**LA TROBE AVAILABLE RESEARCH DATA STORAGE MATRIX**

**Document Control**

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Date** | **Nature of Amendment** | **Distribution List** |
| 1 | 1/09/2016 | Initial release |  |
| 2 | 6/12/2016 | Updated AARNet CloudStor information relating to very large datasets as group account with a default 1 TB allocation are possible. Changed the red cross to a green tick in the appropriate cell.Updated AARNet CloudStor information relating to live datasets. This was showing a green tick but in fact should be a green tick as there is no inherent reason not to store live data on this platform depending upon the individual researcher’s use-case or performance requirements.Updated AARNet CloudStor information relating to sensitive data as this was showing a red cross when it should have been showing a green tick to be consistent with the other cloud storage providers appended with a footnote (9).Updated **Caveats** for the **LTU H:, G:, P:, R: Drives** platform in the last table to reflect deprecated status of H: and R: drive resources. |  |
| 3 | 19/12/2016 | Updated a cell in row 3 of LTU OneDrive column to reflect LTU OndeDrive data is now stored in Australian data-centers thereby improving its suitability for sensitive data storage. Removed reference to H: drive throughout the document as the LTU on-premise storage offering is now deprecated and an active project to migrate user-data is underway. Added a section heading to the last table, “Additional Notes” to provide context for that information. |  |

**Purpose**

The information in this document is intended to provide members of the La Trobe University Research Community with guidance around the available research data storage options and their appropriateness for specific data-sets.

The recommended method for using the matrix is to work your way down the list of use-cases or requirements in the light-green column, and identify all the those that your research project will likely generate (as best as you can); then check under the corresponding cells to the right of each use-case/requirement for the storage options which meet them. By working your way through the list of use-cases and requirements and eliminating those storage options which do not meet them you can shortlist the possible storage options to those that would best suit your project.

Once you have a shortlist, use the detailed information below the matrix to gain a better understanding of the differences between the different options to again clarify which option best meets your requirements specific requirements.

Another option for quickly deciding which storage option to use based on the four key metrics of capacity, collaboration, performance and security is also included using a three-star rating system. Depending upon what is most important, you can quickly determine the options which best meet your key criteria and then review the additional details on the relevant option(s) to ensure it is fit for purpose.

**Available Research Storage Matrix**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **USE-CASE/REQUIREMENT** | **AARNet CloudStor** | **Intersect DeepSpace Archive Storage** | **Intersect SpaceShuttle** | **Intersect SpaceLab** | **Intersect Time10** | **LTU G:, P:, R: Drives(On Premise)** | **LTU OneDrive** |
| **1** | **Will the project generate very large1 datasets?** | ✓ | ✓ | ✓ | ✓ | ✓ | 🗶 | 🗶 |
| **2** | **Will the project generate live2 datasets?** | ✓ | 🗶 | ✓ | ✓ | ✓ | ✓ | ✓ |
| **3** | **Will the project generate sensitive3 data?** | ✓9 | ✓9 | ✓9 | ✓9 | ✓9 | ✓ | ✓9 |
| **4** | **Will the project datasets need to be shared with internal collaborators?4** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓8 | ✓ |
| **5** | **Will the project have external research collaborators?5** | ✓ | ✓ | ✓ | ✓ | ✓ | 🗶 | ✓ |
| **6** | **Will you require remote access?6** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| **7** | **Will the project require High Performance Computing (HPC) capabilities?7** | 🗶 | 🗶 | ✓ | ✓ | ✓ | 🗶 | 🗶 |
| **8** | **Will the project involve datasets generated by instruments/sensors?** | 🗶 | 🗶 | ✓ | ✓ | ✓ | ✓ | 🗶 |
| **9** | **Data Redundancy** | Data stored in multiple locations. | Data stored in multiple locations | Data stored in multiple locations | Data stored in multiple locations | Data stored in multiple locations | Replicated to DR System | Data stored in multiple locations. |
| **10** | **Data Recoverability** | Restore from previous versions | No backups provided | Restore from previous versions available on request at extra cost. | Restore from previous versions available on request at extra cost. | No backups provided | Daily and monthly snapshots | No backups provided |
| **11** | **Cost** | Free | Charged11 | Charged11 | Charged11 | Charged11 | Free | Free |

1 Dataset size (guide only): ‘small’, up to 30GB; ‘medium’, between 30GB and 100GB; ‘large’, greater than 100GB up to 500GB; and, ‘very large’, greater than 500GB.

