Research Repository UCD

Title	Recordkeeping and research data management: a review of perspectives
Authors(s)	Grant, Rebecca
Publication date	2017-07
Publication information	Grant, Rebecca. "Recordkeeping and Research Data Management: A Review of Perspectives" 27, no. 2 (July, 2017).
Publisher	Emerald Insight
Item record/more information	http://hdl.handle.net/10197/8769
Publisher's statement	This article is © Emerald Group Publishing and permission has been granted for this version to appear here (please insert the web address here). Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.
Publisher's version (DOI)	10.1108/RMJ-10-2016-0036

Downloaded 2023-04-07T10:52:29Z

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd_oa)

© Some rights reserved. For more information

Recordkeeping and research data management: a review of perspectives

Author: Rebecca Grant, School of History and Archives, University College Dublin

Introduction

The purpose of this paper is to explore perspectives on the relationship between research data and the record, and recordkeeping and research data management. The perspectives described come from the archival and recordkeeping professions, as well as others writing about the management and preservation of research data.

The relationship between records professionals and research data has been addressed directly by the archival profession in publications and professional fora. The March 2007 *Archival Science* dedicated a special issue to Archiving Research Data, and the International Council on Archives Section on University and Research Institution Archives 2014 Conference theme was research and scientific archives [i]. A 2014 survey of archivists working in American university research libraries indicates that 49% of respondents are involved with research data management and preservation for their institution (Noonan and Chute, 2014). Dooley (2015) argues that archivists have unique skills which should be applied to material with less obvious archival characteristics, including research data. The international Research Data Alliance also facilitates an Interest Group for Archives and Records Professionals [ii] as well as libraries [iii].

It appears that the recordkeeping community is engaging with research data management, and that records professionals are addressing the challenges and opportunities associated with research data. Despite this, no survey or existing review were identified which investigate theory and practice in the area to date. This literature review aims to address this gap by exploring the subject of research data and its management, and how records professionals have engaged with it in the 20th and 21st centuries.¹

¹ In discussing the relationship or perspective of records managers and archivists to research data management, the term “records professional” is used to encompass both roles; and “recordkeeping” to refer to the activities undertaken by records professionals in their management and archiving of records. Referencing the International Council on Archives’ Multilingual Terminology, “records manager” may be defined as “a person, professionally educated, trained and experienced, responsible for the effective and efficient delivery of records management

Methodology

This literature review represents a section of the author's preliminary Doctoral research, undertaken in 2014, on the topic of records professionals and research data. The initial review was conducted in two phases, firstly investigating definitions of research data in academic writing, practical handbooks, textbooks and manuals, and policy documentation. The second phase examined how records professionals engage with research data. This paper focuses on this second topic.

Two main electronic databases were used to identify the literature discussed below: University College Dublin's Library OneSearch, and Google Scholar. As the author's Doctoral research focuses on Irish policy and practice, the Irish Open Access portal RIAN.ie was also interrogated.

The review aimed to investigate literature written by records professionals regarding research data, and also to identify other perspectives on research data and its management, and whether consensus exists as to the professional skills necessary to tackle this area. Initially, peer-reviewed academic journals relating to the archives and records management professions were interrogated. These included *Archival Science*, *American Archivist*, *The Records Management Journal* and *Archivaria*. The compound terms "research data" and "research data management" were used initially to identify articles. This yielded relevant results, but the use of simple terms (for example "research", "data" and "science") produced additional material, for example relating to governmental datasets which are not necessarily referred to as "research data". The use of simple terms, particularly "science" also elicited pre-2000 material, where a discussion of scientific "research data" could be inferred, but the term was not mentioned explicitly in the body of the text. The use of simple terms led to a high volume of results, but those which discussed related concepts, for example Open Data and Big Data, were excluded if research data did not constitute a major component of the paper.

services to meet and organisation's requirements"; "archivist" may be defined as "an individual responsible for appraising, acquiring, arranging, describing, preserving, and providing access to records of enduring value, according to the principles of provenance, original order, and collective control to protect the materials' authenticity and context". Where the term "archivist" or "records manager" was specifically used by the author, these are retained when describing the subject of the literature.

A similar process was used to identify literature which related to the concepts “research data management” and “digital curation”. The International Journal of Digital Curation was targeted specifically as a contemporary source of relevant literature. In order to identify and capture newly published literature relating to research data management, Advanced Twitter Search was also used regularly across the one-year period, using searches such as “research data” [and] “archivist”.

These systematic searches for literature conducted using electronic databases identified an initial set of material. Citations and bibliographic references from this material were then followed and this literature was included where relevant, widening the set further. As literature was identified, it was added to a spreadsheet and coded with thematic labels (for example “data in archival repositories” or “archival methods applied to data”). This categorisation process grouped the literature for analysis in the subsequent review. As additional literature was added, new themes and categories became apparent. Literature was discarded when the primary topic was not found to be relevant to research data and records professionals or information professionals, and these bibliographic references were not retained. 118 relevant sources were identified, of which 72 are cited in this paper. The remaining relevant sources were cited in the longer literature review from which this paper is derived.

