

Data neophytes: first steps into the research data abyss

Patrick Splawa-Neyman
Data Librarian
Monash University

patrick.splawa-neyman@monash.edu
 <http://orcid.org/0000-0003-2388-4104>

Jaye Weatherburn
Information Project Officer
Swinburne University of Technology

jweatherburn@swin.edu.au
 <http://orcid.org/0000-0002-2325-0331>

Abstract:

This paper explores the roles, functions, and possible definitions of data librarians based on two Australian National Data Service (ANDS) case study projects at Monash University and Swinburne University of Technology. The experiences, challenges, and achievements from these research data management projects are examined by discussing the various factors involved, such as liaison with researchers and organisation stakeholders, and the implementation of technological solutions.

What is a data librarian?

There are many ways to think about the role of the data librarian, as data librarianship is a relatively new, emerging field that is constantly in flux.

The label 'data librarian' has gained greater prominence in recent years, suggesting it is a new term. However, data librarians have been written about in academic papers since the mid-1980s. De Vries (1986, cited in Howe 1993), defined the four major functions of a data librarian as data acquisition, data dissemination, data storage and maintenance, and documentation and reference services. This was in a time of magnetic tape storage, microform, machine-readable text files and mainframe software packages (Howe 1993).

By 1997, data librarians were still being discussed; however, 'the Data Librarian' had also become a software package designed to manage large data files produced by laboratory instruments (Liscouski 1997). This software's task was to archive data, prevent data loss and integrate various laboratory systems.

In the 2010s, the term data librarian morphed into how the term is currently (and generally) defined: a merging of traditional librarianship with newer methods of research data management, focussing more on a person, rather than a process.

Ratzek's (2012) profile of a data librarian makes little reference to technical skills and instead focuses on the soft skills such as negotiation skills, compliance management and public relations. He does refer to technical skills, but only in very general terms such as the development of standards, and preservation. Groenewegen (2014) defines a data librarian as a person who uses the traditional skills and strengths of a librarian to facilitate research management, sharing, and preservation in a connected way.

The Australian National Data Service (ANDS) also provides valuable insight into the role of the data librarian. While not explicitly defining a data librarian, ANDS (2015a) advises that librarians bring a wealth of experience and knowledge related to resource description, scholarly publishing, liaising with research staff and delivering programs.

There are many different ways to think about the definition of data librarians, adding to the idea that the role is so varied and covers so many different areas that it is hard to pinpoint an all-encompassing description. In this paper, we explore some of the many skills and competencies required to manage complex research data projects (Patrick Splawa-Neyman at Monash University, and Jaye Weatherburn at Swinburne University of Technology) in order to provide some definition of what a data librarian does. What emerges from this exploration is a difference of opinion.

Patrick defines a data librarian as a person (irrespective of title) who requires highly developed interpersonal skills, and who uses the core knowledge and skills of librarianship to assist researchers (usually in an academic setting) to manage and disseminate their research outputs via a distributed, online network. The 'core knowledge and skills' referred to are contained and expanded in the Australian Library and Information Association's (2014) statement pertaining to the knowledge, skills and attributes of library and information specialists.

This definition was slightly limiting for Jaye, as the Swinburne University project did not involve working in an academic library, but instead required the management of

digital repository resources outside of traditional publishing, and liaison with a wide range of organisations as well as researchers. This role continues to require a combination of traditional librarian skills, IT skills, and an awareness of archivist and knowledge management perspectives (especially for the long-term storage and sustainability of resources), as well as awareness of broader issues such as overarching business practice and governance.

It could be argued that it is limiting to define ourselves with the label 'data librarian', due to the variance of our roles that draw on many different aspects and principles of the information professions. If there is indeed a need for a label, a case could be made that we would be better described using the much broader term of 'information professional' or 'information manager'.

The label 'data librarian' is but one of many titles given to an information professional working in the area of research data management (RDM). The 'librarian' part of this label perhaps rings true for us as recent graduates of the Master of Information Management degree at RMIT University, which includes study of core subjects such as information retrieval, classification, and cataloguing. As a result of our graduate study, we have been able to bring traditional library skills to our projects, extending the role of 'librarian' into the area of RDM, an area that can greatly benefit from these skills.

