

A COORDINATED AND CONSISTENT APPROACH TO MANAGING AND ENABLING SOIL DATA

Project 2.2.005



THE CHALLENGE

Soil CRC projects are generating large volumes of soil data across multiple programs, regions and disciplines. This data has value for both researchers and farmers, such as being used to support predictive modelling and decision-support tools, which are critical for improving soil management and productivity.

Realising this opportunity depends on the data being findable, accessible, interoperable and reusable (FAIR). However, soil data generated across the CRC currently varies widely in format, quality, scale and level of documentation. This diversity makes it difficult to store, share, link and interrogate datasets in a consistent and reliable way.

There are also gaps in data management capability. Skills, knowledge and practices differ across projects and organisations, and many researchers are not familiar with best practice research data management approaches. Concerns around data ownership, licensing, privacy, and long-term storage add further complexity, creating barriers to data sharing and reuse.

Without a coordinated approach, these challenges limit the ability to use soil data effectively post the original intent. A cross-program solution is needed to support secure, reliable storage, improve data discovery, and enable reuse of analysis-ready data. This includes clear guidelines, defined roles and responsibilities, and practical support to build skills and understanding across the Soil CRC.

KEY POINTS

- Soil CRC projects are generating large volumes of valuable soil data, but inconsistent data management practices potentially limit their findability, accessibility and reuse.
- A data maturity survey, the first of its kind for a CRC in Australia, revealed that many researchers are unfamiliar with research data management concepts, and that many projects did not have a formal Data Management Plan or funding in place for management of these data assets.
- Improving coordination and capability in data management would increase the value of soil data and support its use in future projects.
- Training resources on data management were developed for Soil CRC participants to guide those managing data in their project.
- The Soil CRC is well positioned to build an enduring legacy from its research data. A dedicated data steward role has been created and a Virtual Data and Information Catalogue is being developed to help secure the long-term value of Soil CRC data.

OUR RESEARCH

The project began in March 2021 as a collaboration between three universities (Charles Sturt University, Federation University, University of Tasmania) and two government agencies (NSW Department of Primary Industries and Regional Development, Manaaki Whenua Landcare Research). It aimed to develop a coordinated approach to soil data management that could be used across Soil CRC projects, supporting secure storage, discovery and reuse of data.

The project had multiple activities:

Assess the current state of play

The project reviewed how soil data was being managed across partner organisations to establish a clear picture of existing practices. This included examining how data was stored, documented, and shared across universities, government agencies, industry partners, and farmer groups, as well as the capabilities and limitations of the data repositories each organisation used.

Roles and responsibilities for data management were mapped, identifying where accountability was unclear or absent. The project investigated how different data types, including spatial, temporal, and emerging data sources such as near real-time sensor data, move from collection through to storage, analysis, and eventual use in modelling and visualisation tools. This pinpointed where data can be lost, constrained or underutilised along the way.

Key constraints affecting data sharing, including ownership, licensing, privacy, and security, were examined to understand what conditions would need to be met before data could be reliably exchanged across organisations.

Capability and knowledge assessment

Surveys were conducted with Soil CRC participants to assess data management knowledge, skills and practices. This included a data maturity survey, the first of its kind for a CRC in Australia, to establish a baseline of current capability across the organisation. The survey examined familiarity with research data management concepts, the use of Data Management Plans, attitudes toward data sharing, and concerns around licensing, privacy and security. Results were used to identify gaps and inform the development of targeted training and support materials.

Data Management Plan

Existing Data Management Plans from the Australian Antarctic Data Centre, the Digital Curation Centre and Federation University were reviewed to identify a fit-for-purpose approach for the Soil CRC. Researchers and data managers were consulted to understand how data accountability, documentation, and sharing could be structured consistently across the CRC, despite differences in institutional systems and research contexts.

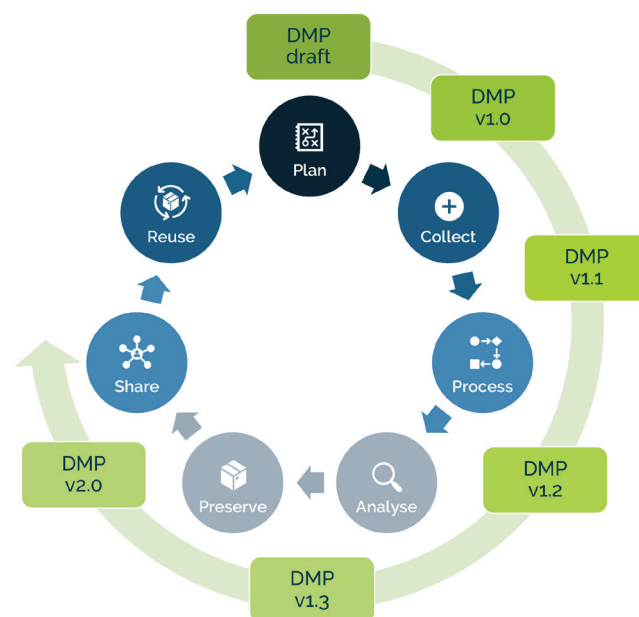


Figure 1. Research data lifecycle, showing how the data management plan (DMP) is revised as a project progresses. Adapted from University of Reading data lifecycle.

