IJDC | *Conference Pre-print*

The Road to Partnership: A Stepwise, Iterative Approach to Organisational Collaboration in RDM, Archives and Records Management

Michelle Harricharan
St George's, University of London

Carly Manson
St George's, University of London

Kirsten Hylan
St George's, University of London

Abstract

Research data management (RDM) sits at the confluence of a number of related roles. The shape an RDM confluence takes depends on several factors including the nature of an organisation and the research that it undertakes. At St George's, University of London, the UK's only university dedicated to medical and health sciences education, training and research, RDM has been intricately interwoven with organisational information governance roles since its inception. RDM is represented on our institutional Information Governance Steering Group and our Information Management Team consisting of information governance, data protection, freedom of information, archives, records management and RDM.

This paper reports on how RDM, archives and records management have collaborated using a stepwise, iterative process to streamline and harmonise our guidance and workflows in relation to the stewardship, curation and preservation of research data. As part of this we consistently develop, conduct and evaluate small projects on managing, curating and preserving data. We present three projects that we collaborated on to transform research data services across each of our departments:

- planning for, conducting and reporting on interviews with wet laboratory researchers
- advocating, building a case for and delivering a university-wide digital preservation system
- ongoing work to recover, preserve and facilitate access to a unique national health database

Learnings from these projects are used to develop our guidance, improve our activities and integrate our workflows, the outcomes of which may be further evaluated. Learnings are also used to improve our ways of working together. Through deeper integration of our activities and workflows, rather than simply aligning aspects of our work, we are increasingly becoming partners on research data stewardship, curation and preservation. This approach offers several benefits to the organisation as it allows us to build on our related knowledge and skills and deliver outcomes that demonstrate greater value to the organisation and the researchers we support.

Submitted 15 December 2019 ~ *Accepted* 19 February 2020

Correspondence should be addressed to Michelle Harricharan, Information Services, St George's, University of London, Cranmer Terrace, London, SW17 0RE, United Kingdom. Email: mharrich@sgul.ac.uk

This paper was presented at International Digital Curation Conference IDCC20, Dublin, 17-19 February 2020

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

It is well established that delivering successful Research Data Management (RDM) services requires cross-departmental collaboration(Pryor, 2014; Rice & Southall, 2016; Verbaan & Cox, 2014; Whyte, 2014). The range of knowledge, skills and expertise required to support data management from the grant application stage through the full lifecycle of a research project and to the curating and archiving stages requires expertise from several related organisational units. RDM is normally led from libraries in close collaboration with university research offices and IT Services(Bradley, 2018; Pinfield, Cox, & Smith, 2014; Söderholm & Sunikka, 2017). There is substantially less in the literature about the important partnership between RDM Services and university Records Management and Archives in documenting, governing, curating and preserving research data. Aligning organisation policies, procedures and standards across these services helps each service to achieve its objectives by sharing knowledge, skills and expertise and addressing shared challenges collaboratively. It also helps to deliver a consistent message about information management and the value of good information management throughout the organisation.

St George's University of London (SGUL) is a specialist health and medical sciences university. Given the kind of research we perform, the bulk of our research data can be considered sensitive. As a result, RDM at SGUL has historically been tightly connected with the organisation's wider information governance structure, including our compliance with the NHS Data Security and Protection Toolkit('Data Security and Protection Toolkit', 2019) which allows our researchers to access data from the UK's National Health Service (NHS). This data is provided by NHS Digital ('NHS Digital', 2019). To ensure that we demonstrate the highest standards of information governance, all of SGUL's information management policies, procedures and standards need to align. Our information management professionals also need to work closely together. As such, we have formed the cross-departmental SGUL Information Management Team which includes information professionals from Data Protection, Freedom of Information, Information Governance, Archives, Records Management and the Research Data. The team is informal, but it was formed out of a shared recognition that our departments need to work closely and collaboratively to streamline our policies and workflows and to achieve our shared objectives.

