Promoting Data Management “best practice” at UNSW

Jacky Cho, Adrian Chew, Cecilia Stenstrom, Luc Betbeder-Matibet
Data Management landscape at UNSW

- Not the first data management conversation we’ve had

ResData
- RDMP
- Data Publishing

UNSW Data Archive

Post-Project Storage

This Photo by Unknown Author is licensed under CC BY
What about Active Data? An example…

Data storage options for Faculty of Science

Internal Data Services

- UNSW (Personal/Group) Shared Drive
- School based Local Storage Servers
  - Group owned storage servers
  - External HDD, Flash Drives

Active data

Computing-linked data

- Katana HPC: Scratch Storage
  - CCRC Transactional Storage

External Data Services

- CloudStor+
- Google Drive
- DropBox
CCRC Storage
Active Data Research System

Data Archive
Personal Email
DropBox or other sharing tool

Collaborations
- X Personal tools used
- X Unsupported Tools with no controls

Data Publishing
- X Immature Tools
- X Limited Open Data Set Support

RDMPS
- X Complex
- X Low Uptake
- X Minimal Integration

Unsuppported Tools
- X Unsupported or Personal Tools to Store and Share
- X Wide range of local systems used
- X Low discoverability of supported Tools

Data Storage
- X Many Systems
- X Lots of data moves

Group and Home Drives
MWAC
CCRC Storage

Faculty System
Active Data Research System
No Research Data Team

Training and Support
- X Low Awareness of Research Data Obligations
- X Minimal Support for Research Data Tools, Data obligations, and Data Moves

My Data?

No Training Capability
From the researcher side…

- ~10% of researchers had RDMPs
- Data Archive was used heavily, but many people did not know about it.
- Those with Data management needs (ethics, funding requirements etc.) sought it out.
- Researchers were:
  - Not sure they’re “managing” “data”
  - Not sure what are they required to do
  - Not sure where they could get information or help
What is Data Management?

Researchers

IT

Library

DVCR

Data Governance

Research Integrity

This Photo by Unknown Author is licensed under CC BY-NC-ND
What is Data Management?

- Ease of use
- Access/sharing
- Cost/effort

What is best practice?

- Storage Capacity/Cost
- Systems Security
- Data Loss Prevention

- Research Integrity
- Data Citations
- Ethics and Policy

- Data Preservation/reuse
- Data Publishing
The Game Plan

People & Policy

- Easy to access information, WIIFM,
- Training, dedicated help/support

Tools & Tech

- Leverage existing platforms, better
discoverability, Streamline/improve
User experience
### People

**Research specific data policies**
- Data governance and handling policies
- Data Classifications
- Data Storage guides

**Websites and promotional material**
- Single source of truth
- Clear defined actions
- Better tools discovery

**Training programs**
- Online and Workshops
- Developed with HDRs/researchers at every step

**Regular RDM Info Sessions**
- Tailored to faculties, schools and research centres

**Dedicated Support Staff/comms channels**

### Tools

**RDMP**
- Redesigned
- Streamlined UX
- Better integration (Ethics etc.)

**Project Storage/Tools provisioning**
- UX focused
- Store: OneDrive, Research Active, DataArchive
  Collaborate: MS Teams, eNotebooks

**eNotebooks**
- General: OneNote
- Advanced: *(In procurement)*
Classifying your Research Data

Data containing information involving minors/at-risk participants or culturally sensitive groups
- The data is classified as **Highly Sensitive**

Data subject to regulatory controls
- The data is classified as **Highly Sensitive**

Data containing personal information
- The data is classified as **Sensitive** or higher

Data containing confidential information
- The data is classified as **Private** or higher

Data collected from human and/or animal participants
- The data is classified as **Private** or higher

Unpublished research data
- The data is classified as **Private** or higher

OneDrive is a UNSW supported research data storage platform for the above data classifications

It is possible that the data may have other requirements that are not covered in the current data classification.

Please contact datagov@unsw.edu.au if you have such unique requirements or doubts about classifying your data.
<table>
<thead>
<tr>
<th>Storage location</th>
<th>SUPPORTED</th>
<th>UNSUPPORTED</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNSW OneDrive &amp; Teams</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Home drive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shared drive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Archive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cloudstor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local storage incl. Internal/External drive, USB, PC, etc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dropbox</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Highest data classification</th>
<th>Highly Sensitive</th>
<th>Sensitive</th>
<th>Sensitive</th>
<th>Private</th>
<th>Public*</th>
<th>Private</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stored in Australia</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>?</td>
<td>✗</td>
</tr>
<tr>
<td>Backup and disaster recovery</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>Syncing with local copy</td>
<td>✓</td>
<td>N/A</td>
<td>N/A</td>
<td>✓</td>
<td>N/A</td>
<td>✓</td>
</tr>
<tr>
<td>External collaborator access</td>
<td>✓</td>
<td>❌</td>
<td>❌</td>
<td>✓</td>
<td>❌</td>
<td>✓</td>
</tr>
<tr>
<td>Storage limit</td>
<td>5Tb per user</td>
<td>Unlimited</td>
<td>Unlimited</td>
<td>100Gb per user</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Version control</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>❌</td>
<td>✓</td>
</tr>
<tr>
<td>Recovery from deletion</td>
<td>60 days</td>
<td>?</td>
<td>Unlimited</td>
<td>❌</td>
<td>❌</td>
<td>Plan dependant</td>
</tr>
<tr>
<td>Post project data retention</td>
<td>7 years</td>
<td>7 years</td>
<td>+ 7 years</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
Acknowledgements:
- UNSW IT: Rizwan Sakkathulla
- UNSW Library: Maude Frances, Alexander Sussman
- University Planning and Performance: Kate Carruthers, Suyash Diwedi
- Researcher Development Unit: Cecilia Stenstrom, Adrian Chew
- Research Technology Services: Luc Betbeder-Matibet, Jake Surman
- Research Integrity: David Do

“Good data management is good research”
- UNSW Research Integrity