The literature reviewed was limited to publication in the English language. The final set of literature included authors and publications based in the United States, Canada, Australia, the United Kingdom, Ireland and Europe. It was not intended to limit the literature to publications from the 20th and 21st century, but no content written prior to 1955 was identified. It is likely that earlier literature may have been excluded due to the variance in terminology used to describe “research data” before the 2000s. The review takes a conservative approach, only including literature where the context of the discussion made it very likely that the subject was equivalent to contemporary “research data”, although this term was not used. The author is currently working in the area of research data management, and some additional resources and projects were identified through her professional experience, for example the Digital Curation Centre’s DMPOnline tool.

The literature is discussed below under the following headings and sub-headings, which reflect the coding and subsequent groupings created during the analysis:

Data, research data and records

- Defining data and research data
- Data and records
- Research data and research records

Data curation and research data management skills

- Researchers, librarians and data curation
- Records professionals and data curation

Data management and records management approaches

Research data in the custody of archives

- Government research data in National Archives
- Archives and the papers of scientists

Data, research data and records

Definitions of “research data” found in the literature are most frequently provided by stakeholders in the research data lifecycle: the funding institutions who fund research activity; the educational institutions who support research activity; and journals which publish the outputs of research activity. The development of research data policy by these stakeholders is linked closely to their support for the principles of Open Access, and subsequently the publication of Open Data. “Open Access” is an initiative which has gained significant traction since the 2000s and the rise of the World Wide Web, promoting the concept that the outputs of publicly funded research (whether in the form of academic publications or data) should be made openly available and accessible for interrogation by other interested parties. The principles of Open Access influence researchers in publishing their research publications in Open Access journals, and publishing their research data as Open Data on the web where possible. Although research data is more commonly associated with academic research, it may also be generated by government departments or other research organisations in receipt of public funding.

Defining data and research data

Borgman (2012) notes that defining the term “data” can be challenging, and references the definition from the National Academies of Science: “Data are facts, numbers, letters, and symbols that describe an object, idea, condition, situation, or other factors”; she concludes that data types and the sources of data can be varied. She also cites Buckland (1991), who makes a distinction between information and “data”, which is defined as “information which has been processed”.

The definitions of “research data” provided to researchers by stakeholders and policy-makers are often broad and inclusive. The academic publisher Elsevier defines research data as “broadly speaking [...] the result of observations or experimentation that validate research findings.” [iv] The EU-wide funding stream Horizon 2020 mandates research data publication, describing such data as “information, in particular facts or numbers, collected to be examined and considered and as a basis for reasoning, discussion, or calculation.” [v] Practical manuals for researchers also provide definitions, for example Managing and Sharing Research Data: Best Practice for Researchers states: “We define research data as any research materials resulting from primary data collection or generation, qualitative or quantitative, or derived from existing sources intended to be analysed in the course of a research project”. Research data can therefore be considered as including information or data which may be the input or the product of research.

Data and Records

The International Standard for Records Management ISO 15489:2016 describes records as “both evidence of business activity and information assets. They can be distinguished from other information assets by their role as evidence in the transaction of business and by their reliance on metadata.” The International Council on Archives Terminology references ISO 15489:2001, defining a record as “information created, received, and maintained as evidence and information by an organization or person, in pursuance of legal obligations or in the transaction of business”. Borglund and Engvall (2014) found that the term “information” was used almost as frequently as “record” in archival discourse, and that “information” was in fact used to mean “record” in most contexts outside of a legal context.

As early as 1976, Miller noted that records and research data both provide primary evidence. Evans et al (2014) suggest that the research data “can encompass just about anything, which suggests that, like

records, research data are a logical construct – a perspective one takes on a recorded information object.” Reed (2005) states that records are inherently transactional in nature, and that they provide evidence that an activity has taken place. Childs et al (2014) suggest that evidence of properly conducted research is an aim of (open) research data (although it is not its only purpose). In the introduction to the 2007 issue of *Archival Science* which focused on the topic of archiving research data, Doorn and Tjalsma (2007) discuss the convergence of research data and the record in the context of the digital age, despite their belief that the two “traditions” developed entirely independently. A discussion of the integration of the Danish Academic Data Service with the Danish State Archives notes that one of the core arguments which was made prior to the integration of the services was that material held by the Data Service and the existing National Archive was essentially the same (Neilsen, 1995). Akmon et al (2011) note that the collection and preservation of digital records are the “raison d’être” of archivists, but believe that to date, the archival community has neglected data.

A potential difference between a records professional’s approach and current research data management practice is the necessity to reflect context when exerting intellectual control over a record, as described in the International Standard for Archival Description (General), or ISAD(G). Elliott (1974), stated that to a records professional, the context of research is more important than the raw data produced, and that correspondence, diaries, reports and publications are therefore a records professional’s priority over data, as they expose the interaction between data and theory.

Shankar (1999) also addresses the similarities between “the record” and research data, comparing scientific lab books to archival records. She notes that electronic lab books allow the capture of scientific data, without taking into account long-term archival and records management needs. Shankar identified a gap in the literature at that time, stating that neither scientific nor archival literature has yet addressed this issue, and considered how records professionals’ concepts of the record map to scientific records. She concluded that a combination of custodial and post-custodial approaches, mediated by archivists, may be appropriate but that more research is required.

