nine grower groups and one government agency. Whilst most participants responded via email to the survey, two participants required telephone calls to assist in filling out the survey. One respondent completed the survey in person at a Soil CRC conference.

The results of the survey are summarised in the following sections as: universities; farmer group, state agency and industry partner (long form); and farmer group, state agency and industry partner (rapid). The results are detailed in Appendix H.

University Responses Summary

Data

Q1. Does your institution have any policy for what data you will and won't accept in terms of who funded the research (e.g. CRC)?

Most respondents (8 out of a potential 10) do have a data repository. Respondents generally indicated that the deposit of data into the repository depended on the affiliation of the researcher with the university, not whether or by whom the research was funded.

Q2. Does the repository accept data covering all disciplines, or is it is restricted to one or more specified disciplines? If the latter, please describe.

All repositories are multi-disciplinary.

Q3. What data limits does your institution have? For example limitations on the volume (maximum size) of the dataset, formats, etc.

Most universities can accommodate various file formats and sizes up to large data sizes (e.g. <200 GB) with an option to add metadata only and link to the dataset from another storage repository. Where PURE is used by one university, this is designed for publications, so only small datasets can be accommodated.

Q4. Does the repository policy allow for the lodging of third-party data associated with a research project (e.g. such as that provided by a grower group organisation)?

Generally load data associated with university researchers.

Q5. Does the repository support capture of links to papers associated with the data?

Metadata can be added, but not necessarily hyperlinked.

Q6. What metrics are available for a researcher to know how frequently the data has been viewed/downloaded?

When the file is attached (all except one repository), file downloads can be monitored; three universities also do page or file views.

Q7. Is the repository suitable for storage of private and/or sensitive data?

Three universities are suitable and used for sensitive/private data; three universities either apply access controls or do not permit metadata for such datasets.

Q8. Who can upload data files into the repository?

For all repositories (except one), university staff /students can load data files. For two universities the data files are reviewed before being published. In one university, the repository team uploads the files.

Metadata

Q9. What metadata standards or languages are being used for institutional records?

Dublin Core was most common (four repositories), RIF-CS and CERIF also mentioned.

Q10. Does the repository support the tagging of metadata records with things such as programme the research was undertaken in, e.g. 'CRC for High Performance Soils'.

This can be done by keywords or notes in all but one repository, which relies on researchers to add metadata.

Q11. Who can load and/or update the metadata or other aspects of the lodged data?

The researcher and/or library staff.

Q12. What metrics are available for a researcher to know how frequently the metadata is viewed/downloaded?

Metrics may include pager views, hits, and/or downloads and are publicly available on all repositories except for one.

Q13. Is the metadata associated with a data set which has been deposited into another repository (either institutional or discipline) added to your institutional repository?

All can do this if the information is available.

Permanent Identifiers

Q14. Does the repository assign (mint) DOIs or other PIDs (please specify) for the data and inclusion in the metadata record?

All repositories mint DOIs.

Q15. Are these automatically or manually generated?

Only two repositories automatically mint DOIs, the rest have a manual process.

Q16. Do the repository and metadata records support CrossRef DOIs, DataCite, both or neither?

All repositories support DataCite DOIs, with two also supporting CrossRef DOIs.

Q17. Does the metadata/repository include and actively link to ORCiDs of researchers associated with the data?

Only two repositories actively link ORCID to outputs/datasets.

Archiving and Disposal

Q18. What is the policy and process regarding long term storage of data and metadata records?

Three repositories retain data indefinitely, the others leave the decision to the researcher and/or used the state public record office guidelines to determine retention periods.

Q19. Is there a disposal policy and procedure (e.g. data in repository is certified as deleted after certain time period) and is a metadata record retained in perpetuity?

Four repositories have no policy or procedure, one does; others would retain metadata and/or be guided by state public record office guidelines.

API

Q20. Does the institute's repository support APIs for publishing/harvesting metadata to a third-party service? If so, are these enabled or being used?

Two repositories use an API, but all enable OAI-PMH feeds an additional option for some harvesting.

Q21. If so, does the API allow requesting any associated links to papers related to the data – if such information is captured. How does this work with respect to publication/harvesting of private data held in the repository – does the requestor have access to metadata which is not publicly available?

No links to associated papers possible, and only public metadata can be harvested. This reflects the current suite of software used by these repositories.

Q22. Does the API support filtering by tags, e.g. can records be published/harvested with a CRC identifier?

Three repositories enable such filtering if the information was included in keywords in the metadata.

Q23. Does your institute they have a technical support person who we could approach for more information?

Availability of technical support contact varies, with some provided by outside vendors. Generally prefer requests to be mediated by the Repository Manager.

Support

Q24. How is access control (e.g. licensing and embargo periods and changes to these) to data and metadata managed, and by who?

Most universities have support staff (e.g. library or repository editorial) that manage licensing and embargo periods. Some universities leave this to researcher who can liaise with support staff. Metadata likewise is mostly managed by support staff with some left to researchers.

Q25. What support is available to researchers to deposit data in the repository?

Library liaison staff are common across the universities to support researchers in depositing data. There are also guides available for universities to assist researchers to deposit data.

Repository Specifications

Q26. What repository software platform is presently in use, and are there plans to replace this software?

Examples include PURE, Vital, FigShare, Cayuse, DSpace, AWS, Esploro. Little planning to forward replace or augment software platforms.

Q27. How is data file version control administered?

Main response was through the repository software. Library team validation and update also occurred and even the creation of a new repository record to reflect updates. A university had no established policy.

Q28. Does your institutional repository have a core trust seal certification, or is this being planned soon?

Overwhelmingly the answer is no although there are universities who have applied for or are considering.

Farmer Group, State Agency and Industry Partner (Long Form) Summary

Below are the results of the grower group and government agency survey (long form). The take-home messages from this set of results are:

- Of the respondents, all institutions have repositories but there was high variability in what they could store, how they were used and how processes were governed
- There are no metrics to allow the researcher to know if their data has been viewed or reused
- High variability in the ability to store sensitive or confidential data
- DOIs are not often minted
- Approach to metadata is highly varied.

Farmer Group, State Agency and Industry Partner (Rapid) Summary

Below are the results of the grower group survey (short form). The take-home messages from this survey with respect to how they store data are:

- Whilst some organisations use SharePoint, many use other methods to share data such as email
- There is considerable variation in how organisations protect their data, with some not protecting their data at all
- Policies and process around all aspects of data management are lacking
- Metadata is rarely, if ever created
- Training in data management is sporadic and unstructured.

ENGAGEMENT

Webinars

Two webinars were delivered on Research Data Management and Data Management Planning (Table 7, Figure 7), attended by researchers, PhD students and managers from universities and government agencies.

Table 7. Data management webinars delivered by the project in November 2023.

Webinar	Date	Number of registered participants
What is Research Data Management?	17/11/2023	18
Data Management Planning and you	30/11/2023	18

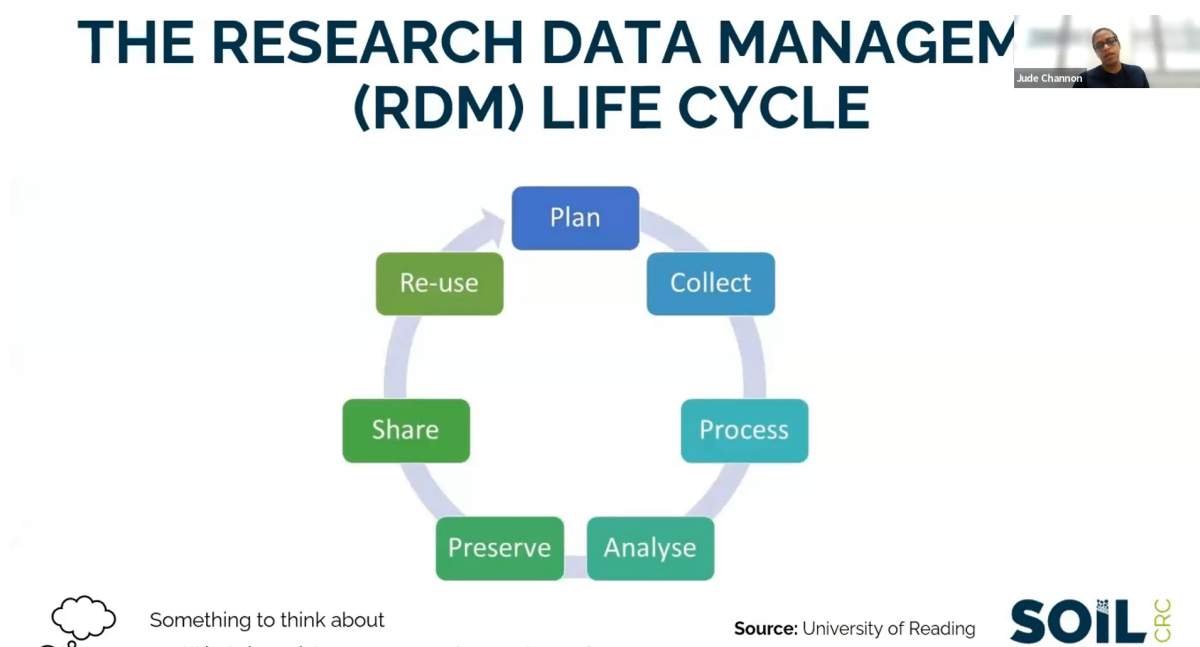


Figure 7. “What is research data management” webinar.

In addition to these webinars, a separate video production of the “What is Research Data Management” has been completed with a version created for Soil CRC members and a separate version for consumption on the public [Soil CRC website](#). This public version was produced as there was interest expressed from those in attendance at the Soil Science Australia Conference (Darwin, June 2023) on learnings and training/educational materials that could be shared with university students and international collaborators.

National Soil Science Conference

Insights on the data maturity survey were shared with this national and international audience (Figure 8). Interest was expressed from audience participants in training and educational material being collated for the project to help students in their RDM. Likewise, for international

researchers in Asia, there was interest expressed in accessing these materials and resources provided by the Soil CRC.

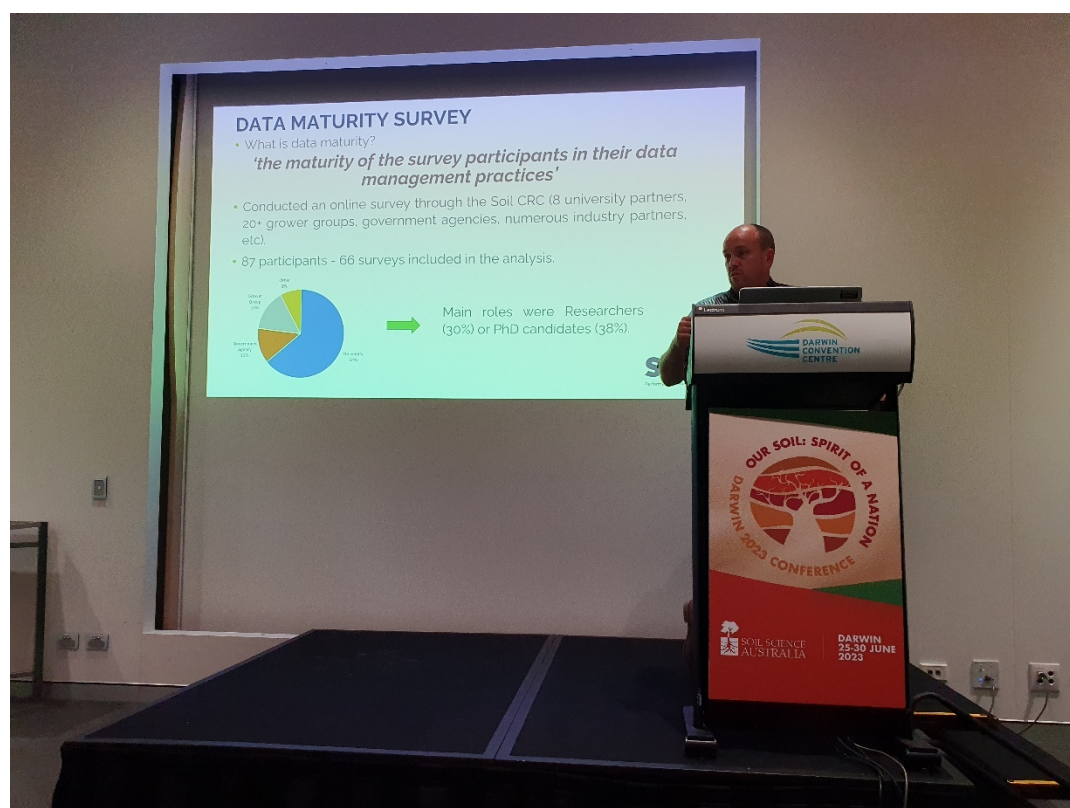


Figure 8. Presentation on the data maturity survey at the National Soil Science Conference, Darwin 2023.

HDR Training and Education Activities

For the 2022 and 2023 Soil CRC conferences, we presented to the HDRs in attendance on data management, the latest tools and developments that can be used to assist their research journey. At the 2022 conference, a short presentation on the project aims and objectives and links to the data maturity survey was shared. This provided some exposure and learnings to those in attendance on data management fundamentals.

At the 2023 conference, an information session on Research Data Management was held with the PhD students. The session comprised a presentation on data management and relevant results from the data maturity survey, Q & A, a demonstration of resources available on the Soil CRC members area of the website (Figure 9) and a pop quiz on learnings from the session. The sessions provided valuable opportunities to share materials and resources for the HDR students to use.

Documents

CEO COMMUNIQUE

AGREEMENTS

RESOURCES

PROJECT MANAGEMENT

SOILCENTRAL

INVESTMENT CALLS

RESOURCES

Research Data Management

-  [Research data management overview](#)
-  [Draft data management plan template](#)
-  [Data management plan guidelines](#)
-  [Glossary of data management terms](#)
-  [Governors of Data song \(YouTube\)](#)
-  [If Broadway songs were about data \(YouTube\)](#)
-  [The Little Red Data Hen - A Cautionary Tale \(YouTube\)](#)

Figure 9. Resources for Research Data Management (<https://soilcrc.com.au/members-login/>)

Use Cases

Use cases provided opportunities to engage with Soil CRC projects and program leaders on data management and the sharing of insights and resources that participants were unaware of. Eight members of the Soil CRC community were engaged directly in this process. Shared input and insights were timely as this helped refine the DMP for the Soil CRC but also introduced the concept and utility of a DMP to these participants.

As part of the exploration of how researchers manage their data, it was decided that providing lived examples through use cases would be a valuable methodology. The University of Tasmania (UTAS), as a Soil CRC partner, was chosen as a use case and a group interview session was held with those involved in the project in December 2022. It was decided that UTAS would make a highly relevant use case as the UTAS members of the project team were part of Sense-T which was ceasing on 31 December 2023. This provided an opportunity to first, test a storing and archiving data management strategy and practices with regards to the data that they had collected in their Soil CRC projects, second, to understand an exit strategy for project partners at the conclusion of a Soil CRC project, and thirdly as an exit strategy for the Soil CRC.

In terms of data and projects (sensor) that UTAS were involved with, these include:

- “Smart shovel” – a probe to measure moisture, temperature, EC and penetration resistance (now the BANDICOOT®)
- Bilby® – node sensor/communication
- eNose (developing a sensor).

The projects included in this use cases are as follows:

Tamar Valley Project

- Five chambers to run tests for calibration/validation of sensors.
- In addition, there are lab trials.

Bilby®

- Field trials were focused on telemetry.
- Six to 12 months for a site – reuse value not clear.

Adoption of practices (Sense-T)

- Cofounded with Data61.
- Business model changed for Sense-T.
- Senapps (own IP in data).

eNose

- Sense-T brought in to assist in the engineering with the aim of developing a commercial product.

Key Findings

- Data management is not often factored into project management costs as a line item in the project budget, and researchers don't often account for this in project funding models.
- Whilst the project mostly dealt with third-party data, the lack of an exit strategy with respect to data retention makes any data collected for this project difficult for the Soil CRC to access.
- IP ownership most likely to revert to those researchers /institutions that created the data; however, IP ownership is usually retained in modelling of data.
- This project is using Dropbox as its data repository (NB. this is considered very poor RDM practice).
- There is clear categorisation of the data.
- Data is password protected.
- Data is anonymised.
- The value of the data is unclear, and its reuse is unlikely.
- Because the data is in DropBox it is not discoverable outside UTAS and only known about by certain researchers within the institution.
- Data is not made openly available but handed back to the industry partner as per contract clauses.
- Some of the Tamar Valley data will be made publicly available.
- The data is linked to journal publications for example for the smart shovel and the Bilby®.

- The data is not on long term institutional data storage or a data repository: what will be lost potentially includes 60,000 points per day for the Tamar Valley when this is switched off, and there is a cost of \$200–400 per month for running of these services.

EVALUATION AND IMPACT

A combination of approaches was used to obtain feedback on the project and the information delivered to participants in the Soil CRC. The key sources of this information include:

- Data maturity survey
- Email correspondence
- Webinar evaluation
- HDR cohort feedback.

Informal feedback was received from webinar participants that highlighted:

- Most participants noted a medium to high level of awareness for data management before the project
- The level of awareness on data management has increased for all participants as a consequence of this project
- Nearly all participants were aware of the education resources now available through the Soil CRC website on data management
- The majority of participants have not accessed these materials, however, they may in the future
- Most participants indicated a moderate level of training in RDM would be required for their needs, and likewise, a moderate level of support would be desired in RDM
- Participants found the two webinars (Research Data Management and Data Management Planning) useful or very useful
- Others (researchers) would benefit from this training or other training on RDM
- To support projects and researchers in the Soil CRC with data management, easy access systems for storage, guidelines on DMP, forward planning for projects, and a “cookbook” on data management were suggested
- Data management is an area of research that one can easily regress in terms of knowledge and skills: having a supportive environment especially at the early stages of research projects with positive reinforcement is critical to set the foundation for good RDM practices
- Systems and storage solutions outside universities are infrequent and poorly understood, making it hard to contribute towards larger programs of work.

DISCUSSION

This project identified that data is already being shared and exchanged amongst trusted partners and researchers in the Soil CRC. However, the practices used, planning and associated documentation via a Data Management Plan are rarely used. The project proposal of a cross-program solution for data sharing, storage, use and integration with existing systems and tools was to provide further value to the Soil CRC and its stakeholders, and was ambitious without the necessary governance and policies as the foundation for these practices. What also became evident through this project was that while data management is flagged as a keystone activity of research, the knowledge and skillset of those operating in this space often do not align with the lofty goals of interoperable data shared through FAIR systems including sensor data streams.

We discuss some of the key results of this project opportunities, challenges and easy wins for the Soil CRC to implement in their next phase towards the completion of the CRC.

DATA MATURITY (PRACTICES AND KNOWLEDGE OF RESEARCHERS IN DATA MANAGEMENT)

In research and problem solving, data management is one of the key pillars of support for collaborative outcomes (von Davier et al. 2017). For “small science” research that is interdisciplinary, the idea of establishing agreed data management practices can be problematic as not all participants in a collaboration require access to this data or data brought by other researchers to the collaboration (Finkel et al. 2020), or they may be collecting or creating data that is totally unrelated and non-standardised (e.g. quantitative soil data vs. qualitative interviews). The complexity of “small science” interdisciplinary RDM is further layered by the needs, requirements and benefits of different fundamental components of data management at different levels (e.g. individual researcher vs. project vs. institutional) (Figure 10).

Research data management is complex and problematic (Pinfield et al. 2014) in science domains using defined standards, practices and protocols that are often applied in “large science”. For research programs and operational settings dealing with a wide diversity of data (types, formats, entities, qualities) as part of interdisciplinary research, it is even more complex and difficult to define common understandings for accepted data management (Finkel et al. 2020). As noted by many (e.g. Fecher et al. 2015; Kratz and Strasser 2015; Tenopir et al. 2011), establishing a community of researchers willing and driven to contribute to a shared data management approach remains difficult to navigate.

Data management practices of researchers and their “data maturity” have rarely been examined in the literature. Most emphasis in existing research is focused on institutional settings, compliance and regulation. A mixed-methods approach combining online survey, focus group and interviews revealed that there is an expansive breadth of understanding on data management practices at an individual, project, institutional and collaborative research setting level. An associated research paper (Robinson et al. 2025) has further details on pathways and priorities to improve research outcomes through better data management.

Questions were answered by respondents in the data maturity survey with 14 interview themes (detailed in Table 1), and for half of these themes, the overwhelming response was uncertainty. Respondents were unsure on IP information, their institutional data management

policy (compliance theme), the cost of data management, data privacy, data retention, data management plans, data licensing and data vocabularies and ontologies. Where questions asked about practices (the “how”), such as data storage, a large percentage were storing/backing-up data (a benefit) but not necessarily in what is considered a suitable location for good data management. Additionally, it was noted from survey responses that respondents were sometimes not clear on the meaning of data management concepts and phrases, with some participants noting “I am not sure what is meant by metadata” and “I don’t know what a data management plan is”.

This project identified that researchers, as part of a collaborative research initiative, are unfamiliar with many of the concepts that are fundamentals of data management. This was reinforced through the survey and qualitative engagement results where considerable interest was expressed in additional training and support. The survey provides a clear direction for the Soil CRC and research collaborations in general to meet the need of improving fundamental data management skills and knowledge of those managing and using data in this collaborative research environment.

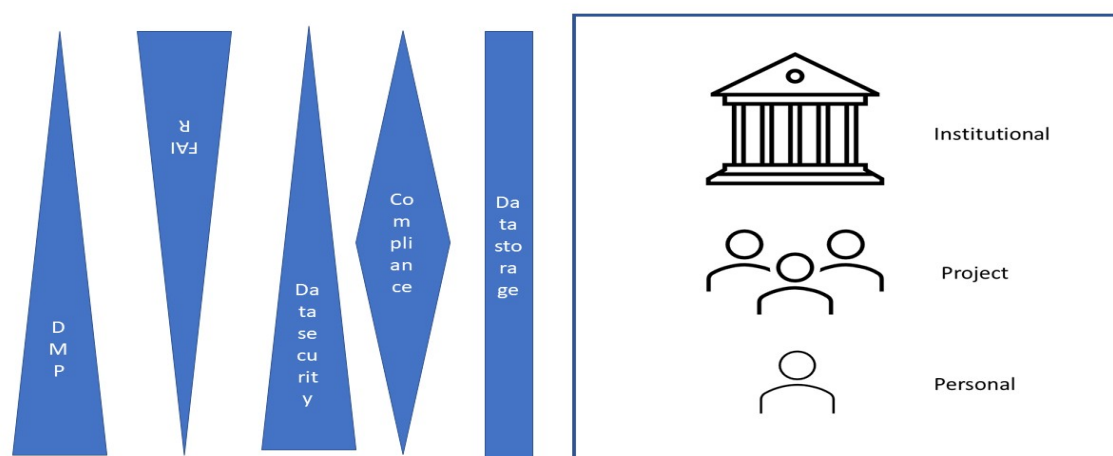


Figure 10. Data management levels. Data management nuances and fundamentals are present at each level. Data management skills affect each level (e.g. good personal data management will provide efficiencies in project and institutional data management).

We suggest that there are three levels of data management responsibilities and practices – institutional and/or funder, project, and personal that contribute to collaborative research. While there are fields of data management that may cross all three levels (Figure 9), there are different requirements for each level (e.g. institutional level – compliance and infrastructure; project requirement – data storage and access, and standards; individual management – time efficiencies) and nuances that require a targeted approach. The outcomes of the survey and interviews highlight the need for a greater understanding of approaches to data management at the individual researcher level (personal) which fulfills data management activities and tasks at the project and institutional levels. Providing the individual researcher with the necessary tools for data management will result in efficiencies and benefits at the project and institutional levels.

It can be argued that data management requires a degree of emotional intelligence because data management is also about people management. For collaborative research (including institutions and projects) in order to support researchers in applying good data management practices, there is a need to understand researcher needs, practices and attitudes to data management to effect meaningful change in data management practices. In a study of researchers at the University of Queensland (UQ), Yu et al. (2017) found that researchers did not regard RDM as an important part of the research process, and instead was often viewed as an administrative burden and not part of doing good science. It was also not always formally recognised that keeping research data ensured research integrity, and on some occasions research data were reported as being neglected once the research output was published. Furthermore, they found that “While there was a lack of understanding about the significance of RDM across the research lifecycle, the feedback indicated that all researchers expected the university to provide and support data storage solutions” (Yu et al. 2017, p 290).

DATA REPOSITORIES AND THE SOIL CRC

For the Soil CRC to advocate for data to be stored and made accessible and findable (fulfilling key requirements of FAIR), data repositories that support partner institutions and organisations were found to be of various qualities, and processes varied considerably between those that managed a repository. For some organisations, they used existing storage filing systems and architectures (e.g. Microsoft) while metadata records for data were infrequent in contrast to research publications including reports and journal papers. The key findings are summarised for universities, and for farmer groups, state agencies and industry partners.