In this group RDM and Records Management, in particular, are actively involved in supporting day to day information management throughout the research information lifecycle. Archives and Special Collections also has a keen interest in ensuring our research information is managed appropriately to facilitate long term preservation of the content. In order to streamline our guidance, activities and workflows in relation to the stewardship, curation and preservation of research data the three departments embarked on an action research influenced process where we consistently and collaboratively develop, conduct and evaluate small projects(Olsson, Wadell, Odenrick, & Bergendahl, 2010). We use our learnings from these projects to develop our guidance and activities, integrate our workflows and improve our ways of working together. This stepwise, iterative process allows us to harmonise our work in an evidence-based way.

In this paper we will briefly describe how we worked together on three such projects and then discuss the outcomes of this approach for our partnership, including the challenges we faced and how we reconciled these. The challenges we faced, and our strategies for overcoming these, are discussed in the context of the individual projects as this best reflects our iterative process and our learning and development over time. Organising the content in this way, however, often ignores some deeper, more complex, roadblocks to achieving greater partnership in our areas. As such, we've ended with a reflection on our progress so far and some of the more top-level, enduring challenges we face and our working strategies for navigating these.

Projects

Case 1: Wet laboratory records and data project

This work was undertaken by the Research Data Support Manager (RDSM) and the Records Manager (RM) to understand current information handling practices among wet laboratory researchers. Earlier surveys with our researchers following the data asset framework ('Data Asset Framework', n.d.) identified a gap in the support offered to our wet laboratory researchers. This project was conducted to address this gap. We sought to understand how wet laboratory researchers were managing their research records and data throughout the research lifecycle. We also explored researcher perceptions, experiences and challenges with managing their records and data to identify opportunities for support.

Open ended, semi-structured interviews were conducted with wet laboratory researchers to inform the development of policies, procedures and services to support this group. The interview questions followed an existing structure (Read et al., 2015) and covered the types of data researchers generated, as well as researchers' retention, archiving and digital preservation strategies. In practice the interviewees led the conversations and determined the priorities. The work was approved by the Head of Research Operations.

The interviews revealed significant information about the kinds of data and file types our laboratory researchers work with, how they record their data, their data documentation practices, their storage, archiving and preservation strategies as well as their perspectives on data sharing. We noted several challenges researchers faced in handing, sharing and archiving their data and records.

Outcomes

We implemented a series of changes to improve research data management and records management out of this project. Changes included developing and promoting best practices; improving knowledge sharing within the information management team and among researchers; conducting a spin-off study on electronic lab notebooks; creating a joint communications plan to perform shared outreach, training, and advocacy; and creating more aligned policies and procedures. Where issues raised by researchers fell outside of our direct remits, we only reported our findings which were escalated to the relevant departments.

Challenges

This was the first project the RDSM and RM undertook together. This project forced us to confront and negotiate some fundamental differences in the way RDM and records management work.

1. Terminology: The very first challenge we faced before even meeting researchers was understanding each other's language - our different terms for the same concepts and even the same terms we use to refer to different concepts. We will not go into the details here as Ogier et al (Ogier, Nicholls, & Speer, 2017) provide a valuable comparison of the terminology used between data curators and records managers. However, we found that fundamental terms such as 'active', 'archive', 'access(ible)', 'information', 'record' and 'data' meant very different things to each of us even if they were terms we had been using together for some time.

2. Lifecycles: We also realised that we had different starting points when it came to managing information. RDM is interested in the entire lifecycle, from the planning stages of a dataset's creation, through to the long-term reuse of that data over many repeated research cycles. Records Management is interested in designating the data a record at the point of creation and applying institutional policy such as retention which isn't always appropriate or easy to apply for research data. Research data can be used,

reused and referenced by research teams over decades, making them of continuing value (and essentially active records) to the research team. There were numerous examples of laboratory data that was collected over 10 years ago which, from a RM perspective could have passed retention but which were the basis of current research and would likely be the basis of future clinical trials. It is difficult to apply standard retention schedules to broad areas of research which have their own individual contexts. This is also acknowledged by Ogier et al (Ogier et al., 2017).

