ARDC Institutional Underpinnings
Element: Active Research Data Management

30/03/2022

CONTENTS

EXECUTIVE SUMMARY 1
DESCRIPTION OF THE ELEMENT COMPONENT 2
DIFFERENCES IN APPROACH AND NEED 3
RECOMMENDATIONS AND ADVICE 4
SETTING EXPECTATIONS 8
WORKING GROUP ACKNOWLEDGEMENTS 8

EXECUTIVE SUMMARY

Institutional Underpinnings is part of the ARDC’s National Data Assets Initiative. In this program, 25 Australian universities are collaboratively developing a national Institutional Research Data Management (RDM) Framework. This Framework is intended to inform institutions’ design of policy, procedures, infrastructure and services, and improve coordination of RDM within and between institutions. This output describes the initial findings of the Active RDM Element of the Framework, providing institutions with guidance to ensure that research practice is efficient and impactful, and that research data is managed according to requirements such as those outlined in the in The Australian Code for the Responsible Conduct of Research. This output outlines some of the infrastructure considerations required to enable effective RDM and highlights the importance of user focused design in ensuring
adoption by researchers. Integration of research data management platforms with existing research administration systems is a recommendation provided. The importance of soft infrastructure elements such as governance and communications is also discussed. Recommendations for institutions and Calls to action are highlighted throughout the Element. Calls to action specifically identify the need for future collective action from institutions and the community. This initial Active RDM Element will be further developed through additional institutional consultation and will be complemented by activities to validate and test the outputs described within.

**DESCRIPTION OF THE ELEMENT**

Providing infrastructure, systems and services that facilitate good active research data management is an important responsibility of the institution. The Australian Code for the Responsible Conduct of Research\(^1\) (The Code) states that:

> Research data controlled by the institution and/or its researchers should be stored in facilities provided by or approved by the institution. These facilities, including information technology, must comply with privacy requirements and other relevant laws, regulations and guidelines, and research discipline-specific practices and standards related to safe and secure storage of data and information.

Although The Code refers specifically to storage here, active data management is bigger than just where data is stored. Active data management refers to the management of research data at any stage during the life of a research project, including selection of what to collect/acquire, collection/acquisition, storage, analysis, visualisation, and collaboration. Active data management ends when the data is either disposed of or moved to long-term storage after project reporting is complete, and does not include sharing data for re-use after the life of the original project.

When providing active data management infrastructure, institutions have a number of drivers that can shape decision-making. Institutions need to ensure that the services they provide can be maintained and supported in a cost-effective and sustainable manner. They also require a good understanding of the quantity and value of data that they hold and sufficient information to make decisions about its retention and disposal, as well as sufficient oversight over that data to ensure compliance with legal and regulatory requirements.

While it is the institution’s responsibility to provide (or endorse) the infrastructure for good active data management, it is generally the researcher who will be responsible for actually managing the data.

---

Researchers have an obligation to manage their data appropriately according to their responsibilities under The Code and as well as other regulatory, legislative and ethical requirements.

Researchers’ decisions about their use of active data management services are driven by many factors, including the need to complete their research efficiently, their knowledge and understanding of the options available to them, and their ability to use those options effectively and receive support when they have difficulty doing so.

A common active data management challenge for institutions is “shadow IT” - the use of non-approved services to capture, store and manage data. The use of shadow IT can open both institutions and researchers to risk, and makes it difficult for institutions to keep track of the data that they are responsible for. Researchers may turn to shadow IT when they are not provided with the services that they need to manage their data, do not understand how to access these services effectively, or find them overly burdensome. These difficulties can be worsened if the institution does not clearly articulate how research data should be managed.

However, even when researchers use institution-supported systems for active data management, institutions can still have difficulty maintaining oversight over their data holdings, particularly where institution-supported systems have originally been provided for corporate or teaching and learning purposes rather than with the research use case in mind. A lack of governance and oversight by the institution can result in competing advice from different areas of the institution, multiple hurdles with the same information having to be provided to various areas of the institution and in the worst cases, a shift of responsibility to make all data governance decisions to the researcher.

DIFFERENCES IN APPROACH AND NEED

When it comes to providing infrastructure for managing active research data, there is no ‘one size fits all’ solution. Before deciding where and how to store and otherwise manage research data, institutions will need to understand the nature of the data their researchers are working with in terms of data types, size, and sensitivity.

Depending on the type of research being conducted at an institution, the profile of active data management needs may be quite different. Factors influencing this might include:

- Very large datasets
- Compute-intensive research
- Research requiring extensive pre-processing or specialised (inc. proprietary) software to handle data
- Data that requires particular technical or procedural protections due to its sensitivity
- Physical (non-digital) data assets such as biospecimens, signed consent forms, soil samples
The profile of researchers themselves will also affect needs. For instance, institutions with large numbers of adjunct positions or higher degree research students will need to consider how to manage research which is conducted on devices that are not owned by the institution.