Research data and research records

A comparison of records and research data has been undertaken more recently by Evans, Reed et al (2014), who consider research data in the context of the ISO 15489 Records Management Standard, concluding that “research data” are in fact “research records”. While the paper considers that records managers and archivists have the skills and knowledge to address the management of research data, it concludes that a “traditional recordkeeping mind-set” would not be appropriate. Research records are also described by Childs and McLeod (2004), and defined as being comprised of records of the research process; its outcomes or products; the management of the process; and the primary research data themselves. The definition of research records as more than just the primary data reflects the perspective noted by Conway et al (2011) that data can only be truly reusable when documentation about its context and history have also been captured.

Comparisons between research data and “records” are therefore apparent in the literature. Records and data both constitute representations of information, but it is also apparent that archival or records management practice require the capture of context, which is not generally the case with research data management practices. The concept of “research records” appears to address the potential benefits of capturing context around scientific processes. In this case, data is a component of evidence generated by scientific research.

Data curation and research data management skills

Contemporary research data management practice has been defined by a variety of stakeholders and practitioners. In the face of growing advocacy from policy makers to make publicly-funded research data publicly accessible, research performing organisations support research data management facilities which may fall within the remit of institutional libraries or archives, research offices, IT departments, or new functional areas. The theory and practice of managing research data has been described by professionals including records professionals, librarians, scientists and administrative staff.

In this context, it should also be noted that “research data management” practices are often referred to as “data curation” or “digital curation” - where the curation may relate to research data, but can also refer to other digital assets. The lack of distinction between digital curation, research data management and data curation may be the reason that discussions of research data management practices often

refer to digital data collections, even though research data may also include analogue material. Palmer et al (2013) note the conflation of the terms “digital curation” and “data curation” over time, and consider it to be a wholly new practice, which should mix traditional theory and practices from library and information science, museum studies, archival science and computer science. Noonan and Chute (2014) suggest that data curation may be defined as “the activity of managing data throughout its lifecycle, appropriately maintaining integrity and authenticity, ensuring that it is properly appraised, selected, securely stored, and made accessible, while remaining usable in subsequent technology environments.” Beagrie (2006) states that the term “digital curation” is relatively new, having first been used at a Digital Preservation coalition seminar in 2001. It is described as “the actions needed to add value to and maintain these digital assets over time for current and future generations of users.” Digital curation was addressed in a special issue of *Archival Science*, with Poole (2015) describing four institutional loci as central to the curation of scientific data: archives, research libraries, institutional repositories, and centres.

The 2009 Digital Curation Centre white paper (Pryor and Donnelly, 2009) provides an overview of roles and responsibilities in research data curation and identifies relevant data-related international education programmes, which include library and information studies, information engineering, data curation and informatics degrees, as well as those in records management and archival studies. Yarmey and Yarmey (2013) discuss the different perspectives towards research data held by library and archives professionals, and consider that “data curation” is a discipline which cuts across the information and library science fields.

Researchers, librarians and data curation

Wilson et al (2010) consider the extent to which researchers should have to devote their time to archiving their own research data. Where researchers undertake data curation, the technical infrastructure provided to support researchers may be more important than the skills of the institution’s information professionals. They note that research data management has much in common with the management of private archives, but that in research data management it is even more important to engage with the data creator at the beginning of the records life-cycle (or researcher’s) career. This reflects guidance given by the JISC-funded Paradigm Project which identified regular engagement with a

record creator as key to the successful transfer and preservation of born digital or hybrid archives. Regular meetings with record creators to develop trust, advice on recordkeeping practices and facilitate the transfer of near-contemporary record are recommended [vi].

Tools are also available to support researchers in creating their own data management plans, which are often mandated by universities or funding agencies. These include DMPonline [vii] and the University of California DMPTool [viii]. From 2010-2011, the UK funder JISC supported the RDMTrain strand, encompassing projects which provided training to researchers in disciplines including the performing arts, social sciences, psychology and archaeology [ix].

While most RDMTrain modules were aimed at training researchers in data management, one (MANTRA) included a module aimed specifically at librarians [x]. Other examples of literature aimed specifically at librarians include the 2014 the Chartered Institute of Library and Information Professionals (CILIP)'s briefing paper, Auckland's 2012 *Re-Skilling for Research* published by Research Libraries UK, and the RDMRose project which specifically targeted university liaison librarians, aiming to provide CPD materials for research data management [xi].

Records professionals and data curation

Although the 2009 DCC white paper cited above does not draw conclusions on the most appropriate career path for a data curator or research data manager, the DCC has produced a life-cycle model which cites ISAD(G) as its suggested metadata schema for research data [xii], as well as an Archival Metadata Manual which gives guidance on archival standards including ISAD(G), EAD, ISAAR(CPF) and EAC. The 2014 Chartered Institute of Library and Information Professionals (CILIP) briefing paper on how academic libraries can support institutional research data management notes that research data management is more closely related to records management and archival thinking than librarianship.

In 2013 the DigCurv project held a conference on the topic of framing the digital curation curriculum [xiii]. The project was funded by the European Commission to establish a curriculum framework for vocational training in digital curation. Bunn and Higgins (2013) explore the skills required for archivists and records managers to engage with digital curation, and whether these skills are taught by established

educational programmes. They note that in University College London a new module was introduced for digital curation which was developed cooperatively across the Department for Information Studies, to include students in records management and archival studies.