Key Findings – University Repositories

- While information about what repositories exist is available, their related processes are not always easy to access or clear to follow for researchers.
- Of the data repositories being used at this level, they accept data from all disciplines.
- The data repositories used vary in terms of their ability to accept third-party data.
- The data repositories used appear to have ample storage space for the current requirement.
- Approach to metadata between the data repositories is highly varied.
- Most institutional repositories appear to support APIs.
- Most institutions mint DOIs used when data is cited in journal articles and reports.
- There is no single standard practice with regards to the data management cycle across institutions.
- The current institutional capacity to effectively manage data repositories well is limited and this affects their use.
- Data security and privacy measures vary significantly across repositories, potentially affecting the reliability and trustworthiness of the stored data.
- Some repositories offer user-friendly interfaces and support for data upload and retrieval, while others are more complex and difficult to navigate.
- There is inconsistency in the level of support and training provided to researchers on how to use the repositories effectively.

- Collaboration features, such as shared workspaces or version control, are present in some repositories but absent in others.
- The integration of data repositories with other research tools and platforms (e.g. lab notebooks, statistical software) varies, impacting workflow efficiency.
- Policies regarding data retention and long-term preservation are not uniformly established across repositories.
- Some repositories have implemented advanced search and data discovery features, while others have limited search capabilities.

Key Findings – Farmer Groups, State Agencies and Industry Partners

- Of those surveyed, all institutions have repositories but there is high variability in what they can store, how they are used and how processes are governed.
- Whilst some organisations use SharePoint, many use other methods to share data such as email.
- There are no metrics to allow the researcher to know if their data has been viewed or reused.
- Policies and process around all aspects of data management are lacking.
- There is high variance in how organisations protect their data, with some not protecting their data at all.
- There is high variability in ability to store sensitive or confidential data.
- Metadata is rarely, if ever, created.
- DOIs are not often minted.
- Training in data management is sporadic and unstructured.
- Data accessibility for external stakeholders is inconsistent, impacting collaboration and data sharing.
- The level of support and resources for data repository management varies widely across organisations.
- Many repositories lack advanced search and discovery features, making it difficult to locate specific data sets.
- The integration of repositories with other data management tools and platforms is inconsistent.

DATA MANAGEMENT PLAN AS AN APPROACH TO IMPROVE DATA MANAGEMENT IN THE SOIL CRC

Data management plans are recognised as a solution to support the management of research data in a sustainable way (Jäckel and Lehmann 2023). While DMPs have become more prevalent today through the requirements of funders, they should be living documents that support those conducting research to manage data efficiently, effectively and to achieve benefits within and beyond the project lifespan. Experiences with DMPs have been mixed as noted in the literature (see Smale et al. 2020), however we would advocate that, for the Soil CRC and those handling or managing data, a DMP is a good introduction to the basics of managing data as a valued asset.

Different disciplines have various conventions with respect to good data management. To facilitate data linkage and interoperability, one should ensure that DMPs used are compatible

with any recognised data management practice and standards in your discipline (e.g. in relation to file formats and the metadata standard to use).

A DMP should be a living document and regularly revisited to ensure its content is updated as necessary as the project evolves and progresses. To support the adoption and use of a DMP, complementary training and development should be prioritised to provide those managing data with skills that will likely be enduring and of value beyond the Soil CRC.

A template DMP provided through this project to the Soil CRC can easily be adapted to suit specific data management needs for research. Not all sections and/or questions will be relevant while other sections can be expanded, and supplementary material attached where relevant. The detail should be clear, concise and proportionate to the scale of the datasets used, collected and generated, and the dataset's anticipated level of value to the research community. Ultimately this DMP is to support those conducting research on behalf of the Soil CRC, and for the Soil CRC to meet data management obligations under the Australian Code for the Responsible Conduct of Research and avoid catastrophes such as the loss of data, violations of privacy or loss of IP from research investments.

A SOIL CRC VIRTUAL DATA AND INFORMATION CATALOGUE

From the benefits and limitations of establishing a VDIC for the Soil CRC, this is a relatively low-cost option to share and promote research and development undertaken by the Soil CRC to its membership and the public. While the delayed construct of a VDIC has lessened the depth of potential benefits, the enduring value of establishing the catalogue and formalising linkages amongst repositories of the institutes in the Soil CRC will ensure longevity of the investment and outcomes the Soil CRC has strived to achieve.

We would strongly advocate that the formalisation of a soil data ecosystem for the Soil CRC is a low-cost and high-return investment, and that a VDIC ensures the longevity of the Soil CRC beyond its lifetime by:

- Enabling data discovery and sharing to further understand and interpret new mechanisms for financially rewarding management of high performance soils
- Ensuring enduring provision of latest scientific discoveries and packing of this information as new integrated soil management solutions that provide greater precision for farmers
- Providing a platform of data discoveries as a foundation of new methods and understandings to measure the performance of soils
- Building foundational science on research data for advanced and innovative products to increase soil fertility and function.

OVERALL OPPORTUNITIES FOR PROGRESS

Below is a summary of the opportunities for progress that have been collected through various research methods outlined in this report. These include:

- Investment: Funders should conduct an accurate cost-benefit analysis and risk management assessment to ensure that adequate resources are available to projects for good RDM and that project data is secure

- Capacity building: Create a role that engages with researchers and assists them in building capacity for good RDM as well as providing advice and guidance on RDM strategies and tools
- DMPs: Require researchers to develop comprehensive DMPs that outline how data will be collected, stored, and shared during and after the project, and should address ethical considerations, privacy concerns, and data sharing protocols
- Legacy planning: Develop a plan for managing project data after the project concludes, potentially involving the transfer of data to appropriate institutions, ensuring data continuity, and preserving valuable datasets for future research purposes
- Centralised data repository: Establish a centralised data repository or platform that is easily accessible, user-friendly and secure, where researchers can store and access project data that has no other home
- Assuming researchers have deposited data in their institutional repositories, and tagged the data as associated with Soil CRC, create a data catalogue that records what soil data and associated data CRC have collected, where metadata is harvested from institutional repositories, and where data is not in an institutional repository, a metadata record should be added to the catalogue describing the soils data and where it is stored (assuming there are no reasons for it not to be made available, e.g. data privacy)
- Training and support: Provide training and support to researchers regarding best practices in data management (e.g. workshops, seminars, or online resources to enhance researchers' skills in managing and organising data effectively)
- Metadata standards: Implement metadata standards to ensure that essential information about the datasets (such as variables, units, and collection methods) is well-documented
- Data quality: Conduct quality checks at various stages of the research process to maintain the accuracy of the data
- Data sharing policies: Develop and communicate to all researchers involved in projects clear policies on data sharing, addressing issues related to intellectual property, confidentiality, and data ownership
- Collaboration and communication: Foster a collaborative environment where researchers can communicate and share their data management strategies
- Continuous evaluation: Regularly assess the effectiveness of data management practices and identify challenges and areas for improvement and adjust strategies accordingly
- This research has unearthed that RDM is relatively immature across the research community. Resourcing is required to further examine the particularities and benefits of managing soil data as a valuable asset. As this research identified data management practices are immature, it's difficult to identify critical issues specific to soil data and its long-term management.

By focusing on these opportunities to improve RDM, projects can enhance their data management practices, ensuring data integrity, accessibility, and usability for current and future research endeavours.

OVERALL CHALLENGES TO PROGRESS

Whilst opportunities exist to improve research data management, there are several limitations and challenges that researchers and their institutions face. These include:

Soil CRC research project final report: A coordinated and consistent approach to managing and enabling soil data | Page 81 of 199

- Resource constraints: Smaller research projects or institutions with limited resources might struggle to invest in data management infrastructure, training programs, or dedicated staff to effectively manage their research data
- Data privacy and security: Ensuring data privacy and security, especially in sensitive research areas such as social science research, can be challenging, as compliance with data protection laws and safeguarding data require fairly robust security measures
- Interdisciplinary collaboration: When considering interdisciplinary research, different fields can have varying data management standards and practices, and changing the way that researchers practice so that there is standardisation can be a complex task
- Rapid rate of technological advances: This can affect data formatting compatibility and what software is used for the long term, so ensuring that data remains accessible and usable over time is a significant challenge
- Data ownership and IP: Determining ownership rights and addressing IP concerns, especially in collaborative projects involving multiple institutions or stakeholders, can be a complex challenge
- Lack of awareness: Researchers, particularly those new to the field, might not be fully aware of best practices in data management, so providing comprehensive training and resources is essential but requires investment in time and money, which may not always be possible
- Ethical considerations: In terms of social research, respecting participant privacy and confidentiality while promoting data sharing requires a delicate balance
- Long-term preservation: Identifying suitable repositories and mechanisms for archiving data in a way that remains accessible and understandable for future researchers is essential
- Data quality: Maintaining and ensuring the quality of data throughout its lifecycle is a significant concern
- Cultural and organisational barriers: Overcoming resistance and fostering a culture of data sharing and collaboration can be challenging.

EASY WINS FOR THE SOIL CRC

The opportunities for progress and the challenges in research data management are numerous and it is unlikely that the Soil CRC has the time or capacity to progress all opportunities or address all the challenges raised above. Currently the soil data ecosystem within the Soil CRC is suboptimal and opportunities exist to leverage and benefit from the data collected so far. However, it should also be stated that the potential impact and benefits of collected data may not be realised until beyond the lifetime of the CRC. However, some bare minimums which would allow for easy wins and add great benefit to its researchers are:

- The creation of a DMP template to assist researchers in planning out their research proposals and help them manage their research data throughout the life of their research project
- The appointment of a Soil CRC data steward (either from within existing resources or new investment) to provide advice and information as well as building capacity of researchers through training and engagement

- Developing and sharing policies and information sheets on soil data sharing, data privacy, data licencing and data for publications
- Developing and testing repository options for data generated through Soil CRC projects
- Creating a virtual data and information catalogue as a current and future legacy resource of the Soil CRC.

On the basis of the research conducted so far through this project, it is recommended that, for the remainder of the Soil CRC, prioritisation be given to the following training and education areas:

- Creating a base level of understanding of data management terminology and vernacular so that researchers are aware of what is expected of them and that they can be supported to manage their data well
- Developing robust DMPs for their projects or signpost researchers to where this training and support can be found, accompanied by a clear DMP process flow chart for the Soil CRC
- Providing researchers with the knowledge and tools (such as a template) to create clear and concise metadata records to enable findability, sharing and collaboration, as well as facilitate “hassle-free” uploading to designated repositories (e.g. This could be done in conjunction with university partners or through online training modules offered by Soil CRC university partners or additional organisations such as ARDC)
- Providing resources to researchers so that they can easily identify licencing options for their work, requiring the Soil CRC to be clear on data for publishing and reuse and have relevant protections in place for this data.

It is important to note that any education and training in RDM will require the Soil CRC to have a robust data governance structure in place, whether that is through its existing committees or through dedicated roles to support good research data governance.

CONCLUSION

The Enabling Soil Data project had lofty ambitions to support the seamless sharing of research soil data from sensor data streams to those in the community looking for data to deliver research goals and objectives set by the Soil CRC. While it has been established that data sharing has been occurring in the cooperative, the processes and practices that underpinned this process were not enacted due to a lack of governance and technical advisory frameworks. This resulted in the project having to pivot towards improved data management practices for those in the Soil CRC as an outcome.

The implementation of a data maturity survey, the first of its kind for a CRC in Australia, revealed that there are many challenges in achieving FAIR data principles. The first and foremost challenge is the training and development of researchers in data management as a key finding of the survey. While FAIR data streams from sensors and other data forms are advocated by this project, it is clear that ambition and reality do not align here in the majority of cases, hence additional research value and insights from collected and collated data are more cumbersome than standard practice.

A thorough review of institutional repositories from universities, farmer groups and industry partners participating in the Soil CRC also revealed that the key systems for enduring storage and provision of data assets are also varied in their services, system configuration and support. Again, while we would advocate for the use of these repositories in the sharing of FAIR data, they remain under-resourced and less than ideal for the lofty aims of FAIR data.

The concept of a Virtual Data and Information Catalogue has been raised as a concept through this project as a resource for the Soil CRC within and beyond the lifetime of the CRC. We suggest the formalisation of a soil data ecosystem for the Soil CRC even at this stage of the Soil CRC is a low-cost and high-return investment option.

Establishing best practice in RDM and its related infrastructure and data carpentry requires a coordinated effort from researchers, institutions, funding agencies, and policymakers. Continued dialogue, training programs, and sharing of best practices can contribute to overcoming challenges and exploring opportunities for progress which will inevitably lead to improved data management practices in research projects for both those funded by the Soil CRC and beyond.

RECOMMENDATIONS

The following are collated recommendations from the key areas of work delivered in this project.

RDM PLAN AND STRATEGY FOR SOIL CRC

1. Soil CRC should establish a comprehensive RDM plan for the CRC for the entire organisation, clearly outlining expectations and guidelines for data management practices across projects.
2. Implement strategies for long-term data storage to ensure accessibility and preservation of soil data beyond the project duration. This could involve forming partnerships with institutions or organisations for data archiving and establishing robust data management protocols as well.
3. Organisations need to consider the cultural change required for researchers to see the value and benefits of using efficient data management systems and processes.
4. Data management strategies should be put in place at the start of the CRC (see Finkel et al. 2020) and the need for an overall approach by the CRC and does this align with the CRA.
5. A traffic light system (compliance and regulation) be implemented to monitor and ensure compliance with data management regulations and standard through projects.

RDM POLICIES AND PRACTICES

6. Encourage researchers to use standardised formats for uploading and storing data in approved repositories. In addition, the data format needs to allow for easy harvesting including metadata.
7. All researchers should be required to develop a data management plan outlining how soil data will be collected, stored and shared during and after the project. DMPs must also address ethical considerations, privacy concerns and data sharing protocols.
8. Develop clear policies on soil data sharing which address issues such as IP, data ownership, and ethical use of shared data.
9. Data management principles need to be integrated into the entire research process to ensure systematic and efficient data handling.
10. The protection and preservation of data needs to be considered and catered for, even legacy data (e.g. Sense T data).
11. Draft policies guidance documents should be created or, where they already exist, be signposted to and provide clear pathways for researchers to access and understand these policies.

RESOURCING FOR RDM

12. Clear budget allocation (and provision) for research data management at both Soil CRC and project levels.
13. Funders need to request research leads provide a clear budget or cost allocation for data management tasks within research proposals.

14. Good data management needs to be incentivised – at the project and individual level – to promote adherence to best practices.
15. A DMP is essential and should be created at the start of the project, and it should also define who pays for and who is responsible for data management.

RDM SUPPORT AND SERVICES

16. A dedicated data steward or similar type role to support researchers in managing their data needs to be established. This role should be supported within the Soil CRC to support and assist partners in managing their research data effectively.
17. Signpost researchers to metadata standards where possible to ensure accurate and standardised information about datasets.
18. The value of good research data management, outside of risk management, needs to be “sold” to researchers. The value proposition and incentive to adhere to good practice needs to be clearly articulated and the enduring benefits explicitly defined.
19. There is a need for a streamlined approach to manage bureaucratic hurdles with respect to data management. It should include a clear position on data ownership and responsibilities for ensuring the long-term stewardship of the data, automated workflows, standardised procedures, and enhanced support services, ensuring consistent, efficient, and user-friendly data management practices across the Soil CRC.
20. For those not aligned to existing institutional repositories, signposting to relevant repositories should be available. This has been established through this project but will require enduring promotion and information sharing.
21. There needs to be more awareness raising amongst researchers of what repositories exist and what processes need to be followed to manage research data better.
22. Assistance is required in designing tools that are user-friendly and accessible to all participants in the Soil CRC for creating metadata records.
23. Information for groups on data management should be created and placed in easily-accessible formats and platforms. This has been met by this project; however, promotion and information sharing will be required for the Soil CRC community.
24. Advice needs to be provided on minting DOIs and supporting APIs for those generating datasets but also on publications and reports.

TRAINING AND DEVELOPMENT OF RESEARCHERS

25. RDM training and development options and opportunities need to be available to upskill researchers as there appear to be scant and infrequent opportunities through institutions and partner organisations.
26. A RDM “jargon buster” resource is required to help researchers understand the RDM space better towards creating efficiencies and benefits from better RDM. Resources have been collated here for sharing amongst the Soil CRC.
27. Training and support should be provided to researchers regarding best practice in data management and the specific requirements of managing soil data and associated data. Additional training and support in various aspects of data management would be beneficial to researchers and projects.

A SOIL CRC DATA AND INFORMATION CATALOGUE

28. Establish an easily accessible, secure and user-friendly data and information catalogue for research data from Soil CRC projects. This catalogue should incorporate an aggregated infrastructure and clear conventions for managing the data, offering enduring value for the Soil CRC partners beyond the projects' lifespan.
29. Develop a virtual catalogue of Soil CRC data (and information) that harvests metadata from institutional repositories to create a unified platform for accessing and managing the Soil CRC project data. This will streamline the process of data discovery and usage.
30. Implement clear conventions and standards within the catalogue to ensure consistency and ease of use. These standards should promote interoperability and ensure that data remains valuable and accessible to partners and stakeholders beyond the lifespan of individual projects.
31. Enhance the catalogue with advanced search and filtering capabilities to allow users to efficiently locate specific data sets and information. This will improve the usability and functionality of the catalogue for researchers and stakeholders.
32. Implement robust data security and privacy measures within the catalogue to protect sensitive information and comply with relevant regulations. This includes encryption, access controls, and regular security audits to safeguard the integrity and confidentiality of the data.
33. Provide comprehensive training and ongoing support for users of the catalogue to ensure they can effectively navigate and utilise the platform. This includes creating user guides, tutorials, and offering support services to address any issues or questions.
34. Regularly monitor and evaluate the performance of the data and information catalogue to ensure it meets the needs of users and aligns with the goals of the Soil CRC. This includes gathering user feedback, analysing usage metrics, and making continuous improvements based on insights gained.

ACKNOWLEDGEMENTS

We would like to acknowledge the engagement and discussion with project team members, including Andrew MacLeod, Alison Ollerenshaw, Jen Corbett and Helen Thompson (Fed Uni); Stephen Cahoon, Simon Edwards and Warwick Gillespie (UTAS); Alistair Ritchie (Manaaki Whenua Landcare Research); Soil CRC representatives (Michael Crawford, Rhona Hammond, Jodi MacLean, Olivia Louis), Soil CRC program leaders (Catherine Allan, Richard Doyle, Megh Mallavarapu and Lukas Van Zwieten) and Cassandra Wardle (HDR coordinator for the Soil CRC). We would also like to acknowledge administrative support though the project provided by Kathy Gamble, Jen Corbett and Kelly Hartwig.

We would also like to acknowledge participants in the two surveys conducted in this project: the data maturity survey and data repository survey and those participants in the focus group and interviews for these surveys. Feedback and responses from presentations to the HDR cohort of the Soil CRC and other parties also helped shape and direct the project to achieve best outcomes possible.

This research is funded by the CRC for High Performance Soils and supported by the Cooperative Research Centres program, an Australian Government initiative.

REFERENCES

- Alhassan I, Sammon D and Daly M (2016) 'Data governance activities: an analysis of the literature', *Journal of Decision Systems*, 25(S1):54–75, doi:10.1080/12460125.2016.1187397
- AIHW (Australian Institute of Health and Welfare) (2022). Data Governance Framework. Public edition, Australian Government, Canberra.
- ARDC (Australian Research Data Commons) (2023). [Planet Research Data Commons program description](#), accessed 25 May 2026.
- Australian Antarctic Data Centre (2015) [Project 1234 data management plan](#), accessed 25 May 2026.
- Bishop B and Borden R (2020) 'Scientists' research data management questions: lessons learned at a data help desk', *Libraries and the Academy*, 20(4):677–692, doi:10.1353/pla.2020.0032
- Bloemers M and Montesanti A (2020) 'The FAIR funding model: providing a framework for research funders to drive the transition toward FAIR data management and stewardship practices', *Data Intelligence*, 2(1-2):171-180, doi:10.1162/dint_a_00039
- Borghi JA and Van Gulick AE (2021) 'Data management and sharing: Practices and perceptions of psychology researchers', *PLoS ONE*, 16(5):e0252047, doi:10.1371/journal.pone.0252047
- Box P and Lemon D (2015) The role of social architecture in information infrastructure: A report for the National Environmental Information Infrastructure (NEII), CSIRO, Australian Government.
- Briney KA, Coates H and Goblen A (2020) 'Foundational practices of research data management', *Research Ideas and Outcomes*, 6:e56508. doi:10.3897/rio.6.e56508
- Brush K (2022). [Use case](#), TechTarget website, accessed 25 May 2026.
- Clobridge A (2010) 'Introduction', in *Building a Digital Repository Program with Limited Resources*, Chandos, Oxford, doi:10.1016/B978-1-84334-596-1.50001-8
- Cox S, Gonzalez-Beltran A, Magagna B and Marinescu M (2021) 'Ten simple rules for making a vocabulary FAIR', *PLOS Computational Biology*, 17(6):e1009041, doi:10.1371/journal.pcbi.1009041
- Dahlhaus P, Imhof M, Stevenson B, Crawford D, Wells N, Grelet G, Mele PM, Wong M, Corbett J, MacLeod A and Bennett JM (2018) *A review of indicators of soil health function: Farmers' needs and data management. Final Report for Project 2.1.001*, Soil CRC.
- Dahlhaus P, MacLeod A, Medyckyj-Scott D, Simons B, Bahlo C, Sexton A, Thompson H, Wong M, Robinson N, Milne R, Neyland A, Gillett H and Ollerenshaw A (2021) *Visualising Australasia's soils: A Soil CRC interoperable spatial knowledge system. Phase 1 (2019 – 2021). Project 2.3.001 Final report*, Soil CRC.
- Soil CRC research project final report: A coordinated and consistent approach to managing and enabling soil data | Page 89 of 199

- DAMA International (2009) *The DAMA guide to the data management body of knowledge*, Technics Publications, New Jersey.
- Digital Curation Centre (2013) [Checklist for a data management plan. V.4.0](#). DCC website, accessed 25 May 2026.
- Fecher B, Friesike S and Hebing M (2015) 'What drives academic data sharing?', *PLoS ONE*, 10(2):e0118053, doi:10.1371/journal.pone.0118053
- Federation University Australia (2015). [Data management planning. Guide and template](#), accessed 25 May 2026.
- Finkel M, Baur A, Weber TKD, Osenbrück K, Rügner H, Leven C, Schwientek M, Schlögl J, Hahn U, Streck T, Cirpka OA, Walter T and Grathwohl P (2020). 'Managing collaborative research data for integrated, interdisciplinary environmental research', *Earth Science Informatics*, 13(3):641–654, doi:10.1007/s12145-020-00441-0
- Fu X, Wojak A, Neagu D, Ridley M and Travis K (2011) 'Data governance in predictive toxicology: A review', *Journal of Cheminformatics*, 3(1):1–16, doi:10.1186/1758-2946-3-24
- Jäckel D and Lehmann A (2023) 'Benefits and challenges: data management plans in two collaborative projects', *Data Science*, 22(1):25, doi:10.5334/dsj-2023-025
- Khatri V and Brown CV (2010) 'Designing data governance', *Communications of the ACM*, 53:148–152, doi:10.1145/1629175.1629210
- Klump J, Lehnert K, Ulbricht D, Devaraju A, Elger K, Fleischer D, Ramdeen S and Wyborn L (2021) 'Towards globally unique identification of physical samples: governance and technical implementation of the IGSN global sample number', *Data Science*, 20:33, doi:10.5334/dsj-2021-033
- Krahe M, Milligan E and Reilly S (2019) 'Personal health information in research: Perceived risk, trustworthiness and opinions from patients attending a tertiary healthcare facility', *Biomedical Informatics*, 95:103222, doi:10.1016/j.jbi.2019.103222
- Kratz JE and Strasser C (2015) 'Researcher perspectives on publication and peer review of data', *PLoS ONE*, 10(2):e0117619, doi:10.1371/journal.pone.0117619
- MacLeod A, Wong M, Gregory L, Schneider D, Williams A, Castleden I, Simons B, Levett K and Box P (2020) [The Agricultural Research Federation \(AgReFed\) Technical and Information Policy Suite. Version 1.0](#), Agricultural Research Federation, accessed 25 May 2026.
- Marchildon P, Bourdeau S, Hadaya P and Labissière A (2018) 'Data governance maturity assessment tool: A design science approach', *Projectics / Proyética / Projectique*, 20(2):155–193, doi:10.3917/proj.020.0155
- Montgomery L, Neylon C and Huang K (3 December 2021) [What can we gain from open access to Australian research? Climate action for a start](#), The Conversation website, accessed 25 May 2026.
- NHMRC (National Health and Medical Research Council) and ARC (Australian Research Council) (2019) [Management of data and information in research: A guide](#)
- Soil CRC research project final report: A coordinated and consistent approach to managing and enabling soil data | Page 90 of 199

[supporting the Australian Code for the Responsible Conduct of Research](#), NHMRC website, accessed 25 May 2026.

