THE GREAT RESEARCH DATA SCAVENGER HUNT

Matthias Liffers, Research Coordinator, Curtin University Library

Presented at ALIA 2016 Conference by Catherine Clark, University Librarian, Curtin University

ABSTRACT

Introduction

Curtin University Library has made the development of research data management (RDM) services a strategic priority. Most of the work is undertaken by a small team and there was not much opportunity to include the rest of the Library in RDM activities. The RDM Coordinator suggested a six-month strategic initiative, the “Great Research Data Scavenger Hunt”, to crowdsource a team of interested Library staff and augment Curtin University’s modest collection of published datasets.

Methods

Expressions of interest were sought from any interested Library staff member – regardless of qualification, expertise or existing duties. 15 participants were drawn from all sections of the Library and placed into 5 teams with colleagues they did not normally work with on a day-to-day basis. Each team was asked to find only one dataset to be published using Curtin University’s data publication system, emphasising that this project should not impact too much on day-to-day duties. Training was provided on a just-in-time basis, with the official kick-off training being delivered by staff at the Australian National Data Service.

Results

Response from researchers to a targeted mail out was poor, which was not unexpected. Thankfully, enough datasets were found using existing networks between library staff and researchers. Each dataset received a DOI and was syndicated to Research Data Australia. Participants gained experience in working with researchers and knowledge in various RDM topics including metadata creation, digital preservation and copyright.

Conclusion

The Great Research Data Scavenger Hunt was a novel method by which to engage with library staff about research data management – especially staff that might not normally have the opportunity to work in this area. As a means by which to augment a library of published data, however, it was very labour intensive for little return.

PAPER

Background

Curtin University is a large technical university with approximately 60,000 students with campuses in Perth, Kalgoorlie, Singapore and Malaysia. In 2014, Curtin introduced a Research Data and Primary Materials Policy that triggered the development of services to assist researchers in managing their data. These services were developed as a collaboration between the Curtin Library, Curtin Information Technology Services and the Office of Research & Development. Primary responsibility for delivering training and support fell upon the Library.

A Coordinator, Research Services (the Coordinator) was assigned to spearhead the Library's contribution to research data management (RDM) support. The first task was to undertake an environmental scan of all Australian universities libraries and how they provide RDM support. This scan involved analysis of publicly-available information on websites and also interviews with representatives from universities perceived as providing good RDM support.

One common theme that arose during this environmental scan was the issue of training library staff in RDM support. A variety of training was discussed, ranging from theoretical lessons such as MANTRA (EDINA and University of Edinburgh Data Library, n.d.) to practical exercises such as working directly with researchers to publish datasets. There was also a middle ground such as Searle's scenarios in introductory RDM workshops for library staff (Searle, 2015). Most of these approaches involved training only staff who were going to be directly involved in RDM support and had the potential to be very time consuming. Barring the Digital Mineral Library project (Liffers, Brown, & McInnes, 2015), RDM support at Curtin

Library was largely undertaken by only the Coordinator and there was no need for in-depth training of any other staff.

During the course of 2014, the first version of RDM support was developed and the Coordinator regularly reported progress to colleagues in information sharing meetings. During one of these meetings, one colleague expressed that she and her team were interested in learning more about RDM, even though the team's day-to-day operational tasks were completely unrelated to the area. In essence, they were motivated by a professional interest in an emerging area of librarianship.

Concurrently, the Library was building a modest collection of datasets published with an open licence – such as Creative Commons – managed with the ReDBox system. Unfortunately, this collection was *very* modest. Apart from 150 datasets that had been published for the Digital Mineral Library, fewer than 10 other datasets had been contributed by researchers, mostly in response to open data mandates from publishers – an indication that the Library's open data publication facility needed much more promotion.

With this relative paucity of datasets in mind, the Coordinator was inspired to bring together the best features of various training programmes to create an introduction to RDM support for Curtin Library staff with the flippant working title of "The Great Research Data Scavenger Hunt", with the equally flippant idea of providing cookies to participants. Unfortunately, the working title stuck and was unable to be changed to something more reasonable during the annual planning process. This project would serve the dual purpose of exposing Library staff to RDM and also bolster the open data collection.

