

# Developing an Effective Research Data Culture

Dr Steve Quenette<sup>1</sup>, Ai-Lin Soo<sup>1</sup>, Dr Andrew Janke<sup>2</sup>, Prof David Abramson<sup>3</sup>, Luc Betbeder-Matibet<sup>4</sup>, Dr Rhys Francis<sup>5</sup>, Dr Stephen Giugni<sup>6</sup>  
1.Monash University, 2.University of Sydney, 3.University of Queensland, 4.University of New South Wales, 5.eResearch Futures P/L, Australia, 6.The University of Melbourne

Research intensive universities in Australia have been working together to articulate the nature of a sustainable and effective research data culture, under the banner of the Research Data Culture Conversation (RDCC).

In the RDCC process, the complexities of university hosted research data have been canvassed in a conversation between those building best practice solutions today, those who look to adopt best practice solutions and those that can facilitate changes in research practice.

The goal is a movement towards metadata-driven and machine-actioned management practices that capture the context of data sufficiently well to support most life-cycle decisions to be made about those data. To this end, a Smart Decision Making (RDMP-2.0 based) process is envisaged, whereby research data management would transition from the current fixed plan based (waterfall) approach to a continuous improvement-based approach.

RDCC Website



## Background

Over the last two years, the RDCC has held regional and national consultations on the basis that endless data volume growth implies an impractical growth in budget support:

- A major Australian University measured data growth of 75% per year, every year, over the last 15 years; and
- There are observations of data growth rates, such as in genomics and imaging, which can be much higher.

In addition, key properties of the situation contributing to the need for a sustainable institutional approach to data management are as follows:

- A. Institutions are increasingly coordinating and investing in large-scale infrastructure to support the growth in the creation and use of research data and to ensure policy compliance is achieved on retention, access and disposal requirements.
- B. The efficient use of such infrastructure depends on information and metadata not readily and systematically available, and that increasing the availability of such metadata depends on the appropriate involvement of researchers.
- C. Research data life cycles span areas of interest and activities traditionally owned by multiple research support pillars, including the library, records, archive and IT functions of universities, and more recently eResearch functions.
- D. Opportunities to gain efficiencies, economies of scale and quality improvements, all require coordinated action by these pillars in relation to each researcher's data and research data overall.
- E. Improved alignment is also needed between institutions, research funders and other third parties, such as those created by national infrastructure investment in data.

## Goals and Principles

Goals related to developing an enhanced research data culture include that data support practices do enable:

- The right people to have access to the right data for the right amount of time acknowledging changes of stewardship and achieving long term preservation.
- Systems to be operated that do reduce the human effort required in curation through reduction of per decision costs in data and data life cycle decision making.
- A cultural change amongst researchers that acknowledges and articulates the kind of data that have end-of-life events that can be implemented by institutions.
- The emergence of a resource aware practice where data naturally migrates towards lower cost but also increasingly compliant retention solutions.

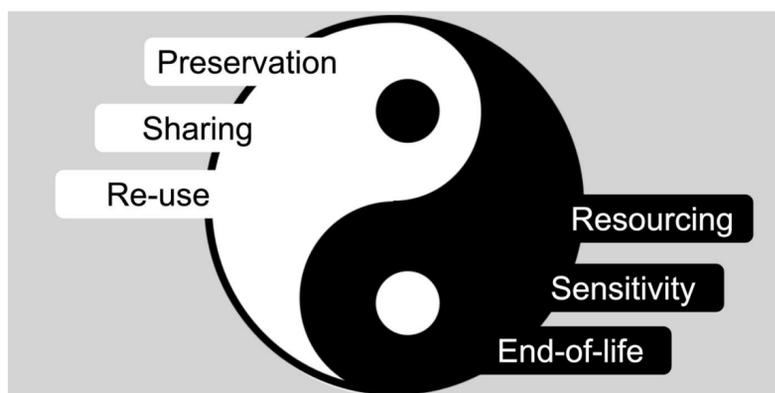
The RDCC process also identified key supporting principles in achieving these goals. They are the need to be:

- Research-domain focused as research domains have different data lifecycles, retention periods and metadata.
- Researcher centric; assisting and motivating researcher contribution and increasing ease of attribution and recognition while providing low barriers to entry.
- Efficient and adding to institutional value, including by facilitating reporting of cost effectiveness, data utilisation and citation.
- Implementable across and within institutions acknowledging that people, policy and process (the P-dimensions) should be considered before tools, technology and transformation (the T-dimensions)

## Process and Observations

Over 105 participants across 34 institutions and organisations including Australia's largest research intensive universities collaborated on articulating the challenges of research data management as well as current examples of good practice.