Skills required

To begin to identify some of the skills required for working in the field of RDM, it is useful to reflect on some of these required for the projects being discussed in this paper. Technical skills needed include:

- Management of digital resources and data, including a working knowledge of metadata standards
- Advocacy for the importance of developing good procedures and policies to manage research data
- Awareness of developments in RDM at national and international levels
- Knowledge about copyright and licensing
- Knowledge of current digital curation and digital preservation practice
- Knowledge of cultural heritage sustainability issues

User experience design and website design are two areas that are becoming increasingly important for information professionals. Better understanding of these areas can aid in the design of tools to educate researchers and organisations about effective RDM practice, may help to increase the uptake of best-practice workflows, and aid advocacy efforts for good management of research data. For example, using engaging online toolkits and tutorials rather than tedious checklists or procedural documents is being encouraged around the world; instances include the University of California's online tool (University of California 2015) and Penn State University's scenario-based online tutorials for RDM plans (Penn State University 2015).

Transferable, or soft, skills are extremely useful in the area of RDM and cannot be underestimated. Coming to our roles from previous careers outside librarianship, we brought a range of skills to our projects, including highly developed advocacy, communication, and multi-tasking skills. These skills enabled us to take on new, cutting-edge projects and ensured that we could begin to develop the policies and

procedures required to deliver good outcomes for our projects. These 'divergent paths of entry' (Galvin 1995) are likely to promote a willingness to learn and the ability to be competent (if not always completely comfortable) working without a defined roadmap. An effective new information professional is willing to seek expertise from other colleagues and institutions, and is an enthusiastic consumer of information in many different areas, enabling them to continually update and enhance their information management practice.

What follows are two case studies and personal reflections from our experiences managing our first ANDS Major Open Data Collections projects. Patrick first describes the management of scientific and medical imaging data at Monash University. Jaye then describes the implementation of procedures and policies for minting Digital Object Identifiers (DOIs) for grey literature research at Swinburne University of Technology.

Monash University MODC Project

Monash University generates an enormous volume of scientific and medical imaging data that is viewed as an asset to the university.

The Major Open Data Collections Project (MODC) was established at Monash University in July 2014, with a vision to showcase imaging datasets in areas of strength and value to Monash University and its collaborative partners (ANDS 2015b).

There were three expected outcomes of the MODC Project:

1. Develop research data management guidelines to encourage Monash researchers using Monash platforms to make their imaging datasets publicly accessible
2. Contribute to a sustainable operational environment to enable improved capture, storage and discoverability of research imaging data (ANDS 2015b)
3. Develop a discoverable Imaging Locus (2015). The purpose of the Imaging Locus was to allow Monash University imaging research to be discoverable

Researcher engagement

In order to achieve these goals, it was imperative to engage with researchers who had completed imaging experiments and who had published in a journal, but who were retaining the datasets. The primary roles in engaging with researchers were to collaborate with them to help them to identify those datasets, describe them so that a researcher outside the discipline could understand them and make them available through the Research Data Australia website (2015).

Engaging with researchers required the creation of close, professional, long-term relationships with them, and one of the best ways to achieve this was through personal and email communication.

Several years ago, the Australian National Data Service ran a workshop for researchers, to ascertain how they prefer to be communicated with regarding their research. The workshop revealed a list of words and phrases that researchers prefer to hear, as well as those that they are not enthusiastic about.

Researchers want to:

- Disseminate their data
- Be able to promote themselves and their research groups
- Collaborate with like-minded researchers
- Increase their citations and leave a legacy at the end of their academic careers
- Make sure that their data is backed up
- Increase their profile which will help with their promotional opportunities
- Create a benefit to the academic and wider communities
- Increase their own opportunities for further research.

Conversely, there are some words that researchers are not enthusiastic about. These words include:

- Share
- Compliance
- Policy
- Governance
- Legislation
- Form, in the sense of 'please fill in this form'
- Metadata
- Should.

Having worked with researchers, Patrick realised that these words need some clarification. Some of these words need to be used with caution, whereas it is best to avoid others. For example, sometimes the best way to ask someone to fill in a form is indeed to say, 'Please fill in this form'.

The next obstacle after developing a communication strategy with researchers was to actually find who were the imaging researchers. A number of very simple strategies were used for this. Previous editions of *Monash Memo* (2015), *Monash News and Events* (2015) and the *Delivering Impact* (2015) magazine were scanned for authors and references to Monash University researchers working with imaging data. The subject librarians who worked in that area were also interviewed, and later, when we finally did speak with researchers, referrals to other researchers were sought.