Enter the Scavenger Hunt

The Great Research Data Scavenger Hunt (the Scavenger Hunt) was devised as a training programme in which participants would form teams, contact researchers and work with them to publish at least one of their datasets. The Coordinator designed the Scavenger Hunt to be:

- Practical;
- Low impact to existing duties;
- Voluntary;
- Open to any member of library staff;

And above all:

- Fun!

These qualities were carefully selected, each for their own reasons.

Practical

Taking active learning one step beyond Searle's approach with scenario-based learning, working with real-world datasets would serve the dual purpose of training staff as well as bolstering the Library's modest collection of datasets.

Minimal impact to existing duties

As with many other cultural institutions, Curtin Library must balance budgetary constraints with its task to provide high-quality information services. In order to get buy-in from Line Managers, they needed to be assured that participation in the programme would have minimal impact on the day-to-day duties of their staff. To support this goal, the entire Scavenger Hunt was scheduled to run over six months.

Voluntary & open to any member of library staff

Very few staff were actively involved in providing RDM support and the intention of the Scavenger Hunt was to give any staff member interested in RDM to the opportunity to learn more. Motivations could include a general professional interest in a new area of librarianship or even the first step of a new career path. The Scavenger Hunt was open to anyone, regardless of level, qualifications or role. All that was needed was their Line Manager's approval and the submission of an Expression of Interest. The Scavenger Hunt would benefit from the diverse experience that staff could bring to the project.

Fun!

Not easy, but certainly fun. The Scavenger Hunt was to stimulate staff and encourage them to engage with something completely new to them. Since participation was entirely voluntary, the Coordinator was aware that a boring or onerous project might drive staff away.

Structured fun

The whole Scavenger Hunt was initially planned to take place over nine months in 2015. Table 1 reflects the plan, which later proved to be somewhat ambitious.

Table 1 - Planned project timetable

What	**When**
Expressions of interest from staff to participate	March
Form teams with wide range of skills and expertise	Early April
Official launch with training provided by Australian National Data Service	Late April

Identify and contact researchers	May
Assign teams to researchers	June-July
Work with researcher to: • Learn about data through a data interview • Determine terms of publication (licence) • Provide advice on file formats for long-term access • Transfer dataset to data store • Describe dataset using Library's metadata hub	June-October
Showcase data collections at finale event for library staff	November

As stated previously, all Library staff were welcome to participate. In the end, a total of 15 participants from across the Library submitted Expressions of Interest, covering a range of levels and expertise. The participating staff roughly represented the relative sizes of each Library section, with the exception of Research Services, where almost all staff chose to participate.

Table 2 - Breakdown of level and expertise

	Corporate services	**Learning services**	**Research services**	**Technical services**
LIS Student	1			
Technician			3	1
Librarian		2		3
Coordinator		1	2	1
Manager			1	

Potential participants were grouped based on existing knowledge and experience in several areas:

- Metadata creation
- Liaison service
- Digital preservation
- Copyright

Ideally, each team had at least one person to provide each one of these skills. Digital preservation was the most poorly understood, as it is not a traditional area of strength for the Curtin Library. No knowledge of research data management was required, but some staff had previously been involved in the Digital Mineral Library project. The Manager and Coordinators participating were told that they were not expected to lead their teams, and were asked to encourage other members to be leaders.

Training was planned to be provided as needed to the participants as they entered each phase of the project. For example, training in required metadata was intended to only be provided after the initial data interview. This was intended to flatten the learning curve and not overwhelm participants with information at the outset of the Scavenger Hunt. Unfortunately, this approach later proved to be a mistake.

The official launch of the Scavenger Hunt was a three hour introduction to research data provided by the Australian National Data Service on their visit to Perth in April 2015. Further training sessions were facilitated by the Coordinator and were generally one hour in duration.

Although Curtin University had already implemented many data management services, the notion of publishing data as an independent research output that supports articles, as opposed to supplemental material to articles, was new to Curtin

researchers. The Coordinator anticipated that researcher response might be poor and contacted data management champions to secure “just in case” datasets.

