

From UNSW Data Archive to UNSWorks – learnings from the ARDC Data Retention project

Ai-Lin Soo, Gerry Devine, Jacky Cho, Jack Surman

The Data Retention Project

“We’re working with research-sector organisations to more effectively manage their data collections. We’ll achieve this by enriching the data collections with consistent international metadata standards to make the collections more FAIR – particularly findable and reusable.

This project focuses on identifying and supporting important data collections that have not been previously supported by federal research infrastructure initiatives” - <https://ardc.edu.au/program/data-retention-program/>

UNSW Data Archive

Non-delete system designed to help researchers meet their obligations for long term management of research data.

Researchers engage with the service by submitting a request through ResToolkit and completing an RDMP.

8,546
Active Records

7 PB
Total Size

March 2014
Oldest Record

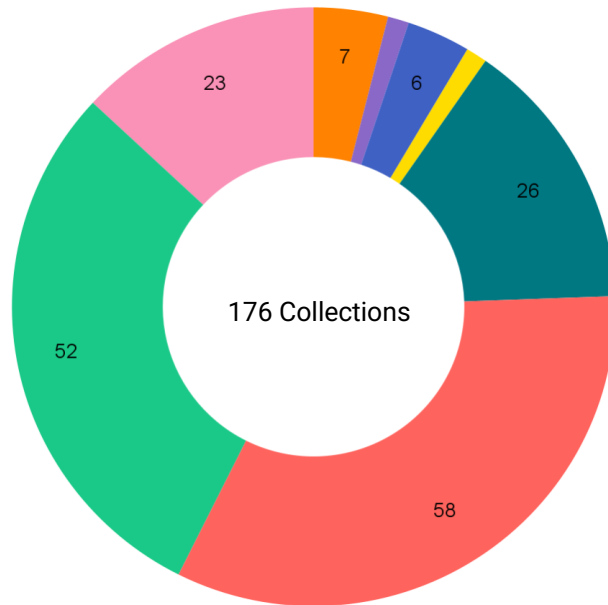
The screenshot shows the UNSW Data Archive website. At the top left is the UNSW Sydney logo. To its right is the text 'UNSW Data Archive'. Below this is a navigation bar with links for 'Home', 'Help topics', 'FAQs', and 'Contact'. The main heading reads 'The UNSW Data Archive. Your research data - safe, shareable, and searchable.' followed by 'Use this site to find out how.' Below this are three numbered steps in blue boxes: 1. 'Apply for storage' with subtext 'update or complete your Research Data Management Plan (RDMP)' and a button 'Apply now / update your RDMP'. 2. 'Learn more' with subtext 'UNSW Data Archive help & information' and a button 'Start here'. 3. 'Access your storage' with subtext 'upload and download your data' and a button 'Access the Data Archive'. The background features a network diagram.

Finding data assets in the UNSW Data Archive

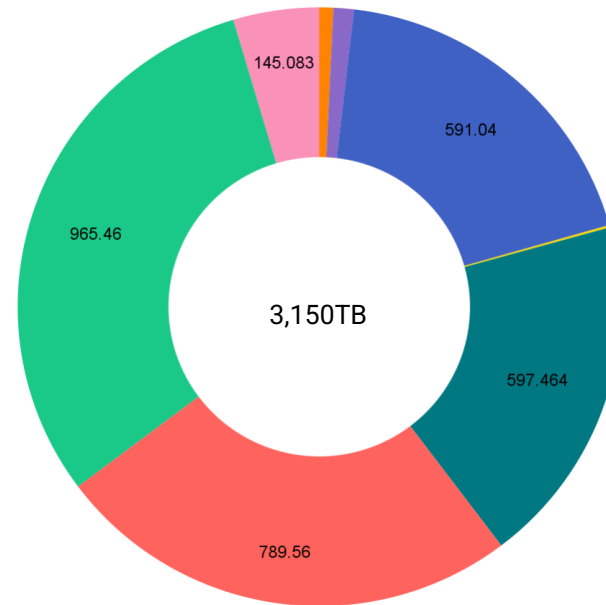
176 collections were sampled from UNSW Data Archive and nominated for inclusion