Maday and Moysan (2014) identify institutional and funder requirements for data management plans as an opportunity for archivists to establish themselves as “data management specialists” who can aid researchers. Poole (2015) also identifies key skills of records professionals in engaging with research data management: provenance, appraisal, authenticity, metadata, risk management and trust. Dooley (2015) argues that archivists should be involved in the management of material with “less obvious archival characteristics”, including research data, and gives examples of ten areas of archival expertise and their relevance to this type of digital content.

Conversely, there are some examples of records professionals who state that data should not be a focus or priority. Elliott (1974) cited above, considered that it was unlikely that any scientist would come to a conventional archives service to consult the data of an earlier scientist. Similarly Brichford (1969) believed that “test and experimental data should be destroyed when the information they contain is condensed in published reports or statistical summaries.” However these views possibly reflect mid-20th century appraisal techniques and the necessity to store paper records. More contemporary objections to records professionals’ involvement with research data could not be identified in the literature.

Discussions of research data management skills often conflate both digital curation and data curation, and the definitions and distinctions between these terms vary. Documentation and training material are available for individual researchers, librarians and records professionals in relation to the new skills needed for these activities. The relevance of the existing skills of records professionals, such as appraisal, and the need for records professionals to begin considering the application of their skills to research data is apparent.

Data management and records management approaches

There are examples in the literature of the application of archival or records management theory and practice to research datasets. These include the consideration of post-custodial approaches and the records continuum model. McDonald (2010) argues that electronic records management and data management are fundamentally the same, and that the only gap between the two occurs due to stakeholder perceptions of the two processes. Maday and Moysan (2014) discuss the potential role of records management for research data, arguing that currently, records professionals cannot influence classification or appraisal of current records, and receive the data only when the scientific research project has ended. Similarly McCarthy and Sherratt (1996) use electronic recordkeeping practices as a comparison to research data management identifying archivists as 'salvagers' of the record, where a closer link between the archivist as record-keeper and the scientist as record creator would be preferable.

Thorpe and Gardiner (2012) provide a case study for this approach, discussing data collected by universities and governmental departments in collaboration with an indigenous community in Australia as part of the Aboriginal and Torres Strait Islander Data Archive. The paper posits not only that records management practice should be applied in managing data after collection, but that the records continuum can provide a framework to allow an appropriate interaction between sociological researchers and an indigenous community. Such an interaction, where data includes anthropological field notes and bark paintings, can allow the community to become active participants in the archiving of their own culture. Wallis et al (2008), discuss the role of the records professional in capturing data generated in ecological research. They provide a case study of a scientific research life-cycle, similar to the records life-cycle, which records professionals may use to ensure that data is captured and published at the appropriate point during the research.

Jones (2012) describes the eScholarship research centre at the University of Melbourne, and the role of the non-custodial records professionals working there. He outlines a collaborative approach, acknowledging that the skills of the records professional are only one aspect of the work required to manage and preserve its datasets. The paper describes the interaction of stakeholders including custodial archives, creators of distributed collections of records and data (akin to the 'post-custodial' model), organisational knowledge managers, data archives, academics and communities in the creation

and archiving of datasets, and notes the specialist skills and input required from each in order to archive these datasets.

The project Managing Primary Research Data and Records for Research in HE Institutions used the records continuum model to map practice in research data and records management across research projects at the University of Northumbria. Different challenges and issues were captured through stakeholder interviews, for example “confidentiality” and “retention management”. The project’s final report suggests that these issues may be addressed through the creation of a virtual records centre, coupled with records management training and support. McLeod and Childs (2003) note that “a logical extension is the need to place these guidelines and training within an institutional framework of records management strategies and policies”.

Two projects, DATUM for Health and DATUM in Action, led by Julie McLeod, Professor in Records Management at Northumbria University, ran from 2010-2012 and addressed records management practices for health research data [xiv]. Childs et al (2014) describe how the two projects examined data management planning for opening health data, and conclude that records management professionals are ideally positioned to facilitate this. Skills in appraisal and the creation of retention schedules are noted as specifically relevant to aid researchers in data management planning.

In Ireland and the United Kingdom, data management falls under records management policy in a number of universities. This is the case in the University of Limerick’s Records Management and Retention Policy [xv], as well as British universities including the University of London [xvi], the University of Oxford [xvii], and the University of Liverpool [xviii].

Projects such as those at the University of Northumbria and the Aboriginal and Torres Strait Islander Data Archive describe the application of records management processes to the capture, appraise and create retention schedules for data. A natural progression of this successful approach to data management is the creation of data management policy with an institutional records management policy, which was visible even in the geographically limited sample of university policies which were examined.

Research data in the custody of archives

Childs et al (2014) note that while records managers are ideally placed to aid researchers in data management planning, archivists have a potential role as data custodians. Since the mid-20th century, datasets generated through research activities have formed a part of the collections transferred to the custody of governmental and scientific archives. Although research data is often thought of in terms of scientific research, datasets can and do include other subjects, for example census data, oral histories and longitudinal social science surveys, and arts and humanities research outputs. This final section of the literature outlines example of research datasets which have been managed and preserved by governmental or scientific archives.

