Towards best practice in research data management in the humanities

by

Catherine Anne Woeber

Submitted to the School of Information Management,
Victoria University of Wellington
in partial fulfilment of the requirements for the degree of
Master of Information Studies

June 2017

Abstract

Research problem: This investigation examined research data management (RDM) in the humanities in UK and Australian institutions to assist academic librarians in New Zealand to develop RDM services that match humanities researchers' needs. The study identified transferable international policies and practices in an under-researched knowledge domain to better support humanities data curation.

Methodology: This investigation adopted a domain analytic approach, and selected information-rich policies, data curation profiles, project reports and technical plans from the UK and Australia for a qualitative document analysis and evaluation. The study was conceived as a systematic review of evidence towards best practice.

Results: Humanities data or "primary materials" consist mainly of collections of digital images, texts, audio and visual recordings, although non-digital (analogue) data are common, especially in the creative arts. Humanities researchers tend to keep their primary materials on their own digital devices, in cloud storage or in physical folders and avoid using networked systems. They are generally prepared to share their data with a plurality of audiences. Humanities data curation benefits from consideration of copyright and intellectual property, curation for very long-term storage and access in a federated system, and digitisation of selected analogue data. Effective humanities RDM begins upstream in the data lifecycle with targeted training, active partnerships, and liaison on the data management plan (DMP), and requires strategic cooperation between researchers, the library and institutional/data repository, and IT services.

Implications: Understanding domain-specific policies and practices in the UK and Australia for curating humanities research data will help New Zealand institutions provide normative and strategic humanities data curation. Further research might include widening the investigation to institutions in the United States and/or Europe, or ascertaining the degree of traction of domain-specific RDM in New Zealand.

Keywords: Research data management, Domain-specific data curation, Humanities, Primary materials, Australia, United Kingdom

Acknowledgements

My sincere thanks to my supervisor, Brenda Chawner, for starting me off on this path and seeing me safely home, and to my husband, John Read, for walking with me and trusting in my conviction that I would eventually reach my destination.

Table of Contents

List of Abbreviations and Acronyms

AHDS – Arts and Humanities Data Service

AHRC – Arts and Humanities Research Council

ANDS – Australian National Data Service

ARC – Australian Research Council

CAiRO – Curating Artistic Research Output

CONZUL – Council of New Zealand University Librarians

CSV – Comma Separated Values

DaaS – Database as a Service

DARIAH – Digital Research Infrastructure for the Arts and Humanities

DCC – Digital Curation Centre

DCP – Data Curation Profile

DIAMM – Digital Image Archive of Medieval Music

DMP – Data Management Plan

FOSS – Free and Open-Source Software

GIS – Geographic Information System

GSA – Glasgow School of Arts

HuNI – Humanities Networked Infrastructure

IT – Information and Technology

JISC – Joint Information Systems Committee

JISCMRD – JISC Managing Research Data

NeCTAR – National e-Research Collaboration Tools and Resources

NEH – National Endowment for the Humanities

OCLC – Online Computer Library Center

ORDS – Online Research Database Service

OTA – Oxford Text Initiative

PI – Principal Investigator

RADAR – Repository of Art Design Architecture Research

RCUK – Research Councils United Kingdom

RDM – Research Data Management

RIN – Research Information Network

SCM – Sydney Conservatorium of Music

SPSS – Statistical Package for the Social Sciences

SUDAMIH – Supporting Data Management Infrastructure for the Humanities

TEI – Text Encoding Initiative

UAL – University of the Arts, London

UCA – University for the Creative Arts, London

UCLA – University of California, Los Angeles

UK – United Kingdom

UNESCO – United Nations Educational, Scientific and Cultural Organization

UNISIST – United Nations International Scientific Information System

US – United States of America

VADS – Visual Arts Data Service

XML – EXtensible Markup Language

> *Data should no longer be abandoned on the woodbench like wood shavings in a carpenter's shop; increasingly it is expected to join the finished assembly of scholarly output as a valued and managed component with an extended life and sustained usability. (Pryor, 2012b, p. 4)*

1. Introduction

1.1 Topic statement

This investigation examined research data management (RDM) in the humanities in UK and Australian institutions to assist academic librarians in New Zealand to develop RDM services that match humanities researchers' needs.