Discovering data assets in UNSW Data Archives

No. Collections by Faculty



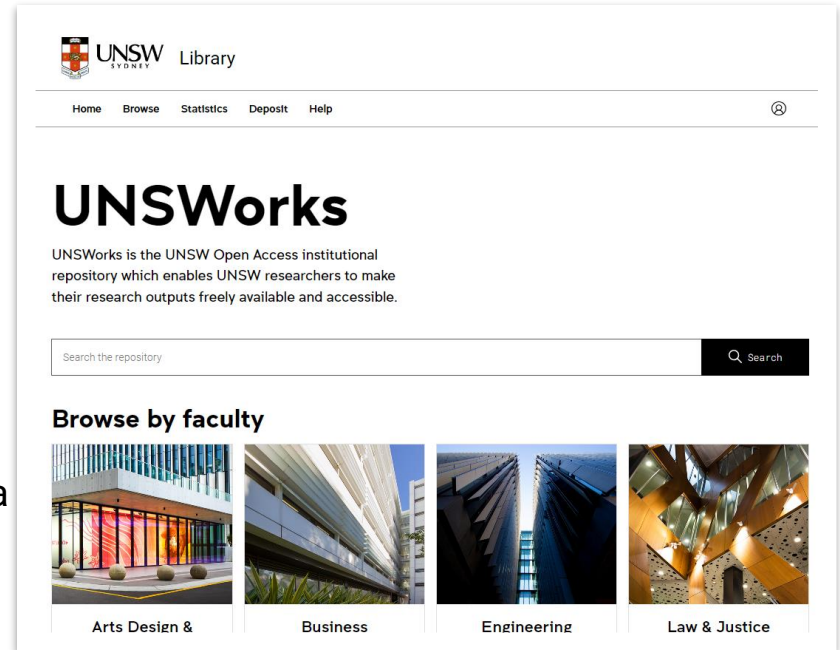
Volume (TB) Collections by Faculty



UNSWorks

- UNSW's Open Access Institutional Repository, enabling UNSW researchers to make their research outputs freely available and accessible
- Ability to deposit and publish (with automated DOI allocation) small **research datasets** up to ~5GB
- Ability to link datasets to publications, grants, researchers, and other datasets
- Record is automatically harvested by Research Data Australia (RDA)