2 For the purposes of this document, “live”, “active” or “working” data are defined as data that require ongoing access for modification, analysis, compilation, etc. Archival storage solutions are more appropriate for ‘end state’ data.

3 For guidelines on potentially sensitive data refer to the following legislative definitions: [Information Privacy Act 2000 (Vic),](http://www.legislation.vic.gov.au/domino/web_notes/ldms/pubstatbook.nsf/f932b66241ecf1b7ca256e92000e23be/4be13ae4a4c3973eca256e5b00213f50/%24FILE/00-098a.pdf) [Privacy Act 1988.](https://www.legislation.gov.au/Series/C2004A03712)  See also the [La Trobe *Research Data Management Policy*](http://www.latrobe.edu.au/policy/documents/research-data-management-policy.pdf)

4 For the purposes of this document, sharing is understood to mean access to specific datasets by users other than the primary or lead researcher such as research assistances or collaborators.
In addition, a distinction is made between access from internal LTU staff and external collaborators, institutions or parties (Also see footnote 9).

5 For the purposes of this document, “external collaborators” refers to non-LTU research or administrative staff, whether from another higher education institution, hospital/medical facility or government/private external research body.

6 For the purposes of this document, “remote access” is defined as access to storage resources when the researcher is physically outside of the LTU network domain.

7 For the purposes of this document, **High Performance Computing (HPC)** is understood to mean a cloud-based platform leveraging the processing power of either a super-computer or massively multi-node cluster. Typically, such platforms enable highly parallel processing of tasks or simply the ability to scale-out computing resources as required.

8 For **LTU G:, P: and R: Drives**, it should be noted that sharing is only possible internally with other LTU Staff as access is predicated on the user having a La Trobe user account.

9 Data is stored at non-La Trobe data-centers though still within Australian jurisdiction. Assess whether this meets your criteria for “sensitive” data storage.

10 Compute platform geared towards High Performance Computing rather than for raw storage capacity.

11 Charges will depend on specific storage allocation and use-cases. For a scheduled of charges, please see <http://www.intersect.org.au/space/rates>.

**Research Storage Matrix By Use-Case**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **USE-CASE/REQUIREMENT** | **AARNet CloudStor** | **Intersect DeepSpace Archive Storage** | **Intersect SpaceShuttle** | **Intersect SpaceLab** | **Intersect Time10** | **LTU G:, P:, R: Drives(On Premise)** | **LTU OneDrive** |
| **Capacity** |  |  |  |  |  |  |  |
| **Collaboration** |  |  |  |  |  |  |  |
| **Performance** |  |  |  |  |  |  |  |
| **Security** |  |  |  |  |  |  |  |