3. Concepts: At times we struggled to understand concepts in each other's fields. For legitimate reasons there are different accepted good practice for managing, sharing, archiving and preserving our different information types. We needed to understand and define the boundaries or scope of our interest, and, importantly, understand why these areas were of interest to each of us before we could really engage with and undertake this project. One example of this is reconciling the concept of 'lifecycle' between our fields (see 2 above). Another was the concept of 'point of creation'. The organisation has a very specific process for archiving regulated research and clinical trials. This concept of 'archiving' for research purposes (and our standard processes for archiving clinical research) was another area we had to negotiate to ensure we had a similar understanding of the issues.

4. Research and corporate: The RM struggled to understand a lot of the scientific language, processes, roles, file formats and technologies that the RDSM had become familiar with in that role. For example, the interviewees mentioned laboratory samples and the inventory management process several times, particularly their relationship to research records and data. The RDSM and RM had to talk though research samples and sample management, sharing and archiving to ensure they both had the same starting point for interpreting this information going forward. The RDSM in return had spent her entire professional life in research. Understanding RDM as part of a big, complex organisational-level governance network was very new and intimidating. The RDSM had to learn to understand, appreciate and talk about legislation, corporate risk and corporate responsibility from a much broader perspective than required for RDM.

5. We also needed to better understand and define our remits in relation to research data and records management for researchers. What is 'data'? What is a 'record'? How do these terms apply to research and research processes? What exactly is the RDSM interested in? What is the RM interested in? Where can we support each other and where is there conflict? What's the best compromise for reconciling areas of conflict? These were important questions about our services that surfaced during and after the project.

Many of these issues were unknowns to us before we started the project. We only discovered them when we started talking to the researchers. We discovered the issues though ad hoc conversations after the first interviews. These ad hoc conversations quickly turned into formal post meeting debriefs to discuss how we understood the interview and to clarify any questions or misconceptions. We also used these debriefs to discuss the interview in the context of the other interviews we had completed. We held regular catch-ups when we were writing up the individual interview notes and while we were writing the report as well. These debriefs and openness for conversation to ensure we understood each other and the content proved essential to the success of the project. Being able to leverage each other's organisational networks and spheres of influence also enabled success in this project.

Case 2: Building a case for a digital preservation system

SGUL was a pilot organisation in Jisc's Research Data Shared Service project(Adams, Goddard, & Macneil, 2018).This project sought to develop a national service to enable researchers to

easily deposit data for publication, discovery, safe storage, long term archiving and preservation. As part of this project we contributed to defining the requirements for a data repository and digital preservation system.

Midway through the project it became clear that the digital preservation system being trialled (Preservica) had value outside of research data. In fact, Preservica could help the entire organisation comply with regulations regarding the retention and preservation of business-critical information as well as the archiving of heritage records to preserve our institutional history. Preservica fit seamlessly in our organisational information management workflows.

RM, archives and RDM joined forces to raise awareness of digital preservation across SGUL. We developed organisation-level requirements for the system to meet standards and regulations in each of our fields. We also developed and executed a programme of advocacy to influence key decision makers and obtain endorsements from stakeholders to reinforce our bid. Activities included presentations and meetings with the senior leadership team including the chief operating officer, information management team, and working groups; identifying information asset owners in order to outline which of their assets were at risk of digital obsolescence and obtain buy-in from them for our work; and creating policy that outlined SGUL's overall responsibility for ensuring digital assets are available for the duration of their lifecycle.

Outcomes

The project culminated in a competitive bid to the organisation for a digital preservation system to be used across the organisation. This plan included:

- Evidence demonstrating the need for digital preservation in the organisation with examples of use cases and regulatory requirements from stakeholders across the university. As a result of our advocacy work these stakeholders agreed to be named (and contacted) in defence of the bid

- A draft protocol with a proposed ownership structure and management responsibilities for departmental level information

- A phased and prioritised implementation plan that identified when research, business critical and heritage records would be on-boarded to the system

- Ongoing education and awareness raising activities

- Developing a network of contacts at other Universities to approach for guidance

Our initial bid was successful and we have been given funding for one year to demonstrate viability of the approach.