- Pinfield S, Cox AM and Smith J (2014) 'Research data management and libraries: relationships, activities, drivers and influences', *PLoS ONE*, 9(12):e114734, doi:10.1371/journal.pone.0114734
- Rafiq M and Ameen K (2022) 'Research data management and sharing awareness, attitude, and behavior of academic researchers', *Information Development*, 38(3):391–405, doi:10.1177/02666669211048491
- Robinson N, Channon J, Clarendon S, and Medyckyj-Scott D (2025) 'Managing data in collaborative research starts with the researcher: a Soil CRC case study', *Accountability in Research*, 33(3): 2525878, doi:10.1080/08989621.2025.2525878
- Rosenbaum, S. (2010). Data governance and stewardship: Designing data stewardship entities and advancing data access. *Health Services Research*, 45(5 Part 2), 1442-1455.
- Smale N, Unsworth K, Denyer G, Magatova E and Barr D (2020) 'A review of the history, advocacy and efficacy of data management plans', *International Journal of Digital Curation*, 15(1):1–29, doi:10.2218/ijdc.v15i1.525
- Srinivasan U, Ramachandran D, Quilty C, Rao S and Nolan M (2018) *Flying Blind: Australian Researchers and Digital Health*, Australian Health Data Series Volume 2, Digital Health Cooperative Research Centre, Sydney.
- Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, Manoff M and Frame M (2011) 'Data sharing by scientists: practices and perceptions', *PLoS ONE*, 6(6):e21101, doi:10.1371/journal.pone.0021101
- von Davier AA, Hao J, Liu L and Kyllonen P (2017) 'Interdisciplinary research agenda in support of assessment of collaborative problem solving: lessons learned from developing a Collaborative Science Assessment Prototype', *Computers in Human Behavior*, 76:631-640, doi:10.1016/j.chb.2017.04.059
- Whyte A and Tedds J (2011) [Making the case for research data management](#), (Digital Curation Centre website, accessed 25 May 2026.
- Wong M, Box P, Epstein J, Lee A, Thompson H, Levett K, Channon J, Wilson P, Taylor N, Hergenhan R, Berger B and Gilliam M (15 June 2021) [The Agricultural Research Federation \(AgReFed\): Steering policies, roles and responsibilities. Version 1.1.](#), Agricultural Research Federation website, accessed 25 May 2026.
- Yu F, Deuble R and Morgan H (2017) 'Designing research data management services based on the research lifecycle – a consultative leadership approach', *Journal of the Australian Library and Information Association*, 66(3):287–298, doi:10.1080/24750158.2017.1364835
- Zhou X, Xu Z and Kogut A (2023) 'Research data management needs assessment for social sciences graduate students: A mixed methods study', *PLoS ONE*, 18(2):e0282152, doi:10.1371/journal.pone.0282152

APPENDIX A: THE SOIL RESEARCH DATA MANAGEMENT LIFECYCLE

Management of soil research data can be considered in terms of a research data lifecycle. The Australian Code for the Responsible Conduct of Research (NHMRC & ARC & UA, 2019) note that there are key responsibilities in the conduct of research including data:

- generation
- collection
- access
- use
- analysis
- disclosure
- storage
- retention
- disposal
- sharing and re-use.

There are general steps in the lifecycle of data from conception through to termination (disposal). The following diagram (Figure A-1) represents seven stages in this lifecycle for research data. For research data, these stages are:

Step 1. Plan – The conception of data that may include reviewing what data already exists, will newly created data satisfy needs to address questions posed, the need for consent and permission, costs of generation and acquisition (all as part of a Data Management Plan).

Step 2. Create – Where data is produced (often as observations or measurements) from experiments, sensors, surveys, etc. Ideally at this stage metadata is also generated that provides provenance and relevant context to the data. A metadata example could be the sensor specifications, mode of operation and procedures used in data creation.

Stage 3. Process – Once data is created, it should be prepared for analytical purposes, Validation, cleaning, combined with other data, transformation and quality control/assurance procedures are undertaken. These procedures in data revision are also described so that they can be repeated.

Stage 4. Analyse – When data is reviewed, explored and interpreted. The analysis is central to research discovery and testing of theories and hypotheses. Generally the findings of analysis are documented and recorded either in reports or publications such as journal papers.

Analytical procedures are also recorded and should be repeatable. Data used in the analysis can be cited in readiness for reuse purposes.

Step 5. Preserve – Once the data has served its primary purpose (original intention or reason for collection), the data can be curated for future reuse purposes. The preservation and archival of this data also fulfills the purpose of substantiating research findings (justification and verification of research). Procedures used here include saving the data in formats that adhere to prescribed standards (e.g. for machine readability in an open file format), creation of a metadata record, derivation of a digital identifier (e.g. DOI), deposition in a suitable and trusted repository for security, linking to other derivatives such as research reports, assignment of licence and permission controls.

Step 6. Share – The data along with appropriate accompanying information (mainly procedures listed under Preserve) should be detailed and this made accessible via a trusted repository. Here the data can be discovered with any conditions on re-use detailed in license and permissions. Access to the data is given when the interested party adheres to detailed conditions. Sharing of the data enables maximum potential for future research and avoids potential repeat data collection (waste of resources).

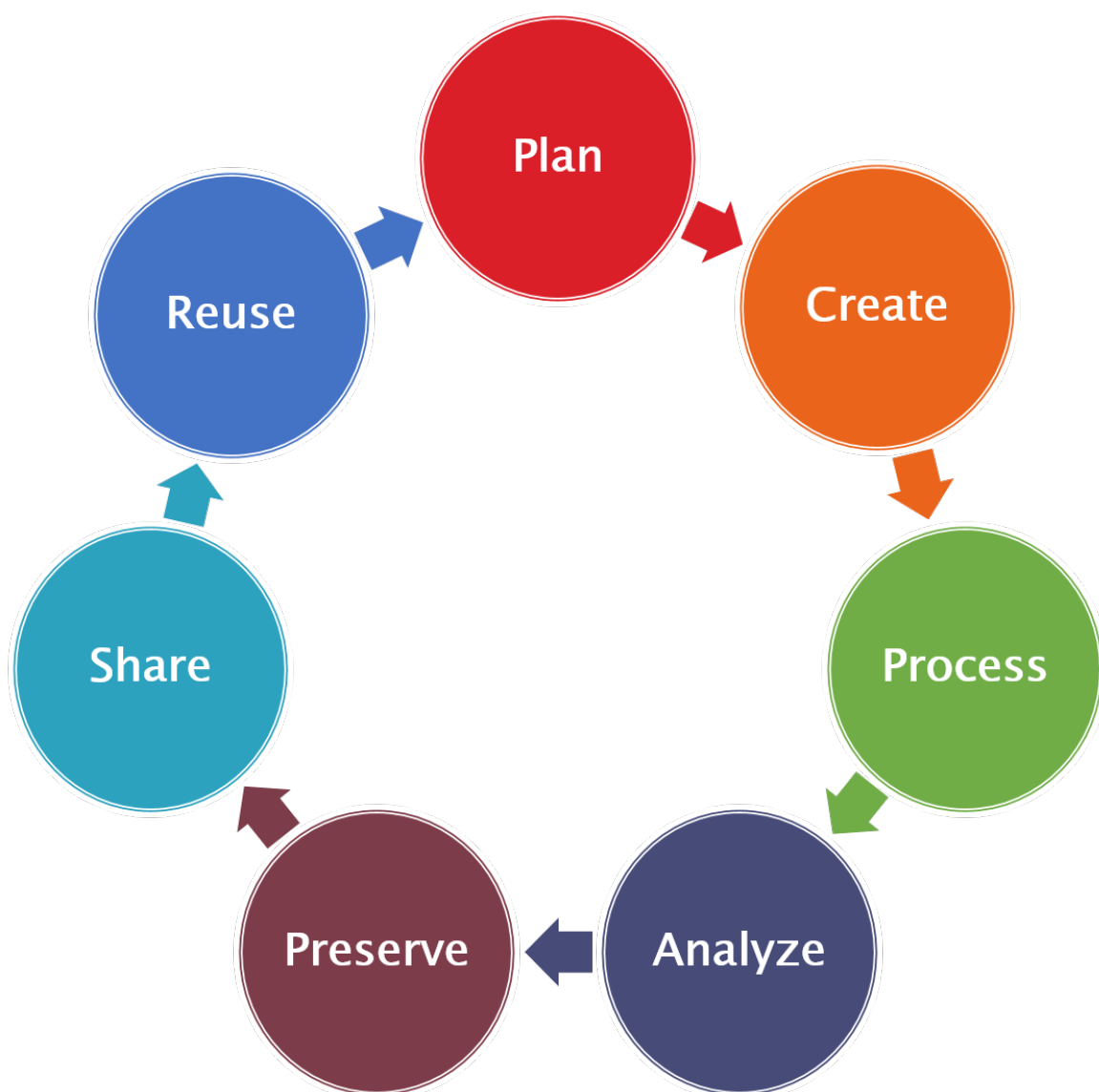


Figure A-1. Research data lifecycle.

Step 7. Reuse – When an interested party deploys the data for a secondary purpose (may or may not been intended in the original collection). Such purposes could include repeating an experimental method to substantiate original findings, to generate new research findings by analysing the data using a new or different method, combine the data with other data for new insights, or for new purposes all together (repurposing). Other secondary purposes could include deployment for commercial outcomes, teaching and for policy and decision making needs. Citation and reuse requirements should be detailed in the metadata record.

Definitions

Dataset: An identifiable collection of data. A dataset can be a smaller grouping of data which, though limited by some constraint such as spatial extent or feature type, is located physically

within a larger dataset. Theoretically, a dataset can be as small as a single feature or feature attribute contained within a larger dataset. [SOURCE: ISO 19115-1:2014, 4.3]

Dataset series: collection of datasets sharing common characteristics. [SOURCE: ISO 19115-1:2014, 4.4]

Data governance: Data governance is the exercise of authority and control over the management of data. Governance is generally concerned with the decisions and who makes those decisions (rights) with respect to the data asset. [SOURCE: DAMA International, 2009.

Data management: Involves the implementation of those decisions in managing the data asset. [SOURCE: (Alhassan et al., 2016; Fu et al., 2011; Khatri & Brown, 2010).

Digital repository: A system that can organise, store, and retrieve digital assets.

[SOURCE: Clobridge, A. (2010) Introduction, Building a Digital Repository Program with Limited Resources, Chandos Publishing, p. 3-11, <https://doi.org/10.1016/B978-1-84334-596-1.50001-8>.]

Information assets: Data that is or should be documented and that has value or potential value. [SOURCE: The Hawley Committee (1995) Information as an Asset: The Board Agenda. KPMG, London.]

Institutional repository: A library of digital objects and associated metadata from a single institution. [SOURCE: Clobridge, A. (2010) Introduction, Building a Digital Repository Program with Limited Resources, Chandos Publishing, p. 3-11, <https://doi.org/10.1016/B978-1-84334-596-1.50001-8>.]

Metadata: information about a resource. [SOURCE: ISO 19115-1:2014, 4.10]

Quality: degree to which a set of inherent characteristics of an object fulfils requirements. [SOURCE: ISO 9000:2015, 3.6.2]

APPENDIX B: SOIL CRC DATA MANAGEMENT PLAN AND GUIDE



Performance through collaboration

Data Management Plan

GUIDELINES



CONFIDENTIAL

Not to be distributed beyond the CRC for High Performance Soils (Soil CRC) Participants and Affiliates without the consent of the CEO.

DISCLAIMER

Any opinions expressed in this document are those of the authors. They do not purport to reflect the opinions or views of the Soil CRC or its partners, agents or employees.

The Soil CRC gives no warranty or assurance and makes no representation as to the accuracy or reliability of any information or advice contained in this document, or that it is suitable for any intended use. The Soil CRC, its partners, agents and employees disclaim any and all liability for any errors or omissions or in respect of anything or the consequences of anything done or omitted to be done in reliance upon the whole or any part of this document.

GUIDELINES TO THE DATA MANAGEMENT PLAN

Sound data management is an important part of the responsible conduct of research. The data that you reuse, collect or create during your research are assets which need to be managed throughout the research project and cared for into the future.

You will benefit from having in place robust and easy-to-use data management practices by being better organised and working more efficiently. For example, you can:

- Find and understand your data when you need to use it
- Avoid unnecessary duplication of effort
- Ensure continuity of the research if project staffing changes
- Maintain the data underlying publications which allows for the validation of your findings
- Ensure your data can be readily shared which can facilitate research collaboration
- Meet data management obligations under the Australian Code for the Responsible Conduct of Research
- Avoid catastrophes such as the loss of data or violations of privacy.

A Data Management Plan (often referred to as a DMP) describes the activities that will be associated with the data other than the direct use of the data and may include: data reuse, creation, storage, retention and disposal, documentation, long-term preservation and provisions for access, data sharing and reuse.

This document is a guide to help structure and inform your planning process in preparation for a written DMP.

Before you start to use the guide and the DMP template, it is recommended that you read and understand these key national and institutional documents:

- Australian Code for the Responsible Conduct of Research (Section Two)
- Your institutional policies concerning data management

If you use the template, **adapt it to suit your specific data management needs** for your research project. Not all sections and/or questions will be relevant to your project. Other sections can be expanded, and supplementary material attached where relevant. The detail should be clear, concise, proportionate to the scale of the datasets used, collected and generated and their anticipated level of value to the research community.

Different disciplines have different conventions with respect to good data management. To facilitate data linkage and interoperability, you should ensure that your data management is compatible with any recognised data management practice and standards in your discipline (e.g. in relation to file formats and the metadata standard to use).

The project's DMP will be a living document so regularly revisit the content of your DMP and update it as necessary as the project evolves and progresses. It should be stored as part of your consolidated research documentation.

SOURCES

Australian Antarctic Data Centre (2015) [*Project 1234 data management plan*](#), accessed 26 May 2026.

Digital Curation Centre (2013) [*Checklist for a data management plan. V.4.0.*](#), DCC website, accessed 26 May 2026.

Federation University Australia (2015) [*Data management planning. Guide and template*](#), accessed 26 May 2026.

ABOUT THE DATA MANAGEMENT PLAN

The DMP template, when completed, will provide your project team and the Soil CRC with the necessary information to ensure that data used and created in the project is managed efficiently and effectively for research and future use purposes.

Before starting, ensure that you have all the information required in the **CHECKLIST**.

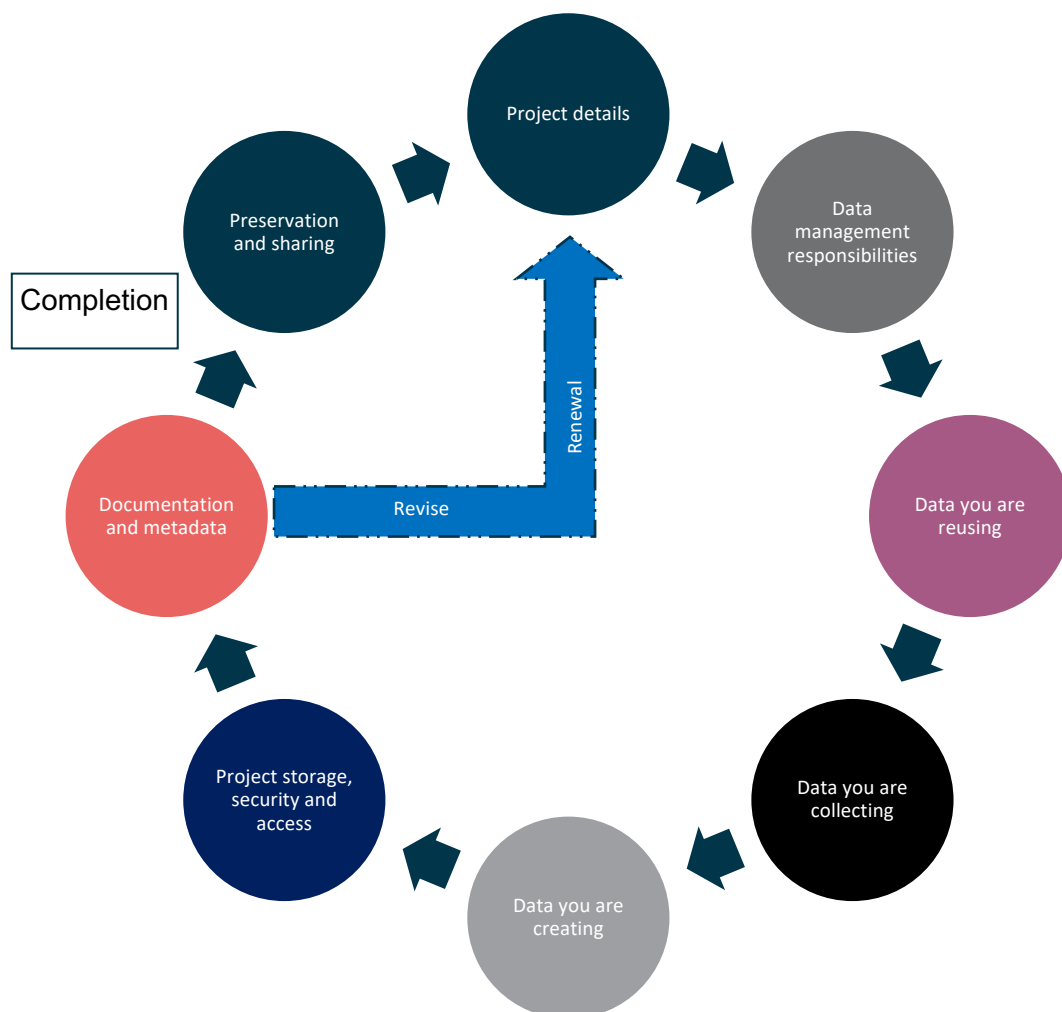
If this is your first time compiling a DMP, this process will enable you to document all aspects of your data management. If this has been done before, then relevant information can be reused provided it is still relevant to the data in your project.

CHECKLIST

Information required to complete this form includes.

☐	Details of your research project (name, duration, project leader, Soil CRC Program, description, contact details)
☐	Who is responsible for data management in your project (operational roles and responsibilities, institutional support, compliance)
☐	The data you are reusing in your project (what, who, terms of use, sharing)
☐	The data you are collecting (what, who, ownership, ethics, copyright, Intellectual property)
☐	The data you are creating (content, types, formats, volumes, supporting software/hardware)
☐	The documentation you have for the data (metadata, version control, associated documents and methods)
☐	The data to preserve, retain, secure and publish beyond the end of your project (retention, disposal, licensing, sharing through repositories)

CHECKLIST STAGES OF A DATA MANAGEMENT PLAN



These are the **CHECKLIST** stages in the creation of a DMP. Note the **Revise** and **Renewal** steps occur as the project evolves with the Soil CRC. As datasets are completed, or the project nears its completion, these datasets are preserved and shared through agreed repositories defined in this plan.

DETAILS ABOUT YOUR RESEARCH PROJECT

This section will detail fundamental details and information about your project including the Soil CRC project: identity, duration, leader, investment round, description and DMP author.

WHO IS RESPONSIBLE FOR DATA MANAGEMENT IN YOUR PROJECT

This section is focused on naming the individual/s who are responsible for key data management activities through the lifetime of your project. It is important that those named individuals who are assigned operational roles and responsibilities (including institutional support, services) are aware, informed and in agreement of these duties.

THE DATA YOU ARE REUSING

Here you will identify existing datasets and data collections that you will be using in your research project. The data you reuse during your project may change and evolve as the project progresses. The table provided to detail these datasets will also assist you in identifying any limitations (noted in reuse terms and conditions) for your research outputs (e.g. onward sharing, commercialisation caveats and limitations). The name of the dataset and its description should be short and concise but with sufficient detail to enable someone outside your project team to obtain a basic understanding of data being reused in your project. This will also be valuable in fulfilling **Pre-existing Material and Third-Party Intellectual Property (IP)** requirements in project contracts and updates.

THE DATA YOU ARE COLLECTING

When your project is collecting data, there may be data collection rules, standards or policies that you must adhere to (e.g. institutional ethics, Soil CRC, legal rules and regulations). It will also help you consider if there are data quality considerations (control, assurance, procedures) required to maintain and ensure the integrity of your data. For projects collecting personal data and information, it is also worth noting this and if there are institutional ethics procedures and permissions required.

You need to consider how data will be stored and secured during the project (storage, security, backup, access and permissions). At a minimum you should deposit data in a system suitable for the management of the data and reuse of the data by the project team members, partners and others. Processes to organise the data (e.g. version control strategies, file and folder structures and naming conventions, and physical filing systems) should also be documented and shared with the project team. You may need to consider data security. Consult the Soil CRC or data specialists in your institution for advice.

THE DATA YOU ARE CREATING

With your project, there will be data you create during the project lifetime. It is important to consider what this is likely to be, and if there will be any institutional requirements to support this (e.g. storage, security, accessibility, file sizes and formats). This section will help you identify if there are likely to be any requirements here to support managing the data your project creates. This will likely evolve as your project progresses and will provide a useful basis from which to engage those in support of institutional data repositories and preservation of data and metadata beyond your project lifetime.

THE DOCUMENTATION YOU HAVE FOR THE DATA

There will be documentation for the data that may include metadata, version control, associated documents and methods. Metadata describing your data should be human- and machine-readable and to a Soil CRC community-endorsed standard. It should also use a controlled vocabulary¹ to assist with findability. It is recommended to use a vocabulary with terms suitable for describing soils data. Provenance information about the collection and/or generation of data and how it was processed should be included in the metadata. The metadata should remain available over time, even if the data is no longer accessible. There should be clear acknowledgement of those involved in collecting, processing and analysing data when data is shared.

THE DATA TO PRESERVE, RETAIN, SECURE AND PUBLISH BEYOND YOUR PROJECT

There will be data to preserve, retain, secure and publish beyond your project (retention, disposal, licensing, sharing through repositories). This research data should be findable, accessible, interoperable and reusable (FAIR) to ensure it is soundly managed. However, there can be legitimate reasons not to share soil data and associated data (e.g. privacy protection, or ethical, legal or commercial constraints). As such, it is your responsibility to be aware of what can be shared, with whom and when, and to take appropriate steps to ensure that the data is as open as possible and as closed as necessary.

Data should ideally be assigned a globally unique persistent and resolvable identifier when shared and when deposited with a secure and trusted long-term data repository. Ideally this should include associated metadata and any relevant ancillary information also. A persistent identifier is a long-lasting reference to the data and associated materials. The data you deposit in a data repository and/or publish should be assigned a globally unique, persistent and resolvable identifier (PID) so that both humans and machines can find it. It is highly recommended that you use open file formats for the data generated by the project, so that others can easily access and reuse it.

Before filling in this section you may want to talk to the Soil CRC and your institutional IT support and data specialists. If there is software associated with the data, you will need to plan how you make the software available.

¹ A controlled vocabulary reflects agreement on terminology used to label concepts in datasets, improving the discovery, linking, understanding and reuse of research data (<https://ardc.edu.au/services/research-vocabularies-australia/>).

GLOSSARY OF DATA MANAGEMENT TERMS

This glossary has been developed by The Centre for e-Research and Digital Innovation (CeRDI) at Federation University Australia. It provides definitions for select terms that might be unfamiliar. This is not meant to be a comprehensive list of all data management terms, rather it is a list of ones that we have found, in our experience, to be most used.

TERMS

anonymised data: Data about individuals that does not reveal their identities nor links to other data that would reveal their identities. This term is not the same as "coded data" or "de-identified data".

archive (verb): The transfer of material to a facility that stores, retains and provides access to that material on a long-term or permanent basis.

attribution: The act of referencing a dataset's original creator, usually signalled by copyright or licensing information accompanying the dataset.

citation: The practice of referencing research. A data citation includes key descriptive information about the data, such as the title, source and responsible parties.

code: This applies to computer codes or technical scripts. In the context of data management, this may include code used in the collection, manipulation, processing, analysis or visualisation of data, but may also include software developed for other purposes.

coded data: Data accompanied by a linkage file that connects unique identifiers about an individual (such as their name, medical number, email address or telephone number) with a unique study ID code not associated with the individuals' personal information.

confidentiality: The right of privacy and security of disclosed personal information. This mainly applies to data collected on human subjects. Researchers are guided by strict legal requirements to prevent the release of private, personally identifiable information provided by research subjects.

copyright: A set of legal rights extended to copyright owners (the author, creator, or organisation to whom the rights have been assigned) that govern such activities as reproducing, distributing, adapting or exhibiting original research.

data: Data is the information used to validate research findings.

data governance: A collection of processes, roles, policies, standards and metrics that ensure the effective and efficient use of data.

data licensing: A legal arrangement from the owner or repository of the data specifying what users can do with the data.

data life cycle: Refers to the sequence of stages that data goes through. It encompasses data creation, data storage, data usage, data archiving and data destruction.

data management plan: A data management plan requirements may be defined differently by different funders, programs or organisations. It is a plan that clearly outlines actions that will be taken throughout the data life cycle.

data quality assurance process: The process of identifying and eliminating anomalies by means of data profiling and data cleansing.

data stewardship: The management and oversight of an organisation's data assets.

de-identified data: Data for which all direct or indirect identifiers or codes linking the data to individual subjects' identities are destroyed.

derivative (of data): Any data, publication, illustration/visualisation or other work that rearranges, presents or otherwise makes use of an existing data set.

digital object identifier (doi): An electronic “tag” created to uniquely identify objects or data.

domain specific data standards: Data standards that are designed and intended for use in a particular domain.

