ANDS Guide

Data storage

Level: Awareness

Last updated: 9 December 2016

Web link: www.ands.org.au/guides/data-storage

This guide is for those who are interested in learning more about options for long term storage of high impact data for the benefit of the nation.

Why Data Storage matters

The choice of storage solutions for research data contributes to how shareable, reusable, discoverable and accessible the data remains over the long term:

- This involves integration with metadata maintenance to continue to provide access and curation
- Different choices of data storage have implications for metadata management and data access

Storage solutions discussed here are not mutually exclusive, but complementary. Each solution is designed to address different needs. In designing a data management strategy, institutions need to harness all the available types of storage. Focusing on just one (such as institutional data stores or repositories) foregoes the advantages of the others, and is unlikely to satisfy all the requirements of researchers themselves.

For example, institutional solutions are more sustainable than discipline based solutions, because they have concrete institutional commitments guaranteeing long-term storage. This solution satisfies the requirement for long-term access. However, discipline solutions are where researchers search for data by default. Because they provide a more discipline appropriate environment to do discovery, they are a better fit for the discovery requirement.

Data storage solutions should enable a researcher's data to be made part of the scholarly record: discoverable, sustainably managed and well described in to order facilitate its reuse (where appropriate).

Obligations and Expectations

Section 2, Management of Research Data and Primary Materials of the Australian Code for the Responsible Conduct of Research (currently under review), states:

"The central aim is that sufficient materials and data are retained to justify the outcomes of the research and to defend them if they are challenged. The potential value of the material for further research should also be considered, particularly where the research would be difficult or impossible to repeat."

In pursuit of this aim, two sections of the Code talk in particular about data storage.

- Section 2.2: "Institutions must provide facilities for the safe and secure storage of research data and for maintaining records of where research data are stored".
- Section 2.6: "Researchers must manage research data and primary materials in accordance with the policy of the institution [and] Retain research data, including electronic data, in a durable, indexed and retrievable form."

Types of Storage options/solutions

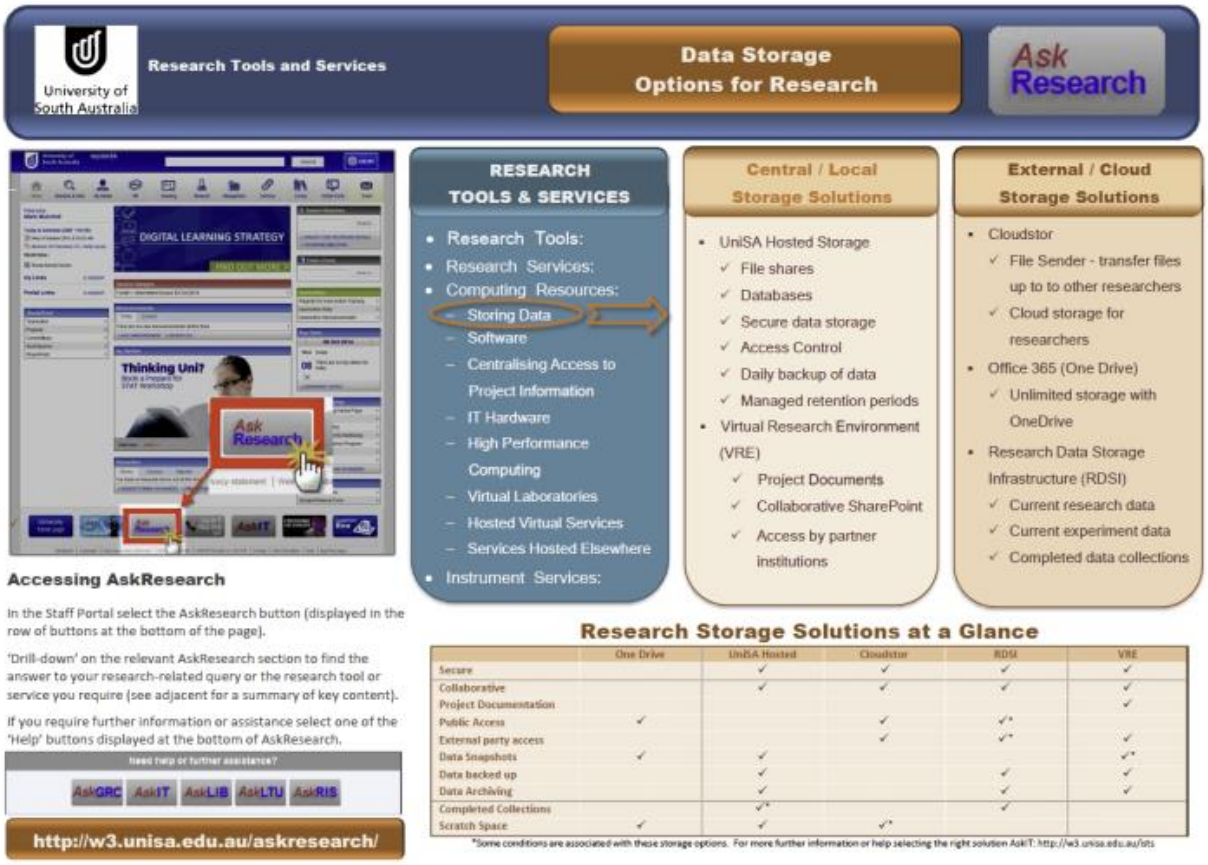
	Description	Example
Individual / Project Data Store	Local storage controlled by researchers and project teams	USB, hard drive on individual laptop, local drives (c:, group drives)
Institutional Repository	Through the ARROW program (2004-07), Australian universities have implemented an institutional repository. Although these repositories have been designed for document objects, e.g. publications, many of them are based on software that allows them to store a range of data objects	Monash University Research Repository contains both cyclone tracking data and ethno-musicology fieldwork recordings
Institutional Data Store	A number of Australian universities and large research institutions have specialized data stores, optimised for large numbers of large objects	Monash University Large Research Data Store (LaRDS), and the University of Melbourne offers flexible solutions
National Data Store Infrastructure	The RDS (2015-) focuses on supporting nine research domains including astronomy, climate and weather science, marine science,	NSW, Vic, Qld, SA, Tas, ACT and WA have storage for merit-