Challenges

When undertaking this work, we faced numerous hurdles(1) between ourselves, (2) amongst the wider information management team and (3) with the broader SGUL community. These included but were not limited to:

- Varying standards and technical requirements for digital preservation between RDM, RM and archiving

- Differing objectives for preserving research with reuse, compliance, and preserving history having a different emphasis in each area. This leads to different conceptual requirements for each area

- Identifying ownership and responsibility for different aspects of the process. For example, when preserving a dataset, at what point does the dataset leave the custodianship of the RDSM and becomes the responsibility of the Archive

- Concerns from the creators of data/information that they would lose control if their content was preserved

- A lack of understanding of what digital preservation is. For example, some people believe that by making a backup they have preserved the digital item

- Contradictory needs and requirements amongst the wider information management team as to where SGUL's risk lay e.g. lack of consensus amongst the information management group as to what our priorities should be and whether digital preservation was a priority

This project was especially important to help us to focus our priorities and objectives. At the start, we very quickly got caught up and lost in the details of implementation and our differing requirements, policies, processes, roles and responsibilities. We quickly got stuck. However, the externally facing nature of the project created an interesting combination of hurdles that, together, helped us to solidify our approach to working together. To face our stakeholders and their uncertainties we needed to resolve our own uncertainties. We needed to decide on a focus and determine objective criteria for assessing our varying approaches to digital preservation.

Firstly, we needed to be clear on our objective. Our objective was to secure funding for digital preservation activities across corporate records, research data and archives. To do this we needed to raise awareness of the vulnerability of digital materials across the organisation and the risks associated with this and acquire buy-in for digital preservation.

Once our objective was in place we determined our priorities around this objective. Agreed priorities provided us with criteria with which to assess our approaches and to make decisions for the benefit of the project and the objective. We decided that our top priority would be corporate records. Many corporate records must be preserved as a legal and regulatory requirement. Corporate records thus presented the greatest institutional risk and had the highest costs attached to them. We decided to take a compliance-focused approach to digital preservation, with our RM 'owning' the digital preservation process on behalf of the organisation. RDM and archives would support the implementation in their areas.

From this process of working out exactly what we wanted to achieve we recognised that much of our uncertainties (requirements, policies, processes, roles and responsibilities) were not relevant to achieving our goal at the time. Our uncertainties were about the strategy and implementation which were several steps ahead. This process gave us the foundations to look forward to implementation with a shared vision. This process also allowed us to informally cement the ways we would work together – through a commitment to collaboration, compromise and a process of small steps. Our objective, focus, vision and approach to working together also allowed us to confront the questions and concerns coming to us from our wider community together, with confidence and an acknowledgement that some questions could only be answered in practice, through future open learning.

Case 3: Recovering, preserving and facilitating access to the Addicts Index

The Addicts Index was created by the Home Office and comprises records from the 1940s to 1994 of individuals seeking treatment for drug dependence, including their personal details, details of those providing their treatment, as well as information on prescribers and inspections of chemists and pharmaceutical companies. Official custody of this data was transferred to St George's by the Home Office on termination of this programme. The database containing these records, including digitised copies of original medical reports is now inaccessible.

The RDSM, archivist, RM and researchers responsible for the Addicts Index were interested in recovering, reusing and facilitating wider access to the resource. The group came together to develop a bid to the Wellcome Trust Open Research Fund to:

- recover the database
- make the database contents findable, accessible, interoperable and reusable (FAIR) (Wilkinson et al., 2016)
- preserve the database for the long term

This work was made possible due to our earlier projects (for example, case 1) and the relationships and trust that we had built with researchers as a result of our work with them.

Outcomes

Work on this bid was initiated after the previous two projects and after much alignment of our policies, activities and workflows. This was a very practical process, rooted in concepts in research data management but demonstrating how good data management succeeds only though collaboration.