FAIR principles: Acronym for the main qualities of a dataset: "Findable, Accessible, Interoperable, and Reusable".

file format: The technical structure used for encoding data in a computer file. File formats are usually identified by the file extension (e.g. .xlsx, .csv, .dbf).

institutional repository: A service hosted by an institution for storing and providing online access to digital content. The institution hosting this service is usually the institution that creates the content.

intellectual property: Legal rights applied to creative works, including (but not limited to) copyright, patents, trademarks and trade secrets.

license: In the context of data management, a legal statement that expresses the terms of use of data.

machine-readable data: Structured data that can be easily processed by a computer.

metadata: Documentation or information about a data set. It may be embedded in the data itself or exist separately from the data. Metadata may describe the ownership, purpose, methods, organisation, conditions for use of data, technical information about the data, and other information.

metadata standards: A requirement which is intended to establish a common understanding of the meaning or semantics of the data.

open access: Refers to online, freely available material that has few or no copyright or licensing restrictions.

open data: Data that can be freely accessed, used, modified and shared. Entails legal and technical permissions for users to access and modify the data.

open-source software: Software for which the source is available under an open license. Allows users to inspect the source code, run their own versions of the program, offer modification suggestions, fix bugs, develop new features, etc.

persistent identifier: A unique and long-lasting reference that allows for continued access to a digital object. Examples of persistent identifier systems include Digital Object Identifiers (DOIs), handles, and Archival Resources Keys (ARKs). Persistent identifiers support interoperability and the reliable citation of digital content.

preservation (of data): Ensuring that data remains intact, accessible and understandable over time. This requires preserving the integrity of digital files themselves.

primary data: Primary data is a type of data that is collected by researchers directly from main sources through interviews, surveys, experiments, etc.

privacy: The protection of personal information from unauthorised access by others, usually governed by legislation.

repository: A facility that manages the appraisal, preservation and accessibility to materials on a long-term or permanent basis.

restricted data or restricted access data: Data which is made available under stringent, secure conditions. Typically, confidential or sensitive data.

secondary data: The data that have been collected for another purpose but has some relevance to your research needs and/or the data is collected by someone else instead of the researcher themselves.

security: Methods of protecting data from unauthorised access, modification or destruction.

standards: Accepted methods or models of practice. In the context of data management, standards typically apply to data or file formats and to metadata.

tabular data: Data that appear in a table format.

usage statement: An expression of the conditions under which a data set may be used. May be formal, such as a license or contract, or informal based on the preferences of the data owner(s).

EXAMPLE DMP

DETAILS ABOUT YOUR RESEARCH PROJECT

Project no.	2.1.006
Project title	A very consistent way to manage your data
Project duration	1/1/2023 – 30/6/2099
Project leader	Dr Nathan Robinson
Soil CRC Program	Program 2 – Soil performance metrics
Soil CRC Investment round	Round 6
Project description	This project will use social research, data analytics and soil indicator domain expertise to determine how commonly used indicators or suites of indicators (“horses”) can best be matched to landscapes, climates, ecosystems and land management goals (“courses”). It will also explore whether alternative indicators might be more successful in matching the expectations of the end-users.
Author details (add ORCID ID if applicable)	Dr Nathan Robinson Senior Research Fellow Soil Science Centre for eResearch and Digital Innovation Federation University Australia Suite 15 Greenhill Enterprise Centre Mt Helen PO Box 663 Ballarat VIC 3353 T +61 3 5327 8436 M 0427 056 585 n.robinson@federation.edu.au

WHO IS RESPONSIBLE FOR DATA MANAGEMENT IN YOUR PROJECT

Who is responsible for the implementation of the Data Management Plan, its revision and updates?	Nathan Robinson
Is ownership and responsibilities for the Data Management Plan part of the agreement between project partners?	No, however there are negotiations underway with project partners due to the complexity around ownership of varying datasets (use cases). The DMP was not part of the original project agreement so applying retrospectively is challenging.
Who (named individual/s) is responsible (and role) for each data management activity?	
Collection rules / standards / policies	Nathan Robinson (Project leader) and Alison Ollerenshaw (Survey leader)
Data reuse (outside the project)	Survey: linked to ethics agreement. Likelihood of sharing aggregated data but Soil CRC owns the data, therefore permission for reuse by the Soil CRC required. Also unclear on reuse rules once the Soil CRC ceases. Query here is the reuse of metadata or data?
Data creation	Survey: Dr Aakansha Chadha; Dr Alison Ollerenshaw; Dr Nathan Robinson Use Cases: In-Young Yeo ; Indishe Senanayake ; Sepideh Jalayer
Storage, security and access for within project use	Survey: Fed Uni SharePoint – Alison Ollerenshaw Use Cases: UoN data servers – Indishe Senanayake
Documentation and metadata	Survey: Fed Uni SharePoint – Alison Ollerenshaw Use Cases: UoN data servers – Indishe Senanayake
Data preservation, retention, security, reuse and managing access to the data beyond the project	Survey: Fed Uni SharePoint – Alison Ollerenshaw Use Cases: UoN data servers – Indishe Senanayake

THE DATA YOU ARE REUSING IN YOUR PROJECT

The name of the reuse dataset	A short description of the data you are reusing	Collection duration	File format/s	Data reuse terms and conditions	Data volume – if applicable (MB, GB, TB)
University of Newcastle soil moisture (SM) probe and carbon collection	The University of Newcastle maintains a network of soil monitoring stations in the Krui River Catchment and Merriwa River Catchment of the Goulburn River. There are 26 monitoring sites (2002–2022) with a number of in-situ soil data collected in the last 20+ years.	1/1/2002 – 31/12/2022	*SM probe proprietary files and csv files for soil data	Free access for the duration of the project has been granted.	10MB
Aggregated social science data from Soil CRC scoping studies	Previous Soil CRC social science research reports by: • Dahlhaus et al. (2018): Survey of 122 respondents with 38% farmers, 30% advisers and 16% researchers • Orgill et al. (2018): Survey of 135 advisers of which 96 estimated they were working with over 3,275 clients in six states • Parsons et al. (2018): Survey of 111 participants from 19 grower groups over five states.	1/7/2017 – 30/06/2018	Word documents	These reports have been made available to this project without condition via the Soil CRC.	10MB
Soil and Landscape Grid of Australia (CSIRO DAP)	National raster datasets for soil properties. These are produced for six depth intervals aligned to GlobalSoilMap.net specifications	1950 – 2021	Geotiff	Attribution 4.0 International (CC BY 4.0)	>1TB

Do the terms of use of the data allow you to onward share data generated from that data? Check data supply agreement.	Survey: Not allowed due to ethics conditions. Use cases: Covered by data supply agreements.
Do the terms of use of the data allow you to retain the data or information derived from that data at the end of your project?	All datasets reused have specified conditions and terms that will be adhered to at the completion of the project.

THE DATA YOU ARE COLLECTING

When collecting data you should abide by any data collection rules/standards/policies mandated by the Soil CRC and your institution as well as any legal requirements.

What data is being collected in this research?	Survey: Individual responses from Qualtrics survey Use Cases: None
Who owns the data (sets) and intellectual property rights arising from your research? (Note: unless otherwise specified in Project Agreement, this is the Soil CRC)	Soil CRC
What data quality assurance processes will you apply?	Survey: In-built Qualtrics QA Use Cases: Internal project team peer review
Does the data you will collect raise any data privacy issues? How will these be handled?	Survey: Yes: Outlined via ethics application Use cases: Data supply and licence agreements
Does the data you collect require ethics approval? How will you obtain this?	Ethics approval granted from Federation University Reference number 2022/213
Where will you store the digital data during the project's lifetime?	Survey: Fed Uni SharePoint – Alison Ollerenshaw Use Cases: UoN data servers – Indishe Senanayake
How (who, where, and how often) will the data be backed up during the project's lifetime?	Data will be backed up according to existing infrastructure procedures
If applicable, how will you manage access restrictions and data security during the project's lifetime?	All access is under institutional login security and permission
How and where will non-digital data be stored, including any security and environmental requirements needed to safeguard non-digital data?	Data storage for this project will be overseen by the principal researcher and the other researchers named on this application. They will ensure appropriate storage of all data and information in a locked filing cabinet, in a locked room, located at Federation University/University of Newcastle. Data collected for this research will be stored on a password-protected computer/laptop or on a password-protected web-based database.

THE DATA YOU ARE CREATING

What data are you creating (content, types, formats, volumes, supporting software/hardware)?

Name of dataset	A short description of the data you are creating	Collection duration	File format/s	Data volume (MB, GB, TB)
Use case 1	Assessment of data management issues with project X collecting and handling soil data	1/11/2023 – 31/01/2024	TXT	<5 MB
Use case 2	Assessment of data management issues with project Y collecting and handling soil data	1/11/2023 – 31/01/2024	TXT	<5 MB
Indicator survey	In development – aggregated dataset according to ethics specifications	1/9/2023 – 30/10/2023	csv file	<2 MB

File formats (others are acceptable):

- Documents: HTML, TXT, XML, RTF, PDF, MS Word
- General observational and measurement data: CSV, NetCDF, MS Excel
- Spatial data: Esri shapefile or geodatabase, GeoTIFF
- Databases: SQLite, MS Access
- Images: EPS, SVG (for vector images), PNG, TIFF (for images with sharp edges), JPEG (for photographs)
- Audio: WAV or formats with lossless compression where possible

THE DOCUMENTATION YOU HAVE FOR THE DATA

What metadata will be created, captured, published (and where) with the data to make them interpretable and reusable?	Records of data collected – shared through Soil CRC, Federation University and University of Newcastle approved repositories
What metadata standard do you propose to use?	ANLIC standard in alignment with RDA, Soil CRC, Federation University and University of Newcastle requirements
Note any other supporting information/documentation that will enhance understanding of the data (e.g. data dictionaries, codebooks, lab notebooks and methodology reports).	Separate reports will be produced to accompany the created datasets
Will there be versions of data created? How will these be managed?	Not anticipated

THE DATA TO PRESERVE, RETAIN, SECURE AND PUBLISH BEYOND YOUR PROJECT

Data sensitivities	
How will you resolve any data sensitivity (e.g. ethical/ privacy/confidentiality) issues that may prohibit sharing some or all the dataset(s)? Will this mean onward use of the data will be restricted in some way?	The survey data will be managed in full compliance of ethics policies and those documented in the Ethics application. There may be onward use of the data restrictions for data provided by the University of Newcastle.
Will your research data be embargoed for any reason? If yes, provide details.	Yes – for research publication purposes in agreement with Soil CRC policies
Access	
Will your data be made accessible: • to other CRC Projects? • to CRC partners and stakeholders? • to the public?	Access only permitted in alignment with survey ethics
Will any access and or usage restrictions will be imposed, and why?	Access only permitted in alignment with survey ethics
How will your data be made accessible and shared?	Access only permitted in alignment with survey ethics
What data sharing license do you think is appropriate for this research data? You may need to consult the Soil CRC or the legal people in your institution for advice on licencing.	Data licensing unlikely due to existing survey ethics
Will the data continue to be made accessible to other Soil CRC projects, Soil CRC partners and the public after the project has ended? If yes, and this is not through the long-term data repository, how will access be provided?	Data access unlikely due to existing survey ethics. Summarised survey results can be used by other Soil CRC projects

Storage and preservation	
When do you expect your research data to be moved to a secure archive for long-term preservation and, usage rights allowing, ongoing access?	This will likely occur in the 6–12 months post the project completion.
Where will your research data be deposited for long-term preservation?	Federation University Figshare – to be in agreement with Soil CRC policies
What physical samples/records associated with the electronic data should or will be stored beyond project completion? Where will they be stored and who will be responsible for them?	No hardcopies to be stored beyond the project. All digital records retained on Sharepoint for a 5-7 year duration after the project finishes.
Who will be the contact person for the data after the project has ended?	Nathan Robinson with the Soil CRC and in adherence of institutional policies.
If permitted, how long will or should the data be kept beyond the life of the project?	5-7 years and in adherence of institutional policies.

APPENDIX C: DATA MATURITY SURVEY - INVITATION

A coordinated and consistent approach to managing and enabling soil data: Data maturity survey

Researchers at Federation University Australia, including Dr Nathan Robinson, Ms Jude Channon, and Ms Jennifer Corbett are conducting research which has been funded by the Cooperative Research Centre for High Performance Soils (Soil CRC).

The aim of this research is to better understand the current state of research data governance principles and management practices in use across Soil CRC projects.

This study involves the collection of data from an online survey. The questions asked to survey participants are solely focussed on the data from Soil CRC projects. The research findings will provide important insights to assist the research team to understand the data governance and management needs of Soil CRC projects to help enable appropriate access and use of the research data produced by the Soil CRC Projects.

Important information about this research is contained in the **Plain Language Information Statement**.

This research has received approval from Federation University Australia's Human Research Ethics Committee (Approval number: 2022-143).

The survey takes approximately 15-20 minutes to complete and can also be **downloaded in hardcopy**. If you agree to participate in this research, please click on the '**next**' button, below, to commence the survey. Alternatively, you can schedule a time to complete the survey over the phone by contacting Jude Channon on 03 5327 9672 or j.channon@federation.edu.au

For further details about this research please contact Dr Nathan Robinson (contact details listed below) or any of the researchers listed in the Plain Language Information Statement (above). The contact details of all researchers are also listed at the end of this survey.

Dr Nathan Robinson

Senior Research Fellow

Centre for eResearch and Digital Innovation

Federation University Australia

E: n.robinson@federation.edu.au

T: (03) 5327 8436 | M: 0427 056 585

Questions

1. Which of the following best describes your affiliation/organisation?

Please choose **only one** of the following:

- State government agency
- University
- Industry partner
- Grower group
- Other

2. Which of the following best describes your role/s across Soil CRC projects?

Click all that apply.

- Project leader
- Researcher
- Administrative assistant
- Technical support
- PhD candidate
- Post-doctoral researcher
- Other – please specify:

3. Which of the following programs aligns to your project/s? Click all that apply

- Program 1 – Investing in high performance soils
- Program 2 – Soil performance metrics
- Program 3 – New products for soil fertility and function
- Program 4 – Integrated soil management solutions
- PhD program
- Prefer not to disclose

4. Please describe the types of soil data and associated data you are collecting and/or using for the research associated with your project/s:

Please write your answer here:

5. Are you bringing any existing data to your project/s?

- Yes
- No

- Not applicable

Please specify what this data is:

6. If you are bringing existing data to your project/s, have you noted this in the Soil CRC intellectual property register?

- Yes
- No
- Not applicable
- Prefer not to answer

7. Do your project/s have a data management plan?

- Yes
- No
- I don't know but am aware of what a data management plan is
- I don't know what a data management plan is

8. What types/file format of data has or will be collected? Click all that apply

- Numerical or statistical (e.g. CSV, MAT, XLS, SPSS, R, etc.)
- Text (e.g. TXT, DOC, PDF, RTF, HTML, XML)
- Geospatial
- Instrument specific
- Models
- 3D imaging
- Images
- Video
- Other – please specify:

9. How is the data generated through your project/s going to be used?

Please write your answer here:

10. Who is primarily responsible for the day-to-day research data management during the active phase of the project/s? Click all that apply

- Myself
- Project leader

- Researcher
- Administrative assistant
- Technical support
- PhD candidate
- Post-doctoral researcher
- Other – please specify:
- I don't know

11. What storage locations do/did you use for the data from the research associated with your project/s? Please choose ALL that apply

- My work computer
- Home computer
- External hard drive/Portable storage/Flash drive (USB)
- Storage provided by my department or school
- Storage provided by my institution or organisation
- Cloud-based data storage e.g. Dropbox, MS OneDrive, Google Drive
- Physical hardcopy (retained in boxes, cabinets, etc.)
- Other – Please specify:

12. How do you ensure the security of your data? Choose ALL that apply

- Data is protected via an authentication process
- Password protected
- Data is de-identified (i.e., only anonymous data stored)
- Data is encrypted
- Data is not secured
- Other – Please specify:

13. Is there a backup or duplicate copy of the data used or produced through your research?

- Yes
- No
- Partial backup of data
- Unsure

14. If you answered 'Yes' or 'partial backup' to the last question, what storage locations do/did you use for the backup?

- Home computer
- External hard drive/Portable storage/Flash drive (USB)

- Storage provided by my department or school
- Storage server, provided by my institution
- Cloud-based data storage e.g. Dropbox, MS OneDrive, Google Drive
- Physical hardcopy retained (in boxes, cabinets, etc.)
- Other – Please specify:

15. How do you deal with different versions of your working data files? Please choose ALL that apply

- I save multiple files and update the files with different names/naming convention
- This is done automatically (e.g., Dropbox)
- I use a versioning control program
- Version management is done by the database I am using
- I am not using versioning control
- I only use one dataset version
- Other – Please specify:

16. How have you published, shared, or otherwise disseminated the data resulting from this research – or how do you intend to do so? Tick all that apply

- I have not released any data and don't intend to
- Personal website or social media
- Via email
- Data server/service/web portal established by the project
- Data server/service/web portal established by a CRC project e.g. Visualising Australasia's Soils
- Institutional data repository
- Disciplinary data repository
- Dataset published in data journal
- I choose not to publish, share or disseminate the data
- Other – Please specify:

17. If you selected "I choose not to publish, share or disseminate the data", please specify your reasons for this?

Please write your answer here:

18. Who will have access to data resulting from your research? Please choose ALL that apply.

- Researchers in other Soil CRC Projects
- Soil CRC partners and stakeholders
- Other researchers, outside the Soil CRC
- Industry organisations
- The general public
- Other – Please specify:

19. Do you know who owns the research data resulting from your research?

- Yes
- No
- Unsure

20. If you answered yes, who owns this data? If your project involves different data, please focus your answer on key datasets only.

Name of dataset	Owner	Additional comments on ownership

21. Thinking about the key data sets, what data licence(s) have you used/will you use when you publish, share, or disseminate the data from your research?

- I have/plan to have data licences established for the data
- I didn't licence the data
- I don't plan on licensing the data
- I haven't decided which licences(s) to use but have access to information to help us make that decision
- I, am unsure which licence(s) to use
- I am unsure what a data license is

22. Who created – or assisted with creating – the metadata for the data that you generated, published, shared, or otherwise disseminated? Choose ALL that apply.

- I haven't created any metadata

- I have yet to create any metadata
- Myself
- Researcher involved in the project
- Administrative assistant
- Data specialist from my institution
- An individual from another CRC project
- PhD candidate
- Post-doctoral researcher
- Other – Please specify:

23. What other supporting documentation do you keep about your data? Choose ALL that apply.

- Description of methodology/protocol (e.g. method and materials)
- Data dictionary/coding manual
- Sources for third party data used in the project (source, terms of use, access restrictions)
- Instrument metadata (manufacture, model, parameters, calibration)
- Additional metadata (e.g. technical, preservation, copyright, access restrictions)
- Other – Please specify:
- I do not record any additional data documentation

24. Will the data you collect or create raise any data privacy issues? Yes/No

If yes, how will these be handled? Please write your answer here:

25. If applicable, how will you manage access restrictions and data security for the project/s duration? Please write your answer here:

26. Are you using domain accepted, standard vocabulary terms (e.g. from lists, glossaries or ontologies) to describe the data within your datasets (e.g. your data column headers)?

- No, I don't
- Yes I do, but descriptions of these terms are not available online
- Yes I do, they are available online (e.g. as webpages or pdf)
- Yes I do, they are available as semantic-web resource (e.g. in CropOntology,

AGROVOC or AgroPortal)

- Unsure what data standards and vocabularies are

27. If you answered yes, which vocabularies did you use? If no, why not?

Please write your answer here:

**28. Where will you store your dataset at the conclusion of your research project?
Choose ALL that apply.**

- Home computer
- External hard drive/Portable storage/Flash drive (USB)
- Storage provided by my department or school
- Storage server, provided by my institution
- Cloud-based data storage e.g. Dropbox, MS OneDrive, Google Drive
- Physical hardcopy retained (in boxes, cabinets, etc.)
- A repository nominated by the Soil CRC
- Other – Please specify:

29. Please provide us with the name(s) or URL of the data management repository(ies) that you are planning to deposit the data in.

Please write your answer here:

30. How long after the completion of your research project must the data be retained for?

- 1 year
- 5 years
- 10 years
- Indefinitely
- I do not know
- Other – Please specify:

31. How long after the completion of your research project would you prefer the data to be retained for?

- 1 year
- 5 years
- 10 years

- Indefinitely
- I have no preference

32. Who do you think is responsible for the ongoing management (preservation) of the research data resulting from your project?

- Myself
- Project leader
- Researcher
- Administrative assistant
- Technical support
- PhD candidate
- Post-doctoral researcher
- The institution you work for/conduct research for
- The Soil CRC
- Other – please specify:
- I don't know

33. Who owns the intellectual property rights associated with the data arising from your research?

Please write your answer here:

34. Have you read your Institution's research data policies about data management, data ownership, and data governance?

- Yes
- No
- I am unsure whether there are any policies

35. If you answered 'yes', to what extent do you believe you adhere to the policies?

- Fully
- Mostly
- Partially
- Not at all

36. Have you heard of the 'FAIR' data principles?

- Yes
- No

37. Did the project funding include provision for data management activities, including data sharing and publication?

- Yes
- No
- Unsure

38. Can you estimate how much funding you have/will spend on activities related to management of data (e.g. storage, supporting sharing of data, creation of metadata records) for this project?

- Yes, - Please specify the amount \$_____
- No, we are not really managing the project data
- No, I/we do not know
- No, we have no idea, but it has probably been quite a lot, *e.g. of people time*

39. If you selected Yes above, is this amount spent:

- more than expected
- less than expected
- unsure - we had no idea how much it would cost to manage the project data

40. Do you have any comments regarding the time and cost, incentives or lack of to managing project data?

Please write your answer here:

41. What role do you think the Soil CRC has in assisting you and your project/s with data management in relation to, for example, policies, guidance on data licences, technical assistance, provision of tools to help with dissemination of data, provision of a data repository for long term archiving?

Please write your answer here:

42. Are there factors relating to soil data and associated data that makes their management different to other types of environmental data?

Please write your answer here:

43. Have you received any training in research data management practices (e.g., creating metadata, file naming, file versioning, data storage)?

- Yes
- No
- Unsure

44. If you answered yes, what training have you received and who provided it?

Please write your answer here:

45. If you answered no, what would you most value getting training on?

- Data management plan
- Metadata
- Data repositories
- How to make your data FAIR (Findable, Accessible, Interoperable, Reusable)
- Data standards (including vocabularies)
- Data security
- Data licenses and sharing
- Data privacy
- Data archiving
- Data identifiers
- Other – Please specify:

46. Please provide any other comments or thoughts on the survey:

Please write your comments here:

For privacy reasons, the information in the following question will be kept separate from the survey.

Would you like to be involved in a focus group or interview for this research?

We propose to run a focus group or interview/s in September/October 2022 and in April/May 2023 to understand how successful the research has been in supporting Soil CRC projects and individuals in achieving better data management practices and processes. This will align with proposed 'use cases' that are to be defined in partnership with Program leaders post the Soil CRC conference.

If so, please provide your contact details below.

Name: _____

Email: _____

Thank you for participating in this survey. Your responses have been successfully recorded.

If you have any questions, further feedback about this project or would like to participate in future research, please contact one of the following researchers:

Dr. Nathan Robinson Email: n.robinson@federation.edu.au	Jude Channon Email: j.channon@federation.edu.au
Jennifer Corbett Email: j.corbett@federation.edu.au	

APPENDIX D: DATA MATURITY ETHICS APPLICATION

PROJECT DETAILS

Project title: (Do not use Acronyms without explanation)

A coordinated and consistent approach to managing and enabling soil data: Data maturity survey

What type of project is this? (Tick as many as apply)

- | | | |
|---|---|---|
| ☐ Funded Consultancy | ☐ Class Research Project | ☐ Postgraduate Diploma |
| ☐ Clinical Trial | ☐ Undergraduate Research | ☐ Masters by Research |
| ☐ Staff Research Project | ☐ Honours Research | ☐ PhD |
| ☒ Other – Funded Research | ☐ Masters by Coursework | |

Through which School/Section is the research to be conducted?

Centre for eResearch and Digital Innovation (CeRDI)

What is your expected **completion** date? (Approval will be granted up until this date)

30/06/2023

RESEARCHERS

Principal Researcher (must be a STAFF MEMBER)

Title & Name:	Dr Nathan Robinson
Position:	Senior Research Fellow
School/Section:	Centre for eResearch and Digital Innovation
Phone number:	03 5327 8436
Email address:	n.robinson@federation.edu.au
Academic qualifications:	Nathan has a PhD and is a Senior Research Fellow at the Centre for eResearch and Digital Innovation. Nathan has been working in the soil science and data analysis domains for over 20 years including work for the Victorian Government and Federation University Australia.
Describe what this researcher will do in the context of this project:	Dr Robinson is leading this project which involves collaboration with five organisations from Australia and New Zealand. Nathan has led the initial engagement with colleagues in

	the drafting of survey questions and will oversee the management of this research. He will contribute to all aspects of this research, including the data collection, data analysis and reporting for this project (including publications arising from this research).
Include a brief summary of relevant experience for this project:	Dr Robinson is a Senior Research Fellow and has managed numerous soil-related research projects including those focused on data and information sharing (e.g. Understanding Soils and Farming Systems - \$5.5M over 5-years for the Victorian Government). Nathan is well placed to coordinate and lead a team with extensive skills and knowledge in data sharing, governance and management. Nathan has previously participated in the development of past Soil CRC surveys including the pilot phase of the Soil CRC on soil data and indicator use by growers, advisors and researchers (project A18-007 led by A/Prof Peter Dahlhaus from CeRDI).