1.2 Research problem

Research today is becoming more data-driven. Data generated by research are "essential scholarly objects to be captured, mined, used, and reused" (Borgman, 2009, para. 23), and therefore require preservation. Research data management (RDM) is the curation of "raw and abstracted material created as part of research processes and which may be used again as the input to further research" (Flanders & Muñoz, n.d.). Research data are increasingly viewed as an economic resource to be organised and controlled, and the sciences, including the human sciences, "have become significant objects of state policies and management" (Whitley, 2000, p. ix). The collaborative management of digital research data is thus high on national agendas, like New Zealand's National Research Data Programme (eResearch2020, 2015). The Council of New Zealand University Librarians (CONZUL) has formed a working group to promote aspects of RDM that would benefit from national oversight, while recognising that other aspects are better supported institutionally (CONZUL Working Group, 2015, p. 8). RDM spans the data lifecycle, from data planning through collection to preservation and disposal, and its stakeholders include the government, funders of research, research institutions, and researchers themselves (CONZUL Working Group, 2015, p. 12). The impact of technology in recent years,

however, has meant that researchers' ability to generate data has outstripped their ability to manage it. Large volumes of data are stored in unknown locations on unstable media, often without contextual documentation or metadata, a situation in which evidence of the scholarly record, previously available for validation and sharing, is likely to be lost (CONZUL Working Group, 2015, p. 7).

Although data in the natural sciences have long been curated to optimise knowledge production and transfer, attention has recently been extended to humanities data. The humanities are regarded by the American Council of Learned Societies (2006, p. 7) to encompass modern and classical languages, linguistics, literature, history, jurisprudence, ethics, philosophy, archaeology, comparative religion, and the history, criticism and theory of the arts. However, despite the potential of humanities research data to validate research outcomes and be reused for new purposes, their place in current RDM policy and decision making is not yet assured. There is thus a need to investigate current international documentation and initiatives in humanities RDM in order to inform best practice policymaking. Best practice is defined by Wallace and Van Fleet as "accepted approaches to action that are based on the shared understanding and collective evaluation outcomes of a professional body" and is developed from research results which inform action and decision making, specifically "approaches and methods for improving professional practice" and "meaningful policy implications" (Wallace & Van Fleet, 2012, p. 22). According to Patton (2002, p. 564), identifying such "lessons learned" has become a popular reason for "cluster evaluations that aim to build knowledge comparatively". International management of the research data generated by humanities scholars with a view to establishing best practice is the topic of this investigation.

1.3 Rationale for the study

The New Zealand research community has fallen behind the rest of the developed world in collecting and curating research data in the sciences (eResearch 2020, 2015, p. 4). Our research institutions have been late investors in new technologies, the research data space today is fragmented, and catch-up investment in e-research infrastructure, which includes academic libraries, is needed to maintain our research impact (eResearch2020, 2015, p. 5). The CONZUL Working Group (2015, pp. 7-8)

tasked with crafting a university-wide position on data curation is convinced that our fragmented approach to all knowledge domains, including the social sciences and humanities, has resulted in RDM practices that are not conducive to competitive research and threaten scholarly conventions in data validation and sharing. A failure to be proactive in this regard has left New Zealand in the wake of developments elsewhere, and a concerted approach to RDM across domains is urgently required. Research data in New Zealand universities tend to be distributed across research groups and faculties, and datasets kept by researchers themselves (CONZUL Working Group, 2015, p. 50), who are typically more concerned with data creation than curation. The CONZUL Working Group (2015, p. 5) recognises, in addition to finding technological solutions to RDM, the cultural shift that must accompany the structuring of research data "appropriate to discipline conventions" to make it available to support publication or benefit the wider community. If this shift is to be achieved, then identifying transferable international policies and procedures in under-researched knowledge domains like the humanities is a necessary first step.

1.4 Importance of the study

Comparatively less research has been done in the humanities than in the natural and social sciences "to understand the work practices and the kinds of digital materials practitioners might generate" (Gilliland, 2014, p. 235). A Research Information Network (RIN) report of an investigation into humanities information practices in the UK today concluded that these practices must be understood "if we are to provide the right kind of support for researchers" (Bulger, Meyer, De la Flor, Terras, Wyatt, Jirotka, Eccles, & Madsen, 2011, p. 6). Moreover, it is the variety of small datasets or collections of digital objects generated by humanities scholars – which are often messier than the natural sciences' big datasets – that academic librarians are most likely to deal with (Akers, 2013, p. 58). According to Borgman (2009, para. 29), "we are only beginning to understand what constitute data in the humanities, let alone how data differ from scholar to scholar", and this investigation aimed to help clarify what defines "the humanities laboratory of the 21st century" (Borgman, 2009, para. 77), and therefore evaluate how best to curate its research data.