Institution size and resourcing can influence the selection of active data management infrastructure. It is important to note that size is distinct from maturity. An institution that deals with a smaller volume of research data may not need the level of automation that is required to manage much larger volumes of data, and so may develop quite different systems and infrastructure as compared to a larger institution. Resourcing will also play a role in decision-making, as institutions must ensure that they can sustainably support the infrastructure that they put in place within their budgets.

Finally, institutions differ in their internal structure and history of research data management solutions. These factors will lead to researchers and service providers having existing cultures, relationships, expectations and workflows that must be taken into account when making changes to how data is managed. A solution that works very well for one institution may therefore be very difficult to implement at another.

RECOMMENDATIONS AND ADVICE

Properties of Successful Approaches

Recommendation 1: Clear Responsibility within the Institution for Decision Making for (governance of) active research data management

Research Data Governance

Effective active research data management relies on a clear understanding of who has the responsibility to make decisions about data and its management, as well as the implementation of systems to record and facilitate the upholding of those decisions. This spans from decisions about high-level data-related policy to decisions about access to particular datasets. If it is not clear who “owns” important functions, they may fall through the gaps.

Poor data governance can affect the ability to provide effective data management services. For example, difficulties in integrating systems for active data management are worsened when it is not clear who has the overall responsibility for developing those integrations and ensuring that business processes are aligned.

Poor data governance can also affect researchers’ interactions with institution-provided data management systems. Researchers may receive conflicting advice from different parts of the institution, or may be unable to find anyone who is able to advise them. If researchers’ trust in the ability of the
institution to manage their data appropriately is damaged, then they may be further incentivized to turn to shadow IT.

One step towards improved research data governance is to map out the decisions and responsibilities related to research data management - for an example, see the Monash Health Data Governance Framework².

**Recommendation 2:** Investment in research data infrastructure should be driven by the three principles of researcher co-design, security by design and (where applicable) privacy by design.

**User-Focused Design**

For an institution’s active data management solution to be effective, it must be used by researchers. It is therefore vital that researchers’ needs are understood. Researchers are likely to become frustrated with a system which does not have the functionality they require, is cumbersome to use, or appears to create unnecessary administrative load. A system which meets the technical and compliance requirements of the institution but does not consider user experience is unlikely to be successful, because researchers will be incentivised to seek shadow IT solutions that better meet their needs.

Researchers must therefore be involved at all stages of the planning process. Local service providers need to understand the profile of needs at their institution, and identify where particular systems might place unexpected burdens on research users. It is also important to recognise where needs vary between different research domains.

**Recommendation 3:** Researchers must be involved at all stages of the Active Data Management solution planning process

Ideally, institution-provided (or institution-approved) systems should make research easier, so that they are more attractive to researchers. One way that this can occur is through “compliance by design”, so that researchers are less burdened by the need to track and meet their compliance requirements.

See attachment [“Case Study - Developing and Implementing Monash Secure eResearch Platform”](https://doi.org/10.26180/5f3335c306a3b) for an example of the co-design process.

**Clarity of Communication**

To be effective, an institution’s active data management systems must not just be used by researchers, but be used correctly. It is therefore important that researchers understand what is expected of them when using these systems.

---

² https://doi.org/10.26180/5f3335c306a3b
It is important to clearly communicate to researchers (and other key stakeholders) what can and cannot be done with a given platform or service, so that obligations around security, privacy etc. can be met. Providing information about why those recommendations/restrictions are in place, so that researchers understand them better is also recommended.

**Recommendation 4:** Clearly outline the appropriate use cases and terms for each platform or service, and disseminate information about endorsed platforms or services through multiple channels.

It is important to understand where researchers seek information about available services within the institution. Researchers may be sceptical of central service providers, and are likely to ask for advice within their local area. Ensuring that information is disseminated through multiple channels will help to improve reach.

Different stakeholders may use the same terms quite differently, which can create problems for communicating about data management requirements. For example, a service provider may have redundancy in a system that allows retrieval of data in case of a catastrophic loss. However, if a researcher is told that the data is “backed up,” they may interpret this to mean that they can retrieve their data if they accidentally delete it. This could lead users to unintentionally misuse services or select inappropriate services for their needs. We recommend that terms are clearly defined during stakeholder consultation and when presenting information about the correct use of institution systems.

**Integration for a Seamless User Experience**

To provide the best user experience, the active data management system should be viewed as a whole, rather than as a set of disparate services. Researchers will have to interact with the system at many different points, and a common complaint is that they are asked to provide the same information over and over again, creating an unnecessary administrative burden. Where possible, information should be passed between parts of the system to reduce double-entry and to allow the automation of processes.

As well as the technical challenge of creating integration between platforms, there are associated procedural challenges. A well-integrated system requires the identification of a set of key metadata about data that supports good management and automation, determination of the points where this metadata will be collected, and a common source of truth for this metadata. This can be challenging when different parts of the system are managed by different business units within the institution with different goals and drivers. The metadata required for retention and disposal decisions should also be considered in these discussions.