<u>Co-Researcher/s</u>	
Title & Name:	Jude Channon
Position:	Research Officer
School/Section:	Centre for eResearch and Digital Innovation
Phone number:	03 5327 9672
Email address:	j.channon@federation.edu.au
Academic qualifications:	Jude has a Master's in Education (RMIT) along with broad knowledge of data governance, policies and practices from an extensive career in healthcare in the UK and Victoria.
Describe what this researcher will do in the context of this project:	Jude will be involved in all aspects of this research including supporting potential participants (recruitment and survey assistance), data analysis and reporting, and preparation of research publications.
Include a brief summary of relevant experience for this project:	Jude has specific skills and knowledge developed through roles in governance, data security, and the implementation of policies and practices in healthcare. These skills will be critical in analysis of survey results and their interpretation to define future strategies and a work program pertaining to the data management and governance focus of this project.

<u>Co-Researcher/s</u>	
Title & Name:	Jennifer Corbett
Position:	Research Coordinator
School/Section:	Centre for eResearch and Digital Innovation

Phone number:	03 5327 9369
Email address:	j.corbett@federation.edu.au
Academic qualifications:	Bachelor of Management (Honours)
Describe what this researcher will do in the context of this project:	Jennifer will assist with specific aspects of this research including creation of the online survey, data collection, data management and analysis.
Include a brief summary of relevant experience for this project:	Jennifer has been employed by Federation University Australia since 2009 initially as a Research Officer and more recently as Research Coordinator. During this time, Jennifer has implemented a range of online evaluations. She also has experience in the analysis of quantitative and qualitative data and the use of Qualtrics (the proposed survey instrument for this project).

RESEARCHER TRAINING

All researchers at Federation University are required to undertake training in responsible research conduct (see *Australian Code for the Responsible Conduct of Research, R16*).

Have all of the above named Federation University researchers completed the online training module provided by the University (Research Integrity, Second Edition)? (Every researcher must complete all of the modules they have been advised to complete by the Research Integrity Office - research.integrity@federation.edu.au)	☒ Yes	☐ No
If no , please list those still to complete the training below and confirm that the relevant researchers will complete the training within 3 months of the date of approval of this application (evidence will be required)	☐ Confirm	

LAY DESCRIPTION

Provide a brief outline of the purpose of the project describing in everyday, jargon-free language the key aspects of the research (e.g., who will be participating, what information will be collected and by what means, what participants will be required to do, etc.) and the key research aims. The lay description must be in everyday, jargon-free language that is comprehensible by the average educated layperson. Define any technical terms or discipline-specific phrases and use the full form of all acronyms the first time they are used. **(300 words max.)**

This project is in response to a call from the Cooperative Centre for High Performance Soils (referred to as the Soil CRC) to improve the strategic management of soil-related research data for current and future projects. This will ensure that institutional responsibilities identified in the Australian Code for the Responsible Conduct of Research are also being

met. The need for improved data management and governance practices for projects collecting and managing soil-related research data was identified as a priority in initial scoping studies of the Soil CRC (Round 1, 2018) and investment designated in Round 3 (2019) to develop a whole of Soil CRC approach to data governance and management.

This research aims to better understand the current state of research data governance principles and management practices in use across Soil CRC projects. Survey data will be collected from research team members of Soil CRC projects. This survey will capture valuable baseline data on research management practices in soil and agricultural science. Analysis of survey data will be used to inform policies, strategies and practices towards achieving the primary objective of improving data management across the Soil CRC.

Federation University Australia's Centre for eResearch and Digital Innovation (CeRDI) is collaborating with Manaaki Whenua Landcare Research, NSW Department of Primary Industries, Charles Sturt University and the University of Tasmania on this project. Federation University is leading the data and information collection phase of this survey project that will then be used to inform and guide a program of work including policy implementation, improved data management strategies and practices that will be delivered by the broader project team.

RESEARCH AIMS & SIGNIFICANCE

State the aims, key research questions, and significance of the project. Where relevant, state the specific hypothesis to be tested. Also provide a brief description of the relevance of your proposed project to current research; (supported by a literature review and references) – refer to section 1 of the [National Statement on Ethical Conduct in Human Research](#), a justification as to why your research should proceed, including an explanation of any expected benefits to the community and its potential to contribute to existing knowledge. (600 words max.)

This project responds to the critical need for a cross-program, whole of Soil CRC approach on data management to ensure secure, reliable storage, sharing, analysis and visualisation for soil-related research data. To meet this need, researchers at Federation University are undertaking research (using an online survey) to explore the understanding and the use of data governance principles and data management strategies by participants involved with and leading Soil CRC-funded research projects.

Participants of Soil CRC projects funded under the four investment programs and the PhD program will be invited to participate in this research and to complete an online survey. The survey comprises a series of questions about the data management processes and policies that participants use within the Soil CRC projects they are involved in or leading. Analysis of the survey data will identify and inform the priorities to support projects in improving the discovery and reuse of research data while making it easier for researchers, farmer groups, growers and advisors to contribute research data (including sensor data streams) into automated and FAIR (Findable, Accessible, Interoperable, Reusable) systems.

The focus of the research is on the management and governance of research data practices in use across many Soil CRC projects. Good data management and governance can improve the visibility and reuse potential of this data for multiple purposes such as research, development and extension in soils and agriculture. Understanding an organisation or individual's level of 'data maturity' enables the prioritisation of activities and policies that lead

to improved data management and increased value of the asset. Importantly, for the data to be a valued asset it needs to be managed and curated using processes and practices that ensure data qualities are retained.

The survey questionnaire is focused on key areas including:

- The types of research data being created and collected;
- The use of data management plans and responsibilities for data management;
- Data storage and security practices;
- Data sharing and access practices;
- Data ownership and licensing practices;
- Intellectual property assignment and rights; and
- Data management practices.

The **key project objective** is to deliver a collaborative solution to data sharing and management that supports the secure, reliable storage, discovery and reuse of analysis ready data, including for use in design ready tools and systems to assist with Soil CRC project related data analytics, modelling, decision support systems and visualisation.

There are three aims of this project:

Aim 1: Develop a fit-for-purpose approach to data management for Soil CRC projects that will enable soil data to be discoverable and reusable to underpin key milestones linked to predictive decision-support tools and applications.

Aim 2: Define solutions for wider sharing and reuse of soil data including sensor data streams across Soil CRC programs to advance access to near real-time and on-farm data for data analytics and future machine learning purposes.

Aim 3: Demonstrate how sharing and reusing soil data between researchers, farmers and advisors through use cases that link research outcomes and data with automated and FAIR systems benefits contributors and users of this data.

This research project will provide new insights about the current data governance and management practices for those involved in soil and agricultural science research projects. Beneficiaries of this research include:

- Soil CRC project researchers that are looking to access relevant research data generated through Soil CRC projects for further insight.
- To increase data value to stakeholders by linking research data to decision support tools, data visualisation and grower ready applications.

Data governance and management solutions developed through this project will be shared with projects across the Soil CRC. A report detailing survey findings and research publications will be made accessible to those interested in improving their practices in data management and the provision of data for future reuse purposes.

References

- Marchildon, P., Bourdeau, S., Hadaya, P., Labissière, A. (2018). *Data governance maturity assessment tool: a design science approach*. <https://www.cairn.info/revue-projectique-2018-2-page-155.htm>
- Soares, S. (2014). *Data Governance Tools: Evaluation Criteria, Big Data Governance, and Alignment with Enterprise Data Management*. Boise, ID: MC Press.

FUNDING & FINANCIAL BENEFITS

Researchers should include any source of funding (e.g., departmental, commercial, non-commercial, governmental) The HREC will consider whether there is a conflict of interest.

Are any of the researchers affiliated with or in receipt of any financial benefit from any of the external organisations involved in your research?	☐ Yes	☒ No
---	------------------------------	--

If **yes**, explain how, how much and for what purpose:

--

Has this protocol received research funding or is this submission being made as part of an application for research funding?	☒ Yes	☐ No
What is the status of the funding application?	☒ Approved	☐ Refused ☐ Pending

If **yes**, what is the source of the funding?

Cooperative Centre for High Performance Soils

What is the project grant title and proposed grant duration?

A coordinated and consistent approach to managing and enabling soil data. Commencement date: 1/11/2021, Completion date: 30/6/2023

What is the registration number of the grant/funding application?

G1783; Soil CRC Project ID 2.2.005

What is the deadline for the granting body?

Project funding has been confirmed.

Does this project require HREC approval before consideration for funding?	☐ Yes	☒ No
---	------------------------------	--

How will participants be informed of the source of the funding?

Participants will be informed of the source of the funding through the Plain Language Information Statement and in the participant recruitment invitation (email)

The copyright and other IP in the data is owned by:

Federation University Australia	☐	In general, Federation University Australia owns intellectual property rights, including copyright, in research data originated by all staff including academic staff, except where the data created by academic staff falls within the definition of a 'scholarly work', Federation will also own copyright in data generated by students as part of a 'collaborative research activity'.
The Researcher	☐	
Federation University Australia (joint ownership) Research conducted by Federation University Australia in collaboration: copyright and IP ownership are documented in an agreement between the organisations	☒	Soil CRC (a) Each Project Participant will make available to the Project its Pre-existing Material as specified in Item 4 of Schedule 1, subject to any conditions or restrictions also specified in Item 4 of Schedule, and otherwise in accordance with MPA Clause 14.

	(b) Upon its creation, all Project IP will be owned by CRC HPS, as further detailed in MPA Clauses 21.1, unless otherwise specified in Item 5 of Schedule 1. (c) The Parties' rights to use Project IP, including any Use Field and Use Field Conditions, are detailed in Item 8 of Schedule 1 and will otherwise be in accordance with MPA Clause 22.
--	--

MULTI CENTRE RESEARCH

Other HREC Approvals

The principal researcher is responsible for informing each HREC of all other sites at which the research is being proposed or conducted; disclosing to each HREC any previous decisions regarding the research made by another HREC; and informing each HREC of whether the protocol is presently before another HREC.

Is this protocol being submitted or has it been previously submitted to another Human Research Ethics Committee?	☐ Yes	☒ No
--	------------------------------	--

If **yes**, give details of other centres involved; the approval status of the study at each centre; and details of any required amendments.

EXTERNAL APPROVALS

If your research involves participants from other organisations (e.g. educational institutions, companies, agencies, collectives), you may need to obtain authorised approval before approaching participants, eg: *Department of Education and Training, School Principals, School Councils* (for research involving Government schools); *Catholic Education Office* (Catholic schools); *School Boards* (Independent schools); *Senior Officers* (Commercial or Government entities); *Elders* (Aboriginal communities); or *Representative bodies* (Collectives). Copies of approval letters must be attached to this application or, if pending at the time of submission, forwarded to HREC when available. Some authorities may decline to provide permission letters until ethics approval has been granted. In such cases, you should submit your application to the HREC for provisional approval pending receipt of the documentation.

Does research involve or impact on participants from external agencies or organisations?	☒ Yes	☐ No
If yes , has required permission been obtained from relevant agencies?	☐ Yes	☒ No

If **yes**, please specify from whom and attach a copy

If **no**, specify from whom, and advise when this will be obtained

Approval in writing will be sought from the Soil CRC and that approval will be forwarded to HREC prior to data collection.

RESEARCH METHODOLOGY

Provide an outline of each step of the proposed methodology, including details of data collection techniques, tasks participants will be asked to complete, the estimated time commitment involved, and how data will be analysed. If the project includes any procedure that is already established and uses accepted techniques, please include a description of the procedure. **(500 words max)**.

Soil CRC project team members from over 75 projects will be invited to participate in this research and to complete an online survey.

Plain Language Information Statement (PLIS): The PLIS outlines the purpose of the study, the researchers involved with it, what participation in the study will involve, and the ethical procedures under which the study is being conducted. The PLIS emphasises that participation in this research is voluntary and specifies how the research data will be used and stored.

Survey: The survey comprises 46 questions that seek insights about key aspects of data governance and management in alignment with principles and responsibilities of the Australian Code for the Responsible Conduct of Research. The survey will take approximately 15-20 minutes to complete.

The online survey will be shared through approved channels (email invitation or phone call from Program leaders or research team members) once HREC approval has been granted. The survey design contains a range of questions including multiple choice, yes/no and short answer responses. The survey will be delivered online using Qualtrics software. Participants will receive an email inviting them to take part in the survey. The survey will be available for completion (once HREC approved) until 2 September 2022. This end date will enable promotion of the survey at the Soil CRC conference (August 23-25).

At the Soil CRC conference, hard copies of the survey will be provided for those participants who would prefer to complete the survey using this medium. For those participants that would like assistance with the survey, we also offer the opportunity to complete the survey over the phone by contacting a research team member at a time that is convenient to both parties (details provided in all documents), or at the Soil CRC conference where iPads are available as a tool to participate in the survey.

The survey includes a final question asking the participant if they would like to be involved in a focus group or interview for this research. We propose to submit an amendment to HREC to enable a focus group or interview/s to proceed in September/October 2022. The primary focus will be to understand how successful the research has been in supporting Soil CRC projects and individuals in achieving better data management practices and processes. This will align with proposed 'use cases' that are to be defined in partnership with Program leaders post the conference.

A list of contact names and email addresses, supplied by the Soil CRC, will be used to invite participants to respond to this research.

Is it likely / possible that any of the data or information collected may be used by you, or others, for any research other than that outlined in this application? See NS Chapter 2.2 and Chapter 3.2 when preparing your response.

☐ Yes	☒ No
------------------------------	--

If YES, describe below and ensure this is outlined in all Plain Language Information Sheets and Consent forms.

- Participants should be fully informed of the possibility of any future use of data and information collected and their 'extended' or 'unspecified' consent gained. Failure to do this may restrict the future use of the data and information.
- Any restrictions on the use of participants' data and information should be recorded and the record kept with the collected data and information. Restrictions must be accessible to researchers who want to access those data and information for research.

National Statement:

2.2.15 – Extended or unspecified consent may sometimes need to include permission to enter the original data or tissue into a databank or tissuebank (see paragraph 3.2.9).

2.2.16 - When unspecified consent is sought, its terms and wide-ranging implications should be clearly explained to potential participants. When such consent is given, its terms should be clearly recorded.

If **YES**, describe below and ensure this is outlined in all Plain Language Information Statements (PLIS) and Consent Forms.

--

RECRUITMENT OF PARTICIPANTS

Participant Details

Describe in detail your proposed recruitment strategy to source target participants. Provide the number and age range, giving a justification of your proposed sample size. Include details of statistical power of the sample where appropriate. To ensure the requirements for consent are met, refer to the [National Statement Chapter 2.2 General requirements for consent](#)

The recruitment strategy is focused on Soil CRC project participants from over 75 approved projects. All participants are aged 18 years and over.

A recruitment email has been created (copy attached) to enable program leaders and research team members to share with project participants to raise awareness of the survey and invite participation. It is anticipated that there will be some self-referred participants through this process. The Soil CRC will provide the research team with the names and email addresses of potential participants. The research team in partnership

with program leaders will contact these individuals (via email and/or telephone) to introduce the research, provide an overview of the research (including a copy of the Plain Language Information Statement) and to invite participation.

The Soil CRC will only assist the researchers to identify potential research participants by providing the researchers with relevant contact names and email addresses. Their involvement in the project will be limited to assisting research team members with these contact details. Participation in the research is voluntary and those contacted about this research will not be compelled to participate. A statement to this effect is included in the PLIS.

Prior to responding to the survey, participants will be provided with information about the research, including the recruitment email and PLIS. The PLIS specifies that participation in the research is voluntary, and involvement is independent of the Soil CRC. It also specifies that choosing to participate in the study (or not) will not be made known to the Soil CRC, nor will survey data be shared with the Soil CRC.

Participants will be required to complete three introductory questions to enable a survey submission, with all remaining questions optional for those that participate.

As this research has a limited sample size (Soil CRC project participants, estimated to be ≤ 200) of a larger community working in soil and agricultural science, we are aiming for > 40 participants (a response rate of 20%). Ideally survey participants will represent a diversity of organisations, project roles and aligned Soil CRC programs, however, we have no control on survey participation and will monitor survey submissions during the survey period commencing late July (subject to HREC approval) until 2 September 2022).

A secondary promotion of the survey will coincide with the Soil CRC conference (August 23-25). The aim here is to promote the survey and aid potential participants should they require any assistance to undertake the survey. A presentation on the project will be provided on 25 August 2022 to those in attendance and hardcopy versions of the survey will be made available to those that would prefer this for survey participation. For those participants that would like assistance with the survey, we will offer the opportunity to complete the survey over the phone by contacting a research team member at a time that is convenient to both parties (details provided in all documents).

Target participants

Who are the target participants? (Tick as many as applicable)

☐ Students or staff of this University
☒ Adults (<i>over the age of 18 years and competent to give consent</i>)
☐ Children/legal minors (under the age of 18 years, with parental consent)*
☐ Children in out of home care (under the age of 18 years – Note: Contact DHS for guidance in this area)
☐ Care Leavers – An adult who spent time in care as a child (under the age of 18) this could have been foster care, residential care, or other arrangements outside the immediate or extended family
☐ Women who are pregnant, Human Foetuses and Neonates

☐ Aboriginal and/or Torres Strait Islander communities
☐ Other collectives where leader/council of Elders may need to give consent
☐ Elderly individuals
☐ Individuals from non–English speaking backgrounds
☐ People in other countries
☐ Pensioners or welfare recipients
☐ Intellectually or mentally impaired individuals unable/with compromised capacity to provide consent
☐ Individuals highly dependent on medical care with a compromised capacity to give consent
☐ Physically disabled individuals
☐ Patients or clients of professionals
☐ Prisoners, parolees
☐ People who may be involved in illegal activities

*Parental/Guardian consent may not be required in some instances - [National Statement Chapter 4.2.8 & 4.2.9](#)

**** Reference:** *Ethical conduct in research with Aboriginal and Torres Strait Islander Peoples and communities: guidelines for researchers and stakeholders and Keeping Research on track II and Guidelines for Ethical Research in Australian Indigenous Studies (GERAIS)*

Proposed Recruitment Method

A copy of all recruitment materials used (e.g., printed advertisements, radio and television advertisement transcripts, posters, letters of invitation) must be attached to this application for review by the committee.

What is the proposed recruitment method? (Tick all that apply)
☐ Mail-out
☒ Email Have you attached a copy of the text of the email that will be sent? Yes ☒ No ☐ If no , please explain:
☒ Telephone
☐ Contact details obtained from public documents (e.g., <i>phone book</i>)
☒ Recruitment by researcher(s)
☐ Participants from a previous study
☐ Snowball (<i>participants suggest other potential participants</i>)

☐ Personal contacts – Provide details:
☒ Other – please explain: At the Soil CRC conference (23-25 August), there will be an opportunity to promote the survey to those attending the conference. A presentation to those in attendance at the conference (25 August) will include details on how to access the survey online. This will include information about the research goals and objectives. This information may be shared through Soil CRC program leaders (recruitment by researchers) as well.
☐ Advertisement (e.g. for a noticeboard or FedNews) Have you attached a copy of the advertisement? Yes ☐ No ☐ If no, please explain:
☐ Facebook Where will you be posting the advertisement on Facebook? _____ Do you have permission to post Facebook advertisement? Yes ☐ No ☐ N/A ☐ Have you attached a copy of the advertisement that will be posted on Facebook? Yes ☐ No ☐ If no , please explain: _____ ☐ Other Social Media platform Name platform: _____ Have you attached a copy of the advertisement that will be posted? Yes ☐ No ☐ If no , please explain:
Target participants Include a geographical target Domestic regions ☒ Tick for all Australia or specify regions: _____ Globally ☐ Tick for all areas, or specify regions: New Zealand _____
☒ Recruitment by a third party (e.g., employer, doctor) Have you attached a copy of the letter requesting their assistance, and/or the letter confirming their willingness to assist? Yes ☐ No ☒ If no , please explain: All Soil CRC Program leaders have been engaged in planning of project directions and will be involved in the distribution of emails to potential participants. A detailed workplan including project scheduling will be shared with Program leaders via email along with a scheduled Microsoft Teams meeting (mid-July) to explain the proposed recruitment process and how the research team can support these leaders in sharing the invitation with potential participants. Research team members will support Program leaders to share the email and survey with participants as directed.
☐ Private sources

Have you attached a copy of the relevant approval letter? Yes ☐ No ☐
If **no**, please explain:

BURDENS OF RESEARCH (RISK & RISK MANAGEMENT)

This section raises the issue of your duty of care toward research participants. To what risks are participants subjected? What will you do should an emergency occur, or should a participant become upset or distressed? What is your risk management strategy? Refer [National Statement Section 2.1 Risk and Benefit](#)

Likely Benefits

Are participants likely to gain direct or indirect benefit from the research?	Yes ☒	No ☐
---	---	-----------------------------

If yes, provide details

The findings of the survey will be used to prioritise support activities and technical developments that enable the Soil CRC to support the data management practices for the research they fund. This will lead to improved outcomes for the Soil CRC and those participants requiring assistance in data management.

How will potential benefits to participants or community outweigh the risks?

There is no risk involved as all survey responses are confidential thereby ensuring that beneficiaries reflect overall priorities identified from survey results.

Research Activities

Which of the following activities will the research involve? (Tick as many as apply)

☒ Use of a questionnaire (<i>attach copy</i>)
☐ Interviews (<i>attach interview questions</i>)
☐ Observation of participants without their knowledge
☐ Participant observation
☐ Audio- or video-taping of interviewees or events
☐ Access to personal and/or confidential data (<i>including student, patient or client data</i>) without participants' specific consent
☐ Administration of any stimuli, tasks, investigations or procedures which may be experienced by participants as physically or mentally painful, stressful or unpleasant during or after the research process
☐ Performance of any acts which may diminish the self-esteem of participants or cause them to experience embarrassment, regret or depression
☐ Investigation of participants involved in illegal activities
☐ Procedures that involve deception of participants

☐ Administration of any substance or agent
☐ Use of non-treatment of placebo control conditions
☐ Collection of body tissues or fluid samples
☐ Collection and/or testing of DNA samples
☐ Participation in a clinical trial
☐ CTN Trial ☐ CTX Trial Please <i>provide Phase number, i.e., either 1, 2, 3 or 4</i>
☐ Testing a medical/diagnostic device

RISK MANAGEMENT PROCEDURES

Identify as far as possible all potential risks to participants (e.g. physical, psychological, social, legal, economic) associated with the proposed research. Explain what risk management procedures will be put in place. Any potential risks should be outlined in the Plain Language Information Statement (PLIS) along with contact details of an appropriately qualified and relevant organisation for participant reference in the event that participants feel distress, eg: Lifeline

Research participants will not be at risk of any emotional or physical harm greater than or in addition to the risks encountered in their normal lives. However, if for some reason participants find the questions asked of them in the survey are emotionally distressing, participants will be directed to contact Lifeline and other support services. This information is outlined in the Plain Language Information Statement.

When/if any off-site visits occur in relation to this research (e.g. Soil CRC conference), the following protocols will be adhered to by the researcher/s. These steps will minimise any risk to researchers when undertaking research off-campus:

- The researcher/s undertaking data collection off campus will provide a schedule of activities and timelines to one of the other members of the research team.
- The researcher/s will carry a mobile phone with them and will have access to the principal researcher, as the emergency contact, at all times during off-site visits.

To ensure researcher/s safety when conducting off-site visits, the researcher will wear appropriate university name tags.

Where will the research be conducted? (*Tick as many as apply*)

☒ Federation University	☒ Other location(s)
---	---

If **other**, please *give details (including the URL for web-based studies)*:

Soil CRC Conference (<https://soilcrc.com.au/events/>).

Are facilities at the research location appropriate for the scientific needs of the research?

☒ Yes

☐ No

If **no**, please elaborate

--

Are the facilities appropriate to meet any physical, emotional or other needs of participants that result from their participation?	☒ Yes	☐ No
---	---	-----------------------------

If **no**, please elaborate

--

Are there any specific risks to researchers?	☐ Yes	☒ No
--	------------------------------	--

If **yes**, please describe the risks identified, and your planned Risk Management protocol for researchers

--

What plans are in place to deal with adverse/unexpected outcomes?

It is unlikely that any adverse or unexpected outcomes will occur during data collection for this research, however, should an incident occur the information outlined in the Plain Language Information Statement will be adhered and the Principal Researcher will be consulted immediately.

Will parts of this project be carried out by independent contractors?	Yes ☐	No ☒
---	------------------------------	--

If yes, please confirm that the independent contractor will receive from the first-named Principal Researcher, a copy of the approved ethics protocol and be made aware of their responsibilities arising from it.

--

If necessary, has the Principal Researcher ensured that the other researchers have undergone a police check and a Working With Children check?	Yes ☐	No ☐	N/A ☒
--	------------------------------	-----------------------------	---

How will the conduct of the project be monitored to ensure that it conforms to the procedures set out in this application, the [University's current human ethics guidelines](#), the [National Statement on Ethical Conduct in Human Research 2007 \(Updated 2018\)](#), and [The Australian Code for the responsible Conduct of Research](#) (In the case of student projects please give details of how the supervisor/s will monitor the conduct of the project, e.g., how often student and supervisor will meet; how meetings will be conducted: email/phone/in person; how efforts will be coordinated if a number of researchers are involved.)