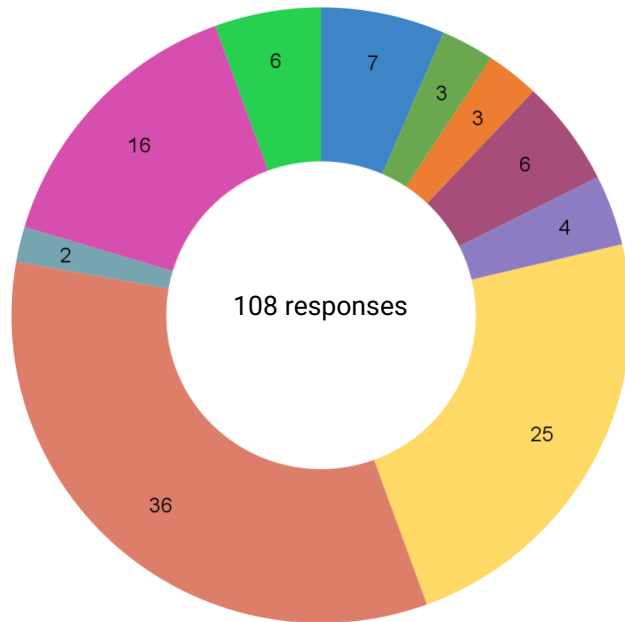
<https://unsworks.unsw.edu.au/>



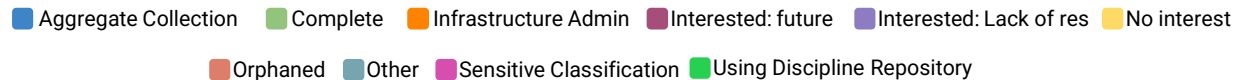
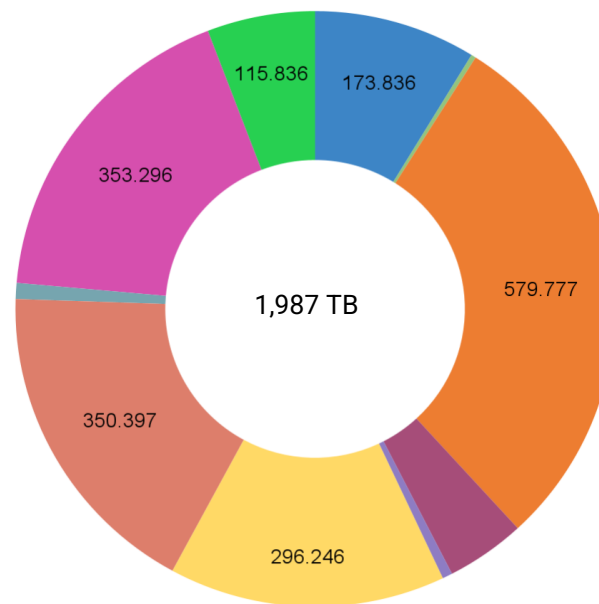
The screenshot displays the UNSW Library website. At the top left is the UNSW Sydney logo and the word "Library". A navigation menu includes "Home", "Browse", "Statistics", "Deposit", and "Help". The main heading is "UNSWorks", followed by a descriptive paragraph: "UNSWorks is the UNSW Open Access institutional repository which enables UNSW researchers to make their research outputs freely available and accessible." Below this is a search bar with the placeholder text "Search the repository" and a "Search" button. A section titled "Browse by faculty" features four columns, each with a representative image and a faculty name: "Arts Design &" (with an image of a modern building), "Business" (with an image of a building facade), "Engineering" (with an image of a building interior), and "Law & Justice" (with an image of a building interior).

Finding data assets in the UNSW Data Archive

No. collections by Responses



Volume (TB) by Response



Finding data assets in the UNSW Data Archive

Aggregate Collection

- Service oriented collection (eg. collection for an instrument)
- Research lab, research centre or large collaboration collection

Interested: Future

- Cannot identify any appropriate datasets at the moment but when they are available would be interested in participating (eg. paper is complete, experiments are finished).
- Follow up at a later point in the year.
- Interested in streamlining the process for future datasets as they arise.

Interested: Lack of resources

- The collection has not had proper data management and no metadata exists at all and hence, no appropriate data can be identified without extensive work.
- Other commitments (eg. teaching, more recent research).

No interest

- No reason provided
- Data within collection will never be suited for publication
- Perceived cost outweighs perceived benefits of data publication
- Person authorised to make decisions is not available/clear
- Unsure what collection was used for.

Major findings

Alignment of UNSW Data Archives and UNSWorks metadata

- Across 34 metadata fields in UNSW Data Archives and UNSWorks, only 4 that roughly matched across the two platforms. These were title, description, authors/contributors or members and start and end dates of the project.

Engagement in meetings doesn't translate into action

- Engagement methods: one on one meetings, dedicated working time and feedback at hackerspaces, pre-filled UNSWorks records
- Only 3 collections completed a record in UNSW works from the 108 collections who engaged

Data publishing is occurring

- About 6 collections indicated that their collections (roughly equal to 100TB) were already published through their disciplines repository.
- Work to understand the full scope of repositories used by researchers to be replicated in UNSWorks would be beneficial.

Method used of this project is not scalable

- Improved integration between ResTech Services capabilities and the Library's capabilities would be most beneficial.

From UNSW Data Archive to UNSWorks – learnings from the ARDC Data Retention project

Ai-Lin Soo ai_lin.soo@unsw.edu.au - ResTech Services
Gerry Devine gerry.devine@unsw.edu.au - The University Library
Jacky Cho - ResTech Services
Jack Surman- ResTech Services