Government research data in National Archives

In 1955 Reingold gave an overview of the types of governmental survey data already available for scientific reuse at the National Archives, such as records of geologic, geodetic, and topographic surveys and records containing meteorological and climatological data. In 1960, the National Archives and Records Service issued General Records Schedule 19 for federal research and development records, identifying records possessing historical and scientific research value, as described by Brichford (1969).

Brown (1996) notes that archivists in the American National Archives had engaged with electronic data since the 1970s. He states that one of the largest early transfers of electronic records to the U. S. National Archives was 865 datasets from the Nautical Chart Data Base of NOAA (National Oceanic and Atmospheric Administration). These datasets were then catalogued in two ways, once as data and once as archival records. Yorke (1997) describes the management of exploration results and seismic data resulting from petroleum exploration, which have been in the custody of the National Australian Archives since 1974. The challenges noted relate to digital preservation concerns rather than those concerning datasets specifically.

Garrod (2000) discusses the National Digital Archive of Datasets which has been managed by the National Archives in the UK, focusing on indexing using the UNESCO thesauri. The necessity to add elements to ISAD(G) to more appropriately describe a dataset is mentioned only briefly. Sleeman (2004)

describes the genesis of the National Digital Archive of Datasets (NDAD) at the UK National Archives in the context of archivists' responsibilities to preserve the databases of government departments. Sleeman touches on the appropriateness of the archivist in preserving these datasets, believing that ISAD(G) is the obvious choice for cataloguing data, and that the profession has relevant expertise in presenting information in context. She also suggests that the definition of records should perhaps cover not only business transactions, but evidence of human activity.

O'Neill Adams (2007) describes how the U.S. National Archives and Records Administration acquires collections from federal agencies that are generated by data processing applications, comprised predominantly of quantitative data, specifying that governmental records can and do constitute research data when they are the by-product of governmental research. While administrative datasets also exist, they are considered to be a separate category to research datasets. She states that "The collections of some data archives, especially those with national or international clientele, include selected governmentally produced data that may, depending on the country, also be within the purview of the nation's traditional archives."

Archives and the papers of scientists

Grover (1962), third archivist of the United States, gave explicit reasons why an archivist is the most appropriate professional for scientific recordkeeping. Grover considered the archiving of the scientific record as a gap which must be closed, and stated that as much of the work of scientists today is sponsored by government, it is governmental archives that have responsibility for these records. This reflects contemporary Open Access policy which suggests that publicly funded research should be deposited in an appropriate repository and made publicly accessible, a perspective seen for example in Ireland's National Principles for Open Access (2013).

Advocacy for records professionals' engagement with scientific data, both from scientists and from records professionals themselves, can be identified as early as the 1950s. Harvard Professor Gerald J. Gruman (1958) made a plea in a letter to *Science* that scientific papers should be archived in order to provide context for research, giving the example of a key figure in American science from the turn of the century whose papers had already disappeared. Reingold (1955) also described the absence of an

archival tradition for scientists, in comparison to the relationship between other depositors to the U. S. National Archives such as military or federal agencies. Brichford's 1969 *Scientific And Technological Documentation Archival Evaluation And Processing Of University Records Relating To Science And Technology* was a manual for American university archivists engaging with technical and scientific collections. It anticipated that the creation of "data storage files" by scientists would be an outstanding characteristic of the second half of the 20th century, but also that a unique characteristics of records relating to science and technology at the time was that they had not received due attention from historians and archivists.

The Carl Woese Papers at the University of Illinois Archives include computer printouts, reports, drafts, notes, photographs, slides and transparencies concerning microbiology, evolution, the genetic code, bacteria, translation, ribosomes and ribosomal RNA sequencing amongst other, as discussed by Anderson (2014) in her paper for the International Council on Archives Section on University and Research Institution Archives conference. The conference theme was Research and Scientific Archives and topics included genetic data, neurological data, marine data and hospital records, as well as the application of EAD to scientific data [ixx].

The role of the records professional in preserving scientific papers is also addressed by the National Cataloguing Unit for the Archives of Contemporary Scientists, which aims to preserve and make accessible the archives of distinguished contemporary British scientists and engineers. Harper (2005) describes collections which may include data and experimental results, records of discussions and information, reflections on work, drafts for publications, correspondence, bibliographical references and annotated background material. However he notes that in the United Kingdom, the papers of academic scientists are generally considered to be private- rather than governmental archives.

Two concerns are apparent: the requirement to preserve datasets which are part of governmental records, and the accessioning of the papers of scientists which may also include data. Thus archivists who work in these areas have necessarily needed to address the challenges of research data archiving since the 20th century.

Findings

The literature discussed were divided broadly into the following thematic areas: comparisons of research data and the “record”; the skills required to manage research data; records management approaches to research data management; and research data in the custody of archives. An underlying theme also emerged across the thematic areas of the literature: that the involvement of records professionals with research data can or should include the capture of context along with data.