This project is being undertaken by researchers who understand and conform to the Human Research Ethics guidelines. Should any issues of concern arise during this research, the Principal Researcher will be notified, and issues will be addressed as a matter of priority. Regular project meetings will be scheduled, and the research team will also be in regular contact with each other, ensuring consistent and frequent monitoring of the project.

Will there be support provided for participants? (*You may need to consider having additional support for participants during or after the study, depending on risks to participants. Consider whether your project would require additional support and what support would be available.*)

It is not envisaged that participants in the survey will require any extra support, however, if for some reason participants find the surveys emotionally distressing, they will be directed

to seek assistance through Lifeline or similar services. This information is outlined in the Plain Language Information Statement.

Deception

If this research project includes deception what debriefing will participants receive following the study and when? (*Attach a copy of any written material or statement to be used in such a debriefing. Participants may need to talk with the researchers about the experience of being involved in the study as well as learn more about the aims of the research.*)

There is no deception associated with this research. All information about the project is presented to participants in the Plain Language Information Statement.

INCENTIVES FOR PARTICIPATION

Note that while participants may, in certain circumstances, be paid or reimbursed for their inconvenience and time, the payment should not be of an amount that risks inducement to participate, thus potentially biasing the project results. If rewards are to be used, all participants are to receive the reward.

Are financial or other rewards proposed to be given to participants?	☐ Yes	☒ No
--	------------------------------	--

If **yes**, describe how much and in what form the payment/incentive will take (e.g., money to reimburse travel costs, vouchers for movie tickets, chocolate frogs).

CONSENT

Dependent or Unequal Relationships

The consent of a person to participate in research must not be subject to any coercion. Research involving those in dependent or unequal relationships (e.g., teacher/student, manager/employee, parent/child, doctor/patient) may compromise a participant's ability to give consent that is free from any form of pressure (real or implied) arising from this unequal power relationship. The HREC therefore recommends that, where possible, researchers should choose participant cohorts where no dependent relationship exists. However, if the researcher believes that research involving people in dependent relationships is purposeful and methodologically defensible, the HREC will require additional information explaining why this is so and how any risks inherent in the dependent and unequal relationship will be managed. The HREC will also need evidence to show that participants have been reassured that refusal to participate will not result in any discrimination or penalty. Applicants should note that reasons of convenience will not normally be considered adequate justification for conducting research in situations where dependent relationships exist.

* [National Statement on Ethical Conduct in Human Research 2007 \(Updated 2018\) Chapter 4.3](#) for information on unequal relationships before answering the following question.

Does a dependent or unequal relationship exist between any participant and researcher, particularly those involved in recruiting?	Yes ☐	No ☒
---	------------------------------	--

If yes, please explain the relationship and the steps to be taken by the researchers to ensure that the participant's participation is purely voluntary and not influenced by the relationship in any way.

Informing Participants – Plain Language Information Statement (PLIS)

The potential participant must be provided with information *at their level of comprehension* about the purpose, methods, demands, risks, inconveniences, discomforts, and possible outcomes of the research (*including the likelihood and form of publication of research results, and whether their data may be made available for future research projects*) so their consent is fully informed. Download the current template for the [PLIS](#) from the HREC website.

Have you attached a copy of the PLIS for participants?	☒ Yes	☐ No
--	---	-----------------------------

If **no**, please explain

Does the PLIS comply with the following guidelines?	YES	N/A
It is presented on the Federation HREC approved template, downloaded from the website	☒	*
It has clear identification of the University, the School(s) involved, the project title, the Principal and Other Researchers (including Federation contact details).	☒	*
It details what involvement in the project will require (e.g., involvement in interviews, completion of questionnaire, audio/video-taping of events), estimated time commitment, any risks involved.	☒	*
It advises how participants' contact details were obtained and/or how potential participants were selected	☒	☐
If staff or students of the Federation University Australia are to be involved as participants, it advises that the project has received clearance by the HREC	☒	☐
It advises that if the sample size is small this may have implications for privacy/anonymity.	☐	☒
It states clearly that if participants are in a dependent relationship with any of the researchers involvement in the project will not affect ongoing assessment, grades, employment, management or treatment of health (as relevant).	☐	☒
It states clearly that involvement in the project is voluntary and that participants are free to withdraw their consent to participate at any time, and to withdraw any unprocessed data previously supplied.	☒	*
It states that arrangements will be made to protect confidentiality of data, including that confidentiality of information provided is subject to legal limitations (e.g., subpoena, freedom of information claim, or mandatory reporting in some professions).	☒	☐

It advises whether or not data will be destroyed after a minimum period.	☒	☐
It advises the de-identified data collected may be used in future research projects	☒	☐
It provides any other relevant information.	☒	*

* Required

Obtaining and Documenting Consent

How will informed consent be obtained/recorded?
☐ Signed consent form
☐ Recorded verbal consent
☒ Implied by return of survey <i>*NB If consent is to be implied by return of survey, all information that would normally be presented on the consent form must be included in the PLIS</i>
☒ Other (Please specify): For those who would prefer a hardcopy of the survey.

The correct template for the consent form can be found at:

<https://federation.edu.au/research/internal/support-for-students-and-staff/ethics/human-ethics/human-ethics3>

Is a copy of the consent form attached to this application form?	☐ Yes	☒ No
--	------------------------------	--

If **no**, please explain how consent will be documented:

Consent details and information are detailed in the PLIS.

Does the consent form comply with the following guidelines?
☒ It is presented on the Federation HREC approved template, downloaded from the website
☒ It states the title of the project and names of the researchers
☒ It confirms that the project is for research
☒ It confirms that involvement in the project is voluntary and that participants are free to withdraw at any time or withdraw any unprocessed data previously supplied
☒ It details specific requirements of participants (e.g., interviews will be audio-/video-taped)
☒ It advises of any legal limitations to data confidentiality
☒ It advises that if the sample size is small this may have implications for privacy/anonymity
☒ It provides any other information relevant to obtaining participant consent

DISCONTINUING PARTICIPATION

Are participants advised as part of the informed consent process that they have the right to withdraw at any time or withdraw any unprocessed data previously supplied?

☒ Yes

☐ No

If **yes**, please detail how participants are informed of this right.

Participants are advised in the Plain Language Information Statement that they have the right to withdraw at any time from this research, or withdraw any unprocessed data previously supplied.

If **no**, please explain why this advice has not been given

DATA AND INFORMATION MANAGEMENT PLAN

Confidentiality

Please give attention to implications for compliance with legislative requirements including, for example [Guidelines Approved under Section 95A of the Privacy Act 1988](#), produced by NHMRC, and [Statutory Guidelines on Research Issued for the Purposes of Health Privacy Principles](#) produced by the Office of the Health Services Commissioner.

What are Data? [National Statement on Ethical Conduct in Human Research 2007 \(Updated 2018\) Chapter 3.1](#) Data are pieces of information, eg:

- What people say in interviews, focus groups, questionnaires, personal histories and biographies;
- Analysis of existing information (clinical, social, observational or other);
- Information derived from human tissue such as blood, bone, muscle and urine.

(Note: Where the sample size is very small, or information is obtained through a focus group, it may be impossible to guarantee anonymity or confidentiality of participants' identity, and participants involved in such projects need to be advised of this limitation.)

Tick which method will be used to guarantee confidentiality/anonymity?

- | | |
|-------------------------------------|---|
| ☐ | Individually identifiable data , where the identity of a specific individual can reasonably be ascertained. Examples of identifiers include the individual's name, image, and date of birth or address. |
| ☐ | Re-identifiable data , from which identifiers have been removed and replaced by a code, but it remains possible to re-identify a specific individual by, for example, using the code or linking different data sets. |
| ☒ | Non-identifiable (anonymous) data , which have never been labelled with individual identifiers or from which identifiers have been permanently removed, and by means of which no specific individual can be identified. A subset of non-identifiable data are those that can be linked with other data so it can be known that they are about the same data subject, although the person's identity remains unknown. |

Tick all that apply from the boxes below:

☐	Participants will have the option of being identified in publications arising from the research.
☒	Participants will be referred to by pseudonym in publications arising from the research.
☐	Personal information will be obtained from a Commonwealth department or agency? (If yes , you may need to comply with the requirements of the Privacy Act 1988).
☒	Any other method of protecting the privacy of participants (e.g., use of direct quotes with specific, written permission only; use of real name with specific, written permission only). Please describe: Participant feedback to open ended questions in the survey may be used for reporting purposes. This information has been provided to participants in the PLIS as well.

Security and Storage

Does the Principal Researcher accept responsibility for the security of the data collected?	☒ Yes
---	---

Who will have access to data and information during the project?	
☒ Access by named researchers only	☐ Access by other(s) than named researcher(s)

If others have access to data and information, identify who, at which storage site, for what purpose, and their connection to the project.

--

Which of the following methods will be used to ensure data and information security?	
☒ Physical, eg. Hard copy (please describe)	Where participants don't have online access to the survey, hard copies can be provided (as noted in the survey and PLIS). These survey responses will be secured in locked cabinets.
☒ Network, eg. Internal University drive, ECM, Sharepoint (please describe)	All collected and analysed data pertaining to the research will be stored on a password protected file accessible only by the named researchers.
☐ System or other technological security measures eg. Cloud, Figshare etc. (please describe)	
☐ Other (please describe)	

--

Does data storage comply with the National Statement on Ethical Conduct in Human Research 2007 (Updated 2018) .	☒ Yes	☐ No
---	---	-----------------------------

If **no**, please explain

--

If data and information is to be kept elsewhere during fieldwork, please explain how and where data will be held, including arrangements for data and information security

Data storage for this project will be overseen by the Principal Researcher and the other researchers named on this application. They will ensure appropriate storage of all data and information in a locked filing cabinet, in a locked room located at Federation University. Data collected for this research will be stored on a password protected Federation University computer/laptop or on a password protected web-based survey database tool (Qualtrics).
--

Post Project

Will the data or information potentially be disclosed or shared and, if so with whom?

Aggregated and deidentified data will be included in the report of survey findings and may be used in a research journal publication. The report will be shared with the Soil CRC community (pending Soil CRC publication approval). There may be specific findings relevant to programs of the Soil CRC (e.g. PhD program) that will be used to support these programs in improved data management practices. There is also the potential for aggregated findings to be shared publicly in the production of research papers and conference presentations.

How will the data or information be shared?

Aggregated and de-identified data from this research will be included in the final report and in any publications prepared. Data may also be used in the preparation of communications and scholarly outputs such as journal publications and conference papers. Source data will be retained and secured with only listed researchers having access to this.

Please confirm that any data collected will be kept for a <u>minimum</u> of 5 years from date of research publication.	☒ Yes
---	---

Will the data be destroyed at some point after being kept for the minimum 5 year period? (Data may be kept indefinitely, but must be appropriately secured)	☒ Yes	☐ No
--	---	-----------------------------

If **yes**, how and when will data be disposed of?

The physical data will be disposed of by the Principal Researcher through approved processes of the Data Disposal Officer at Federation University Australia. The physical data will be disposed after five years. However, data may be retained for longer, as needed.

Please confirm that any data collected will be disposed of by the Principal Researcher, who is responsible for the data.

☒ Yes

N/A

DISSEMINATION OF RESULTS

Explain when, how, where and to whom results will be disseminated, including whether participants will be provided with information on the project's findings or outcomes.

A report consolidating the research findings from the survey, using de-identified data, will be prepared and provided to the Soil CRC community (pending Soil CRC publication approval). This report is anticipated in late September – early October 2022 to enable planning of activities for the remainder of the project. Journal publications and conference presentations are additional, anticipated outcomes for this research that are likely in 2023.

How will results be made available to **participants**? (Tick as many as apply)

☒ Written summary of results

☐ Copy of final manuscript (*thesis, article, etc.*)

☒ Verbal presentation (*info session, debriefing, etc.*)

☐ Presented to all participants

☒ Presented if requested

☒ Presented to representative participants (*e.g. CEO, school principal*)

☐ Other - *Please explain:*

☐ None - *Please explain:*

How will results be made available to **peers and colleagues**: Tick as many as apply

☒ Conference papers

☒ Journal article(s)

☐ Thesis

☐ Book

☒ Other - *Please explain:* In public communications (as identified above); project report

☐ None - *Please explain*

LEGAL ISSUES

Does the project involve subject matter or conduct that may give rise to legal vulnerability of participants or researchers?

☐
Yes

☒ No

If **yes**, please give details

--

Are adequate precautions to be taken?	☐ Yes	☐ No	☐ N/A
---------------------------------------	------------------------------	-----------------------------	------------------------------

If **yes**, please give details

--

Confidentiality of information provided can only be protected within the limitations of the law. Depending on the research proposal, you may need to state these limitations specifically (<i>subpoena, freedom of information claim, mandated reporting by some professions, etc.</i>) <i>Have you included appropriate information on the legal limitations of protecting confidentiality in the PLIS and consent form?</i>	☒ Yes	☐ No	☐ N/A
---	---	-----------------------------	------------------------------

If **no**, please advise how participants will be advised

--

CHECKLIST OF ATTACHMENTS

Please check that the following documents are attached to your application. Applicants should note that where questionnaire or interview questions are submitted in draft form, a copy of the final documentation must be submitted for final approval when available.

Are the following documents attached?	Yes	No	N/A
Recruitment advertisement (e.g. for noticeboard or FedNews)	☐	☐	☒
Plain Language Information Statement	☒	☐	*
Consent form	☐	☐	☒
Evidence of external approvals related to the research	☐ ☐ Pending	☐	☒
Questionnaire	☒ ☐	☐	☐
Interview Schedule	☐ ☐	☐	☒
Debriefing material	☐	☐	☒
Other: A copy of the email that will be sent to potential participants	☒	☐	☐

* Required

DECLARATIONS

Researcher Declarations:

The information contained herein is, to the best of my knowledge and belief, accurate. I have read the [University's current human ethics guidelines](#), and accept responsibility for the conduct of the procedures set out in the attached application in accordance with the guidelines, the Australian Government's [National Statement on Ethical Conduct in Human Research 2007 \(Updated 2018\)](#), [The Australian Code for the responsible Conduct of Research](#) and any other condition laid down by the Federation University's Human Research Ethics Committee or its sub-committees. I have attempted to identify all risks related to the research that may arise in conducting this research and acknowledge my obligations and the rights of the participants. I and my co-researchers and supporting staff have the appropriate qualifications, experience and facilities to conduct the research set out in the attached application and to deal with any emergencies and contingencies related to the research that may arise. supporting staff have the appropriate qualifications, experience and facilities to conduct the research set out in the attached application and to deal with any emergencies and contingencies related to the research that may arise.

.....
Principal Researcher

Dr NATHAN ROBINSON

.....
Date: 24/06/2022

.....
Other Researcher

JUDE CHANNON

Date: 27/06/2022

.....
Other Researcher

JENNIFER CORBETT

Date: 27/06/2022

Declaration by authorised signatory:

Deputy Dean, Dean, Associate Dean (Research) of your School or School/Discipline Ethics Coordinator (approved appointment)

I have reviewed this project and consider the methodological/technical aspects of the proposal to be appropriate to the tasks proposed and recommend its approval.
I consider the Researcher(s) to have the necessary qualifications, experience and facilities to conduct the research proposed and to deal with any emergencies and contingencies that may arise.

Comments/Provisos

Signature:

Signatory Name: Associate Professor Helen Thompson

☐ Associate Dean (Research), or

☐ Dean, or

☒ Centre Director, or

☐ School/Discipline Ethics Coordinator
(approved appointment)

Date 30/06/2022

APPENDIX E: DATA MATURITY SURVEY PLIS

Centre for eResearch and Digital Innovation

PROJECT TITLE:	A coordinated and consistent approach to managing and enabling soil data: Data maturity survey
PRINCIPAL RESEARCHER:	Dr Nathan Robinson
OTHER RESEARCHERS:	Jude Channon Jennifer Corbett

What is this research about?

Researchers at Federation University Australia (Dr Nathan Robinson, Ms Jude Channon and Ms Jennifer Corbett) are conducting research to better understand the current state of research data governance principles and management practices in use across Soil CRC projects. This research is in response to a research priority call from the Soil CRC to improve the strategic management of soil-related research data for current and future projects. The need for improved data management and governance practices for projects collecting and managing soil-related research data is focused on achieving a whole of Soil CRC approach to data governance and management (i.e., data maturity).

Data management and governance can improve the visibility and reuse potential of soil-related research data for multiple purposes such as research, development and extension that can support farmers in their pursuit of optimised soil productivity.

The aim of this research is to explore current research data management practices across Soil CRC projects. The Soil CRC has funded this research. You are invited to participate in this research project being conducted by researchers at Federation University (the lead organisation), together with Manaaki Whenua Landcare Research, NSW Department of Primary Industries, Charles Sturt University and the University of Tasmania.

Your contact details have been supplied to us by the Soil CRC.

This research involves the collection of data using an online survey from participants on research projects funded by the Soil CRC. The research findings will provide important insights to assist the research team to understand the needs of current and future Soil CRC projects in terms of data governance, data management strategies and data management tools. This research will inform future data governance work in the Soil CRC to enable better access and use of the soil data that is produced by other projects funded by the Soil CRC.

What is involved in this research?

If you agree to participate in the research, you will complete an online survey that takes approximately 15-20 minutes to finish. The survey will include questions about data management, security, publishing and access to your research data.

Your survey responses will remain anonymous and confidential. All data collected through the survey will be aggregated and survey data will be non-identifiable. Survey data will not include any information that can be used to identify individual participants.

How will my information be used?

All information collected as part of this research will be used to identify priorities for enabling better research data management. No individuals will be identified by researchers for specific quotes, comments or responses in the reporting of research findings. To ensure your anonymity and confidentiality, all potentially identifying information will be removed from these.

A report that synthesises key findings from the consolidated data collected through this research will be prepared by the researchers for the Soil CRC. This report may be distributed publicly. Some research findings may also be developed as publicly released communications, and for scholarly purposes, including journal publications and conference presentations in alignment with Soil CRC publishing guidelines.

Please be aware that in participating in this research, your de-identified data may be used to inform future research. Only consolidated research findings will be compiled in the report for distribution to the Soil CRC and/or publicly.

Participation

Participation in this research is voluntary. You may choose to respond to all or only some of the interview questions. Please note that Questions 1 to 3 (your affiliation group; role in research project and Soil CRC Program that your project/s aligns to) are compulsory to enable submission of the survey. Please just answer the questions you are comfortable with.

You are also free to withdraw from the research, without explanation, at any time. There are no consequences for withdrawing from this research, or not taking part in it. If you withdraw before, during or immediately after completing the survey, the information you have provided can be deleted. However, if you withdraw from participation at a later stage, after your information has been combined with that of the other participants, we may no longer be able to identify your information and it may be unable to be deleted.

Please note that participation (or non-participation) in this research, will not be shared with the Soil CRC. Details about your involvement (or not) in the study will not be supplied to the Soil CRC.

How is data stored?

All named researchers for this study have extensive experience in conducting research and abiding by ethical research principles. Data confidentiality will be maintained. Published results will not include information that identifies any participants. All data will be stored on secure password protected systems and retained for a minimum of five years.

Further information

It is unlikely that you will experience any distress by participating in this research. However, should you have concerns arising from the interview we advise that you access professional support through your workplace or contact a medical health practitioner. You may also access information via Beyond Blue (1300 224 636) and Lifeline (131 114),

providing expert advice, support and information to deal with personal challenges and stress.

If you have any questions, or you would like further information regarding the research project please contact the Principal Researcher, Dr Nathan Robinson: **EMAIL:**

n.robinson@federation.edu.au

PH: (03) 5327 8436.

Ethical approval

This project has been approved by the Federation University Australia's Human Research Ethics Committee (Project number: 2022-143), in accordance with the guidelines provided by the National Health and Medical Research Council for research involving humans.

Should you have any concerns about the ethical conduct of this research, please contact:

Federation University Coordinator Research Ethics, Research Services, Federation University Australia,

P O Box 663 Mt Helen Vic 3353;

Telephone: (03) 5327 9765;

Email: research.ethics@federation.edu.au

Research team members

Dr Nathan Robinson

Senior Research Fellow, Centre for eResearch and Digital Innovation, Federation University

E: n.robinson@federation.edu.au

T: (03) 5327 8436

Jude Channon

Research Officer, Centre for eResearch and Digital Innovation, Federation University

E: j.channon@federation.edu.au

T: (03) 5327 9672

Jennifer Corbett

Research Officer, Centre for eResearch and Digital Innovation, Federation University

E: j.corbett@federation.edu.au

T: (03) 5327 9369

APPENDIX F: FOCUS GROUP/INTERVIEW AGENDA

A coordinated and consistent approach to managing and enabling soil data:
Focus group questions.

Agenda

10.20	Zoom meeting start	
10.30	Introductions & context to today	
10.35	Participant introductions and welcome - Your role - Your project	Participants
10.43	Q1. What data and data management related activities have you undertaken <u>during</u> the project? - Who is responsible for the data management related activities? - Why or why not?	Participants
	Q2. How are you storing data <u>during</u> the project?	Participants
	Q3. Describe the practices you will follow <u>at the end</u> the project with respect to your institution's data storage requirements? - Who is responsible for putting the data into storage? - Where are you putting the data? - Have you considered a data repository and if so, which one? - Should the Soil CRC have any say over where this data is stored at the end of the project?	Participants
	Q4. What do you currently do around data management that could be improved on?	Participants
	Q5. What do you find hard to do around data management practices?	Participants

11.45	Q6. What do you believe are the benefits of creating a data management plan for a research project?	Participants
	Q7. What is the value beyond the project e.g., for your institution, funders of having and implementing DMP's?	Participants
	Q8. What is the benefit to <u>other researchers</u> (<u>researchers not involved in your project</u>) of them following a data management plan? (this can be teased out in Q6 and Q7)	Participants
	Q9. Have you given any consideration to the costs of data management? How do you think one could price such a thing? • How much of this is BAU (business as usual) and is absorbed into your current project? • What is the impact of this on the amount of research you can do?	Participants
12.20	Any comments/feedback from participants	Participants
12.25	Wrap up and next steps	
12.30	Thanks & Close	

APPENDIX G: DATA REPOSITORY SURVEY

A coordinated and consistent approach to managing and enabling soil data: Data repositories survey.

DATA

Q1. Does your institution have any policy for what data you will and won't accept in terms of who funded the research, e.g. CRC?

Q2. Does the repository accept data covering all disciplines, or is it restricted to one or more specified disciplines? If the latter, please describe.

Q3. What data limits does your institution have? For example limitations on the volume (maximum size) of the dataset, formats, etc.

Q4. Does the repository policy allow for the lodging of third party data associated with a research project (e.g. such as that provided by a Grower group organisation)?

Q5. Does the repository support capture of links to papers associated with the data?

Q6. What metrics are available for a researcher to know how frequently the data has been viewed/downloaded?

Q7. Is the repository suitable for storage of private and/or sensitive data?

Q8. Who can upload data files into the repository?

METADATA

Q9. What metadata standards or languages are being used for institutional records?

Q10. Does the repository support the tagging of metadata records with things such as programme the research was undertaken in, e.g. 'CRC for High Performance Soils'.

Q11. Who can load and/or update the metadata or other aspects of the lodged data?

Q12. What metrics are available for a researcher to know how frequently the metadata is viewed/downloaded?

Q13. Is the metadata associated with a data set which has been deposited into another repository (either institutional or discipline) added to your institutional repository?

PERMANENT IDENTIFIERS

Q14. Does the repository assign (mint) DOIs or other PIDs (please specify) for the data and inclusion in the metadata record?

Q15. Are these automatically or manually generated?

Q16. Do the repository and metadata records support Crossref DOIs, Datacite, both or neither?

Q17. Does the metadata/repository include and actively link to ORCiDs of researchers associated with the data?

ARCHIVING & DISPOSAL

Q18. What is the policy and process regarding long term storage of data and metadata records?

Q19. Is there a disposal policy and procedure (e.g. data in repository is certified as deleted after certain time period) and is a metadata record retained in perpetuity?

API

Q20. Does the institute's repository support APIs for publishing/harvesting metadata to a third-party service? If so, are these enabled or being used?

Q21. If so, does the API allow requesting any associated links to papers related to the data – if such information is captured. How does this work with respect to publication/harvesting of private data held in the repository – does the requestor have access to metadata which is not publicly available?

Q22. Does the API support filtering by tags, e.g. can records be published/harvested with a CRC identifier?

Q23. Does your institute they have a technical support person who we could approach for more information?

SUPPORT

Q24. How is access control (e.g. licensing and embargo periods and changes to these) to data and metadata managed, and by who?

Q25. What support is available to researchers to deposit data in the repository?

REPOSITORY SPECIFICATIONS

Q26. What repository software platform is presently in use, and are there plans to replace this software?

Q27. How is data file version control administered?

Q28. Does your institutional repository have a core trust seal certification, or is this being planned soon?