	minerals and exploration data	allocated and commercial storage
Cloud Store (Corporate Data Store Infrastructure)	Cloud storage is networked enterprise storage through virtualized pools of storage which are generally hosted by third parties	CloudStor (AARNET) combines storage and rapid file sending capabilities commercial cloud storage providers include: Dropbox, Amazon Simple Storage Service (Amazon S3), Google Drive, FigShare
Discipline Repository	A number of disciplines have well-established locations for storing and sharing data, managed typically by a consortium of institutional members. This storing and sharing often occurs in combination with the publication process	A list of 1,200 data repositories including many discipline repositories is available at re3data.org Australian Data Archive (for social science data) Aboriginal and Torres Strait Islander Data Archive

Which solution is right for my data?

No matter which storage solution is chosen the creation of rich metadata to accompany the underlying data ensures that the data can be easily found and reused as required, both by the researcher and their team, as well as their institution and the wider research community.

The University of South Australia has developed a one page storage options guide for their researchers to aid them in their data storage choices.



The various storage solutions can be compared against two main criteria:

1. The value of the data and its potential for reuse
2. The types of components which give value to data, such as its discoverability, curation, and whether the storage is reliable, large and sustainable

The only type of solution which does not satisfy the goals of reuse and sharing data are the individual and project level solutions. The reason they remain so widely used is their convenience, low cost and that they can be set up rapidly to answer some of the research needs during a research project.

To steer researchers away from such quick fix solutions, institutions need to encourage researchers to plan at the beginning of a project for how they will store data, and to budget for it. Such planning takes place in a data management plan, and is described in other ANDS Guides; Data management plans is one starting point.

Storage Solutions	Suitable for Working Data	Suitable for High Value, High Reuse Data	Weaknesses	Strengths
Individual / Project Data Store	Yes	No	May not meet access guidelines for publishers and funders, fragile sustainability, no DOIs	Cheap, fast, easy, convenient
Institutional Repository	No	Yes	May not be optimised for data	Reliable, well curated, sustainable, supports bibliometrics (e.g. DOIs)
Institutional Data Store	Yes	Yes	In some cases not supporting discovery	Reliable, sustainable
Cloud Store	Yes	Depends on sufficient data description	CloudStor: few disadvantages Commercial: May be subject to bandwidth and file security concerns, less control over the hardware where your data is stored	Automated backup, sharing and access to cheap compute to process stored data
National Data Store (RDS, etc)	Yes	Yes		Highly reliable, supports large datasets

Discipline Repository	No	Yes	Quality of long term funding support is highly variable	Supports researcher discovery

Interaction between storage solutions and with metadata stores

Discipline stores are often the repository of record for a discipline, and can follow several models. Two examples of models of discipline repositories interacting with other repositories are:

- International Virtual Observatory Alliance is a federated data store: data policies and metadata are managed centrally, but storing the data itself is shared among members of the consortium.
- Worldwide Protein Data Bank is a consortium whose members are themselves collaboratories, and is funded by a variety of sources, with a single archive mirrored between its members.

Research Data Services (RDS) Project (2015-) enables researchers to easily store, discover, access and share their data for better research outcomes. The RDS project is strategically aligning priorities with other federally funded e-infrastructure investments in:

- Computational infrastructure in supercomputing through National Computational Infrastructure (NCI) and Pawsey Supercomputing Centre
- Cloud, Virtual Labs, tools and services through National eResearch Collaboration Tools and Resources project (NeCTAR)
- Data management through Australian National Data Service (ANDS)

More information

- Metadata ANDS Guide
- Storing metadata
- Data management plans

Feedback?

We welcome your feedback on this guide. Please email contact@ands.org.au with any comments or questions.

About ANDS

The Australian National Data Service (ANDS) makes Australia's research data assets more valuable for researchers, research institutions and the nation.

ANDS is a partnership led by Monash University in collaboration with the Australian National University (ANU) and the Commonwealth Scientific and Industrial Research Organisation (CSIRO). It is funded by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS).

This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to reuse and republish this work, or any part of it, with attribution to the Australian National Data Service (ANDS).