

Research Libraries, Data Curation, and Workflows

*Ingrid Mason*¹

Ingrid Mason¹, Roxanne Missingham²

¹ AARNet, Canberra, Australia, ingrid.mason@aarnet.edu.au

² The Australian National University, Canberra, Australia, roxanne.missingham@anu.edu.au

INTRODUCTION

Libraries and indeed the GLAM (Galleries, Libraries Archives and Museums) sector have been actively engaged in exploring opportunities to create digital collections to take advantage of the opportunities of the digital environment. In seeking to expose our rich collections and contribute to new forms of scholarship the need for effective workflows and digital processes is been a critical area for development. Digitisation has been a vital step in providing collection access, but processes have often been clunky and focused on delivering a digital equivalent of the physical object, rather than exploring the how unpacking content can deliver solutions for digitally literate researchers. Improving workflow and processes will enable greater productivity and result in a greater number of collection items being available digitally within the limited budgets available for this work. In addition, collection material will be available in a form that creates a true research infrastructure for research and study.

This presentation reviews data curation work undertaken in two projects with digital cultural heritage material maintained by two research libraries in Australia, designed to meet the needs of Humanities, Arts and Social Science (HASS) researchers. The aim of this review is to share the (1) outcomes of data curation work enabled through expertise sharing and collegiality; (2) the iterations of a Data Curation Framework, and, (3) lessons learned in the work of developing a data curation framework to operate in different contexts, but with the same purpose and audience.

RESEARCH LIBRARIES

Two Australian research libraries, the Australian National University (ANU) Library and the National Library of Australia (NLA), worked on projects targeted to meet the needs of HASS researchers, by exploring (1) data curation and workflows, and (2) developing a data curation framework, with AARNet as a partner in 2018.

The project work with the ANU commenced in 2018 (and is ongoing) and focuses on a joint exploration of:

- Data packaging and publishing (drawing material from heritage collections and building a corpus)
- Data access and movement (using CloudStor storage for a digitisation workflow and computation services)
- Researcher engagement (enabling the use of new technologies and new research opportunities)
- Working within a data curation framework (drawing together digital curation and data science expertise)

The project work with the NLA was completed in 2018 as a key deliverable (in the data curation work package) for the HASS data enhanced virtual lab development. This required project partners and others on the lab project team to:

- Develop a data curation framework that would make data available for research interrogation
- Establish data flows and make data accessible to the research community in partnership
- Identify, document and publish workflows for data flows and curation
- Publish reusable data processing workflow examples for text with, analysis, transcription, and geo-spatial elements
- Establish workflow examples as a standards for the research community to follow when reusing the tools

DATA CURATION FRAMEWORK

The first version of the data curation framework was aimed at experienced research support and infrastructure professionals in universities or collecting organisations, such as libraries (and other custodians of digital scholarly resources or data) interested in making cultural heritage material (as data) more accessible to researchers. The framework was an output of the Research Data Service programme in 2016, via the Access to Culture and Community Data for Research project. The framework provided ideas for guidance on what might be done to extend or provide new research data support services for Humanities, Arts and Social Science (HASS) researchers in an institutional context with heritage collections. The idea of his framework was drawn into Australian Research Data Commons (ARDC) funded Tinker Virtual Lab project work in 2018, with a view of being reframed and positioned to guide data curation and interoperability in a national ecosystem of platforms and to aid researchers. The framework was further refined and simplified in early 2019 to aid engagement with HASS researchers on the curation of their research data.

LESSONS LEARNED

The idea of data curation and the act of data curation are related but not the same, as the act of curation operates with a unique set of factors and forces that impact upon the nature of the data and its “life” as a research resource.

Workflows are vital to optimise the digitisation process and deliver access to collection items in an effective manner.

Lessons learned from these two projects relate to data and what happens when attempting to make data FAIR (Findable, Accessible, Interoperable, Reusable) and the range of technical and social challenges encountered in doing such as:

- digitisation techniques, selection limitations and workflows [1]
- republishing data from published works, incompleteness in data sources, and ethical consequences
- dealing with different senses of time and timelines or place and place names
- working with existing and new infrastructures
- integrating research processes with a 360° data sharing model

Lessons learned from these two projects relate to people and what occurs when data curation involves working as a team with a mosaic of multi-disciplinary expertise and professional skills and experience, and drivers. The composition of project teams and where they operate, in what’s referred to as the “third space” [2], needs to be better understood.

Working effectively across organisational and jurisdictional boundaries on data curation, and the gains that can be made are a means for building trust and establishing successful partnerships in research data infrastructure development.

REFERENCES

1. Sköld, Olle., et al., Moving Forward with Digital Scientific Images: a Study of Infrastructure, Digitisation Work, and Digital Research Practices. In: *Proceedings of the Digital Humanities in the Nordic Countries 4th Conference (DHN 2019)*, Copenhagen, Denmark, March 5-8, 2019. / [ed] Costanza Navarretta, Manex Agirrezabal, Bente Maegaard, Copenhagen, 2019, Vol. 2364, p. 415-425. Available from: http://ceur-ws.org/Vol-2364/38_paper.pdf accessed 18 June 2019.
2. M Veles, Natalia, and Carter, Margaret-Anne (2016) Imagining a future: changing the landscape for third space professionals in Australian higher education institutions. *Journal of Higher Education Policy and Management*, 2016, 38 (5). p. 519-533. Available from: DOI: 10.1080/1360080X.2016.1196938