Thank you for participating in this survey. If you have any questions, further feedback about this project or would like to participate in future research, please contact one of the following researchers:

Dr. Nathan Robinson Email: n.robinson@federation.edu.au	Jude Channon Email: j.channon@federation.edu.au
Jennifer Corbett Email: j.corbett@federation.edu.au	Kay Steel Email: kay.steel@federation.edu.au

APPENDIX H: DATA REPOSITORY ETHICS APPLICATION (AMENDMENT)

A condition of your original ethics approval was that any variations or modifications you might wish to make to your project must be formally notified to the ethics committee for further consideration and approval.

If you are proposing a major change to your project you may need to submit a new application to a meeting of the HREC.

Please phone the Ethics Office on 5327 9765 if you require further information.

Please Note: Requests for Amendments will not be approved if there are overdue reports for this project.

1) ADMINISTRATION:	
HREC approval No:	2022-143
Project name:	A coordinated and consistent approach to managing and enabling soil data: Data maturity survey
Current Approval dates:	19/07/2022 to 30/06/2023
Principal Researcher	Nathan Robinson
Other Researchers	Jude Channon Jennifer Corbett Simon Clarendon David Medyckyj-Scott

Complete each category requiring assessment and approval:

PROPOSED CHANGES TO THE PROJECT	
☒ Amendment (go to Category 2)	☒ Extension (go to Category 3)
☒ Change of Personnel (go to Category 4)	

2) AMENDMENT to Approved project
(Note: authorisation by Dean, Deputy Dean, Associate Dean Research or School/Discipline Ethics Coordinator is required in addition to authorisation of the Research team – see final page)

2.1 Project Description:
Provide a one paragraph lay summary of your original project
This research aims to better understand the current state of research data governance principles and management practices used across Soil CRC projects. A survey (known as the 'data maturity' survey) has been used to collect preliminary data from research team members of Soil CRC projects about their research management practices in soil and agricultural science.

In the original ethics application, approval was also granted to conduct a focus group meeting and/or interviews to explore data management practices providing insights that will be incorporated into 'use cases' defined in partnership with Soil CRC program leaders. This ethics amendment includes information the questions to be used in the focus group and/or interviews and the recruitment strategy to be adopted.

Ethics approval is also sought to conduct an additional survey to further consolidate the research findings focused on institutional data repositories. This data will provide evidence about the suitability of these repositories as long-term storage solutions.

2.2 Nature of and reason/s for amendment/s:

Provide details of the changes you propose to make to the project and explain why they are necessary.

Data maturity focus group and/or interviews

As specified in the original ethics application, a focus group and/or interviews was intended to explore how successful the project had been in supporting Soil CRC projects and individuals in better data management practices. The focus group and/or interviews (with interviews offered to interested participants who are unable to attend the focus group) will be used to capture insights about data management costs, data retention and reuse, data storage and repositories, data management plan, perceived benefits and costs of better data management, policies and procedure adherence and training and education requirements. This will offer new insights and understandings about data management practices, enabling the research project team to prepare a report, which will be supplied to the Soil CRC, proposing recommendations to improve management and sharing of data across the Soil CRC. Invitations to participate in a focus group and/or interview will be emailed to the 16 participants who expressed an interest in being further involved in this research following their completion of the data maturity survey.

The email and formal letter of invitation, the PLIS, consent form, and details on the focus group (date, time, online meeting location, questions) are appended to this application.

Data repositories survey

In this amendment application we also seek approval to deploy an additional survey (the 'data repositories' survey) to examine the use of institutional repositories and their application for long-term research data storage. The survey comprises 28 questions focused on the use or permanent data identifiers, storage and allowance for 3rd party data lodging, metadata, archival and disposal processes, production of the application programming interface (API's), support for those sharing data and repository specifications.

Members of the repository sector at the project partner institutions will be invited to complete the data repositories survey. University's data management expert Kay Steel will join the research team (a personnel amendment to include Kay in the research is included below) and provide support to engage and invite the most appropriate members from within the university repository sector to participate in the survey using the recruitment methods outlined in the original ethics application, e.g. voluntary participation, assistance to participants where required, opportunity to complete the survey over the

phone). The Soil CRC partners (farmer groups, industry partners and state agencies) will also be invited to participate in this research. Kay will assist the research team with interpreting the data relevant to national data management strategies. Dissemination of the research findings remain consistent with the outputs listed in the original ethics application.

A separate, covering email and formal letter of invitation, the PLIS and consent form, and the survey are included with this amendment application.

2.3 Possible inconveniences or risks to subjects:

Outline any inconvenience or possible risks that the changes you propose may create for participants eg. changes to confidentiality provisions, physical or psychological risks, increased time commitments etc

The proposed extension to this research remains consistent with the original ethics application - it's research aims and objectives. The proposed research outlined in the amendment does not divert from the original research intent and scope for which ethics approval was originally granted. Therefore, the changes proposed in this amendment will not incur any further risks or inconvenience than those already outlined in the original ethics application. The activities associated with this amendment have minimal risk and impact on participants and this remains a low-risk research project.

2.4 Actions to be taken by researchers to reduce risks:

Advise details of any additional actions and/or support that you will need to provide to subjects as a result of the proposed changes

No additional supports are required as there are no greater risks to participants beyond those outlined in the original application. The updated PLIS will be provided to all focus group/interview participants, and those participating in the data repositories survey. The PLIS includes information to participants about where to seek support should they require it.

2.5 Expected date of implementation of amendments to research:

As soon as approval is granted.

2.6 Will any funding arrangements for the research been affected by the changes?

No

2.7 Implications for compliance with legislative requirements:

Please check current legislation and related requirements, if appropriate - e.g. Privacy Act 1998 / National Statement / Australian Code for the Responsible Conduct of Research

N/A

If no other change is required, please go to Section 5 (if applicable), then Sections 6 & 7

3) EXTENSION OF THE PROJECT

3.1 Provide the proposed new completion date.

The requested new completion date for the project is 22 December 2023.
--

3.2 Provide the reason for the extension of the project.

The request for the extension to the timelines for this project will enable researchers to prepare research publications (e.g. journal publications) and conference presentations from the research data collected for this research.

If no other change is required, proceed to Section 6

4) CHANGE OF PERSONNEL

4.1 Removing personnel from the project: (Copy and paste for multiple changes)

Name:	
Date of Departure:	
Signature of Principal Researcher:	

4.2 Adding new personnel to a project (Copy and paste for multiple changes)

Note: A Principal researcher must be a staff member:

New researcher:	
Name	Kay Steel
Phone	+61351226536
Email (Use Federation details for internal personnel)	kay.steel@federation.edu.au
Position on this project: (eg: Principal Researcher, Co Researcher)	Co researcher
Researcher Qualifications, Experience & Skills 4.2.1 Describe what the researcher will do in the context of this project.	Kay will be support the data collection relating to the data repositories survey for organisations in the Soil CRC. This audit will establish a foundation to advocate support for repositories as part of federated 'virtual data catalogue' for the Soil CRC. Kay has exemplary skills and is a domain specialist in Research and Research Collections and is perfectly placed to support this project activity. Kay will be involved in all stages of data collection, analysis and preparation of report and dissemination of research findings.
4.2.2 List the researchers academic qualifications, and outline the experience and skills that the additional researcher/s has in carrying out the research.	Kay has experience in the compilation, distribution, analysis, and reporting on large survey as part of MLib (Monash University). Member' lead of national working groups compiling, analysing and reporting on institutional repository-related surveys through Council of Australian University Librarians during 2017, 2018 and 2019.

	Drafted and finalised questionnaire, distributed questionnaire, led analysis of results, drafted, finalised and approved final report of 2020 Researcher Skills Needs Survey in 2020 for distribution to University Research committee and RHD, and publication under CC licence on Federation.figshare; leading development of 2023 Researcher Skills Needs Survey. Kay is in a substantive position at the University and is, responsible for development of University programs, services and infrastructure for best practice research data management, including those related to Federation University data and publications repositories.
Declaration of the new researcher: I confirm I have read the originally approved application and subsequent amendment/s and agree to comply with the standard conditions of approval and the National Statement	12/01/2023

New researcher:	
Name	
Phone	
Email (Use Federation details for internal personnel)	
Student ID:	
Position on this project: (eg: Principal Researcher, Co Researcher)	

Researcher Qualifications, Experience & Skills	
4.2.1 Describe what the researcher will do in the context of this project.	
4.2.2 List the researchers academic qualifications, and outline the experience and skills that the additional researcher/s has in carrying out the research.	
Declaration of the new researcher: I confirm I have read the originally approved application and subsequent amendment/s and agree to comply with the standard conditions of approval and the National Statement	
RESEARCHER TRAINING	

All researchers at Federation University are required to undertake training in responsible research conduct (see *Australian Code for the Responsible Conduct of Research, R16*).

Have all of the added named Federation University researchers completed the online training module provided by the University (Research Integrity, Second Edition)? (Every researcher must complete all of the modules they have been advised to complete by the Research Integrity Office - research.integrity@federation.edu.au)	☒ Yes	☐ No
If no , please list those still to complete the training below and confirm that the relevant researchers will complete the training within 3 months of the date of approval of this application (evidence will be required)	☐ Confirm	

4.3 Changing the status of a researcher on the project (Copy and paste for multiple changes)

Researcher changes:	
Name	
Email (Use Federation details for internal personnel)	

Current position on this project: Principal Researcher, Co Researcher	
New position on this project: Principal Researcher, Co Researcher	
Reason for change	

Proceed to Section 6

5) ATTACHMENTS: (if applicable)

- ☒ Copies of amended surveys, questionnaires or interview questions
- ☒ Copies of the amended advertisement, plain language statement, and consent form: Email invitation
- ☐ Details of other permission or approvals required as a result of your proposed changes - NA

6) DECLARATIONS: (for all proposed Amendments, Extensions & Changes to Personnel)	
PRINCIPAL RESEARCHER (STAFF MEMBER): I accept responsibility for the conduct of this project and for the supervision of all associated personnel listed.	
Name	Dr Nathan Robinson
Signature:	
Email:	n.robinson@federation.edu.au
Date	11/01/2023

OTHER RESEARCHER/S: I will perform all tasks/activities in compliance with the National Statement, under the direction of the principal researcher			
Name	Signature	Email	Date

Jude Channon		j.channon@federation.edu.au	11/01/2023
Jennfier Corbett		j.corbett@federation.edu.au	11/01/2023
Kay Steel		kay.steel@federation.edu.au	12/01/2023

7) ADR or Dean or Deputy Dean or School/Section Ethics Coordinator *NB: Amendment requests (only) will not be considered without this additional authorisation:		
Name	Signature	Date
A/Prof Helen Thompson		18/01/2023

For assistance contact the Ethics Office, Mt Helen:

Ph: +613 5327 9765

Email: research.ethics@federation.edu.au

APPENDIX I: DATA REPOSITORY SURVEY

PLIS

Centre for eResearch and Digital Innovation

PROJECT TITLE:	A coordinated and consistent approach to managing and enabling soil data: Data repositories survey
PRINCIPAL RESEARCHER:	Dr Nathan Robinson
OTHER RESEARCHERS:	Jude Channon Jennifer Corbett Kay Steel Dr David Medyckyj-Scott (Landcare Research - Manaaki Whenua) Dr Simon Clarendon (New South Wales Department of Primary Industries)

What is this research about?

Researchers at Federation University Australia (Dr Nathan Robinson, Ms Jude Channon, Ms Jennifer Corbett and Dr Kay Steel), Dr David Medyckyj-Scott (Landcare Research) and Dr Simon Clarendon (NSW DPI) are conducting research to better understand the current state of research data governance principles and management practices in use across Soil CRC projects. This research is in response to a research priority call from the Soil CRC to improve the strategic management of soil-related research data for current and future projects. The need for improved data management and governance practices for projects collecting and managing soil-related research data is focused on achieving a whole of Soil CRC approach to data governance and management (i.e., data maturity).

Data management and governance can improve the visibility and reuse potential of soil-related research data for multiple purposes such as research, development and extension that can support farmers in their pursuit of optimised soil productivity.

The aim of this research is to explore current research data management practices across Soil CRC projects. The Soil CRC has funded this research. You are invited to participate in this research project being conducted by researchers at Federation University (the lead organisation), together with Manaaki Whenua Landcare Research, NSW Department of Primary Industries, Charles Sturt University and the University of Tasmania.

This research involves the collection of data using a survey from repository managers of partners in the Soil CRC. The audit of research repositories will provide foundation information to support current and future Soil CRC projects in terms of data governance, data storage and access, data management strategies and data management tools. This research will inform future data governance work in the Soil CRC to enable better access and use of the soil data that is produced by projects funded by the Soil CRC.

What is involved in this research?

If you agree to participate in the research, you will complete a survey that takes approximately 10-15 minutes to finish. The survey will include questions about: data, metadata, permanent identifiers, archiving and disposal, APIs, support and repository specifications.

Your responses will provide foundational information to engage the Soil CRC on recommended data management practices. Survey data will not include any information that can be used to identify individual respondents unless written permission is granted by the respondent.

How will my information be used?

All information collected as part of this research will be used to identify priorities for enabling better research data management. No individuals will be identified by researchers for specific quotes, comments or responses in the reporting of research findings. Every effort will be made so that the data does not directly identify you. To facilitate this, the researchers will use data that is de-identified, and aggregated prior to analysis.

Please note that the sample size for this research is small and the questions asked during this research are specific. It is possible that your feedback could be identifiable. While every effort will be made to present the data so that it does not directly identify you, the uniqueness of the sector, the small participant numbers and grower groups involved in this research makes it possible that someone familiar within the sector and/or from other organisations, such as the Soil CRC, could identify survey respondents. Please be mindful of this in your answers and do not to include any sensitive information or information that you do not wish to be included in the research.

To ensure your anonymity and confidentiality, all potentially identifying information will be removed from these.

A report that synthesises key findings from the consolidated data collected through this research will be prepared by the researchers for the Soil CRC. This report may be distributed publicly. Some research findings may also be developed as publicly released communications, and for scholarly purposes, including journal publications and conference presentations in alignment with Soil CRC publishing guidelines.

Please be aware that in participating in this research, your de-identified data may be used to inform future research. Only consolidated research findings will be compiled in the report for distribution to the Soil CRC and/or publicly.

Participation

Participation in this research is voluntary. You may choose to respond to all or only some of the questions. Please just answer the questions you are comfortable with. Should you require any assistance with the survey, a time can be scheduled to complete the survey over the phone by contacting Kay Steel on 03 5122 6536 or email kay.steel@federation.edu.au, or Jude Channon on 03 5327 9672 or email j.channon@federation.edu.au.

You are also free to withdraw from the research, without explanation, at any time. There are no consequences for withdrawing from this research, or not taking part in it. If you withdraw before, during or immediately after completing the survey, the information you have provided can be deleted. However, if you withdraw from participation at a later stage, after your

information has been combined with that of the other participants, we may no longer be able to identify your information and it may be unable to be deleted.

Please note that participation (or non-participation) in this research, will not be shared with the Soil CRC. Details about your involvement (or not) in the study will not be supplied to the Soil CRC.

How is data stored?

All named researchers for this study have extensive experience in conducting research and abiding by ethical research principles. Data confidentiality will be maintained. Published results will not include information that identifies any respondents. All data will be stored on secure password protected systems and retained for a minimum of five years.

Further information

It is unlikely that you will experience any distress by participating in this research. However, should you have concerns arising from the survey we advise that you access professional support through your workplace or contact a medical health practitioner. You may also access information via Beyond Blue (1300 224 636) and Lifeline (131 114), providing expert advice, support and information to deal with personal challenges and stress.

If you have any questions, or you would like further information regarding the research project please contact the Principal Researcher, Dr Nathan Robinson: **EMAIL:**

n.robinson@federation.edu.au

PH: (03) 5327 8436.

Ethical approval

This project has been approved by the Federation University Australia's Human Research Ethics Committee (Project number: 2022-143), in accordance with the guidelines provided by the National Health and Medical Research Council for research involving humans.

Should you have any concerns about the ethical conduct of this research, please contact:

Federation University Coordinator Research Ethics, Research Services, Federation University Australia,

P O Box 663 Mt Helen Vic 3353;

Telephone: (03) 5327 9765;

Email: research.ethics@federation.edu.au

Research team members

Nathan Robinson

Senior Research Fellow, Centre for eResearch and Digital Innovation, Federation University

E: n.robinson@federation.edu.au

T: (03) 5327 8436

Jude Channon

Research Officer, Centre for eResearch and Digital Innovation, Federation University

E: j.channon@federation.edu.au

T: (03) 5327 9672

Jennifer Corbett

Research Officer, Centre for eResearch and Digital Innovation, Federation University
E: j.corbett@federation.edu.au
T: (03) 5327 9369

Kay Steel
Associate Librarian, Research and Research Collections, Federation University
E: kay.steel@federation.edu.au
T: (03) 5122 6536

APPENDIX J: MINIMUM METADATA ELEMENTS

(AgReFed Collection Descriptions in Research Data Australia).

Information type (field)	Meaning
Metadata publisher	The organisation that is contributing the metadata record
Identifier	A unique identifier for the resource, i.e. DOI
Metadata source	The primary/authoritative source of truth for the metadata record, as represented by a URI.
Collection Type	The type of collection being described, i.e. collection, dataset, software, etc
Title	The name or title of the collection, should be descriptive and unique, avoid acronyms.
Parties	A related person or organisation linked to the collection (include ORCID if possible) e.g. creator, owner, manager.
Location	Online location (DOI, Handle or URL) of the metadata record OR to download the resource
Related Service	Include a link to the AgReFed portal RDA record (workflow TBA); or to other Services.
Citation	The preferred form for citing a collection to enable data to be referenced.
Access Rights	Collection access conditions. Specify one of: open, conditional or restricted.
Licence	License conditions associated with the collection; a standard licence, e.g. creative commons is preferred.
Description	A summary description of the collection. Provide sufficient information to enable a user to assess suitability of the data for reuse for their purpose.
Subject	Keywords or terms to describe the topic of the resource. Include at least one ANZSRC-FOR code. Additionally, AGRIVOC terms should be used.
Spatial coverage (if relevant)	The geometry for the location the resource relates to.
Temporal coverage (if relevant)	The time period the resource relates to, in W3C Date/Time Format.

APPENDIX K: DATA REPOSITORY SURVEY RESPONSES

SURVEY RESPONSES – UNIVERSITIES (UN)

Data

Question	UN1	UN2	UN3	UN4	UN5	UN6	UN7	UN8
Q1. Does your institution have any policy for what data you will and won't accept in terms of who funded the research, e.g. CRC?	No. The relevant policy states that we will manage data if the creator is affiliated to us.	as long as datasets are associated with us, we are able to create a record in the Institutional Repository regardless of funder.	Nothing specific that I am aware of in our policy whether data will be accepted according to who funded the research.	**I have not responded to the questions re Data storage as we do not have a repository for storage of data. Researchers store data elsewhere and are encouraged to create metadata only records in the repository.**	Yes – foreign interference, data management procedure, <u>deposit license</u> and management procedures.	For the repository we accept any final state research data outputs associated with grant-funded research and with manuscripts and journals submissions that have a data policy requirement. Research data associated with ARDC or NHMRC is a high priority target however any research data associated with a funding stipulation can be published in the repository.	Not that I am aware of. I believe we can include data from any research funder	No, there is no policy.
Q2. Does the repository	Our repository is an institutional	The repository accepts data	All disciplines no restrictions.		Yes – multi-disciplinary.	The repository accepts final	All disciplines	The institutional repository does

Question	UN1	UN2	UN3	UN4	UN5	UN6	UN7	UN8
accept data covering all disciplines, or is it is restricted to one or more specified disciplines? If the latter, please describe.	one and as such accepts data from all disciplines.	covering all disciplines.				state research data from any discipline.		not accept large datasets, rather a metadata record is captured only, The responsibility for where the data is stored/ archived sits with the researcher/ research team. We are aware that this is not best practice and are planning to change this.
Q3 What data limits does your institution have? For example limitations on the volume (maximum size) of the dataset, formats, etc.	Our repository is able to store files up to 2GB. If the files are larger than this, then the Library has an alternative mediated option to larger files available. Most formats are able to be uploaded.	We can facilitate access to most data files.	All users are allocated 5gb initially we can increase if necessary. Figshare documentation states Figshare for institutions also offers larger file size uploads up to 5TB. If the dataset is large it is recommended, researcher links to it rather than store in Figshare.		For the institutional RISE repository, recommended max size is 100MB. Alternative storage options are considered for larger file sizes such as provided by central IT services.	To accommodate data limits, structures, formats and mediated data, research data is not stored in the repository. A metadata record is created within the repository with a link either to an open access dataset, stored on Research Space, or a link to a mediated access form, for	We can accept up to 2GB file size. With regards to formats, we will accept anything that is machine-readable, non-proprietary, unencrypted etc. as per general RDM standards	Researchers' choice of where the dataset is deposited – ad hoc choice. Small datasets preferred - Preferred not to accept large files, as we don't have scope. Pure is primarily a RIMS and for trad research outputs (articles etc)

Question	UN1	UN2	UN3	UN4	UN5	UN6	UN7	UN8
						users to request access. For large datasets, our eResearch Services team assists with storing the data and providing an access link.		
Q4. Does the repository policy allow for the lodging of third-party data associated with a research project (e.g. such as that provided by a Grower group organisation)?	No our research integrity policy which covers research data management, and the repository is only for data created by us.	Where a project spans several institutions, at the outset of the project the research team must determine which institution or University Researcher has custody and responsibility for the storage and eventual disposal of the related research data and primary material.	Yes, if we have permission of the third party to load. There isn't anything in our policy not allowing it. Would need to be lodged by a staff member or student as only they have access to Figshare.		Yes – <u>deposit license</u> , DMPs.	No. As a general rule we would only include data affiliated with the university.	No, given that we only include on the Portal research conducted by our researchers, we would not include research exclusively conducted by non-affiliated parties. We could, however, include a link out on a research data record to the associated third-party data.	As noted in Q2, we currently do not have a data repository. The IR – Pure – is primarily for research outputs i.e. articles and monographs, conference proceedings etc.
Q5. Does the repository support capture of links to papers associated with the data?	Yes, related links to publications can be captured in our repository.	In Metadata but not currently hyperlinked	There is a field for resource title and Doi. It needs to be manually entered by the researcher.		Yes.	Yes, the metadata form has a dedicated field to capture and make public links to supporting publications. The connection to the data also is captured in the repository	Yes, and will be included if the researcher provides them.	See above.

Question	UN1	UN2	UN3	UN4	UN5	UN6	UN7	UN8
						metadata form for any corresponding articles or papers held in the repository. At this stage, there isn't a dedicated metadata field like with data; the link is being made available in the public notes field.		
Q6. What metrics are available for a researcher to know how frequently the data has been viewed/downloaded?	Currently there are no metrics showing for datasets. However, we are able create reports on dataset downloads on request.	Visitors -The number of unique sessions created during this time period. Hits - The total number of requests made to the repository. Downloads	Downloads		Within the repository, metadata records will capture views and downloads where applicable.	As no files are attached to the research data records within the repository, the only statistical data collected is page views. For mediated access research data, we can manually monitor requests.	Portal includes metadata record views and file views/downloads	Available in the portal and anyone can view it – number of views, downloads, citations However, this is not monitored.
Q7. Is the repository suitable for storage of private and/or sensitive data?	While datasets can be restricted we do not allow highly sensitive datasets to be stored in the repository as its primary purpose is to publish data and make it accessible.	It can be used in that way but is not currently.	Yes can store private and sensitive data which cannot be viewed by others.		Yes, access controls can be applied.	Although the repository is not used for storing data, we do not accept private or restricted data into the repository.	Yes, it is possible to restrict access/embargo data.	The current recommended active data storage option – now that CloudStor has been decommissioned – is OneDrive. OneDrive is recommended

Question	UN1	UN2	UN3	UN4	UN5	UN6	UN7	UN8
								for private and/or sensitive data. Archival data storage -there is no recommended solution yet.
Q8. Who can upload data files into the repository?	Data may only be deposited by our staff members or our enrolled students or by a staff member's nominated authorised delegate ("Submitter" or "Trusted User"). Library staff can also upload files on request.	Our Researchers and Research students can create a record in our Data Management Dashboard for their dataset. This record is then reviewed by Repository staff before publication to the repository.	Our staff and students can login onto Figshare.		All our staff and students.	Data files are not uploaded into the repository but into a storage area on Research Space. They are uploaded into Research Space by the repository team.	Our permanent and fixed-term Researchers (will be reviewed by Repository staff before making live), and Repository staff.	We currently do not have a dedicated data repository. Researchers can upload their datafiles to the IR – CRO/ Pure. However, there is low awareness that this is a requirement or an option.