Following the identification of imaging-data researchers, we then developed a two-part strategy to make contact with researchers. The first hurdle was to try to meet with researchers face-to-face. This contact was usually made by email where it was important to state the value that the library could create for researchers. The second part of the strategy, after having met initially with the researchers, was to follow up with him or her regarding outstanding tasks. The primary outstanding task was for researchers to complete a form to allow their research data to be made publicly available on ANDS' Research Data Australia website.

We had some very good responses initially, from researchers who were interested in opening their data. Researchers talked about the MODC Project, using phrases such

as being 'very exciting' and 'greatly interested', even going so far as to say 'I would love to be a part of this'. This positive initial contact with researchers was from early career researchers, such as lecturers, through to established researchers, such as full professors.

1. Research data management guidelines

The first major identified outcome of the MODC Project was the creation of an open access policy. This is aimed at researchers at Monash who use instruments such as transmission electron microscopes and other imaging instruments, and promotes an understanding that there is an expectation to make their datasets publicly accessible.

This policy was called the Monash Technology Research Platforms Guidelines - Data Management. A great deal of collaboration and consultation was sought to draft the Guidelines. Initially Monash University looked at what similar academic institutions were doing for open access and sought to amalgamate those guidelines into a similar document that would suit the purposes of Monash University. Then began the process of culling the parts that did not fit in with the Monash research culture and adding other parts, making sure the language accurately reflected the needs of Monash University researchers.

The Monash Technology Research Platforms Guidelines - Data Management have since been circulated numerous times amongst the research staff for comment, and have been discussed at several eResearch Committee meetings. When these Guidelines pass the strict standards set by the Monash University research community, they will be endorsed by the eResearch Committee and become a formal open access policy.

The most important issue regarding the Guidelines is to encourage researchers to make research datasets of their choosing publicly available. Researchers are the experts in their fields of research and as they are closest to the research are in the best position to determine what datasets can be publicly released.

Of course, there are instances when datasets cannot be made publicly available, including the requirements of legislation, ethical, contractual or confidentiality obligations, or the potential for further research or publications.

2. Sustainable operational environment

Researchers are concerned about data storage, data management including sharing with co-workers, and publishing their datasets. Therefore, the MODC project needed to do three things. It needed to ensure that the data was diligently safeguarded, that researchers could safely collaborate within and external to their research teams, and that when they chose to publish the datasets, that they could do so as widely as possible.

Monash University, like most research institutions, conducts groundbreaking research and produces vast amounts of data that must be managed. Monash software developers worked in collaboration with researchers from the Australian Synchrotron, as well as from within the University, to develop MyTardis, an open-source highly automated web application that allows the acquisition and

management of large datasets at the point of creation in a private and secure manner (Splawa-Neyman et al 2015).

Figure 1 shows how MyTardis fits in with the data lifecycle. At the top of the figure are instruments such as the Australian Synchrotron and the various microscopes at Monash University. MyTardis allows for the capture of images and management of the data at the time the data is created. This is achieved in a number of ways: via long-term data storage, by allowing the sharing of data with other researchers either through a time-limited link or by adding the authentication details of a researcher from another institution, and by allowing researchers to publish their data when they choose to.

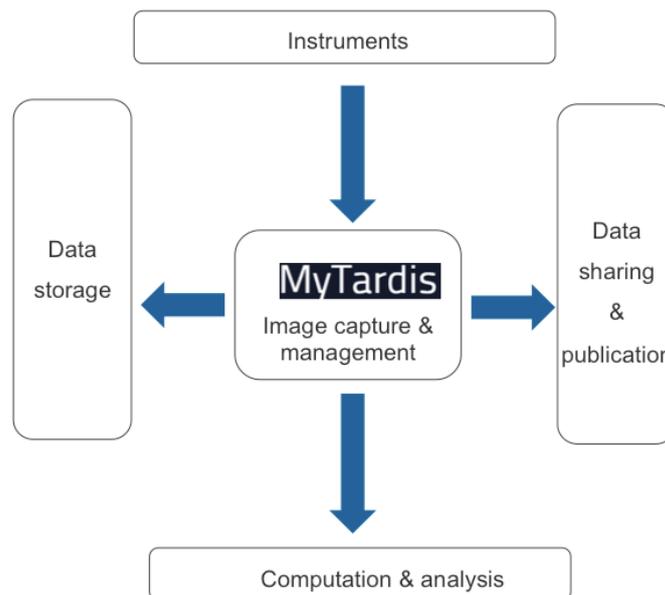


Figure 1: MyTardis and the data lifecycle

Within this research data lifecycle, MyTardis allows for the high performance computation and analysis of the research data, which is from where the real value in the data comes.

