IJDC | *Conference Pre-print*

Recordkeepers Managing Research Data: a Survey of Irish Organisations

Rebecca Grant

University College Dublin

Abstract

This paper describes a survey undertaken in 2017 to establish which research data management policies and practices were in place at Irish organisations; the extent to which archivists and records managers were employed to manage research data at those organisations; and the impact that archival skills have on research data management at an organisation. The paper describes the survey methods and data analysis, and presents findings including the presence of archivists and records managers at more than half of the surveyed organisations. Next steps for the research are also outlined.

Submitted 16 December 2019 ~ *Accepted* 19 February 2020

Correspondence should be addressed to Rebecca Grant, UCD School of History, Newman Building, University College Dublin, Belfield, Dublin 4, Ireland. Email: beck.grant@gmail.com

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

This paper summarises a chapter of the author's doctoral thesis, which explores the connections between research data management, and the theory and practice which underpin the work of recordkeeping professionals (archivists and records managers), with a focus on Irish organisations. Archivists and records managers are concerned with the stewardship, preservation and accessibility of records and their work generally focuses on records of enduring evidentiary value and historical archives. However a 2017 literature review indicated that recordkeepers have tested the application of records management approaches to research datasets, and data have been accessioned, catalogued and held in the custody of archives since the mid-20th century (Grant, 2017).

In this paper a survey undertaken at Irish organisations is described, which assessed the extent to which recordkeeping professionals are involved in research data management. A brief overview of current professional roles in data management is provided to contextualise the study, and the survey methodology and data analysis are summarised. Findings from the survey are presented, and the paper concludes by briefly outlining next steps for the research.

Roles in Research Data Management

Those working in digital curation or data curation may have job titles including archivist, librarian or simply data curator (Oliver & Harvey, 2016). Data management-related educational programmes are associated with library and information studies, information engineering, data curation and informatics degrees, as well as records management and archival studies courses (Pryor & Donnelly, 2009). A 2018 report described contemporary data management as "a cottage-industry with incompatible and fragmented data stewardship approaches" however, indicating inconsistencies in perceived skills, competences and roles (Hahnel et al., 2018). Within universities, the library is seen as having a key role in supporting researchers and research data management (Brown et al., 2015 and Yu, 2017). Support for librarians taking on research data management is widely available, including the training modules and courses,[1] briefing papers and reports by professional bodies (CILIP, 2014), reviews of necessary competences (Federer, 2018 and LIBER, 2018), and practical guidance.[2] One of the first major outputs of the global Research Data Alliance (RDA) was *23 Things: Libraries for Research Data,* which gathers resources and guidelines for librarians aiming to support research data management at their institution (Witt, 2019). The RDA Libraries for Research Data Interest Group has nearly 500 members as of 2019, making it the most subscribed of the RDA's working and interest groups.[3]

Although much of the available literature is written for librarians managing research data, the relevance of archival and records management theory to data management is also acknowledged. Dooley (2015) gives examples of ten areas of applicable archival expertise including donor relations, context, appraisal and authenticity. The CILIP Research Data Management briefing paper states that data management is "more closely tied to records management and archival thinking than librarianship," and that this poses a challenge for librarians (CILIP, 2014). Ramírez (2011) notes that professional archivists have considerable expertise in handling volumes of research data, and that archival methods can add efficiency to

[1] Do-It-Yourself Research Data Management Training Kit for Librarians: http://datalib.edina.ac.uk/mantra/libtraining.html. RDM For Librarians: ttp://www.dcc.ac.uk/training/rdm-librarians. RDMRose: www.shef.ac.uk/is/research/projects/rdmrose

[2] Role of Libraries in Data Curation: https://www.oclc.org/research/themes/research-collections/datacuration.html

[3] Libraries for Research Data IG: https://www.rd-alliance.org/groups/libraries-research-data.html

digital data management. She urges greater collaboration between university librarians and archivists, and observes that librarians would be well served to embrace input from archivists to remain relevant in data stewardship. Records management strands have been added to the Library and Information Studies Master's Degree options at UCL Qatar to prepare students to work with both records and research data (Straube, 2018). Additionally, the Digital Curation Centre's Digital Curation Life-Cycle Model which describes the lifecycle stages required for successful data curation cites the records continuum model (which underpins records management practice), and reflects dimensions of the continuum such as "creation" and "capture" (Higgins, 2008).

