



Transforming Library Collection Data for eResearch

Investigating how to make the David Scott Mitchell collection at the New South Wales State Library available for modern digital humanities research.

The David Scott Mitchell collection (DSM) is the State Library's most renowned collection. Whilst these digital resources are available via catalogue for individual viewing and download, the Library lacks the expertise to standardise the data for modern digital humanities research, which involve bulk access, ensuring compliance with emerging community and global standards for transcription, data citation, licensing, linked open data, national and international data aggregation, and improvements to support the use of computer vision, machine learning, and automation.

The key questions the project will address are

- What are the user requirements from digital humanities researchers to effectively use the DSM collection in their research work?
- What are the practical data standardisation and data management steps the Library must make for the DSM collection to meet the FAIR principles expected by the research community?
- What new policy decisions (such as scope, access, terms of use, and citation) and impact to cataloguing and collection management practice must the Library consider for this service offering which now gives bulk, 24x7 and high-speed access to collections never available before?

Start date

18 June 2019

Expected completion date

21 October 2019

Investment by ARDC

\$49,999

Co-investment partners

[Library Council of New South Wales](#)

Lead node



STATE LIBRARY®
NEW SOUTH WALES



1. Workshops

Information gathering sessions with key stakeholders, documenting and analysing requirements. Hack sessions with externals to pilot the extract, transformation and publishing of collection content to select eResearch tools/platforms



2. Development of scripts and tools

improving bulk access to collection data via APIs, extract and ransformation steps will be documented in notebooks.

Notebooks will demonstrate the conversion of library records and associated digital items to eResearch platform datasets. All scripts, tools, Jupyter notebooks and datasets will be published and made publicly available.



3. Presentation

Project outcomes will be presented at an ARDC Data and Services Summit in October 2019.

Core features



Scripts and tools

All scripts, tools, Jupyter notebooks and datasets will be published and made publicly available.

Who is this project for?

- Research organisations
- Researchers
- Historians.



What does this project enable?

Projects in this area will contribute to the discoverability and use of cultural collections within research communities. Using FAIR principles, organisations can assess current state of usability, whilst also identifying and addressing cases where cultural protocols for access are required.



Handy resources

- [Final Report](#) [PDF 94KB]
- [FAIR Assessment](#) [PDF 147KB]
- [Presentation](#) [PDF 1MB]
- [Jupyter Notebooks](#)
- [OCFL](#)
- [Portland Common Data Model](#)
- [Schema.org CreativeWork](#)
- [Voyant tools](#)



STATE LIBRARY[®]
NEW SOUTH WALES

Library Council of New South Wales

VISIT



Australian Research Data Commons

[HOME](#)

[ABOUT US](#)

[SERVICES](#)

[COLLABORATIONS](#)

[NEWS & EVENTS](#)

[CONTACT US](#)

[PRIVACY POLICY](#)

[TERMS & CONDITIONS](#)

The Australian Research Data Commons is enabled by NCRIS.



Stay up to date with our latest news and events and subscribe to our newsletter.

SUBSCRIBE