Taking my lead from the CONZUL Working Group (2015, p. 8) and the OCLC's call to action for library directors, *Starting the conversation: University-wide research data management policy* (Erway, 2013, p. 5), I have chosen in this short project to examine institutional management of research data (sometimes known as the "long tail" of RDM) rather than *infrastructural* research data management ("big data"), although both are necessary for development of a complete humanities RDM strategy. Since the outcomes of evaluation research are transferable "when the results of a project conducted in one environment have direct implications for decision making or action in a different environment" (Wallace & Van Fleet, 2012, p. 21), the results of this study will assist academic librarians in New Zealand to develop best practice RDM policy and services for researchers in the humanities, and contribute to the process of catch-up investment in e-research infrastructure.

2. Literature review

This investigation has used the domain analytic paradigm (Hjørland & Albrechtsen, 1995) in evaluating international documentation relating to institutional management of humanities research data with a view to discovering principles that could inform best practice. Domain analysis, advocated by Hjørland and others as an alternative to individual behavioural and cognitive paradigms (Hjørland & Albrechtsen, 1995, p. 400), has been described by Bawden and Robinson (2013, p. 103) as "a framework for studying information communication within subject areas and user groups and for the provision of information services to such groups". Domain analysis is a socio-cognitive approach in that it considers knowledge sharing within groups or fields (Bawden & Robinson, 2013, p. 92). To justify using the domain analytic approach, literature had to be found demonstrating that the approach is valid for RDM, that humanities data are part of RDM, and that effective practices in RDM exist.

The literature selected for this review is representative in its coverage, supports my choice of theory and topic of investigation, and provides a background to issues which are among those my research hopes to investigate further. These include (1) applicable theory, (2) humanities data and practices, (3) policies and initiatives, (4)

services in academic libraries, and (5) immersion or partnership models. The literature was discovered through searching Te Waharoa (Victoria University Library's discovery portal), scanning references cited within relevant literature and following citations of relevant articles in Google Scholar – known as "retrospective" and "prospective" reference list checking (Gough, Oliver, & Thomas, 2012, pp. 125-126) – and browsing through recent issues of relevant e-journals.

2.1 Applicable theory

The domain analytic paradigm has previously been applied to the natural sciences in order to analyse the distinguishing features of the domain and how these features influence scientists' data practices. **Weber, Baker, Thomer, Chao and Palmer (2012)** examined five sub-disciplines within earth sciences to find out how data change in value as they are re-used, whether data can *gain* in value and under what conditions this could happen, and how such changes would affect the design of e-research infrastructures. They drew on Hjørland and Albrechtsen (1995) and subsequent domain analytic studies to support their study's focus on the sub-disciplinary level. Interviews with principal investigators revealed the importance of capturing the context of data production, that data gained value when used in different contexts, and that infrastructure design must thus be capable of managing ever more robust data. **Roos and Hedlund (2016)** applied domain analysis theory to data practices in biomedicine. Their study used Richard Whitley's work on the intellectual infrastructure of scientific domains, which are influenced by the "degree of mutual dependence" between researchers and the "degree of task uncertainty" (Whitley, 1984, 2000). Many natural sciences are characterised by high mutual dependence (e.g. teamwork) and low task uncertainty (e.g. standard procedures). By applying this theory to previously collected data in biomedical sub-disciplines, the authors concluded that data practices need to be studied in their (sub-)disciplinary context, and domain-specific information tools and services need to be developed. Their application of domain analytic theory (and Whitley's work) to the sciences provides justification for my approach. The low mutual dependence and high task uncertainty which characterises humanities disciplines (typically lone researchers pursuing idiosyncratic studies) affects data management and sharing practices.

2.2 Humanities data and practices

A significant difficulty in any investigation into the humanities is the complexity, indeed amorphousness, of humanities data, which requires nuanced exploration. **Borgman's (2015)** chapter on humanities data scholarship in her important book on the data deluge has tried to grapple with what characterises humanities data. In two extended case studies, she considers ancient sites, artefacts, and scriptural texts and recognises that it is the sheer diversity of the objects of study and research methods in the humanities that constitute the challenges of characterising their data. In his examination of the design for the Humanities Networked Infrastructure (HuNI), a 'virtual laboratory' built under Australia's NeCTAR programme, **Burrows (2013)** views humanities data as the "annotations, tags, links, associations, ratings, reviews, and comments produced during the humanities research process", together with the semantic entities (such as person and place) to which these interventions refer. Although my study has not focussed on infrastructural initiatives like HuNI, Burrows' account of the project provides useful information about the variety of humanities data requiring curation. More relevant for my purposes, **Henry (2014)** describes a successful collaboration in humanities digital scholarship at Rice University in Texas. Her case study furnishes extensive examples of types of humanities data curated in partnership with researchers, acknowledging both the richness and the challenges.