The first area described how data and research data are described as comparable to records in terms of their role as evidence and their informational value. The skills required for managing such data were then examined, and found to cut across the professions of librarians, records professionals and IT professionals. The individual who should attain these skills may be a researcher, a librarian, a records professional, or a specially trained data curator. The relevance of the theories and practices employed by records professionals are referred to repeatedly.

Case studies which apply records management approaches to research projects were also reviewed. The records life-cycle or records continuum have been used with success to underpin data management practices. The connections between research data management and records management are further emphasised through the inclusion of data management as a component of university records management policies.

Perhaps the most compelling finding is that research datasets have been accessioned as part of archival collections since at least the mid-20th century. In the examples described data is transferred, catalogued and preserved by National Archives using techniques and methods which are often no different to the treatment of other records. Since the 1950s there has been a call for archivists to prevent a gap in the scientific record by preserving the papers and “data storage files” of researchers, and in some cases this gap is addressed by scientific archives.

The overall findings of the literature review can therefore be summarised as follows:

1. Research data is described as comparable to the record in terms of informational and evidential value.

2. Research data management is currently being undertaken by a variety of professionals, including librarians and records professionals.
3. The skills and professional expertise of records professionals are acknowledged to be relevant to the management and preservation of research data.
4. Case studies and research projects have successfully tested the application of records management approaches to managing research data.
5. Research datasets have been accessioned, catalogued and held in the custody of archives since the mid-20th century.

Conclusions

Literature discussing records professionals and the management and preservation of research data was found to be extensive. Despite the volume of literature on the topic, there is no discernible consensus as to whether records professionals are the most appropriate managers or custodians of data however. In fact, in reviewing guidelines, reports and online training modules, librarians or individual researchers are more often the target audience for information and support regarding research data management. Despite this, the skills, expertise and professional practices associated with the work of records managers and archivists are frequently cited as being relevant, and even necessary, to manage research data.

In the context of the case studies and examples seen in the literature, it seems likely that research datasets will continue to be transferred to the custody of national or scientific archives. Although the application of “traditional” archival arrangement and description have been applied by projects such as the National Digital Archive of Datasets in the UK, it is not clear how this impacts the reuse of data by researchers. Internationally, the publication of research data is beginning to align with specific standards such as the FAIR principles [xx] and processes such as the application of persistent identifiers are not generally part of an archivist’s workflow. Additionally, as funding agencies increasingly mandate the publication of research data, universities must develop research data services with the support of institutional librarians, records managers or archivists, amongst others. Given the experience and expertise of records professionals, there is a gap in the literature regarding a potential role of records professionals as leaders in the development of research data management. Additionally there is a lack of

guidance aimed specifically at supporting records professionals tasked with implementing research data management practices in their institutions.

In the context of the findings of this literature review the following are suggested:

- Advocate for professional bodies to publish guidelines and recommendations to support records professionals in addressing the challenges of managing research data.
- Develop guidelines and documentation aimed at users outside of the records management and archival professionals, demonstrating the application of records management and archival theory and practice to research data curation.
- Records professionals already working in research data management roles can engage with the scientific and data science communities through channels such as the Archives and Records Professionals for Research Data group of the Research Data Alliance. This in turn can raise the profile of records professionals as data experts in the broader scientific community.

These activities could serve to crystallise the role of records professionals as experts in research data management, and support records professionals already engaged in the management and preservation of research data.

URLs cited:

- [i] <http://icasuv2014.univ-paris-diderot.fr/conference-13/conference-program/article/speakers>
- [ii] <https://rd-alliance.org/groups/archives-records-professionals-for-research-data.html>
- [iii] <https://rd-alliance.org/groups/libraries-research-data.html>
- [iv] <https://www.elsevier.com/about/company-information/policies/research-data>
- [v] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
- [vi] <http://www.paradigm.ac.uk/workbook/collection-development/snapshot.html>
- [vii] <https://dmp.cdlib.org/>
- [viii] <https://dmponline.dcc.ac.uk/>
- [ix] <http://www.dcc.ac.uk/training/train-trainer/disciplinary-rdm-training/disciplinary-rdm-training>
- [x] <http://datalib.edina.ac.uk/mantra/libtraining.html>
- [xi] <http://www.shef.ac.uk/is/research/projects/rdmrose>
- [xii] <http://www.dcc.ac.uk/sites/default/files/Ingest%20and%20Store%20Checklist.pdf>

- [xiii] <http://www.digcur-education.org/eng/International-Conference>
- [xiv] <https://www.northumbria.ac.uk/sd/academic/ee/work/research/clis/dlar/datum/>
- [xv] <http://www2.ul.ie/pdf/803890985.pdf>
- [xvi] http://www.london.ac.uk/fileadmin/documents/foi/RM_guidance_-_Research_data.pdf
- [xvii] <http://researchdata.ox.ac.uk/university-of-oxford-policy-on-the-management-of-research-data-and-records/>
- [xviii] https://pcwww.liv.ac.uk/csd/records-management/guidance-notes/research_records.pdf
- [ixx] <http://icasuv2014.univ-paris-diderot.fr/conference-13/conference-program/article/contributions>
- [xx] <https://www.force11.org/group/fairgroup/fairprinciples>

References

Akmon, D., Zimmerman, A., Daniels, M., Hedstrom, M. (2011), "The application of archival concepts to a data-intensive environment: working with scientists to understand data management and preservation needs", *Archival Science* Vol. 11 Iss. 3, pp. 329-348.

