Institutional Role in an Effective Research Data Culture

Rhys Francis[1], Steve Quenette[2]

[1]eResearch Futures P/L, Australia, rhys.francis@icloud.com
[2]Steve Quenette, Monash University, Australia, steve.quenette@monash.edu

Description

Prefer a 60 or 80 Minute Session

Outcomes of the activity set out below are expected to be reported as part of the ARDC Data Summit. This session will socialise the activity with the attendees of the conference, and invite a broader participation in its continuation into 2020.

A workshop drawn from archive, research management, eResearch , ITS, library and records functions from the Australian National University, Monash University, the University of Melbourne, the University of New South Wales, the University of Queensland and the University of Sydney met in January this year to explore their joint interest in the following motivation.

In response to over a decade of exponential growth in demand, institutions understand that they will increasingly coordinate large-scale infrastructure to support research data. The efficient use of such infrastructure depends on the availability of appropriate information and metadata, and on the recognition of constraints, that depend on the involvement of researchers. The solutions will need to:

- Deliver cost effective management and curation processes that support a concurrent set of societal, institutional and researcher based data goals;
- Focus curatorial investment towards the data prioritised by institutions to enhance outcomes, and the data prioritised by researchers to enhance their research impact;
- Hold a larger share of the overall data in less expensive services; and
- Contain the total cost, by containing the total amount of data stored and curated.

The emerging concept concerns the redevelopment of Research Data Management Plans (aka RDMP2.0) to better support decision making on the different trajectories that data may take during its lifetime.

During 2019, two further workshops are being planned around the following topics;

Topic 1	What pathways can data traverse during its lifetime that need tailored approaches and what are the key decisions that need to be made often and routinely in supporting each pathway?
	
	How do these pathways intersect with the Archives, DVCR, eResearch, ITS, Library and Records functions of the University?
	
	Pathways include:
	    ● working & archive (ie no decision to move down a pathway)
	    ● open pathway
	    ● disposal pathway
	    ● commercial pathway
	    ● sensitivity / data protection pathway
	    ● integrity pathway

Topic 2	For each pathway, how do we imagine we can achieve a situation where: 1. Researchers assist with the decision making as part of normal research practice; 2. Data Management Plans assist automate decision making; and 3. For which appropriate Data Management Plans could be more directly generated?

The ARDC has expressed an interest in engaging the discussion as part of a broader consideration of the role of universities in a data commons. The BoF will explore how to widen participation in activities expected to be undertaken in 2020.

Summary Of Research Data Culture - National Discussion - Workshop Held In January 2019

The discussion was kicked off by two observations explored in a BoF at eResearch Australasia 2018:

A. Due to the continuing exponential growth in data creation and use, institutions will increasingly coordinate large-scale infrastructure to support research data; and

B. The efficient use of such infrastructure depends on the availability of appropriate information and metadata, and on the recognition of constraints, that depend on the involvement of researchers.

This meeting and its breadth of attendance was triggered in response to the further observations that:

C. Research data lifecycles involve activities traditionally owned by library, records, archive and IT and more recently eResearch activities; and

D. Efficiencies, scalability, quality improvements and opportunities for strategic prioritisation require a coordination of a researcher's data through these pillars.

Key Discussion Outputs

- **Research Data Management involves a Focus on the P-Dimension (People - Policy - Process) as much as, or possibly more than, the T-Dimension (Technology - Tools - Transformation)**

- **Research Data Management is a Multi-Group, Multi-Responsibility, Multi-Support, activity within the university:** Includes collaboration from eResearch, IT, Library, Records Management and Archives from the most research intensive institutions. There is a need for encouraging the leadership within these areas and enabling the leverage between these areas (this is a horizontal activity).

- **Governance first / best practice:** The purpose is to accelerate the adoption of the change in practice across many universities, as led by those who are facing the problems at scale.

- **Researcher centric:**
    - Has low barriers / no barriers to entry.
    - Has low / no effort to perform.
    - Understands how to motivate the researcher to contribute, on research terms.
    - Is tolerant and enabling to those researchers doing something new / not well understood.
    - Is efficient and enabling to the majority of researchers.
    - Is actively and iteratively tested on real research use cases.
    - Provides potential for individual curation.

- **Automate where possible:**
    - Focus the use of human capital (to support research / domain specific needs).
    - Reduce / limit the amount of orthogonal human capital to achieve the goals.

- **Plan for capture and access:**
    - Improve the capacity to capture from the outset (instead of at end of life).
    - Request retention parameters from the outset and automate capture of metadata.
    - Provision access to the right people for the right amount of time, addressing changes of stewardship and long term preservation.

Concept Arising

It is envisioned that an outcome may be a best practices reference model for implementing Research Data Management Plans - 2.0 (RDMP-2.0). Specifically the (Web-)2.0 denotes a shift to an ecosystem that is researcher centric, just as Facebook is yourself centric, and where, the economics isn't consumer pays but rather based around those investing in the data. RDMP-2.0 would be automatically generated as a result of research being done in the university environment.