Very little academic literature could be identified which described data management practice in Ireland, or the role of recordkeepers in this work. The intention of the survey was therefore to capture insights into Irish research data management practice and the work of recordkeeping professionals, which were not available in the existing literature.

Methodology

The survey questions were designed using the Digital Curation Lifecycle model[4] as a framework to assess the extent to which surveyed organisations were fulfilling standard processes for data management. The survey also included questions on the respondents' legal or other obligations surrounding data management, the policies in place at their organisations, and their opinions on organisational policy and practice. Respondents were also asked to indicate whether a recordkeeping professional had ever been employed at their organisation, and to what extent this person had contributed to research data management policy and practice.

The target population for the survey was defined as Irish publicly funded bodies which were likely both to generate research data, and to be subject to legislative requirements for data archiving and data publication. A list of publicly funded Irish organisations was drafted using governmental websites and the Central Statistic Office's "2014 Register of Public Sector Bodies (including General Government Bodies) in Ireland."[5] Purposive non-probability sampling was used to narrow the list from over 100 organisations to 23, selecting participants based on the body's involvement with either the Public Bodies Working Group (which steers the development of the Open Government Data portal, data.gov.ie) or the Cultural Heritage Working Group (which aims to manage and publish Irish cultural heritage datasets). Snowball sampling was used to add an additional 5 organisations for a total of 28.[6]

[4] DCC Curation Lifecycle Model: http://www.dcc.ac.uk/resources/curation-lifecycle-model

[5] 2014 Register of Public Sector Bodies (including General Government Bodies) in Ireland: https://www.cso.ie/en/media/csoie/methods/nationalaccountsoutputandvalueaddedbyactivity/Regofpublicsectorbodies2015April.pdf

[6] The Central Statistics Office; Fingal County Council; The Department of Arts, Heritage and the Gaeltacht; The Office of Public Works; Transport Infrastructure Ireland; The Digital Repository of Ireland; Programmable Cities Project, Maynooth University; Office of the Revenue Commissioners; The Marine Institute; Department of Environment, Community and Local Government; Ordnance Survey Ireland; Department of Public Expenditure and Reform; The National Museum; The Heritage Council; Dublin City Council; The Environmental Protection Agency; Irish Social Science Data Archive; Irish Qualitative Data Archive; Institute of Public Health; Centre for Support and Training in Analysis and Research, University College Dublin; Health Research Board; Teagasc; University College Cork; University College Galway; Dublin City University; Trinity College Dublin; University of Limerick.

Data Collection and Analysis

The survey was hosted using Qualtrics survey software, and an invitation and link were emailed to representatives of the 28 organisations on 1 October 2017, along with an information sheet detailing the objectives of the study, its risks, the rights of participants, and how the data from the survey would be stored and used. Organisations were asked to respond within four weeks, and a reminder email was sent after two weeks. At the end of the final survey period, of the 28 invited respondents 13 did not reply, while 4 declined. Of those who declined to participate, 2 stated that their organisation did not create or archive data, and 2 stated that they did not have the necessary expertise to respond to the survey. Eleven completed responses were received, demonstrating a response rate of 39%, which aligns with the expected response rate for online organisational surveys (Baruch & Holtom, 2008).

The survey data were exported from Qualtrics to Excel and subject to univariate and bivariate analysis, and qualitative text responses were coded in Word. Due to the sampling method used and the small size of the respondent group, the results of the survey are not intended to be generalisable or presented as statistically significant.

Data Management Practices at Irish Organisations

The Digital Curation Lifecycle model was used to define the survey questions which relate to organisational data management practices. The model is a graphical, high-level overview of the lifecycle stages required for successful curation, but is not intended to be prescriptive (Higgins, 2008). However an organisation which undertakes very few of the key elements (for example which receives and stores data, but which does not describe, appraise, or preserve them) was not considered to be performing comprehensive data management in the context of this study. Key elements of the lifecycle underpinned the survey questions which were used to assess data management practice: Create or Receive Data; Appraise Data; Ingest Data; [Maintain] Description and Representation Information; Access, Use and Reuse Data; and [Undertake] Preservation Actions.