Anderson, B. G. (2014), "The Structure of Scientific Archives: Digital Research Data and the Nature of Scientific Documentation", paper presented at International Council on Archives, Section on University and Research Institution Archives, July 9, 2014, Paris, France.

Auckland, M. (2012), *Re-Skilling for Research: An investigation into the role and skills of subject and liaison librarians required to effectively support the evolving information needs of researchers*, RLUK, London.

Beagrie, N. (2006), "Digital Curation for Science, Digital Libraries, and Individuals", *International Journal of Digital Curation*, Vol. 1, pp 3-16.

Borglund, E. and Engvall, T. (2014), "Open data: Data, information, document or record?", *Records Management Journal* Vol. 24 Iss. 2, pp. 163-180.

Borgman, Christine L. (2012), "The Conundrum of Sharing Research Data", *Journal of the American Society for Information Science and Technology*, Vol. 63 Iss. 6, pp. 1-40.

Brichford, M. J. (1969), *Scientific And Technological Documentation Archival Evaluation And Processing Of University Records Relating To Science And Technology*, University of Illinois at Urbana-Champaign, Illinois.

Brown, T. E., (1996), "Myth or Reality: Is There a Generation Gap Among Electronic Records Archivists?", *Archivaria*, vol. 41, pp. 234-243.

Buckland, Michael (1991), "Information As Thing", *Journal of the American Society for Information Science*, Vol. 42 Iss. 5, pp. 351-360.

Bunn, J. and Higgins, S. (2013), "Mainstreaming Digital Curation: An overview of activity in the UK archives and records management profession" in Casarosa, V (Ed.), *Proceedings of the Framing the Digital Curation Curriculum conference*, CEUR, Florence.

Chartered Institute of Library and Information Professionals (2014), *Research Data Management Briefing Paper*, available at:
http://www.cilip.org.uk/sites/default/files/documents/Research%20data%20management%20briefing%20July%202014_0.pdf (accessed 29 September 2016).

Childs, S. and McLeod, J. (2004), "Sharing Research Records and Research Data: Findings from a Research Project in Higher Education", *New Review of Information Networking*, Vol. 10 Iss. 2, pp. 131-145.

Childs, S., McLeod, J., Lomas, E. and Cook, G. (2014), "Opening research data: issues and Opportunities", *Records Management Journal*, Vol. 24 Iss. 2, pp. 142 – 162.

Conway, E., Giarretta, D., Lambert, S., and Matthews, B. (2011), "Curating Scientific Research Data for the Long Term: A Preservation Analysis Method in Context", *The International Journal of Digital Curation*, Vol. 6 Iss. 2, pp. 38-52.

Dooley, J. (2015), *The Archival Advantage: Integrating Archival Expertise into Management of Born-digital Library Materials*, Dublin, Ohio.

Doorn, P. and Tjalsma, H. (2007), "Introduction: Archiving Research Data", *Archival Science*, Vol. 7 Iss. 1, pp. 1-20.

Elliott, C. A. (1974), "Experimental Data as a Source for the History of Science", *The American Archivist*, Vol. 37 No. 1, pp. 27-35.

Evans, J., Reed, B., Linger, H., Goss, S., Holmes, D., Drobik, J., Woodyat, B., Henbest, S. (2014), "Winds of change: A recordkeeping informatics approach to information management needs in data-driven research environments", *Records Management Journal*, Vol. 24 Iss. 3, pp. 205 - 233.

Garrod, P. (2000), "Use of the UNESCO Thesaurus for Archival Subject Indexing at UK NDAD", *Journal of the Society of Archivists*, Vol. 21 No. 1, pp. 37-54.

Grover, W. C. (1962), "The Role of the Archivist in the Preservation of Scientific Records", *Isis*, Vol. 53 No. 1, pp. 55-62.

Gruman, G. J. (1958), "Preserving the Stuff of History", *Science*, Vol. 127 No. 3313, p. 27.

Harper, P. (2005), *Preserving scientific archives: the work of the National Cataloguing Unit for the Archives of Contemporary Scientists*, Museu de Astronomia e Ciências Afins, Rio de Janeiro.

International Council on Archives (2000), *ISAD(G): General International Standard Archival Description Second Edition*, ICA, Ottawa.

International Council on Archives (2015), "Multilingual Archival Terminology: Archivist", available at: <http://www.ciscra.org/mat/mat/term/65> (accessed 29 September 2016).

International Council on Archives (2015), "Multilingual Archival Terminology: Record", available at: <http://www.ciscra.org/mat/mat/term/60> (accessed 29 September 2016).

International Council on Archives (2015), "Multilingual Archival Terminology: Records Manager", available at: <http://www.ciscra.org/mat/mat/term/298> (accessed 29 September 2016).

ISO/TC 46/SC 11 (2016), "ISO 15489-1:2016 Information and documentation -- Records management -- Part 1: Concepts and principles", available at: http://www.iso.org/iso/catalogue_detail?csnumber=62542 (accessed 29 September 2016).