A researcher who finds a record in MyTardis (2015) will discover a great deal of information before even looking at the data. He or she can determine the title of the experiment, and which researchers conducted the experiment. There is also a short description and information identifying under which Creative Commons licence the data has been released. The number of datasets, number of files and the total size of the experiment can also be found here.

MyTardis is undergoing further development, but has been embraced enthusiastically by the research community.

3. Monash Imaging Locus

The Imaging Locus is a web presence that allows research data at Monash to be discoverable, and to display the infrastructure and the expertise at Monash.

The purpose of the Imaging Locus is to showcase imaging research at Monash University and to demonstrate to researchers around the country and around the globe the high standard of research, as well as the support provided to researchers at Monash University.

The Imaging Locus (2015) sits under Research Infrastructure on the Monash University website. This location situates it alongside the technology research platforms, Monash-partnered national research infrastructure, and Monash technology capabilities.

The Imaging Locus allows public users to search for imaging datasets created by Monash researchers. Much of the content currently in the Imaging Locus has been harvested from experiments completed by Monash researchers, where the data is stored by MyTardis on Monash University long-term storage.

All data in the Imaging Locus have been released by researchers under Creative Commons licences. The results displayed in the Imaging Locus show who created the data, the title of the experiment, the Field of Research codes (2012), and, where appropriate, the link to the Protein Data Bank (2015).

The longer-term plan, already under development, is for monash.figshare (2015) to become the data repository for Monash University. This customised version of figshareⁱ allows Monash researchers to upload, manage and describe their own research data without the intervention of library staff.

There are a number of advantages to monash.figshare including:

- The creation of a Digital Object Identifier
- Secure data storage
- Data release under a Creative Commons licence
- The ability to embargo data
- Greater visibility of research.

Patrick's reflections

One challenge was learning how to interact with researchers while ascertaining how to complete the project. I felt great pressure to learn as much as I could as quickly as possible, so that I could make a significant contribution as soon as possible. Researchers are busy and want to work with people who can make their jobs easier, therefore I had to put aside self-doubt, throw myself into the role and have faith in my own abilities.

Working on the project while simultaneously learning about the research culture was another huge challenge. My first career was in social work, my second was in pharmaceutical sales, and so in addition to working on a project in a new environment, I was learning about an entirely new and at times seemingly strange culture.

My greatest challenge, however, was advocating to researchers for open data. Researchers work very hard to write proposals, obtain grants, conduct research and publish their results. It was very difficult at times to firstly meet face to face with

researchers, and secondly promote the idea of making their data accessible to anyone with an internet connection.

When I was studying the Master of Information Management at RMIT University, I knew I wanted to work in an academic library and I always had my sights set on Monash University. However, I was not familiar with research data management until I applied for this role. I learned very quickly on this project how much I enjoyed working with researchers. Researchers are the most intelligent, passionate and hard-working people I have ever had the pleasure of working with. There are high expectations from researchers, but great rewards when working alongside them, such as helping make publicly funded research data publicly available.

I also discovered how quickly I can learn under pressure. There was a clear expectation that, as the Data Librarian, I was the expert on certain topics such as Creative Commons licences and metadata, and whether or not I was the expert, I had to be. It was important to be intrepid and throw myself into as many tasks as I could. Even though I learned a great deal in those twelve months, it is clear to me how much more there is to learn. I will never conquer this field; I will just keep learning and growing.

Swinburne University MODC Project

The Swinburne Institute for Social Research (Swinburne University of Technology) established Australian Policy Online (APO) in 2002, as an online cataloguing and dissemination service for policy related research resources. It has since grown in scope to become a diverse platform model – while it still embraces a core mission to disseminate the latest policy research resources in a broad range of subject areas, it is also an open access repository housing 13 years' worth of content (over 30,000 research resources). In addition, it is a collaborative user-focused platform encouraging interaction with various tools, and allows user-generated cataloguing of content from an audience of:

- Government (policy analysts and advisors, administrators, researcher-practitioners, politicians)
- Education (university researchers, lecturers, students)
- Non-governmental organisations (research analysts, charity and community sector workers)
- Health and medical practitioners, legal and business companies
- Journalists, librarians.