The application process demonstrated our better understanding of each other's work and how we can support each other in delivering outcomes in each of our areas in the stewardship, curation and preservation of research data. It also reflected our emerging approach to working together - through a commitment to collaboration, compromise and a process of small steps.

The methods advanced in the bid would have been the first practical test of the structures we have put in place to transform research data services at St Georges. However, the bid for funding was unsuccessful and negotiations are under way to transfer the Addicts Index to the Archive for possible recovery and preservation.

Challenges

This project presents several challenges to sharing, preserving and curating digital research data that are less specific to SGUL and more community focused. Firstly, the funding landscape for the recovery, sharing and preservation of digital assets of scientific value is not clear. The research team had been searching for a suitable source of funding for some time and the Wellcome Trust Open Research Fund was the grant that came the closest to helping the researchers to achieve their objectives in both recovering and sharing the Addicts Index with the wider research community. Opportunities to recover valuable research data need to be made more widely available.

Secondly, where funding might be available from archival communities, the argument for recovery falls flat if an organisation does not have the means to look after/preserve the asset once it is recovered. Digital preservation (our second case study) is thus very closely related to funding to digitise historical data or recover inaccessible content.

Lastly, both archival and RDM communities have long advocated for earlier engagement with stakeholders as a key ingredient to successful data curation and stewardship. In this case the data was made known to the archivist when it was already inaccessible. There are several legitimate reasons for this, but once data gets to this stage it is very difficult to reverse. Our learning from this is that the archivist and RDSM need to engage with researchers earlier in the project and start preservation planning early – recognising that datasets will change stewards several times in their lifetimes and they need to be robust enough to survive those changes for the long term. There is also the popular perception that content should go to the archive when they are no longer considered viable. Earlier engagement in the stewardship process, and advocacy to change the perception of archives in research communities, is needed to better look after research data for the long term.

However, this approach also comes with related challenges. Archives face significant challenges in the context of research data. If some research can be considered to be continuous, building on previous research grant after grant, then handing data over to archives for

preservation can be contentious. From researchers' perspectives, they will lose control over data that may be of later value to them. However, from an archival perspective, the archivist's role is to actively look after research outputs until they are required again. So, for researchers, an archive is a place while for archivists archiving is a process. For archivists to reach researchers earlier in the stewardship process we believe a compromise may need to be made between archives as place and process. Complicating the matter further, for some kinds of research 'archiving' is understood as a very formal and standard process to meet regulatory compliance (see case 1). This interpretation of 'archiving' is not related to maintaining legacy and heritage. This concept of archiving is something that special collections archives may need to reconcile. To this end, the RDSM, RM and archivist will continue to work closely with our research community, building trust in small steps.

Reflections

In this paper we have demonstrated how RDM, archives and records management have collaborated on three projects at St George's, University of London. We then discussed the outcomes of each of these projects, including the challenges we faced and how we reconciled these. Underlying all three projects though, was a method; a stepwise, iterative process to streamline and harmonise our guidance and workflows in relation to the stewardship, curation and preservation of research data. The learnings from each of these projects are used to develop our guidance, improve our activities and integrate our workflows, the outcomes of which will be further evaluated and applied to new projects. Through deeper integration of our activities and workflows, rather than simply aligning aspects of our work, we are increasingly becoming partners on research data stewardship, curation and preservation.

This iterative, stepwise approach provided several benefits to the team. This process:

- allows us to test and revise our solutions; there is no pressure to get it 'right', only to be aware, actively learn and improve
- helps us to set realistic goals
- has provided us with an agreed vision/structure to working together
- helps us to identify genuine, practical challenges
- allows for openness and flexibility
- allows us to engage in skills sharing and leveraging each other's networks

This last point is very important as it allows us to boost our individual/collective value and impact to the organisation and to the researchers we support.