The best laid plans

As with many projects, the Scavenger Hunt encountered several hurdles that delayed its completion.

Lack of researcher uptake

The Scavenger Hunt participants identified 52 researchers to contact based on previous interactions with the Library, such as providing manuscripts for green open access to espace, Curtin’s institutional publication repository. These researchers were contacted by the Library’s Faculty Librarians, who have well-developed networks within the Faculties. Of the 52 researchers, approximately 20% responded and only half of those expressed a desire to learn more about the Scavenger Hunt. When provided with more information, only two researchers agreed to take part.

This lack of response was mitigated by one participant already knowing a researcher who had some data to publish, and the “just in case” datasets that the Coordinator had already secured.

To compound the lack of researcher uptake, the two researchers who initially expressed interest pulled out due to their own research and teaching commitments. Thankfully, the champions’ datasets organised previously by the Coordinator helped mitigate this outcome.

The greatest impact felt by lack of researcher uptake was on the Scavenger Hunt’s timeline. Although originally scheduled to be complete by November 2015, the end-

of-project celebration was not until March 2016, and the final showcasing to other Library staff did not take place until April 2016.

Drawn out project

The Scavenger Hunt was planned to have minimal impact on existing duties, and therefore the Coordinator's original timeline was quite generous – six months from start to finish. This timeline certainly did give participants the opportunity to work the project around their existing commitments, but conversely the drawn-out nature meant that different teams completed the Hunt at vastly different speeds, with one team completing the entire process before many of the others even had a chance to meet with their researcher. This meant that the Coordinator's original idea of gradually introducing concepts to all teams simultaneously, during monthly or bi-monthly training sessions did not achieve the participants' training needs as participants either needed the information earlier, or would forget the content as they did not apply it until a month or two later.

Furthermore, the drawn-out process may have also contributed to researchers' commitments getting in the way of contributing datasets to the Scavenger Hunt – they had spare time at the time of the original contact, but the delay in mobilising to meet them may have meant lost opportunities.

Bigger and better things

During the course of the Scavenger Hunt, some staff movement was experience as participants changed jobs or left the University. Fortunately, this did not have a great impact on the other participants. Only one team was missing a team member for most of the Scavenger Hunt, but the remaining members were more than able to complete the work between them.

Apart from staff leaving the University, every single other participant remained committed to the Scavenger Hunt until the end, despite the timeline extending beyond initial estimates.

Mission complete

In March 2016, the remaining participants got together to celebrate the end of the Scavenger Hunt, and the Coordinator provided cookies that were promised when the Hunt was first proposed as a Strategic Initiative.

The published datasets

The datasets that were published for the Scavenger Hunt represent a range of the disciplines at Curtin University in health science, humanities and physical sciences.

- Clinical Guidelines for Management of Bone Health in Rett Syndrome Based on Expert Consensus and Available Evidence.
  DOI: http://dx.doi.org/10.4225/06/5697464CB9EE4 (Jefferson, Leonard, Fyfe, & Downs, 2016)
- Content Analysis of Indonesian Higher Education Institutional Repositories.
  DOI: http://dx.doi.org/10.4225/06/565BACB4C3CBA (Liauw, 2015)
- Photos of public open spaces in the City of Swan.
  DOI: http://dx.doi.org/10.4225/06/564405E287439 (Middle, 2015)
- SIESTA input and output files for calculations on γ-glycine.
  DOI: http://dx.doi.org/10.4225/06/564AA162170CB (Rohl, Carter, & Kahr, 2015)
- Unscrambler X input and selected output files for chemometric analysis of blue ballpoint inks.

DOI: http://dx.doi.org/10.4225/06/56CFB3F39A3C2 (Sauzier, Giles, Lewis, & Bronswijk, 2016)

Just 23 more (research data) things…

Although the Scavenger Hunt gave participants the opportunity to work through the process of publishing a dataset from the start to the very finish, it was not as in-depth as many of them would have liked. Discussions around some topics, such as copyright, metadata and long-term preservation were generally limited and skirted around some of the more difficult questions.