Metadata

Question	UN1	UN2	UN3	UN4	UN5	UN6	UN7	UN8
Q9. What metadata standards or languages are being used for institutional records?	The repository metadata is based on CERIF.	Dublin Core RIF-CS Metadata MARC	Dublin Core – But can use other languages or standards if requested.	I have not responded to the questions re Data storage as we do not have a repository for storage of data. Researchers store data elsewhere and are encouraged to create	Some DC fields, other fields are proprietary to the software but could be mapped to RIFCS.	The repository uses a Dublin Core based metadata schema that has been modified for local needs.	The Research Portal uses the dedicated Esploro metadata schema, which is based on standard formats such as DataCite and Dublin Core, but with additions to	Dublin Core

Question	UN1	UN2	UN3	UN4	UN5	UN6	UN7	UN8
				metadata only records in the repository.			include more types of metadata.	
Q10. Does the repository support the tagging of metadata records with things such as programme the research was undertaken in, e.g. 'CRC for High Performance Soils'.	Yes this could be applied via keywords.	We can add a keyword to group relevant metadata records.	Could put in author field if a contributor is a research centre or enter this information in the keyword's fields.	No	Yes capability exists.	The repository supports the inclusion of metadata to capture the language of the content, the medium or method of delivery, the format of the content, sponsorship and faculty type information. Program information is captured in the Public notes.	Can include 'Additional Subjects' (keywords/phrases)	Affiliation is added. Using standard fields set in the Data Set template, Elsevier Pure set fields. No free tagging is available.
Q11. Who can load and/or update the metadata or other aspects of the lodged data?	Library staff or the creator of the dataset or a trusted user.	Scholarly Publishing staff, based in the library, are the only staff able to update Repository metadata.	The researcher who loaded the metadata or file or if part of a project the other contributors - Repository admin staff also have permissions to edit metadata or load files.	All staff and students	All our staff and students can load and updates can be performed by those if their name is on the record.	Only the repository team can load and/or update metadata or other aspects of the lodged data.	Our permanent and fixed-term Researchers (will be reviewed by Repository staff before making live), and Repository staff.	The metadata team in the Library add metadata; if research output is already published try to link data up to the publication.
Q12. What metrics are available for a researcher to know how frequently the	This is restricted to library staff running reports.	Visitors -The number of unique sessions created during this time period.	Views	View and download metrics are displayed on every record summary page	Within the repository, metadata records will capture views and downloads	the only statistical data collected is page views – all time usage, top item views etc.	Portal includes metadata record views.	Available in the portal and anyone can view it – number of views,

Question	UN1	UN2	UN3	UN4	UN5	UN6	UN7	UN8
metadata is viewed/downloaded?		Hits - The total number of requests made to the repository.		(both in the staff client and in the Public repository)	where applicable.	For mediated access research data, we can monitor request.		downloads, citations.
Q13. Is the metadata associated with a data set which has been deposited into another repository (either institutional or discipline) added to your institutional repository?	Yes, we pull in our research datasets records from external repositories via a search tool call Data monitor.	Dependent on other Institutional requirements. Data underpinning publications should be archived in the Institutional Repository	No not at present only if researcher adds it.	Yes	Yes.	Yes, if the dataset is affiliated with the university the metadata can be added with a link to another repository.	Yes, we can do that if the researcher provides us with this information.	Yes, if identified

Permanent identifiers

Question	UN1	UN2	UN3	UN4	UN5	UN6	UN7	UN8
Q14. Does the repository assign (mint) DOIs or other PIDs (please specify) for the data and inclusion in the metadata record?	Yes DOIs are minted for all datasets.	DOI & Handle assigned for Open or Mediated data ; Handle only for Closed data.	Mints DOI when item is loaded.	DOIs are minted for metadata records in the repository and included in the metadata record.	DOIs.	Yes, DOIs are minted and included in the metadata record for all research data published in the repository.	We can mint DOIs in the Portal; platform also assigns unique record identifier.	Yes, DOIs by DataCite System assigned PID – Pure handle and CRO RUL
Q15. Are these automatically or manually generated?	Currently Library staff manually generate them but there is an option to automate this.	Currently the DOI is manual & Handle automatic. However we have an upgrade of our	Yes these are automatically created.	Currently manually generated, but the functionality to automatically generate these is currently	Semi-automatic.	Automatically generated.	DOIs - manually generated within the platform; record identifier – automatically generated by the platform.	CRO and System PID are system-generated. DOI is manual.

Question	UN1	UN2	UN3	UN4	UN5	UN6	UN7	UN8
		Data Management Dashboard that will automate the DOI minting		being implemented in our new repository.				
Q16. Do the repository and metadata records support Crossref DOIs, Datacite, both or neither?	DataCite.	Datacite	Datacite, Crossref-	Datacite	Any DOI can be added, but DOI generation is through DataCite.	Datacite.	DataCite DOIs	Both – Crossref DOI and Datacite
Q17. Does the metadata/repository include and actively link to ORCIDs of researchers associated with the data?	No, our repository only stores ORCIDs on the person record not on research output records.	We do not currently have this functionality.	Yes, it does link to ORCIDs. When you login to Figshare, for the first time, complete My Profile and link your <u>ORCID</u> . Login to ORCID and activate the DataCite link. This will allow your records for datasets and NTROs to be added automatically to your ORCID record.	ORCIDs are currently only displayed on the Author/Researcher record in the Repository and are not linked to the research data record	Yes, once authenticated.	No. The data record is harvested from the repository to the university's research publication system (Symplectic Elements) and attached to the researcher's ORCID account and to their university profile.	The records on the Portal link to the author's academic profile which will include their ORCID if they have added it.	Data Monitor works of affiliation

Archiving and disposal

Question	UN1	UN2	UN3	UN4	UN5	UN6	UN7	UN8
Q18. What is the policy and process	The repository is not a preservation	Researchers can identify how long their data	Retain data and metadata long term.	No current policy	The dataset and metadata records remain	As the data is stored in Research	As per our <u>university's Policy</u> (see	No as yet, in development

Question	UN1	UN2	UN3	UN4	UN5	UN6	UN7	UN8
regarding long term storage of data and metadata records?	system however it is expected that once the data is published it will remain accessible alongside the metadata.	should be retained in line with the State Records Act 1998 or their funding.	Data is held long term not actively deleted unless there is a requirement to do so.		in the institutional repository under the researcher's profile unless the researcher leaves the institution, at which point the researcher's profile will no longer be visible in the public view of the repository. However, the dataset and metadata record remain discoverable in the public repository and if there are co-authors / co-owners affiliated with our institution, then the dataset and metadata record will remain visible in the public view of the institutional repository under the profiles for those co-authors / co-owners. Storage of data as per University policy as contained in the <u>Research</u>	Space it falls under the university's data storage requirements and standards and those of State Archives. Depending on the data type, we would be required by to retain data between 5 and 15 years after the last action. As the last action could include being accessed on the repository, in practice data would likely be retained permanently.	No.15-21 on pg. 4-5), data must be retained for a minimum of five years post-publication, in accordance with the State Sector Disposal Authority, Recordkeeping Policy, and other relevant standards; agreements; and ethical protocols. The Policy also outlines that longer retention periods may be required for different types of data. <u>Our University's policy</u> . See No. 11-18 on pgs. 3-4) outlines the process regarding storage of research data during and at the conclusion of the project.	

Question	UN1	UN2	UN3	UN4	UN5	UN6	UN7	UN8
					<u>Data and Primary Materials Management Procedure.</u>			
Q19. Is there a disposal policy and procedure (e.g. data in repository is certified as deleted after certain time period) and is a metadata record retained in perpetuity?	There is no disposal policy and procedure for data in the repository however there are disposal procedures for data stored in other university data storage systems. Metadata will be retained as long as necessary (and after the creator has left the University).	Not at this stage	Metadata is retained in perpetuity.	No	See answer above for the public view of the institutional metadata record. The University adheres to the legislative authority of the <u>University Sector Retention and Disposal Schedule – QDAN601</u> Section 8 RESEARCH, to determine retention requirements. Metadata elements to be retained after disposal are recommended as per the State Recordkeeping Metadata Standard and Guideline	Once data is published by the University with a DOI, we consider it to be part of the permanent scholarly record as it could be cited by other researchers and professional groups. If disposal were requested by the custodian, this would require manual intervention by staff. A tombstone metadata record would be required to meet Datacite guidelines, and we would likely include in that metadata record the rationale for disposal/withdrawal of the dataset.	As per our University's Policy and Procedure, disposal of research data must be in accordance with State or any relevant procedures as determined by the University. Policy and Procedure do not address retaining metadata and no decisions have been made on this with regard to the Research Portal at this stage.	No as yet, in development

API

Question	UN1	UN2	UN3	UN4	UN5	UN6	UN7	UN8
Q20. Does the institute's repository support APIs for publishing/harvesting metadata to a third-party service? If so, are these enabled or being used?	Yes, we harvest metadata to Research Data Australia using an API.	Metadata can be harvested to approved services via OAI-PMH feed.	Yes it does support it. We are currently using OAI-PMH rather than API. Yes we have scheduled harvests to third party services such as RDA, Trove, Library discovery layer.	Don't know.	Yes, Scopus, WOS, ORCID (early stages of implementation) . Also harvested via OAI-PMH by the National Library of Australia.	The repository supports APIs, and these are used to support data flows for internal systems, such as Symplectic Elements. External services such as Trove and Unpaywall harvest metadata using OAI-PMH. We have a strong preference for OAI-PMH harvesting over direct API access for third parties, for security and server performance reasons. API access would need to be negotiated and would require access credentials.	Yes, the Portal data is harvested by Trove, and we are still in the process of adding to other third-party service providers.	Unsure
Q21. If so, does the API allow requesting any associated links to papers related to the data – if such	There are API endpoints for research outputs as well however we have not set up a harvest that	N/A	The requestor will only have access to the metadata if it is publicly available.	No	No.	Requestors do not have access to private metadata fields and no private data is held in the repository.	The metadata can always be harvested whether the research data is restricted or not. And we put on	Closed datasets – metadata is always visible.

Question	UN1	UN2	UN3	UN4	UN5	UN6	UN7	UN8
information is captured. How does this work with respect to publication/harvesting of private data held in the repository – does the requestor have access to metadata which is not publicly available?	pushes datasets and associated linked papers into another system. Security on our APIs is quite strict and so access requests are managed on a case by case basis.						the Portal record that the researcher/auth or can be contacted regarding the data, if appropriate.	
Q22. Does the API support filtering by tags, e.g. can records be published/harvested with a CRC identifier?	Yes, through a keyword.	Not at this stage – a specific set would need to be determined depending on requirements.	If this information is included in the metadata as an author or keyword it should be able to be searched.	Don't know.	Not currently, but under consideration.	No, the API doesn't support filtering by tags at time of harvest.	I think the best we could do with this is put this in the Notes/Comments or as an additional subject field which we could then filter.	Unsure
Q23. Does your institute they have a technical support person who we could approach for more information?	Yes, however at this stage all requests should come via Library Manager, Research Publications and Data Services	You can start enquiries with me, and I can find the appropriate contact for you.	These were setup by Figshare support or the support at the third party site.	No	Yes.	Repository staff are responsible for local administration and can be contacted directly. More advanced technical support for our repository is provided by a commercial company on a fee-for-service basis. We would need to be	With regards depositing into the Research Portal, we have a generic repository email. We do not have a dedicated technical support person, but contact can be made via email.	Yes, via email.

Question	UN1	UN2	UN3	UN4	UN5	UN6	UN7	UN8
						involved in any technical conversation that your team is having with our provider.		

Support

Question	UN1	UN2	UN3	UN4	UN5	UN6	UN7	UN8
Q24. How is access control (e.g. licensing and embargo periods and changes to these) to data and metadata managed, and by who?	Via Library staff in the Research Publications and Data Services team	Licensing is determined by the researcher sometimes with consultation with Library staff. Embargoes in the repository are managed by the Scholarly Publishing Team	Figshare is unmediated so access managed by whoever enters the record they determine the embargo period, add the license details and can amend any metadata. They can amend this data anytime. Once the embargo date is entered the system automatically makes it available at the appropriate time.	Application of licenses and embargo periods are applied manually by repository editorial staff. Embargos expire automatically.	The system enables the provision of license, embargo, access controls. Managed by repository editorial staff.	Access control to the data and metadata is managed by the repository team. For mediated access, the repository team manages the request and access process between data custodians and users.	As per our Policy, we recommend the data is as open as possible via a CC-BY license, but will apply more restricted access/embargo periods to data that is sensitive etc. This is applied in the Research Portal upon submission by the researcher.	Set by the system and then managed by metadata team
Q25. What support is available to researchers to deposit data in the repository?	The Research and Academic Engagement team in the Library provides support.	Scholarly Publishing staff – system support & Doi minting; advice for researchers	There is a library guide with information, Liaison Librarians conduct regular webinars on	User guides Advice and assistance from Library staff	SharePoint, Knowledge articles, Help videos, drop-in sessions, one-on-one zoom support.	Our Academic Engagement Services - Researcher Services team provide general advice,	Currently holding data initiative workshops, general advice from repository staff regarding	Library Metadata team - support adding metadata, dataset description,

Question	UN1	UN2	UN3	UN4	UN5	UN6	UN7	UN8
		Research Liaison Librarians - specialist support for researchers in the management of research data.	how to use Figshare. It is also possible to book individual appointments with a liaison librarian for individual assistance.			consulting and training in research data management. The repository team works directly with researchers to support the creation of high quality metadata and upload of data to the repository.	depositing into the Portal, Library's Research Data Management research guide	dataset report, DOI, Librarian team preliminary training and awareness of service

Repository specifications

Question	UN1	UN2	UN3	UN4	UN5	UN6	UN7	UN8
Q26. What repository software platform is presently in use, and are there plans to replace this software?	We use Pure (Elsevier CRIS system). No plans to move off this platform.	Vital	Figshare for institutions.	Cayuse – new software implemented in Dec 2022 for metadata only. Discussion around implementing a solution for data storage.	Cayuse Repository, currently investigating augmenting this particularly in relation to storage of datasets.	The repository is on a DSpace platform with an upgrade to DSpace 7 planned for this year.	We currently use Ex-Libris Esploro Research Portal. We migrated to this software from EPrints this year.	We are investigating the use of AWS.
Q27. How is data file version control administered?	By the library team. All changes and new files are validated by this team.	Metadata records are persistent - additional data file versions may require a new Doi / repository record—depending on	The Repository software.	NA	Version control functionality provided within the system.	Versioning is controlled through the repository platform by the repository team.	We haven't encountered this yet.	No policy yet established.

		the level of change.						
Q28. Does your institutional repository have a core trust seal certification, or is this being planned soon?	We have applied for this and are waiting for the outcome.	Not at the moment	No we don't have core trust seal certification . It is not something we would do in 2023 but could look at in 2024.	NA	No, not at this stage.	Our repository is not currently certified and we have no plans to do this.	Not that I am aware of	unsure

SURVEY RESPONSES –FARMER GROUPS (FG), STATE AGENCIES (SA) AND INDUSTRY PARTNERS (IP)

Data

Question	FG1	FG2	SA1	IP1
Q1. Does your institution have any policy for what data you will and won't accept in terms of who funded the research, e.g. CRC?	No, all data is accepted.	No policy. But we mainly create data, rather than accept it. If we are given data relevant to a project, it might be stored on someone's personal computer rather than being uploaded onto our repository	No – there is no mention in the Data Management Policy of what data is accepted.	Not to my knowledge but all research has to acknowledge the source of the material. If a paper was undertaken by the CRC we would acknowledge the CRC as the source of the research material.
Q2. Does the repository accept data covering all disciplines, or is it restricted to one or more specified disciplines? If the latter, please describe.	All disciplines – our data is primarily generated from our research, demonstration, extension activities.	Our SharePoint system contains mainly soil test data but also paddock histories where applicable. We have also collected financial information related to farming systems and what farmers normally apply. Research data in terms of response to products. Monitoring data	We don't have a specific repository for all data. We do have an Information Asset Register providing metadata and location of stored data which has a few datasets and associated metadata	.Accept all disciplines as pertinent to the nature of the business
Q3 What data limits does your institution have? For example limitations on the volume (maximum size) of the dataset, formats, etc.	No limit, we just store all our project data, which is not very big.	We don't tend to have limitations, other than photos. The limitation is the staff member, whether they upload it onto SharePoint. This comes down to whether it needs to be shared or stored or time constraints. Six years of data is usually stored on SharePoint and then its archived and to access you need to contact the site administer (CEO). Data is just stored in terms of the soil test result, survey form or however collected or reported.	None	Have not exceeded limits as yet,
Q4. Does the repository policy allow for the lodging of third-party data associated with a	We hold data from collaborative projects from other grower groups.	We don't really have a policy. We could put it there if we	There is no official data repository. Some groups within the department have their own.	No policy

Question	FG1	FG2	SA1	IP1
research project (e.g. such as that provided by a Grower group organisation)?		thought it was worthwhile and worth storing.		
Q5. Does the repository support capture of links to papers associated with the data?	Increasingly, we are storing the publications that are generated as part of projects with the data.	It could, but we don't tend to produce formal research papers, but certainly info for our trial results booklet.	No	Yes have the ability to capture links
Q6. What metrics are available for a researcher to know how frequently the data has been viewed/downloaded?	Not available	None.	N/A	Internally no
Q7. Is the repository suitable for storage of private and/or sensitive data?	Probably not. It is just in SharePoint and so all the organisation has access to it.	We can limit access to people but only people I think with administrative rights can do this.	No. Additionally, the data management policy states the policy..." also does not apply to Third Party Data made available to us or held by us under agreements or arrangements with third parties or terms of consent."	Yes
Q8. Who can upload data files into the repository?	All employees upload data into their various areas of SharePoint.	Every team member who has been given access rights.	There is no formal repository. Some groups have their own which is generally restricted to staff within their group.	Some sites are restricted

Metadata

Question	FG1	FG2	SA1	IP1
Q9. What metadata standards or languages are being used for institutional records?	None	We don't use metadata standards currently, but will be under advice from our Uni partners for data we submit to VAS	Nil	Not sure
Q10. Does the repository support the tagging of metadata records with things such as programme the research was	No	It would in most cases. The trial name is generically named, with the funding body, location and year the data is relevant for	The information asset register would be capable of doing this	Company would be supportive

Question	FG1	FG2	SA1	IP1
undertaken in, e.g. 'CRC for High Performance Soils'.				
Q11. Who can load and/or update the metadata or other aspects of the lodged data?	Any employee, although this is not practice at moment.	Everyone in our organisation with access rights.	Assigned data custodians, staff who have write access to the Information Asset register	
Q12. What metrics are available for a researcher to know how frequently the metadata is viewed/downloaded?	Data is not available to outside users. No metrics.	None	Unsure	
Q13. Is the metadata associated with a data set which has been deposited into another repository (either institutional or discipline) added to your institutional repository?	NA	Probably not.	It can be.	

Permanent identifiers

Question	FG1	FG2	SA1	IP1
Q14. Does the repository assign (mint) DOIs or other PIDs (please specify) for the data and inclusion in the metadata record?	No	I don't know what these stand for, so probably not.	No	
Q15. Are these automatically or manually generated?	NA	They would be manually generated if we did them.	N/A	
Q16. Do the repository and metadata records support Crossref DOIs, Datacite, both or neither?	No	I don't know what these stand for, so probably not	No- I don't think so	
Q17. Does the metadata/repository include and actively link to ORCiDs of researchers associated with the data?	NA	No	No	

Archiving and disposal

Question	FG1	FG2	SA1	IP1
Q18. What is the policy and process regarding long term storage of data and metadata records?	No archiving and disposal policy at present.	We don't have a policy.	The data management policy states: "Research Data will be stored in a secure location, in an approved data repository, as defined by our data architectural requirements, in machine-readable format, and in such a way that enables future access."	
Q19. Is there a disposal policy and procedure (e.g. data in repository is certified as deleted after certain time period) and is a metadata record retained in perpetuity?	NA General organisation data is destroyed after 7 years in line with ATO.	No policy, but it is kept.	No	

API

Question	FG1	FG2	SA1	IP1
Q20. Does the institute's repository support APIs for publishing/harvesting metadata to a third-party service? If so, are these enabled or being used?	No	Don't know what APIs stand for.	I don't think so.	
Q21. If so, does the API allow requesting any associated links to papers related to the data – if such information is captured. How does this work with respect to publication/harvesting of private data held in the repository – does the requestor have access to metadata which is not publicly available?	NA		No, although the policy does state that data must be stores "in machine-readable format, and in such a way that enables future access."	
Q22. Does the API support filtering by tags, e.g. can records be published/harvested with a CRC identifier?	No		I don't think so	

Question	FG1	FG2	SA1	IP1
Q23. Does your institute they have a technical support person who we could approach for more information?	No	CEO would be a good starting point.	Yes, I think so although it may take some sleuthing. Please contact me for more details.	

Support

Question	FG1	FG2	SA1	IP1
Q24. How is access control (e.g. licensing and embargo periods and changes to these) to data and metadata managed, and by who?	Data available	Probably CEO and those he provides administrative rights to, which would be cropping manager.	Data is assigned to a controller and that person it the contact in the information asset register.	
Q25. What support is available to researchers to deposit data in the repository?		Support would be sort from CEO.	Limited. There is no official repository although metadata can be entered into the Information Asset Register.	

Repository specifications

Question	FG1	FG2	SA1	IP1
Q26. What repository software platform is presently in use, and are there plans to replace this software?	SharePoint is our only repository software.	SharePoint. No plans to replace.	There is no specified repository	
Q27. How is data file version control administered?	Date stamped file name	Generally by the year the data was taken. This works ok for one year projects, but not really for projects which run over multiple years. So we tend to copy project master files into the current year and add to them	N/A	
Q28. Does your institutional repository have a core trust seal certification, or is this being planned soon?	No	No and No.	No	

RAPID SURVEY RESPONSES –FARMER GROUPS (FG), STATE AGENCIES (SA) AND INDUSTRY PARTNERS (IP)

QUESTION	FG3	FG4	FG5	FG6	IP2	SA2
Q1. Where do you put your data from your research? Does it stay there once your project is over or does it get stored elsewhere?	In folders on SharePoint. When the project finished, the folders get moved to a different folder.	Don't tend to collect data, this is mainly through research partners. There is some data for demonstration sites – this is mainly Excel files. Online Farm Trials is an option for the storing of data.	Local server within our organisation. Unless the project is a partnership, in which case data will be stored by other organisations in their system. Mostly it stays here – although we also share the data with participating farmers (if we are using a farmer's property to do the trial/research), in which case they will also have the data	On a Shared Drive. We have a 'Completed' projects folder and a 'Current' projects folder. When a project is completed, the project folder that contains all project information is moved to the 'Completed' projects folder indefinitely.	Do not house any data so cannot respond to questions.	The only data that we have is that contained in the Final Reports that they receive. They feel they can't help with this survey.
Q2. Who has access to it?	All staff	Our staff	Typically all staff – however if there is sensitive info (e.g.: pricing/financial), it may be locked to only a few individuals having access	All Staff		
Q3 How do you share your data?	Via email, if needed	Via email – mainly as Excel files.	One on one extension, grower group/shed meetings, newsletters, workshops, Facebook videos, industry seminars	Internally by sending links. Externally (e.g., to project partners) by sending the files via email.		
Q4. How do you protect your data?	Don't	NA	Office 365 2FA, also IT system managed by external provider	Staff computers/laptops require a password/PIN when they are turned on.		

QUESTION	FG3	FG4	FG5	FG6	IP2	SA2
Q5. How do you version control your data?	Poorly. Name – everyone does it different.	NA	Office 365, so either autosave updates, or create new file for updated version.	We don't normally. Occasionally we will have e.g. _V2 or a date at the end of a file name if we update a file and want to keep the original.		
Q6. How would you know if someone else has accessed, shared, or downloaded your data?	Wouldn't know. Only if they've asked where it is.	NA	Rely on IT service provider if unauthorised access occurs	I wouldn't, unless I saw them in the word document or excel spreadsheet at the same time (Microsoft 365 shows person's name accessing the file up the top)		
Q7. If you have a storage policy, does it allow for the lodging of third-party data associated with a research project (e.g. such as that provided by another organisation)?	Don't have one.	NA	No strong policy around 3rd party, however our privacy policy states any data will remain confidential unless all participating parties agree to it being shared.	We don't have a policy.		
Q8. Does your organisation have a data repository? If so, what software is used?	No.	No. Use of OFT if required.	SharePoint	No		
Q9. Is the repository suitable for storage of private/sensitive data?			Yes – layered access. Some folders only have management access, some only individual staff access for private data			
Q10. Who can upload files into repository?			Anyone who has access. IE Management only for management files, all staff for all staff access folders			

QUESTION	FG3	FG4	FG5	FG6	IP2	SA2
Q11 What data limits does your organisations repositories have for example limits on volume, size of dataset etc?			Not sure – I don't believe we will have an issue with storage size unless we start doing a lot more (as in 1000 x more) drone or other imagery.			
Q12. Do you record metadata? Is so. How?	Not well. It's in spreadsheets or word docs	NA	Not really, although we try and store everything logically, so that it is obvious. E.g.: grower data in named grower folders, variety trials in variety trial folder etc.	No		
Q13. Who can upload and/or update the metadata or other aspects of the lodged?	All staff		N/A – see above			
Q14. What is the policy and process regarding long term storage of data and metadata records?	Don't have one.	NA	No policy – up to management's discretion.	No policy or process. All data is retained on the server indefinitely.		
Q15. Is there a disposal policy and procedure (e.g.) data in repository is certified as deleted after a certain time period) and is a metadata record in perpetuity?	Don't have one.		No policy – up to management's discretion	No disposal policy and procedure. All data is retained on the server indefinitely.		
Q15. What support is available to employees to deposit data into the repository?	I would love to have a data repository set up.	NA	Typically one on one training, or external IT service provider can support and troubleshoot	It's the responsibility of senior staff to teach junior staff about where and how to store data on our server.		

PAGE DELIBERATELY LEFT BLANK