Sherratt (2015) writes about the National Library of Australia's cultural information resource Trove as a platform: 'For me the term 'platform' speaks of something unfinished – an invitation and an opportunity. Trove is permanently under construction, constantly improved through the labours of its developers and community'. Similarly, APO has been constantly improved by its editors and managers since its formation in 2002, and today is sustained and continually reimaged by both its staff and an engaged community from the public policy sector in many different fields of research. The community-contributed model that allows users to register and upload research or data in many different subject areas for moderation and subsequent online publication is one of the strengths of the APO platform, as user engagement and participation drives interest areas and enhances

the collection. As with Trove, where each added user-created tag can aid discoverability, add context, and create new connections (Sherratt 2015), APO creates new connections with organisations, government, and public policy researchers to enhance and promote the significant value of grey literature research.

The APO project

The Swinburne University Major Open Data Collections (MODC) project was supported through one of the Australian National Data Service's (ANDS) core missions: 'to build research data capability across the research and scholarly communications lifecycle in organizations, systems, services and people' (Treloar 2012, p. 175).

Just as ANDS is 'working to enable the transformation of research data that are unmanaged, disconnected, invisible and single-use into structured collections that are managed, connected, findable and reusable', (Treloar 2012, p. 173), the APO repository exists to provide sustainable, long-term access to public policy research and data, and maintains that it is essential to find better ways to 'access, control, evaluate, collect and preserve grey literature' (Lawrence 2012, p. 8). Throughout its existence, APO has seized opportunities to create new partnerships and develop new services, and this move into the research data management field is an example of this willingness to continually improve the platform.

The Swinburne MODC began in 2014, involving a project to mint Digital Object Identifiers (DOIs) for the APO collection in order to increase citation of important and influential policy research, contribute to managing the ongoing stability of resources, and also increase the credibility of grey literature research data, which has often not been fully recognised as being a valuable part of the citable scholarly record, due to not being associated with recognised research publications. Another aim was to increase researcher and organisation awareness of the importance of collecting and managing associated research data for publications.

As well as making APO resources more discoverable, citable, and usable, this project also aimed to:

- Contribute to eResearch competency at Swinburne University
- Promote collaboration with both internal and external researchers and organisations
- Enable the wider public policy community to discover, share, and reuse high quality research
- Increase the visibility of Swinburne University research.

The project involved the implementation of a DOI-minting lifecycle for research data at APO, beginning with the development of two separate policies: one in-depth procedural manual for APO to mint and manage DOIs for its resources, and one developed at the broader institution level for all Swinburne University departments considering DOI minting in the future. Prior to this project, Swinburne had not implemented an ANDS DOI minting service, therefore it was necessary for this project to incorporate both the construction of a university-wide consistent approach to the minting of DOIs for resources, and develop local APO repository processes.

The project involved several stages, from writing the policies to building the infrastructure to support DOI minting on the open source Drupal system that APO uses. Various questions required answering, such as what resources should have DOIs, and at what level of granularity, and long-term versioning and change management guidelines for resources had to be developed.

Identifying project scope and boundaries

One of the biggest challenges for this project was defining a 'roadmap' or best-practice guide for implementing a complete DOI minting lifecycle for grey literature research data.

In order to identify the basic elements that would be required for an APO-specific framework to use to structure the different stages required for the project, various case studies and data citation experiences from other universities around Australia and internationally were consulted. These included Griffith University's data citation project in 2012-13 and the management guide for DOIs (Simons, Searle & Lee 2013), a case study from the University of Bristol (Gray & Duke 2014), and other initiatives at the CSIRO and the Australian Antarctic Data Centre.

The project scope and boundaries were identified early in the process in order to provide a framework:

- The need to develop policies and procedures for identifying research in the repository suitable for DOI minting
- The need to liaise with data providers and producers to confirm permission to mint DOIs
- Researching how best to implement the ANDS Cite My Data service and the DataCite (DataCite 2015) metadata schema with the APO Drupal system.

In the creation process for the APO policy document, it was determined that to enhance long-term management capability such as providing persistent access to content, APO would be the caretaker platform and would require permission to host the full text for any resource to be given a DOI. The size of data and sensitive data were not issues for this project due to the nature of the resources, and the APO requirement that full open access to the content must be granted before minting DOIs.