Appetite amongst participants for continued training in research data management is evidenced by fully 75% of them going on to participate in ANDS' *23 (research data) Things* programme (Australian National Data Service, n.d). In response to this demand, the Coordinator is facilitating a *23 (research data) Things* community group for Curtin University. This group is not just limited to Curtin Library staff, but is also open to staff and students of other parts of the University, such as the Department of Information Studies.

So you think you can Hunt

As research data management becomes a higher priority for research funders and publishers around the world, academic libraries are responding by developing data management and curation services. Research data management is a relatively new field for many academic libraries and there is little in the way of formal training – although at least one Australian university offers a unit (Charles Sturt University, 2016). Thanks to the efforts of organisations such as the Australian National Data Service, there is much material available online that can be used for self-paced

training at little or no cost. Unfortunately, these offerings do not provide much in the way of practical, hands-on experience.

For anyone wishing to run a Scavenger Hunt at their own institution to supplement online training offerings, the Coordinator learnt these lessons:

Build relationships first

Do not expect cold-calling researchers to work, even if you think that they might already be interested. Instead use existing professional networks and leverage those contacts to find datasets. If those networks do not exist, spend some time to build them for the benefit of all research services offered by a library.

Don't dawdle, but be flexible

The Scavenger Hunt at Curtin took a year, from start to finish, which is arguably too long. The pressure of a shorter time frame could encourage participants to work swiftly. Conversely, the timeline must be flexible enough to allow for researchers' competing priorities as they may not have time to respond in a timely fashion during busy parts of the academic calendar.

Take advantage of existing experts

Organisations such as the Australian National Data Service and the Digital Curation Centre in the United Kingdom contain experts that want to share their knowledge and build capability in their respective countries. Do not feel obliged to write all the training material yourself – use what is already on offer.

Share your experiences

Let your colleagues in other organisations know about how you are raising awareness of research data management within your own organisation. Ideally, provide details so that they can adapt, remix and built upon your work.

Acknowledgements

The author would like to acknowledge the participants of the Great Research Data Scavenger Hunt and the hard work they put into the initiative over the year.

Special thanks go to the staff at the Australian National Data Service for the training and support they provided at the official kick-off.

REFERENCES

Australian National Data Service. (n.d). 23 (research data) Things - ANDS. Retrieved March 31, 2016, from http://ands.org.au/partners-and-communities/23-research-data-things

Charles Sturt University. (2016). INF449 Research Data Management (8). Retrieved March 31, 2016, from http://www.csu.edu.au/handbook/subjects/INF449.html

EDINA and University of Edinburgh Data Library. (n.d.). Research Data MANTRA - Library Training. Retrieved March 4, 2016, from http://datalib.edina.ac.uk/mantra/libtraining.html

Jefferson, A., Leonard, H., Fyfe, S., & Downs, J. (2016). Clinical Guidelines for Management of Bone Health in Rett Syndrome Based on Expert Consensus and Available Evidence. Retrieved from http://dx.doi.org/10.4225/06/5697464CB9EE4

Liauw, T. T. (2015). Content Analysis of Indonesian Higher Education Institutional Repositories. Retrieved from http://dx.doi.org/10.4225/06/565BACB4C3CBA

Liffers, M., Brown, A., & McInnes, B. (2015, October). *Major Open Data: The Digital Mineral Library at Curtin University*. Presented at the eResearch Australasia Conference 2015, Brisbane. Retrieved from https://dx.doi.org/10.6084/m9.figshare.1574004.v1

Middle, I. (2015). Photos of public open spaces in the City of Swan. Retrieved from http://dx.doi.org/10.4225/06/564405E287439

Rohl, A., Carter, D., & Kahr, B. (2015). SIESTA input and output files for calculations on γ-glycine. Retrieved from http://dx.doi.org/10.4225/06/564AA162170CB

Sauzier, G., Giles, P., Lewis, S., & Bronswijk, W. van. (2016). Unscrambler X input and selected output files for chemometric analysis of blue ballpoint inks. Retrieved from http://dx.doi.org/10.4225/06/56CFB3F39A3C2

Searle, S. (2015). Using Scenarios in Introductory Research Data Management Workshops for Library Staff. *D-Lib Magazine*, *21*(11/12). http://doi.org/10.1045/november2015-searle