

# CSU Framework for Research Data Storage

---

Enterprise Architecture  
Division of Information Technology (DIT)  
September 2015 Version 0.1

## Contents

Purpose .....	3
Scope.....	3
Audience .....	3
Related Documents and Resources .....	3
Research Storage Principles.....	4
Research Data Services .....	4
Primary Data Storage.....	4
CSU Storage - Active Research.....	4
CSU Storage - Finalised Research.....	4
AARNET Cloudstor.....	4
Intersect Storage.....	5
Compute and Storage Services .....	5
NeCTAR .....	5
Data Sharing and File Transfer Services.....	5
AARNET Cloudstor.....	5
Commercial Cloud Solutions .....	5
Storage Decision Matrix.....	6
Digital Data Storage options at CSU.....	6

## Purpose

This document describes the current framework for the storage of CSU research data to help ensure the University, is well positioned to support Researchers in improving Research Data Management (RDM) practices and the implementation of the RDM Policy. It provides information relevant to decisions on appropriate storage solutions throughout the research data lifecycle.

This document describes the current storage environment provided by CSU as well as services gained through subscription to partner organisations and commercial service providers and is bound by the limitations of those technologies and CSU policy. The technology and contractual arrangements negotiated by DIT is continuously evolving to improve the university's storage capabilities which will substantially change this framework over time.

## Scope

This document relates to the storage of data that is:

- Generated by CSU Faculty, Centre and Institute research projects.
- Sourced externally for use as part of CSU research projects
- Analysed, modelled and stored as part of a published research output
- Retained to meet records management requirements and CSU Research Data Management Policy

It does not relate to:

- Data that is not part of a CSU research activity and therefore not under the purview of the CSU Research Data Management Policy
- The publishing, description and citation of research data

## Audience

The audience is the CSU Research Community

## Related Documents and Resources

The following documentation and resources are related to the Framework and provide additional context.

Data Management Policy	<a href="https://policy.csu.edu.au/view.current.php?id=00328">https://policy.csu.edu.au/view.current.php?id=00328</a>
Research Authorship Policy	<a href="https://policy.csu.edu.au/view.current.php?id=00272">https://policy.csu.edu.au/view.current.php?id=00272</a>
Code of Conduct for Research	<a href="https://policy.csu.edu.au/view.current.php?id=00256">https://policy.csu.edu.au/view.current.php?id=00256</a>
Research Policy Page	<a href="http://www.csu.edu.au/research/policy">http://www.csu.edu.au/research/policy</a>
Intellectual Property Policy	<a href="https://policy.csu.edu.au/view.current.php?id=00155">https://policy.csu.edu.au/view.current.php?id=00155</a>
Research Office	<a href="http://www.csu.edu.au/research">http://www.csu.edu.au/research</a>
Research Support Services	<a href="http://www.csu.edu.au/research-csu/research-support-services">http://www.csu.edu.au/research-csu/research-support-services</a>
Research Storage Page	<a href="http://www.csu.edu.au/research-csu/research-support-services/repository-and-data-storage">http://www.csu.edu.au/research-csu/research-support-services/repository-and-data-storage</a>
NeCTAR Research Cloud website	<a href="http://nectar.org.au/research-cloud">http://nectar.org.au/research-cloud</a>

## Research Storage Principles

Research storage solutions referred to in this framework:

- Are centrally supported by the Division of Information Technology
- Support the CSU Research Data Management Policy.
- Are located on approved infrastructure and technologies.
- Ensure appropriate security, retention and management mechanisms are available.
- Positively contribute to research efficiency.
- Foster collaboration, data sharing, re-use and citation where appropriate.
- Leverage CSU partnerships and sector-wide government services established to support research activities.

## Research Data Services

The following describes the services available to researchers for storage, processing and collaboration around research data and provides guidance around which services are the most appropriate depending on various parameters of the Research Activity.

Additional Information about storage options can be found at the [Research Storage](#) web site

### Primary Data Storage

*In the context of this document “Primary Data Storage” refers to enterprise class disk storage that is optimised for high availability, provides comprehensive data protection through backup mechanisms and facilitates secure access control and flexible permission schemes that can be controlled by the data owner. While data can be stored on other locations and on other devices for transfer, and collaboration purposes there should always be a copy kept on primary storage to ensure data is protected and well managed.*

### CSU Storage - Active Research

This storage is provided as part of the CSU S:\ drive folder structure under “S:\Research\Research Data” and has the same levels of service support as the P:\ and S:\ drives. Refer [Service Catalogue](#) for additional information.