Collaboration and outreach

As part of the DOI project, APO aimed to increase awareness and educate about the role of data citation for important policy research, and advocate for the importance of assigning persistent identifiers to research data. The project involved advocacy for the benefits of grey-literature research data citation to a number of external organisations that provided their resources for DOI minting (see Table 1). Working out how best to consult and liaise with organisations with different organisational cultures and attitudes about the importance of good curation and citation of research data (NGOs, think tanks, education and research centres) was challenging. This was achieved by creating various flyers and educational blog posts (APO 2014, 2015a, 2015b), frequent email contact with organisational representatives, and by seeking 'research champions' – individuals within organisations who understand the benefits of having persistent access to material, and who do not want their important

resources to remain buried in the digital deluge of information often found online on organisation websites.

Name of organisation	Sample DOI resource
Australia and New Zealand School of Government	http://dx.doi.org/10.4225/50/55821631B89F5
Australian Academy of Technological Sciences and Engineering	http://dx.doi.org/10.4225/50/555C25C88E7A3
Australian Research Alliance for Children and Youth	http://dx.doi.org/10.4225/50/557E0CB86CB5D
Australian Research Centre in Sex Health and Society, La Trobe University	http://dx.doi.org/10.4225/50/557E59BB7EC17
Australian Workplace Innovation and Social Research Centre, University of Adelaide	http://dx.doi.org/10.4225/50/557F8A642DD5C
Centre for Community Child Health, The Royal Children's Hospital Melbourne	http://dx.doi.org/10.4225/50/5578DA99168A5
ClimateWorks Australia	http://dx.doi.org/10.4225/50/555BCE11E7318
Committee for Economic Development of Australia	http://dx.doi.org/10.4225/50/557639AC46698
Deeble Institute for Health Policy Research, Australian Healthcare and Hospitals Association	http://dx.doi.org/10.4225/50/557E6B7686933
News and Media Research Centre, University of Canberra	http://dx.doi.org/10.4225/50/557F89CA26B7F
Regional Development Australia	http://dx.doi.org/10.4225/50/557104A88F61B
Swinburne Institute for Social Research, Swinburne University of Technology	http://dx.doi.org/10.4225/50/557F89CA26B7F

Table 1: External organisations and sample DOIs

Collaboration and networking with other university groups running data citation projects around Australia through monthly meetings organised by ANDS proved invaluable for sharing experience and outcomes, particularly issues associated with assigning granularity of research data in the policy and procedure documents,

getting institutional support, and governance and caretaker issues around ensuring the persistence of DOIs.

Project outcomes

As the project ended in June 2015, the total number of DOIs minted for APO resources was 288, from 20 organisation contributors (see Table 2).

1	Anglicare Australia
2	Australia and New Zealand School of Government
3	Australian Academy of Technological Sciences and Engineering
4	Australian Research Alliance for Children and Youth
5	Australian Research Centre in Sex Health and Society, La Trobe University
6	Australian Workplace Innovation and Social Research Centre, The University of Adelaide
7	Benevolent Society
8	Centre for Community Child Health, The Royal Children's Hospital
9	Centre for Social Impact (UNSW Australia, Swinburne University of Technology, The University of Western Australia)
10	ClimateWorks Australia
11	Committee for Economic Development of Australia
12	Deeble Institute for Health Policy Research, Australian Healthcare and Hospitals Association
13	Evidence Base (journal)
14	Media International Australia (journal)
15	Mission Australia
16	News and Media Research Centre, University of Canberra
17	APO, Swinburne University of Technology
18	Regional Development Australia
19	Salvation Army Australia
20	Swinburne Institute for Social Research, Swinburne University of Technology

Table 2: Organisation contributors to DOI project

The project enabled the bridging of existing gaps between Swinburne University library, Swinburne Research, and APO, particularly as a result of developing the Swinburne-wide policy guide for minting DOIs, which is continuing to influence how DOIs for research data are managed at the university.

This project has provided many ongoing benefits, including:

- Encouraging organisations to provide better metadata for their publications and to know about the importance of citation for research
- Giving permanent, persistent access to resources in an open source system
- Implementing research data management policies and procedures to guide decision making for ongoing management of resources
- Contributing to the establishment of an institutional culture of citation and sustainable access to information
- Incentivising organisations and implementing processes and platforms that make collaboration easier.

Jaye's reflections

Through my role as Information Project Officer on this ANDS-funded project, I have worked and shared ideas with numerous different collaborators, such as IT professionals, business analysts, knowledge managers and archivists. At the beginning of this project, I considered myself a neophyte data librarian because I faced many challenges armed only with an almost-complete Masters degree in information management, and some surprisingly useful transferable skills from my previous career in the film and television industry. I found that highly developed communication skills are essential for managing research data projects, and the new information challenges that data present mean that collaborative skills sharing is essential for implementing good processes that allow resources to remain retrievable, searchable, and accessible for the future.