All of the 11 surveyed organisations reported that they created or received data, and all provided some form of access to their data to facilitate their reuse; 10 transferred data to an archive, repository or data centre; 8 kept metadata or other technical information alongside the data; and 7 appraised the data or received data after appraisal. Only one organisation stored technical metadata in addition to descriptive metadata, and one indicated that they requested additional contextual documentation from depositors which was stored alongside descriptive metadata. For data sharing, most respondents (8 of 11) indicated that their data were shared through publication in reports, white papers, research publications or other documentation. Respondents were also asked why they managed and stored their data, and whether there was legislation in place which compelled them to do so, or other factors. The most common responses were that respondents' organisations were required to store data due to their obligations under PSI (the Public Sector Information regulations) and FOI (the Freedom of Information Act). Only two respondents stated that funder requirements obliged them to manage and publish data.

The survey also asked respondents to describe data preservation actions at their organisation, defined as "actions taken with the aim of ensuring that data will still be available in 5 years' time." Respondents were asked for their opinions on these preservation actions, and whether they believed that their organisation's data would remain authentic, reliable and usable over time. More than half of the respondents stated that their organisation had an archive with appropriate environmental and access controls which was used to store their analogue records. For digital data, 8 respondents reported copying data to a reliable digital storage system, while 4 stored multiple copies of the data. Only 2 organisations reported the use of contextual technical records to ensure that data would remain meaningful over time. The majority (10 respondents)

stated that they managed the data in accordance with good IT practices, for example by backing up data and ensuring data security.

Seven respondents believed that their preservation practices ensured that data remained authentic, reliable and usable over time; 3 believed that their practices ensured this for some datasets but not others, or that improvement was required; and only one did not believe that appropriate practices were used. Respondents were asked to explain why they believed their preservation practices would (or would not) allow the data to be accessed in 5+ years. One stated: "We have policies and procedures for long-term preservation of data, and we store preservation and provenance metadata, as well as descriptive and technical metadata. We calculate checksums and check for file tampering or corruption periodically. The descriptive metadata is generally quite rich and allows data to be understood." Another noted that their preservation activities must be effective because an archivist was employed for this purpose: "Yes for that data that we wish to preserve, as we have an archivist whose role this is."

Respondents were asked which policies they had in place to support research data management activities and were provided with a list to choose from (Figure 1). The policies on the list represent those which are likely to be present both in an archives service, and to support data management and publication. The most common policy at the responding organisations was an access policy, which reflects the number of respondents who reported that they provide some form of access to their datasets. A total of 5 organisations (nearly half) had a records management policy in place; the same number had a collection policy or a retention/disposal policy.

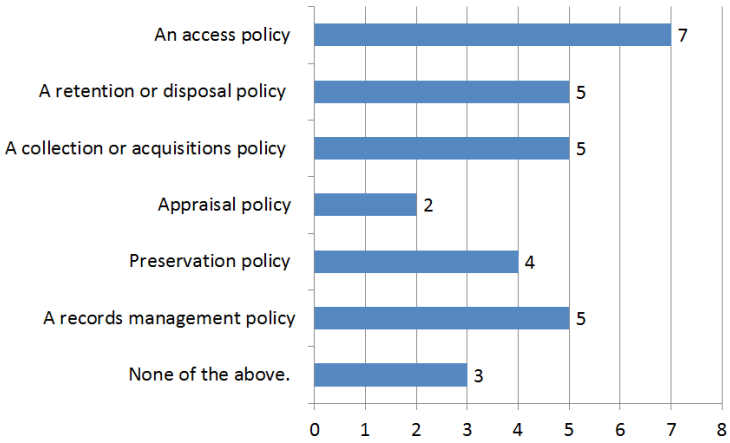

Figure 1. Data management-related policies at responding organisations.

Three of the respondents reported that their organisation had no policies at all, although all noted that they were in the process of developing policies or planned to do so in the future. When asked "What, if anything, would you change about your current data management policies and/or practices?" these respondents also noted that they were aware that they should have standard policies in place. One, an organisation which only provided an access policy, also stated that they intend to introduce a records management policy in the future. Respondents were asked which resources had been used to aid the development of their policies and practices, and 3 stated that they did not use any external resources to do this. The resources mentioned by other respondents include the Data Seal of Approval (now the Core Trust Seal) which certifies Trusted Digital Repositories; the Digital Preservation Coalition's documentation; the Digital Curation Centre's documentation; and OAIS (the Open Archival Information Systems Reference Model).