The storage is centrally provided and supported by the Division of Information Technology (DIT) and provides a responsive, safe and secure location exclusively for research data. A Researcher is able to manage the space and access to the data can be configured to allow either private access or collaboration with other CSU staff and Higher Degree Research (HDR) Students. This data cannot be accessed by people that do not have a CSU login and access outside the CSU network is currently restricted.

### CSU Storage - Finalised Research

Currently this is the same storage location as described above. Finalised data sets however should be clearly separated from Active data sets so that when future long term retention solutions are implemented this data can be identified. The permissions on the folders should be changed appropriately so data cannot be changed if it is required for verification purposes etc.

### AARNET Cloudstor

This service offered by the Australian Academic and Research Network provides both storage and collaboration features that allow sharing of data with people that are affiliated with other member organisations of the Australian Access Federation. Researchers receive 100 Gb of data by default.

## Intersect Space (RDSI Node Storage)

The Research Data Storage Infrastructure (RDSI) was established by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS) to support research collaboration and support research data storage.

Intersect ([www.intersect.org](http://www.intersect.org)) has 12 NSW based university members including CSU and is the Sydney Primary Node<sup>1</sup> operator. CSU as an Intersect member have a storage allocation that can be utilised by CSU researchers.

The storage can be accessed via the Australian Access Federation (AAF) and access to specific data sets is managed by individual researchers. Intersect Space allows collaboration with external colleagues that do not have a CSU login.

## Compute and Storage Services

*Compute and Storage Services refer to solutions that allow research data sets to be attached to compute services used to conduct intensive computational analysis.*

*Note: This storage should not be used as a Primary Storage service.*

## NeCTAR

The NeCTAR Research Cloud allows researchers to create Virtual Servers through a self-service portal that can be used in conjunction with storage for analysing data. These servers can then be configured with research software tools. Researches can trial the service by creating up to two small servers for a 3 month period. If additional time or resources are required a submission must be made and resources allocated on the basis of merit.

Access to the Research Cloud is controlled through AAF login.

## Data Sharing and File Transfer Services

*Data Sharing and File Transfer Services refer to solutions that allow research data sets to be shared among collaborators to conduct analysis.*

*Note: This storage should not be used as the Primary Storage service.*

## AARNET Cloudstor

Cloudstor can be used to share files and data sets even if it is not used as the Primary storage location.

## Commercial Cloud Solutions

There are many solutions that allow the ability to share files in a simple and intuitive way. Many of these offer advanced feature sets for sharing data and come with free storage options. Most of these services have Web-based and desktop interfaces and can be freely downloaded and used.

These services should only be used to transfer and collaborate on data that is not sensitive or does not compromise Intellectual Property or privacy policies. Many of these services are hosted outside Australia and may not meet data sovereignty or records management requirements.

DIT continues to work with researchers to better understand their requirements and how commercial cloud solutions may provide a viable solution.

---

<sup>1</sup> Nodes are data centres with large data storage facilities that constitute the RDSI. The nodes are managed by eResearch support agencies.

## Storage Decision Matrix

### Digital Data Storage options at CSU

The matrix below indicates the appropriate storage location, and related services that should be used based on different aspects of the data.

Phase	Active				Finalised					
	Restricted		Open		Restricted		Open			
	External	Internal	External	Internal	External	Internal	External	Internal		
CSU Storage - Active Research		✓		✓					Primary Data Storage	
CSU Storage - Finalised Research						✓		✓		
AARNET Cloudstor <sup>2</sup>	✓		✓		✓		✓			
NECTAR	✓	✓	✓	✓						Compute & Storage
Commercial Cloud Solutions			✓				✓			Data Sharing
AARNET Cloudstor	✓		✓		✓		✓			

Primary Data Storage	The location where the primary copy of data set should be stored – parts or all of the data can be held in other locations for compute and sharing purposes
Compute and Storage	If data analysis requires computational resources not provided by a PC, virtual servers can be created and co-located with a copy of the data in close proximity
Data Sharing	Copies of data can be held on these services to facilitate collaboration and sharing of data

Data Aspects		
Phase	Active	Data sets that are changing frequently due to data collection or data analysis
	Finalised	Data sets that are finalised and static/immutable for verification of research outputs
Sensitivity	Restricted	Data that contains information that could compromise privacy, intellectual property or is subject to other confidentiality agreements
	Open	Data that does not contain sensitive information and is available for use by fellow researchers or for public consumption.
Sharing	External	Collaborating on data with colleagues that do not have a CSU login or do not have access to the CSU Research Storage Services
	Internal	Collaborating on data with colleagues that have a CSU login and have access to the CSU Research Storage Services

<sup>2</sup> Where possible data should be copied to CSU Research Storage