During the course of the project, I have been able to implement new research data management practice for APO, and new standards and better practice for managing grey literature resources. Implementing policy and procedures for DOI minting enabled the creation of new workflows for other APO processes previously lacking cohesion, such as the workflow for fixing dead links. This project also laid the groundwork for the implementation of related identifiers and a better system for inputting organisations as 'creator' entities.

The future brings new challenges and projects. New projects include more advocacy work for the benefits of grey literature research data citation to external organisations, implementing ways to track citations and measure the impact of DOIs for the collection, and working out how to explore deeper levels of granularity for resource cataloguing in the future, especially when linking data to publications.

I am developing a new 'resource type' taxonomy structure for the APO grey literature resources, and am keenly observing the Drupal 8 development cycle (Drupal 2015) to help contribute to the management of resources during the technical upgrade of the database in the coming months. I am also monitoring the now embedded DOI-minting system and continuing to solicit organisational interest in minting DOIs, and am learning principles of information architecture in order to improve accessibility and facilitate serendipitous browsing of the APO repository. I believe that my

'Information Project Officer' role will be one without a defined roadmap (or a pithy definition) for some time to come.

Conclusion

As neophyte data librarians being involved with such challenging projects, we feel that there is a need for institutional training and professional development programs in the area of RDM. Auckland's (2012) research determines nine areas in RDM showing a high skills gap that should be addressed by implementing training and development, including knowledge to:

- Advise on data management and curation, including ingest, discovery, access, dissemination, preservation, and portability
- Support researchers in complying with the various mandates of funders, including open access requirements
- Advocate, and advise on, the use of metadata.

The two case studies presented in this paper are testament to the fact that the RDM skills described in Auckland's research can be developed through project work; a research data manager does not necessarily need to possess all of the skills thought to be required at the beginning of projects such as these. Certainly these projects also illustrate that while it may currently be necessary for organisations to grow their own talent through various on-the-job project work and training, it is important that training opportunities be put in place to allow research data managers to improve and learn new skills. While a 2015 Jisc report examines the idea of development on-the-job being necessary to a certain extent in such a rapidly changing field, it also maintains that 'almost all staff involved with RDM will need training' in the future (Brown, Bruce & Kernohan 2015, p. 12).

Any new definition of the term 'data librarian' needs to take account of the many and varied personal and technical skills that encompass the work of today's information professionals. However, a significant difficulty in defining this term is that while data librarianship is a subset of librarianship, it is still a vast sub-discipline. Any definition that includes the whole of data librarianship must necessarily be indeterminate.

Regardless of what label this role is given, it is clear that it encompasses a wide range of skills and abilities, borrowing from traditional librarianship, IT, archival principles and practice, and knowledge management. The information revolution demands a new breed of information professional able to display these abilities, as well as advanced research and analytical skills.

References

- APO 2014, *Announcing the ANDS funded project - Digital Object Identifiers (DOIs)*, viewed 3 November 2015, <http://apo.org.au/APO-News/announcing-ands-funded-project-digital-object-identifiers-fois>
- APO 2015a, *DOI project update*, viewed 3 November 2015, <http://apo.org.au/APO-News/doi-project-update>
- APO 2015b, *Policy Online now minting DOIs*, viewed 3 November 2015, <http://apo.org.au/APO-News/policy-online-now-minting-fois>
- Auckland, M 2012, *Re-skilling for research: an investigation into the role and skills of liaison librarians required to effectively support the evolving information needs of researchers*, Research Libraries UK, viewed 1 August 2015, <http://www.rluk.ac.uk/wp-content/uploads/2014/02/RLUK-Re-skilling.pdf>
- Australian Library and Information Association (ALIA) Canberra 2014, *The Library and Information Sector: Core Knowledge, Skills and Attributes*, viewed 16 August 2015, <https://www.alia.org.au/sites/default/files/The%20Library%20and%20Information%20Sector%20Core%20Knowledge%20Skills%20and%20Attributes%20December2014%20%282%29.pdf>
- Australian National Data Service (ANDS) Canberra 2015a, *Research data and librarians?*, viewed 14 August 2015, <http://www.ands.org.au/guides/dmframework/dmskills-information.html>
- Australian National Data Service (ANDS) Canberra 2015b, *Monash University Imaging Locus Project*, viewed 17 July 2015, <https://projects.ands.org.au/id/MODC01>
- Brown, S, Bruce, R & Kernohan, D 2015, *Directions for research data management in UK universities*, Jisc, viewed 1 August, 2015 http://repository.jisc.ac.uk/5951/4/JR0034_RDM_report_200315_v5.pdf
- DataCite 2015, *DataCite Metadata Schema Repository*, viewed 3 November, 2015 <https://schema.datacite.org>
- De Vries, R 1986, 'Providing local data services', *IASSIST Quarterly*, vol. 10, no. 2, pp. 46-47.
- Delivering Impact 2015, viewed 20 August 2015, <http://www.monash.edu.au/monashmag/>
- Drupal 2015, *Drupal 8 development cycle*, viewed 15 August, 2015 <https://www.drupal.org/core/dev-cycle>
- Galvin, T 1995, 'Convergence or divergence in education for the information professions: an opinion paper', paper presented at *ASIS Annual Meeting*, 9 October, Chicago, viewed 16 August 2015, <https://www.asis.org/Bulletin/Aug-95/galvin.html>
- Gray, S & Duke, M 2014, *Assigning Digital Object Identifiers to research data at the University of Bristol*, Digital Curation Centre, viewed 3 August 2015, <http://www.dcc.ac.uk/resources/persistent-identifiers>

