

# Monash Large Research Data Store (LaRDS)

## Overview

Neil Clarke – Director, ITS Research Support Services  
v1.0.3 rev 18-Aug-2009

### Overview

The Monash *Large Research Data Store* ([LaRDS](#)) is Monash's central '*petascale*' research data store. *LaRDS* provides thousands of terabytes (TB) of capacity for storage of Monash research data. A petabyte (PB) is 1,000 terabytes and a terabyte is 1,000,000 megabytes (MB). *LaRDS* is available for all Monash researchers, including HDR students, research collaborators at other institutions and research support staff, for storing research data of all types and all sizes, big and small.

*LaRDS* provides a wide range of interfaces suitable for different applications and modes of use – *LaRDS* is not just a single application, database or server. *LaRDS* can be used to store all types of research data: from raw experimental data; data downloaded directly from scientific instruments; backup of research data held on your c: drive; databases; simulations; collaborative works; cultural archives; audio, video and multimedia digital content; long-term preservation of archival material; through to published research results via the [ARROW](#) repository. A range of services is available for different applications, some being suitable for each client workstation environment:

- Windows; Mac; Linux

### LaRDS is available to all Monash staff and HDR students

Some LaRDS services are based on open internet standards, some use vendor proprietary methods. Some are end-user services, some are 'connectors' to other applications and systems. Some are restricted to within the Monash network for security reasons, some can traverse externally. Some are restricted to Monash users, some can be accessed by external collaborators.

### LaRDS is secure and reliable

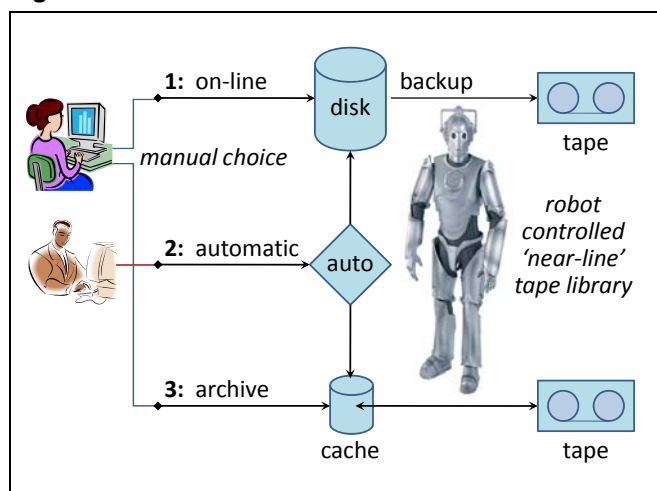
All data held in *LaRDS* is secure and reliably backed-up to tape across multiple geographically diverse data centres. Data in *LaRDS* is not automatically published to the world. Users have complete control of the level of access to the data: from data accessible only by the individual, through to data shared within a research team/workgroup (the most common case), or shared more broadly to a community of interest, or (in a subset of cases) made fully public.

LaRDS services are generally aimed at *workgroup* applications, i.e. not accessible by *unauthorized* people outside your research team. Services can be configured for *individuals* (i.e. the case of a workgroup of size 1). While some LaRDS services can be configured to provide public access, that is never the default.

### On-line and archival modes of use

LaRDS provides '*on-line*' disk storage for instantaneous retrieval of works in progress or frequently accessed reference materials, and low-cost '*near-line*' tape archive storage for less frequently accessed materials (Figure 1). All data is backed up to tape. All tape is held on-site in robot controlled tape libraries for rapid retrieval. All modes 'look like' disk to the end-user/application, even where the data is held on tape.

Figure 1: On-line and archive modes



LaRDS provides the following grades of service (Figure 1):

1. **On-line:** All data guaranteed to be held on disk, and backed up to tape.
2. **Automatic:** Reasonable probability that recently accessed data will still be on disk. System automatically moves less used data to tape only.
3. **Archive:** Low probability that data will still be in disk cache. All data held on tape. Retrieval typically takes less than a couple of minutes.
4. **Streaming:** First couple of minutes of media content held on disk, remainder held on tape only. Tape is retrieved in parallel with playing from disk.

Generally a user elects for either:

- **manual** control: user has explicit access to modes 1 and 3; or
- **automatic:** all data is held in an appropriately optimized mode for the intended use (e.g. modes 2 or 4).