Jones, M. (2012), "'Encountering the Stranger': Working digitally to connect records and data for communities", paper presented at the ICA Congress, 20-24 August 2012, Brisbane, available at: <http://ica2012.ica.org/files/pdf/Full%20papers%20upload/ica12Final00278.pdf> (accessed 23 September 2016).

Knight, G. and Pennock, M. (2009), "Data Without Meaning: Establishing the Significant Properties of Digital Research", *The International Journal of Digital Curation*, Vol. 1 Iss. 4, pp. 159-172.

Maday, C. and Moysan, M. (2014), "Records management for scientific data", *Archives and Manuscripts*, Vol. 24 No. 2, pp 190-192.

McCarthy, G. and Sherratt, T. (1996), "Mapping Scientific Memory: understanding the role of recordkeeping in scientific practice", *Archives and Manuscripts*, Vol. 24 Iss. 1, pp 78-85.

McDonald, J. (2010), "Records management and data management: closing the gap", *Records Management Journal*, Vol. 20 Iss. 1, pp 53-60.

McLeod, J. and Childs, S. (2003), *Managing Primary Research Data and Records for Research in HE Institutions: Final Project Report*, Northumbria University, Newcastle.

McLeod, J. and Hare, C. (2010), "Development of RMJ", *Records Management Journal*, Vol. 20, Issue 1, pp 9-40.

Miller, W. E. (1976), "The Less Obvious Functions of Archiving Survey Research Data", *The American Behavioural Scientist*, Vol. 19 Iss. 4, p409.

National Standards Authority of Ireland (2004), "Information And Documentation, Records Management. Part 1: General" available at:
http://www.taoiseach.gov.ie/attached_files/Pdf%20files/30%20ISO.15489-1%20-%20IRISH%20VERSION.pdf (accessed 29 September 2016).

National Steering Committee on Open Access Policy (2013), "National Principles for Open Access Policy Statement", available at:
<http://www.dri.ie/sites/default/files/files/National%20Principles%20on%20Open%20Access%20Policy%20Statement%20%28FINAL%2023%20Oct%202012%20%29.pdf>, (accessed 29 September 2016).

Nielsen, Per (1995), "Merging Cultures: Danish Integration of Academic Data Service into Traditional Archive System". *IASSIST Quarterly* 19 No. 2, pp. 27-35.

Noonan, D., and Chute, T. (2014), "Data Curation and the University Archives", *The American Archivist*, Vol. 77 No. 1, pp. 201-240.

O'Neill Adams, M. (2007), "Analyzing archives and finding facts: use and users of digital data records", *Archival Science*, Vol. 7 Iss. 1, pp. 21-36.

Palmer, C. L., Weber, N. M., Munoz, T. and Renear, A. H. (2013), "Foundations of Data Curation: The Pedagogy and Practice of "Purposeful Work" with Research Data", *Archive Journal*, Vol. 3.

Poole, A. M. (2015), "How has your science data grown? Digital curation and the human factor: a critical literature review", *Archival Science*, Vol. 15, pp. 113-119.

Pryor, G. and Donnelly, M. (2009), "Skilling Up to Do Data: Whose Role, Whose Responsibility, Whose Career?", *International Journal of Digital Curation*, Vol. 4 Iss. 2.

Reed, B. (2005), "Records", McKemmish, S., Piggott, M., Reed, B., Upward, F. (Eds.), in *Archives: Recordkeeping in Society*, Centre for Information Studies, Charles Sturt University, Wagga Wagga, New South Wales, pp. 101-130.

Reingold, N. (1955), "The National Archives and the History of Science in America", *Isis*, Vol. 46 Iss. 1, pp. 22-28.

Shankar, K. (1999), "Towards a framework for managing electronic records in scientific research", *Archival Issues*, Vol. 24 Iss. 1, pp. 21-36.

Sleeman, P. (2004), "It's Public Knowledge: The National Digital Archive of Datasets", *Archivaria*, Vol. 58, pp. 173-200.

Thorpe, K. and Gardiner, G. (2012), "Creating, preserving and connecting data in the digital domain", paper presented at the ICA Congress, 20-24 August 2012, Brisbane, available at: <http://ica2012.ica.org/files/pdf/Full%20papers%20upload/ica12Final00278.pdf> (accessed 29 September 2016).

Wallis, J. C., Borgman, C. L., Mayernik M. S., and Pepe, A. (2008), "Moving Archival Practices Upstream: An Exploration of the Life Cycle of Ecological Sensing Data in Collaborative Field Research", *The International Journal of Digital Curation*, Vol. 3 Iss. 1, pp. 114-126.

Wilson, J. A. J., Fraser, M. A., Martinez Uribe, L., Patrick, M., Akram, A., and Mansoori, T. (2010), "Developing Infrastructure for Research Data Management at the University of Oxford", *Ariadne*, Vol. 65.

Yarmey, K. and Yarmey, L., (2013), "All in the Family: A Dinner Table Conversation about Libraries, Archives, Data, and Science", *Archives Remixed Critical Perspectives and Pathways: Data Curation*, Iss. 3.

Yorke, S. (1997), "Management of Petroleum Data Records in the Custody of Australian Archives", *Archives and Manuscripts*, Vol. 25 Iss. 1, pp. 62-73.