Trialling data management approaches

To ensure that proposed data management approaches were grounded in real project needs, the approaches were trialled with four active Soil CRC projects:

1. Matching soil performance indicators to farming systems (2.1.006)
2. Measuring soil microbes (2.1.008)
3. Optimising soil constraint management through computer-based learning and modelling (4.3.006)
4. In-paddock variability of plant available water (a PhD research project).

A fifth case, focused on the University of Tasmania's involvement in the Soil CRC, examined how to manage and transfer data at the end of a project lifecycle.

These projects were selected because their data feeds directly into modelling, analytics, and decision-support tools, where the demands on data quality and consistency are highest.

Engagement and refinement

Throughout the project, workshops, webinars, and targeted consultations were held to engage researchers and partner organisations in data management practices. These activities served to:

1. Build awareness and capability among those managing data day-to-day.
2. Gather feedback to refine the guidelines, tools and training materials being developed.

RESEARCH FINDINGS

The project originally aimed to support the seamless sharing of soil data across the Soil CRC. While data sharing was occurring, the governance and technical advisory frameworks needed to underpin a coordinated approach were not in place. As a result, the project pivoted to focus on improving data management practices and capability across the CRC as a foundational step toward that longer-term ambition.

Current state of play

The review of data management practices across Soil CRC partner organisations revealed significant inconsistency in how soil data is stored, documented and shared. Institutional repositories used by universities, government agencies and industry partners varied considerably in their structure, metadata standards, and capacity to accept third-party data. In many cases, information about how to access or contribute data was not readily available, limiting discoverability across the CRC.

Mapping data flows highlighted multiple points where data was being lost, constrained, or underutilised, particularly in the transition from collection to storage and from storage to use in modelling or decision-support tools. For example, when trialling data management approaches, one project revealed that data from active sensor networks was being stored in a cloud file-sharing platform, making it undiscoverable outside the institution and invisible to the broader Soil CRC. As the project neared its end, 60,000 data points per day faced permanent loss, with no clear pathway for transferring or archiving the data, no budget allocated to data management, and unclear IP ownership.

Near real-time sensor data presented challenges more broadly, with few clear pathways for integrating these emerging data sources into existing Soil CRC datasets.

Ownership, licensing, privacy, security and data misuse concerns were consistently identified as barriers to data sharing across organisations. These constraints reflected a broader lack of agreed processes and governance structures to guide how data should be handled and exchanged.

Capability and knowledge assessment

The data maturity survey revealed wide variation in data management knowledge and practices across the Soil CRC. Many participants were unfamiliar with research data management concepts and terminology, and most projects did not have a formal Data Management Plan in place. Where data management was considered, it was often viewed primarily as a means of reducing the risk of data loss, rather than a way to enable broader reuse and long-term value.

There was also a strong and consistent desire for formal training in research data management, with participants identifying a need for practical guidance on data standards, licensing, and sharing practices.

Data Management Plan

In response to the survey findings, a Data Management Plan template, accompanying guide, and glossary were developed and made accessible to all Soil CRC participants through the online members' portal. A Data Asset Register, built on the Data Management Plan template, was also developed to help projects identify and record key datasets, and is now linked to final project reporting requirements across the Soil CRC.

An aspirational data governance policy was developed for the Soil CRC, defining roles and responsibilities for those generating, managing, and maintaining soil data, along with standards and processes to support consistent, FAIR data practices.

Training materials

A suite of training and development materials was produced to assist Soil CRC researchers in improving their data management practices. These cover defining metadata fields, data repositories, FAIR data principles, data standards, security, licensing, privacy, archiving, digital identifiers, file naming conventions, publisher requirements, and a top ten tips guide. A set of 13 frequently asked questions on research data management, based on feedback from workshops, interviews and surveys was also developed. All resources are accessible to participants through the Soil CRC members' portal.

SIGNIFICANCE OF THE FINDINGS

The data collected across Soil CRC projects represents a significant investment, with unrealised value that improved data management practices could unlock. Improved coordination and capability would make this data more accessible for use in modelling, visualisation, and decision-support tools, strengthening the research outcomes available to farmers and land managers.

The training and engagement delivered through this project has prepared Soil CRC participants to manage data assets more effectively, with flow-on benefits including improved research visibility, stronger funder confidence, and greater efficiencies in future research.

The groundwork laid by this project positions the Soil CRC to build an enduring legacy from its data. The data already collected has value that extends well beyond individual projects and beyond the CRC's ten-year funding cycle. Realising that value requires sustained commitment to the roles, systems, and practices this project has identified and begun to establish.

NEXT STEPS

Following this project, future Soil CRC projects were asked to define data management resourcing as a line item from the outset, ensuring it is treated as a core component of research rather than an afterthought.