**Additional Notes**

|  |  |  |
| --- | --- | --- |
| **STORAGE PLATFORM** | **BEST FOR** | **CAVEATS** |
| **AARNet CloudStor**<https://cloudstor.aarnet.edu.au/> | * Suitable for storing and sharing files with external parties. (max tested file size for upload is 100 GB). Be aware that web browser may have a maximum upload size limit of 2 GB. An ownCloud or WebDav client can be used to upload files larger than 2 GB.
* Storage Quotas: Researcher: 100GB (free) Student: upon application Research Group: upon application
* Additional capacity can be requested by emailing cloudstor@aarnet.edu.au. A charge may apply to increases in quota.
* Individual researcher or share owner is responsible for who can access the data and with what privileges.
* Access from a Unix command line is possible using either a Davfs2 or Cadavar client.
* ownCloud client can be used to synch files from your local workstation to the cloudStor location.
* Mobile apps for iOS, Android and Blackberry available.
 | * AARNet is not backing-up the data stored in the system, however it is stored in a geographically distributed file system. The system builds on the infrastructure used for AARNet's Mirror and CloudStor services, which have very high reliability and availability. AARNet monitors all its systems on a 24x7 basis.
* Data is encrypted in transit using TLS tunnels between the client and the ha-proxy off-loaders within the CloudStor environment. This includes data sent via the web browser, ownCloud client and Webdav.
* Data at rest, however, is NOT currently encrypted by CloudStor. If you require encryption at rest, please use client-side encryption.
 |
| **Intersect DeepSpace Archive Storage** | * Suitable for static data or files that are not accessed frequently
* Subscription based model.
 | * Subscription based model mean costs for this service will need to be factored into the research grant proposal.
* Data is stored outside LTU domain boundary.
 |
| **Intersect SpaceShuttle** | * Suitable for live datasets or files actively being accessed or modified frequently.
* Good for securely transferring, storing and managing large amounts of active research data.
* Ingest data directly from instruments and sensors.
* Automate workflows using scriptable command line tools.
* Uses a combination of disk and tape with two copies stored on tape.
* Transfer data via the dedicated high-speed [Australian Academic and Research Network](https://www.aarnet.edu.au/) optical fibre network (with no ingress or egress charges between AARNet destinations).
* Transfer data at very high speed using IBM Aspera FaspTM technology.
* Transfer securely using AES-128 data encryption.
* Scale to petabytes of storage.
* Manage access to your data safely and securely, using your existing credentials.
* Share and collaborate with other researchers from universities and research organisations.
* Manage your own data using a convenient web browser interface.Share files or folders of any size using URLs.
* Share broadly using a URL or limit access to your data, using single-use access tokens.
* Collaborate actively using a web interface.
 | * Subscription based model mean costs for this service will need to be factored into the research grant proposal.
* Data is stored outside LTU domain boundary.
 |
| **Intersect SpaceLab** | * Suitable for individual researchers or groups needing to build and manage their own hosted virtual laboratory, using their own research software without the need to buy, operate and maintain hardware and operating systems.
* Suitable for live datasets or files being accessed or modified frequently.
* Subscription based model
 | * Subscription based model mean costs for this service will need to be factored into the research grant proposal.
* Data is stored outside LTU domain boundary.
 |
| **Intersect Time** | * Suitable for High Performance Computing (HPC) and/or computational processing of active datasets in either massively parallel or scale-out computational models.
* Choice of two HPC platforms (Orange and Raijin at NCI).
* NeCTAR Research Cloud for Cloud Computing also an option for creating your own virtual lab of Linux based computers to host your research project.
* Subscription based model.
 | * Access to Orange and Raijin HPC platforms is merit based and judged by a Resource Allocation Committee and by Intersect HPC experts once a year, however, requests for small amounts of compute time by new researchers or for smaller research proposals are actively sought/encouraged at any time.
* Subscription based model mean costs for this service will need to be factored into the research grant proposal.
* Data is stored outside LTU domain boundary.
 |
| **LTU G:, P:, R: Drives** | * Suitable for sensitive or critical data where the data cannot leave the LTU domain.
* Small to medium sized datasets.
* Datasets generated manually or by instruments.
* Access only to staff with LTU email address.
* Multiple snapshots per day retained for 30 days with the first snapshot of the month retained for 6 months.
* For G: drive data, users are able to recover data via Previous Versions (snapshots) which run three times a day (6 AM, 12 PM and 6 PM). The 6 AM and 12 PM snapshots are retained for 7 days with the 6 PM retained for 31 days. In addition, a monthly 3 AM snapshot on the first of the month is retained for 12 months.G: drive data is also replicated to Disaster Recovery (DR) data-center at the Bendigo campus.
* P: and R: drive data is replicated to a standby (DR) system which has daily snapshots at 7 AM retained for 31 days and a monthly snapshot on the first of the month at 1 AM retained for 12 months with recovery requested via ICT.
 | * Access from outside the university requires LTU VPN client connection
* H: and R: drives are deprecated. Migration of user’s data from their H: Home drive to their LTU OneDrive space is underway.
* Similar project to move researcher’s R: Personal Research Drive to their AARNet CloudStor space is being planned.
 |
| **LTU OneDrive** | * Cloud based storage provided to all LTU staff for personal, i.e. not group, datasets.
* Access to external person’s by email invitation of OneDrive owner.
* Access from outside LTU network does not require LTU VPN client.
* Best for personal or individual user files/datasets.
* Suitable for work-in-progress
* Synchronization of file updates occurs in the background but only while connected to the internet.
 | * No backups or previous versions
* Data is stored outside LTU domain boundary.
 |