Each of the categories of LaRDS services are introduced overleaf, and their main capabilities are summarized in Table 1.

### Monash Campus Grid

The [Monash Sun Grid](#) (MSG) and [Monash Green SPONGE](#) high performance computing (HPC) services use **LaRDS NFS** as their sole means of storing all user data. Thus, these HPC systems provide a Unix **login** style of access to LaRDS.

## LaRDS services – by category

### LaRDS Network Drive (LaND) services ('R:\' drive)

LaRDS provides a variety of basic Network Drive (LaND) services, as appropriate to your requirements and types of workstation (Windows, Mac or linux) – a LaND service 'looks' just like a local disk within your workstation, except that it is actually located in a central Monash data centre, data is backed up to tape, and the data can be shared between members of your research team / workgroup. Some LaND services also provide web-based access, for access from anywhere in the world (inc. internet café). Please refer to the separate [LaRDS LaND brochure](#) for further details of options suitable to you.

### LaRDS Data Management services

Creating order out of data chaos – LaRDS Data Management services assist researchers to organize, catalogue, search and retrieve research data sets, materials and documents, and to add comprehensive descriptive information ('*metadata*') about the data.

- [Mediaflux Desktop](#) is the only LaRDS interface that allows the addition of structured *metadata* (descriptive information and annotations to data collections). Mediaflux is an open standards (XML) based digital asset management system for storing audio, video, image, document or any other file types. Mediaflux Desktop is a web-based application accessible from any web browser (Firefox, IE or Safari). Metadata recorded in Mediaflux expedites publication of datasets via the [Monash ARROW institutional repository](#).
- [M-eNotes](#) is a MS-Word based digital asset management system, equally suitable for laboratory and non-laboratory data management alike, and for all text-oriented research data management situations, whenever the researcher wants a familiar MS-Word based user interface. Data can be digitally sealed for non-repudiation. All data are stored in their original formats, and all text-based information about the data is stored in open XML standard format, not proprietary MS formats.
- [Subversion \(SVN\)](#) is a popular version management system, primarily intended for computer software source code management, but also suitable for general document version management.
- SQL database platforms to support applications that require Oracle, MySQL, PostgreSQL or MS-SQL.

### LaRDS Data Mover and Replication services

LaRDS provides a full range of data mover services, including

- FTP, SFTP, GridFTP, SCP and SSH file transfer protocols,
- rsync data backup, and
- iRods (previously SRB) data replication services.

### LaRDS Collaboration services

While all the above LaRDS services allow data sharing between collaborators, LaRDS also provides some additional applications tailored for collaboration purposes, including:

- Sakai all-in-one collaboration environment
- Confluence and Twiki wikis
- WordPress blog

Each application stands-alone (no data sharing between each application), all store data in proprietary (e.g. SQL) databases and in 'unstructured' mode (i.e. no associated formal descriptive metadata schema).

**Table 1: LaRDS service categories comparison – current status (August 2009)**

service / category:	LaND [2]	Mediaflux Desktop	M-eNotes	SVN	SQL	Data movers	Sakai VRE	Confluence wiki
<i>service attributes:</i>								
Sharing btwn apps [1]	(some)	✓				✓		
Status	<b>prodn</b>	pilot	pilot	pilot	<b>prodn</b>	<b>prodn</b>	pilot	pilot
Web interface	(some)	✓			N/A	✓	✓	✓
On-line store	✓	✓	✓	✓	✓	✓	✓	✓
Near-line archive store	(some)	✓				✓		
Network drive	✓	via WebDav					via WebDav	
Open standards	(some)	✓	✓					
Structured metadata		✓						
ARROW connector		<b>soon</b>						
Version management	(informal)	(informal)	(informal)	✓	✓	(informal)	(informal)	✓
External users	(some)	✓	(manual)	(manual)	(manual)	(manual)	✓	(manual)
Federated authentication	(SSH-FS)	<b>soon</b>						<b>soon</b>
External access	(some)	✓	(via VPN)			✓	✓	✓

Notes:

[1] Some LaRDS services can access the same data as each other, others cannot.

[2] Please refer to the separate [LaRDS LaND brochure](#) for further details.