- Groenewegen, D 2014, 'Becoming a (Research) Data Librarian: Managers perspective', viewed 11 August 2015, <https://www.youtube.com/watch?v=SSoitBXbzpM>
- Howe, SR 1993, 'Challenges to the Data Librarian', *Social Science Computer Review*, vol. 11, no. 1, pp. 7-14.
- Imaging Locus (Monash University) Melbourne 2015, viewed 1 September 2015, <https://www.monash.edu/research/infrastructure/imaging-locus>
- Lawrence, A 2012, 'Electronic documents in a print world: grey literature and the internet', *Media International Australia*, vol. 143, pp. 122-131, viewed 11 August 2015, <http://dx.doi.org/10.4225/50/5580B0D739AF5>
- Liscouski, J 1997, '2. The Data librarian: introducing the Data Librarian', *Journal of Automatic Chemistry*, vol. 19, no. 6, pp. 199-204.
- Monash.figshare, Melbourne 2015, viewed 20 August 2015, <https://monash.figshare.com>
- Monash Memo 2015, viewed 20 August 2015, <http://monash.edu/news/memo>
- Monash News and Events 2015, viewed 20 August 2015, <http://monash.edu/news/>
- MyTardis, (Store.Synchrotron) Melbourne, 2015, viewed 1 September 2015, <https://store.synchrotron.org.au>
- Penn State University 2015, *Data management plan tutorial*, viewed 15 August 2015, <https://www.e-education.psu.edu/dmpt>
- Protein Data Bank (RCSB Protein Data Bank) California 2015, viewed 1 September 2015, <http://www.rcsb.org/pdb/home/home.do>
- Ratzek, W 2012 'The Digital Library: A Multi-Faceted Information and Communication System', *International Journal of Digital Library Systems*, vol. 3, no. 4, pp. 62-70.
- Research Data Australia (Australian National Data Service) Canberra 2015, viewed 28 August 2015, <https://researchdata.ands.org.au>
- Sherratt, T 2015, 'Seams and edges: dreams of aggregation, access & discovery in a broken world', paper presented at *ALIA Online 2015*, 3 February, Sydney, viewed 16 August 2015, <http://discontents.com.au/seams-and-edges-dreams-of-aggregation-access-discovery-in-a-broken-world>
- Simons, N, Searle, S & Lee, S 2013, *Digital Object Identifiers (DOIs): introduction and management guide*, Griffith University, viewed 1 August 2015, http://ands.org.au/cite-data/griffith_doi_guidelines-4.pdf
- Splawa-Neyman, P, Dickson, N, Kannan, A & Groenewegen, D 2015, *Australian Open Research Data Showcase*, Australian National Data Service, Canberra, 19 June 2015, viewed 18 August 2015, <https://dx.doi.org/10.4225/03/55BABDC03C1E1>
- Treloar, A 2012, 'Contrasting national research data strategies: Australia and the USA', in G Pryor (ed.), *Managing Research Data*, Facet Publishing, London, pp. 173-203.

University of California 2015, *DMP tool*, viewed 15 August 2015, <https://dmp.cdlib.org/>

Endnotes

ⁱ <https://figshare.com/>