Finally, respondents were asked what they would change about their organisation's current data management policies and practices. Of the nine who responded to this question, only one stated that they would not change anything. This respondent believed that their policies are already sufficiently robust to support good practice, stating: "We review our policies regularly and update them in light of changing requirements." Four respondents intended to introduce additional policies, while one also planned to introduce a data strategy, and one aimed to implement staff training. Two suggested technical improvements, including an integrated data lake; the implementation of repository infrastructure; and a system for converting data to open formats.

Recordkeeping Professionals and Data Management

The survey was sent to 28 organisations without prior knowledge of whether they employed a recordkeeping professional (defined in the survey as an archivist or records manager with formal qualifications). More than half of the responding organisations (6 respondents) stated that they employed either an archivist or a records manager to support research data management or had done so in the past. Respondents who confirmed that they employed a recordkeeping professional or had previously done so were asked several additional questions to assess the involvement of the recordkeeping professional in data management activities.

Every one of the Digital Curation Lifecyle actions was performed by a recordkeeping professional at one or more of the responding organisations (Table 1). They were most frequently involved with the development of policies and managing the transfer of data to long-term storage (5 of 6 respondents stated that they did so). They were involved least often in activities supporting data storage, data security, or data preservation. This may reflect the technical skills required to plan and manage data storage, which may be more likely to involve an IT professional at larger organisations. In relation to six suggested policies, on average organisations which did not employ recordkeeping professional had 3.5 of these policies in place while organisations which employed a recordkeeping professional had fewer, at 2.3 policies. Two organisations which employed recordkeeping professionals did not have any of the listed policies in place. The presence of a recordkeeping professional did not increase the presence of data-related policies, even though nearly all of the recordkeeping professionals were reported to contribute to policy development in the course of their role. Additionally, many of the suggested policies would be standard in an archives service or to support records management functions, for example a records management policy, an appraisal policy or a preservation policy. Organisations with a recordkeeping professional in place were slightly more likely to have a records management policy. This may reflect the commitment to records management which is exemplified by the organisation's choice to employ a recordkeeping professional. Other than this, there were no identifiable trends connecting the presence of a specific type of policy to the presence of a recordkeeping professional. Across all respondents, organisations fulfil an average of 4.2 out of 6 of the Data Curation Lifecycle actions. Organisations with or without a recordkeeping professional undertook the same average number of key elements, so no difference was identified.

When asked for the job title of the person with overall responsibility for data management, it was not clear whether that person was a recordkeeping professional or not. Two organisations stated that the Chief Information Officer had overall responsibility; 2 that the data manager was responsible; 3 indicated that more than one person had responsibility; and 5 that there was either no one responsible at all, or that the person or people couldn't be identified by the respondent. It is not clear from the survey data whether the recordkeeping professionals involved take on data management as their full-time role, or as a part of it.

Table 1. Data management activities undertaken by recordkeeping professionals at responding Irish organisations.

Data management activity	Number of recordkeeping professionals undertaking each activity (n=6)
Developing policies relating to records management, appraisal, access, preservation, acquisition and disposal of data.	5
Managing transfer of data to a repository, archive or data centre, or another form of long-term storage	5
Providing access or managing access to the data	4
Evaluating and/or selecting data which will be moved to long term storage	4
Selecting data storage solutions or influencing decisions on data storage	3
Having full or partial responsibility for data security and preservation	2

A key finding of the survey was therefore the proportion of organisations which employed a recordkeeping professional for the management of research data. It appeared that there was no specific impact on data management due to the presence of a recordkeeping professional however, for example the likelihood of having certain policies or practices in place. The data gathered could not be used to assess the level of responsibility assigned to recordkeeping professionals in relation to data management, or the other organisational factors relating to data management strategy and staffing which impact on a recordkeeping professional's role.

Conclusion and Next Steps

The survey tool was effective in capturing information on practices, policies and perspectives relating to research data management at Irish organisations. As anticipated by the sampling method, the responding organisations were all undertaking research data management in some form. A limitation of the method was the self-selection of the survey respondents however, with 17 organisations choosing not to respond.

As a snapshot, the survey provides granular information on data types generated and stored by Irish organisations, organisational infrastructure and strategies for data management, and how their representatives feel about their data management practices. While the results cannot be extrapolated to draw conclusions regarding organisational practices across Ireland, they do provide insight into the quality and completeness of data management practice where it is being undertaken by Irish organisations. The analysis also indicates that although recordkeeping professionals are involved in data management at their organisations, they are not generally involved in data preservation and have not impacted on the creation of additional policies, even those policies which are standard at archives services.