We aim to build on this work going forward, constantly reviewing what is working and what is not working and improving our practice, our communications and our procedures. Out of the work described in this paper we have identified the following as future steps for us:

1. Creating a common language for our use going forward
2. Developing standard operating procedures for digital preservation across records, archives and RDM
3. Determining and applying appropriate metadata to different kinds of digital content for preservation
4. Consulting on the value of formal institutional data pipelines for different stages of the research lifecycle
5. Exploring provision for electronic laboratory notebooks

6. Applying our learning to other projects planned at the organisation for greater uptake, for example, integrating conversations about what data to keep and safe places to store research may be more acceptable in anticipation of estates work when researchers are thinking about clearing out their offices
7. Considering appropriate mechanisms to communicate the value of archives/archiving to researchers, including consulting on new processes for looking after data that balances place and process interpretations of archives (see case 3).

As the amount of research data grows and the number of formats and locations it is stored in increases we are all concerned with the need to robustly manage our records and data, and ensure that our researchers are able to access their research for as long as required.

References

Adams, R., Goddard, N., & Macneil, R. (2018). *Jisc Research Data Shared Service: Looking back and looking forward*. Retrieved from https://era.ed.ac.uk/handle/1842/31312

Bradley, C. (2018). Research Support Priorities of and Relationships among Librarians and Research Administrators: A Content Analysis of the Professional Literature. *Evidence Based Library and Information Practice*, *13*(4), 15–30. https://doi.org/10.18438/eblip29478

Data Asset Framework. (n.d.). Retrieved 6 December 2019, from https://www.data-audit.eu/

Data Security and Protection Toolkit. (2019). Retrieved 6 December 2019, from NHS Digital website: https://digital.nhs.uk/data-and-information/looking-after-information/data-security-and-information-governance/data-security-and-protection-toolkit

NSH Digital. (2019). Retrieved 6 December, from https://digital.nhs.uk/

Ogier, A., Nicholls, N., & Speer, R. (2017). Open exit: Reaching the end of the data lift cycle. In L. Johnston (Ed.), *Curating research data: Practical strategies for your digital repository* (Vol. 1, pp. 235–250). Chicago: Association of College and Research Libraries.

Olsson, A., Wadell, C., Odenrick, P., & Bergendahl, M. N. (2010). An action learning method for increased innovation capability in organisations. *Action Learning: Research and Practice*, *7*(2), 167–179. https://doi.org/10.1080/14767333.2010.488328

Pinfield, S., Cox, A. M., & Smith, J. (2014). Research Data Management and Libraries: Relationships, Activities, Drivers and Influences. *PLOS ONE*, *9*(12), e114734. https://doi.org/10.1371/journal.pone.0114734

Pryor, G. (2014). Who's doing data? A spectrum of roles, responsibilities and competencies. In G. Pryor, S. Jones, & A. Whyte (Eds.), *Delivering research data management services: Fundamentals of good practice* (pp. 41–58). London: Facet Publishing.

Read, K. B., Surkis, A., Larson, C., McCrillis, A., Graff, A., Nicholson, J., & Xu, J. (2015). Starting the data conversation: Informing data services at an academic health sciences library. *Journal of the Medical Library Association : JMLA*, *103*(3), 131–135. https://doi.org/10.3163/1536-5050.103.3.005

Rice, R., & Southall, J. (2016). *The data librarian's handbook*. London: Facet Publishing.

Söderholm, M., & Sunikka, A. (2017). Collaboration in RDM activities – Practices and development at Aalto University. *Septentrio Conference Series*, (1). https://doi.org/10.7557/5.4247

Verbaan, E., & Cox, A. M. (2014). Occupational Sub-Cultures, Jurisdictional Struggle and Third Space: Theorising Professional Service Responses to Research Data Management. *The Journal of Academic Librarianship*, *40*(3), 211–219. https://doi.org/10.1016/j.acalib.2014.02.008

Whyte, A. (2014). A pathway to sustainable research data services: From scoping to sustainability. In P. Graham, S. Jones, & A. Whyte (Eds.), *Delivering Research Data Management Services: Fundamentals of Good Practice* (pp. 59–88). London: Facet Publishing.

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 1–9. https://doi.org/10.1038/sdata.2016.18