The role of a dedicated data steward was created within the Soil CRC operations team to provide the ongoing support, oversight and capability building needed to maintain momentum beyond this project. Equally, a Virtual Data and Information Catalogue is being created so that there is an enduring, searchable record of Soil CRC data and information resources, ensuring the value of the CRC's research investment is accessible within and beyond its ten-year funding cycle. Figure 2 illustrates the proposed data ecosystem such a catalogue would need to support, highlighting the complexity of relationships and dependencies across partner organisations.

Data Management Plans are most effective when treated as living documents, regularly revisited and updated as projects evolve. Complementary training and development should be prioritised to ensure those managing data have the skills to use them effectively, skills that will remain valuable well beyond the Soil CRC.

Since this project began, generative AI has emerged as a potentially significant tool for data management, including the coordination and linkage of datasets. Exploring its application within the Soil CRC data ecosystem warrants consideration, alongside the governance frameworks needed to ensure its responsible use with soil data.

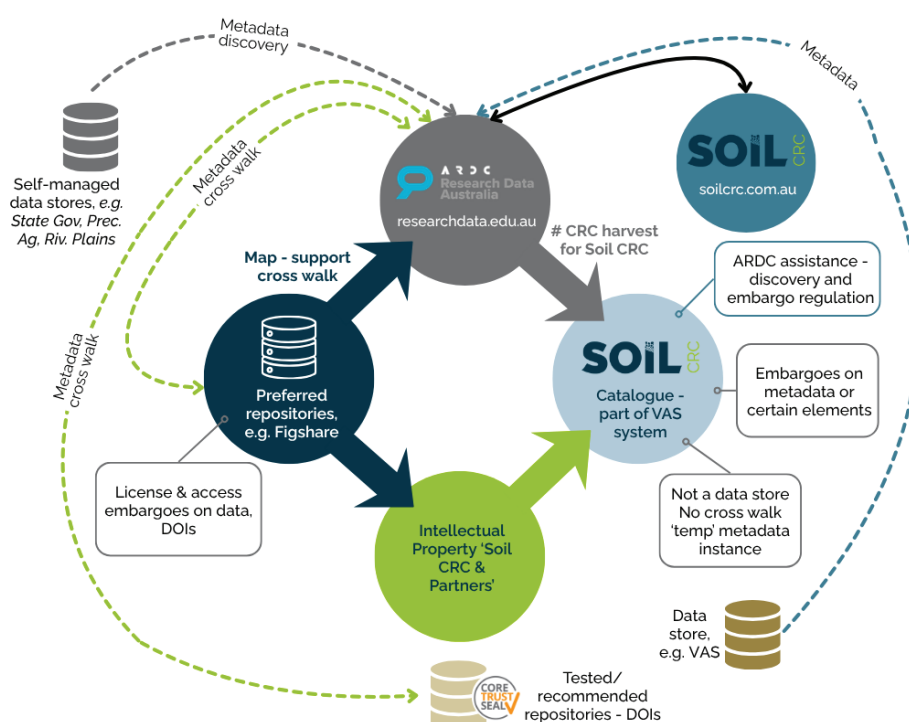


Figure 2. Soil data ecosystem model for the Soil CRC, showing the relationships and dependencies that a Virtual Data and Information Catalogue would need to support.



Project team

- Nathan Robinson, Jude Channon, Megan Wong, Peter Dahlhaus, Helen Thompson, Andrew MacLeod, Jennifer Corbett (Federation University)
- David Medyckyj-Scott, Alistair Ritchie (Landcare Research New Zealand, now Bioeconomy Science Institute)
- Simon Clarendon, John Friend (NSW Department of Primary Industries and Regional Development)
- Xiaodi Huang (Charles Sturt University)
- Stephen Cahoon, Warwick Gillespie, Simon Edwards (University of Tasmania)

RESEARCH PUBLICATION

Robinson, N., Channon, J., Clarendon, S. & Medyckyj-Scott, D. (2026). 'Managing data in collaborative research starts with the researcher: A Soil CRC case study', *Accountability in Research*, 33(3). <https://doi.org/10.1080/08989621.2025.2525878>

PROJECT WEBINAR

Watch the Soil CRC's research data management webinar for an overview of the data management lifecycle, data storage, metadata, and data sharing and publishing. <https://soilcrc.com.au/resources/soil-crc-research-data-management-overview-webinar-2023/>



This fact sheet and related publications are available on the Soil CRC Knowledge Hub (soilcrc.com.au/resources).

The CRC for High Performance Soils (Soil CRC) brings together scientists, industry and farmers to find practical solutions for Australia's underperforming soils. Our aim is to enable farmers to increase their productivity and profitability by providing them with knowledge and tools to improve the performance of their soils. The Soil CRC is the largest collaborative soil research effort in Australia's history, with funding until 2027. We have attracted more than \$167 million in cash and in-kind resources over 10 years from our 39 participants and the Australian Government.