As the level of responsibility of these recordkeeping professionals was not assessed by the survey, the next phase of research will further investigate whether the presence of recordkeeping

professionals can impact positively on research data management at their organisations using a comparative cross-case analysis. A comprehensive review of the research data policy landscape in Ireland will also be undertaken to triangulate the survey responses.

Acknowledgements

The author would like to thank her Doctoral supervisor Dr. Elizabeth Mullins (University College Dublin) as well as her Doctoral Studies Panel, Dr. Sandra Collins (National Library of Ireland), John McDonough (Dublin City University) and Dr. Julie Brooks (University College Dublin).

Data availability

An anonymised version of the survey responses dataset is available in the figshare repository at https://doi.org/10.6084/m9.figshare.11365598 (Grant, 2019).

References

Auckland, M. (2012). Re-skilling for research. RLUK. Available online: https://www.rluk.ac.uk/portfolio-items/re-skilling-for-research

Baruch, Y. & Holtom, B.C. (2008). Survey response rate levels and trends in organizational research. Human Relations 61, 8, 1139-1160. doi: https://doi.org/10.1177/0018726708094863

Brown, R., Wolski, M. & Richardson, J. (2015). Developing new skills for research support librarians. The Australian Library Journal, 64, 224-234. doi: https://doi.org/10.1080/00049670.2015.1041215

CILIP (2014). Research Data Management Briefing Paper. Available online: https://archive.cilip.org.uk/sites/default/files/media/document/2017/research_data_management_briefing_july_2014_0.pdf

Digital Science, Hahnel, M., Fane, B., Treadway, J., Baynes, G., Wilkinson, R., Mons, B., Schultes, E., Olavo Bonino da Silva Santos, L., Arefiev, P., Osipov, I. (2018). The State of Open Data. London: figshare. Available online: https://figshare.com/articles/The_State_of_Open_Data_Report_2018/7195058

Dooley, J. (2015). The Archival Advantage: Integrating Archival Expertise into Management of Born-digital Library Materials. Ohio: OCLC. Available online: https://www.oclc.org/content/dam/research/publications/2015/oclcresearch-archival-advantage-2015.pdf

Federer, L. (2018). Defining data librarianship: a survey of competencies, skills, and training. Journal of the Medical Library Association, 106, 3. doi: https://doi.org/10.5195/jmla.2018.306

Grant, R. (2017). Recordkeeping and research data management: a review of perspectives. Records Management Journal 27, 2, 159-174. doi: https://doi.org/10.1108/RMJ-10-2016-0036

Grant, R. (2019). Data from a 2017 survey of Irish organisations relating to their data management policy and practice. figshare. doi: https://doi.org/10.6084/m9.figshare.11365598

Higgins, S. (2008). The DCC Curation Lifecycle Model. The International Journal of Digital Curation 1, 3, 134-140. doi: https://doi.org/10.2218/ijdc.v3i1.48

LIBER Europe (2018). Data librarian, expert on research data management, description, archiving and dissemination. Available online: https://libereurope.eu/blog/2018/05/14/data-librarian-expert-on-research-data-management

Oliver, G. & Harvey, R. (2016). Digital Curation. London: Facet Publishing (2nd ed.).

Pryor, G. & Donnelly, M. (2009). Skilling Up to Do Data: Whose Role, Whose Responsibility, Whose Career? International Journal of Digital Curation 4, 161-163. doi: https://doi.org/10.2218/ijdc.v4i2.105

Ramírez, M. L. (2011). Opinion: Whose role is it anyway? A library practitioner's appraisal of the digital data deluge. Bulletin of the American Society for Information Science and Technology. doi: https://doi.org/10.1002/bult.2011.1720370508

Straube, A. (2018). Records management and data management quite literally mean the same thing: the Integration of archives, records and data management into the MLIS Programme at UCL Qatar. In Aparac-Jelušić, T., Casarosa, V. & Macevičiūtė, E. (Eds.) FEIS – International Symposium organized by EINFOSE 10–11 September 2018 Pisa, Italy, (pp. 237-239). Available online: http://einfose.ffos.hr/feis-2018/proceedings

Witt, M. (2019). 23 Things: Libraries for Research Data. Research Data Alliance. doi: https://doi.org/10.15497/RDA00005

Yu, H. (2017). The role of academic libraries in research data service (RDS) provision. The Electronic Library 35, 4. doi: https://doi.org/10.1108/EL-10-2016-0233