**Ongoing Investment is Essential**

Active data management systems can’t be “set and forget” - they require active development, support and investment. Any IT infrastructure will require ongoing maintenance. User needs will shift over time.
as research practices advance, and may change dramatically in a short time if there is a change of strategy that alters the research profile of the institution. The regulatory environment is not static, and there are also constant advancements in technology and the associated best practice. Commercial providers can also make changes to agreements and technical specifications that can require a quick response to maintain the effectiveness and compliance of active data management systems. It is therefore very important that institutions recognise the need to factor ongoing investment into their budgeting around active data management.

Consideration of Institutional and Individual Legacy

Any active data management solution will involve not just technology but also people and processes. The legacy of past research data management solutions will affect each of these, and must be considered carefully. Institutional factors go beyond just the migration of data and metadata into the new system - there may also be existing business processes that need to be adjusted.

Similarly, researchers will have existing workflows based on the systems that they are used to (often using shadow IT) that may be seriously impacted by a shift to a new active data management system. It should also be remembered that for some researchers, their unapproved solution for research data management represents a considerable personal investment of time and effort, and there is an emotional cost to moving away from this solution (particularly if the new solution does not have the ease of use or customisation of the old). Careful change management is therefore important.

It is also important to consider the incentive structures operating for researchers. Under current incentive structures, time-poor researchers may invest less in research data management activities, which offer less reward than activities such as publication.

Platform Selection

The specific platforms that are selected to underpin active data management will vary according to institutional needs. However, we outline some key considerations below:

- Disaster recovery
- Ability to support data discovery
- Interface with instruments that collect data
- Ease of local collaboration
- Ease of external collaboration (particularly the management of authorisation and authentication)
- Ease of use with required research tools, analysis environments, or analysis workflows
- Ability to maintain central control/oversight
- Location of storage
- Ability to manage access
Call to action 1: Institutions are encouraged to share with one another their active data management platforms in use, including their benefits and limitations

Complexity of Requirements
Designing active data management systems requires an understanding of the types of data that must be managed and the requirements for those data types. We recommend the creation or adoption of a clear research data classification scheme to help with this task. Platforms can then be assessed against these requirements, and appropriate procedures developed to ensure that platforms are used appropriately.

Institutions often begin by designing an active data management solution for basic use cases, and then treating other data types as an exception. We recommend that the management of data with complex requirements (e.g. sensitive data, big data) be considered from the beginning, as it is almost inevitable that the solution will need to handle these data types.

Call to action 2: Institutions are encouraged to share with one another examples of the standards in use for their data classification

SETTING EXPECTATIONS
It is unrealistic to expect an institution to provide or support every platform that might be requested by its researchers. There is a careful balance to be struck between supporting diversity of research and maintaining a sustainable service. Institutions should instead aim to understand what the majority of use cases require, and to provide or approve a minimum set of platforms and systems that meet these needs. Research support capability should be available to assist researchers who have needs that go beyond the standard offering.

WORKING GROUP ACKNOWLEDGEMENTS

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gnana Bharathy (Chair)</td>
<td>ARDC</td>
</tr>
<tr>
<td>Nichola Burton (Editorial Committee)</td>
<td>ARDC</td>
</tr>
<tr>
<td>Matthew Bellgard</td>
<td>Queensland University of Technology</td>
</tr>
<tr>
<td>Dianne Brown</td>
<td>Monash University</td>
</tr>
<tr>
<td>Jac Charlesworth</td>
<td>University of Tasmania</td>
</tr>
<tr>
<td>Racheal Chesters</td>
<td>University of Southern Queensland</td>
</tr>
<tr>
<td>Jacky Cho</td>
<td>University of New South Wales</td>
</tr>
<tr>
<td>Name</td>
<td>Institution</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Juan Cooper</td>
<td>Western Sydney University</td>
</tr>
<tr>
<td>Kate Croker</td>
<td>University of Western Australia</td>
</tr>
<tr>
<td>Gerry Devine</td>
<td>Macquarie University</td>
</tr>
<tr>
<td>Kate Elder</td>
<td>Charles Darwin University</td>
</tr>
<tr>
<td>Cameron Fong</td>
<td>University of Sydney</td>
</tr>
<tr>
<td>Iftikhar Hayat</td>
<td>Charles Darwin University</td>
</tr>
<tr>
<td>Kyle Hemming</td>
<td>University of Canberra</td>
</tr>
<tr>
<td>Michael Lynch</td>
<td>University of Technology Sydney</td>
</tr>
<tr>
<td>Helena Lynn</td>
<td>University of Melbourne</td>
</tr>
<tr>
<td>Ryan McConville</td>
<td>University of Sydney</td>
</tr>
<tr>
<td>Kay Steel</td>
<td>Federation University</td>
</tr>
<tr>
<td>Jake Surman</td>
<td>University of New South Wales</td>
</tr>
<tr>
<td>Catherine Tan</td>
<td>Griffith University</td>
</tr>
</tbody>
</table>

The outputs of this working group were edited for public release by Frankie Stevens, Lyle Winton and Nichola Burton (ARDC)


Licence: Creative Commons Attribution 4.0 International