A simplified summary was developed during the consultations, in which the main institutional data challenges are seen to have both Yin and Yang dimensions.



This image depicts the idea that data management goals and practice need to encompass Preservation, Sharing and Re-use as well as Resourcing, Sensitivity and End-of-Life, if research institutions and research communities are to more effectively manage research data. Importantly, national and international discussions around research data need to devote energy to strengthen our understanding in all six dimensions.

In addition to this observation, the RDCC process also noted a number of key findings.

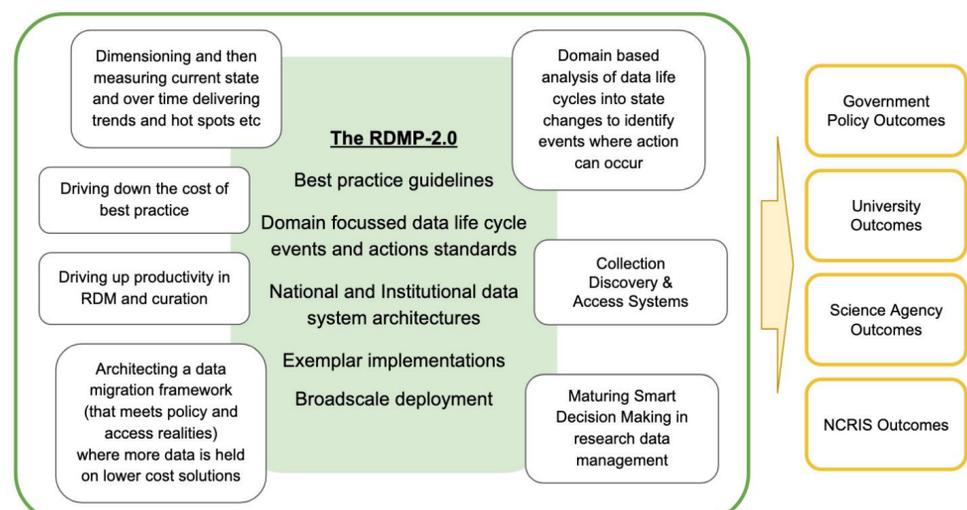
- Data is a new asset class requiring new investment.
- The ever increasing expansion of data holdings is driven by research opportunities and the ever more powerful observational capabilities technology is producing.
- Data growth in research is now so rapid, if continued unabated forever, it threatens an impossible to realise budget shift across a university - something about data needs to change.
- Research Data Management Plans should inform decision making (and in current practice don't) and must deliver the productivity gains needed as data grows.
- Institutions, which are a primary location where data ambitions and obligations meet budgets, need nationally-coherent discipline-sensitive responses to be codified (as a national conversation, the RDCC process limited itself to 'nationally-coherent').

## Next Steps

The RDCC process identified three key themes for future implementation that would support the broad ambitions of the research data culture consultations.

1. **Research data practice:** Improvement is dependent on the implementation of systems/processes in institutions and between institutions; including dimensioning and measuring the current state, better information capture around data and data migration into services that better meet policy, access and cost realities over time.
2. **Research data purpose:** Improvement is dependent on working with researchers to form community agreement on the events that cause data to have a life cycle and the articulation of the normal practice to apply when data undergoes such events.
3. **Research data investment:** Improvement is dependent on a better understanding of the value of research data, clarifying the goals of research data management with respect to the value and cost of data, and cross-informing and connecting 1 and 2.

The two reports of the RDCC (see RDCC website) provide more details of the proposed way forward, scoping out the need for a 'Research Data Management Plan 2.0'.



In response to the RDCC process, the Australian Research Data Commons (funded under Australia's National Collaborative Research Infrastructure Strategy) has established a program of work, see this link on the ARDC website ([Institutional Underpinnings](#)) for details. This new program, working through 2021 and 2022, aims to develop a joint framework between Australian Universities addressing the issues set out here.