

Data and Services Discovery projects - Transformative Data Collections

Title

An Australian historical criminal justice data repository

Approach

The proposal aimed to develop an application within the Prosecution Project to publish curated datasets to the Australian Data Archive. The approach built on a number of recent developments in the research field of criminal justice history, the primary domain of the research stakeholders. These developments included:

1. Production of the Prosecution Project (2013-continuing) web-based research infrastructure for accessing, recording and storing historical records on Australian criminal trial proceedings, principally drawing on relevant archival holdings.
2. Development in 2016-2017 of the NCRIS-funded HASS DEVL project for sharing Prosecution Project data with the archive repositories that store the original data (successful implementation of data-sharing using an API, with Tasmanian Archives and Heritage Office and the National Library of Australia Trove search engine).
3. Increasing number of requests from external researchers (national and international) for access to subsets of Prosecution Project data. These requests were typically supplied by generating a flat file of data (csv format) to the requestor under informally agreed usage conditions.
4. Development by NCRIS-funded HASS DEVL in 2018-19 of Tinker (tinker.edu.au), a platform for humanities and social science research training, to which the Prosecution Project contributed an exemplar dataset for the Tinker toolbox.
5. Increasing use of IT systems in historical research, with researchers building bespoke databases that have limited or no sharing orientation or capability.

The opportunity provided by the ARDC Transformative Data Collections initiative was taken up by this project with the primary aim of securing a workflow for archiving datasets from the Prosecution Project to the Australian Data Archive. Consultation with the historical research community also confirmed support for use of the Prosecution Projects platform as a medium through which data generated by other (non-Prosecution Project) researchers might be ingested, curated and archived in ADA, thus expanding greatly the research capability of this domain.

Implementation was achieved through the following steps:

1. Analysis of project requirements by Griffith University Business Analyst (Michael McGuinness)
2. Discussion of project aims and outcomes in a meeting at the Australina Historical Association Conference at Toowoomba, 11 July 2019, largely attended by about 30 researchers who provided strong support as well as suggestions for the project.
3. Planning of workflow through communication with Australian Data Archive (Steve McEachern, Marina McGale)
4. Development and testing of the application (Jeroen van den Muyzenberg)
5. Successful production of the application (16 Oct 2019), resulting in publication on the ADA Dataverse of a number of datasets (5 to 19 Oct 2019), and enabling future production of multiple datasets.

Outputs

The key outputs are the datasets, examples including (at 21 Oct 2019):

NTSC 1873-1899	Northern Territory Supreme Court trials.	http://dx.doi.org/10.26193/U3ZZGY
TASSC 1824-1849	Trials on indictment at Tasmanian Supreme Court from 1824-1849.	http://dx.doi.org/10.26193/BJCVE0
VICSC 1855-1942 bigamy	Trials on indictment for offence of bigamy,	http://dx.doi.org/10.26193/5S8LLB

	Victorian Supreme Court 1855-1942	
NSWCCJ 1788-1810	Trials in Court of Criminal Jurisdiction NSW 1788- 1810.	http://dx.doi.org/10.26193/0SOO15
VICPG Prisoners discharged 1864-1869	Prisoners discharged from Victorian Prisons, 1864-1869	http://dx.doi.org/10.26193/GCTWXW

As outlined above an important aim of the project was to demonstrate the functionality of the archiving workflow for ingesting data produced by researchers from outside the Prosecution Project research team. This has been successfully implemented – examples include two datasets in the list above (NSWCCJ and VICPG) both of which datasets originate in other projects, with the data subsequently processed to Prosecution Project standards ready for archiving in the ADA Dataverse.

FAIR

As indicated in the attached FAIR assesment spreadsheet the project has met major objectives of improving access to a range of datasets for the advancement of research in a significant research community. The creation of the Australian Historical Criminal Justice Data Dataverse (https://dataverse.ada.edu.au/dataverse/australian_historical_criminal_justice_data) on the ADA website is a major advance in the Australian historical research community, enabling the sharing of data collected for publicly-funded research of national and international importance.

The research rationale for the Prosecution Project already met FAIR standards to at least some degree – and its primary objective of archive recovery and tranformation into machine-readable data in particular made this kind of data both more findable and accessible. The project’s public website also addressed accessibility for the general as well as research communities through a search engine interacting with the constantly upgraded database. Research potential of the project data was however limited by the research design and culture (orientation to an individual project with conventional publication outputs) as well as technical limitations – this especially limited interoperability and reusability of the data.

The successful archiving of data into the Dataverse now makes possible this data being accessed by other future researchers and projects. The use of the Prosecution Project platform to enable other researchers to contribute their data to an archive that meets consistent data standards builds the potential of this data being used by new projects that lie beyond the criminal justice history domain, eg demography, social history of communities, urban history.

Collaboration and coverage

Collaboration is national and international. The application drew together researchers from 9 other Australian and 2 UK universities. The data coverage of the archive created will be national (all Australian State jurisdictions) and longitudinal. Archiving in the ADA Dataverse makes the data available to bona fide researchers, including research students, in Australian and internationally.

Sustainability

Griffith University has committed to maintaining the Prosecution Project platform for at least another 5 years (to 2024). The archiving workflow is embedded in a robust infrastructure supporting the Prosecution Project. The original research team is now distributed through four universities and this ARDC project has consolidated research partnerships with at least another dozen leading criminal justice researchers in Australia and the UK. Beyond the immediate interests of this distributed community of research stakeholders, the datasets archived in these structured formats will remain usable for the foreseeable future in the Australian Data Archive.

Learnings

Key learnings:

- Embedding the understanding and use of consistent data standards in research training
- Expanding knowledge of the FAIR principles in the HASS research communities
- Wider knowledge among researchers of the role and structure of metadata
- Importance at project planning stage of design and budgetary requirements for archiving data
- More proactive institutional support for data management at the project level, in effect changing the research recognition and reward culture to assist this.

Impact

This project has aimed to establish a more secure, sustainable and shareable database for the conduct of criminal justice history as well as related projects. Assessment of impact must await evaluation of the future usage of the data. But the following features of the datasets and their accessibility in the ADA Dataverse point to the following potential impacts:

- Expansion of the number of quantitative historical studies of , crime, social relations, policing and prosecution.
- Increasing number of comparative studies (multi-jurisdictional)
- Access of the data by other research domains including demography, economic and social history
- Increased use of high quality Australian historical datasets by international researchers
- Tracing of impact and usage of these shareable datasets will be facilitated by the attachment of persistent metadata describing the sets (doi citations) and the attribution of researchers responsible for collecting and curating the data through use of ORCID ids.

Report prepared by: Mark Finnane, Griffith University

Date: 22/10/2019