Useful data about the research data types and practices of humanities scholars (in comparison with other domains) have been collected in several surveys of tenured academics at different US universities (**Akers & Doty, 2013**; **Weller & Monroe-Gulick, 2014**; **Kutay, 2014**; **Whitmire, Boock, & Sutton, 2015; Mohr, Bishoff, Bishoff, Braun, Storino, & Johnston, 2015**). Among common findings were that data practices vary by research methodology and academic discipline; manuscripts, photographs and videos are common types of humanities data; and effective RDM services require discipline-specific approaches. Humanities scholars, for example, tend to prefer the terms "research materials" or "primary materials" to "data", and **Mohr et al. (2015)** devised their methodology to take this into account, using in their survey the terms researchers used to describe their own research practices. These studies are complemented by **Kennan and Markauskaite (2015)**, who in 2011 surveyed the research data types, storage and sharing practices of academics at ten

New South Wales universities, but did not analyse the data across domains, which limits the usefulness of the study for my purposes. New research by **Trace and Karadkar (2016)**, however, shows how data are captured and used by humanities researchers. The authors interviewed researchers who were using archives in the US, Germany and Canada, to investigate how they collect and organise primary materials (for personal collections constitute a form of research data), so that mobile software can be developed to support their practices. **Rath's (2016)** participant observer case study, in which she tested two free and open-source software (FOSS) data sharing tools, TAMS Analyzer and Viewshare, recognises the "expansiveness" of humanities data, and the economic and technological barriers faced by humanities scholars. Attention to what constitutes humanities data and digital information practices in my study will help to define the humanities laboratory of today.

2.3 Policies and initiatives

Policy documents from academic institutions are the first outcome of any new RDM services, reflect prevailing trends, concerns and attitudes, and are revised in the light of experience, so they are important to any investigation towards best practice. Research funders' data policies are also instructive. **Jones (2012)** provides an overview of UK research funders' data policies and requirements, and shows that the rhetoric of data policies emphasises data sharing. The other chapters in **Pryor (2012a)**, although a bit dated, still offer a useful point of comparison with any current assessment of how trends and drivers in the UK landscape have shifted. **Pryor, Jones and Whyte (2014)** provide case studies of three universities across the US, the UK and Australia, as well as two UK national bodies, to show the variety in RDM policies and practices. Case studies of four US universities are also included in **Ray (2014)**, and the book as a whole makes a foundational contribution to the field. Two anonymous UK case studies were used by **Higman and Pinfield (2015)** to deepen the findings of a content analysis of 37 RDM policies from UK academic institutions, which showed that virtually all policies deal with data security, funders' requirements, data sharing, data preservation and data management plans. They used interviews with staff involved in RDM policy and services to interpret the findings, which they then viewed through the lens of Actor Network Theory, in which documents become agents for change. Case studies are very useful because they show an institution

engaging with the challenges of humanities RDM, and they allow for comparison with recent documents in my study to see which policies have stood the test of time.

2.4 Services in academic libraries

Surveys or interviews with librarians have become common methods to assess the state of institutional RDM programmes. The results clearly show that significant development has taken place in academic libraries in the past five years. The first examination of New Zealand academic libraries' involvement in RDM was **Brown (2010)**. She surveyed librarians in late 2009, and the results of the snapshot study showed that RDM at New Zealand institutions was still in the early planning stages. That New Zealand research organisations are now responding to the requirements of e-research is evident from **Ohaji's (2016)** exploration of the data librarian role. Data from interviews with New Zealand librarians, researchers, and repository and IT managers in 2014 was triangulated with analysis of data librarian advertisements from Australia and the US, and two Australian institutional e-research profiles, to interpret the findings. In an influential study, **Corrall, Kennan and Afzal (2013)** surveyed 219 academic libraries in Australia, New Zealand, Ireland and the UK to investigate what RDM services were being offered or planned, what constraints existed, and what training and support were required. Their findings showed that data curation competencies and understanding of the research environment were now required, and that point-of-need service models were planned for delivery.

Some of their data is supported by the studies of **Cox and Pinfield (2014)** and **Pinfield, Cox and Smith (2014)**. The first reported on a quantitative survey to find out what RDM services UK university libraries offered, which of these were a priority, and what the key issues and attitudes were. The second aimed to gain a fuller picture of developments by interviewing volunteers from among the original survey respondents. The survey results showed limited RDM services in an increasingly complex arena, although policies would soon be in place since RDM was regarded as a core library activity for the future. However, when the authors viewed the findings through the lenses of the 'hype cycle' model and Andrew Abbott's work on professional jurisdiction (the competitive-cooperative tension between roleplayers), they concluded that the library's role in RDM is still unclear. The interview data

assisted the authors in capturing the complexity of RDM in a library-oriented model of an institutional RDM programme. The up-to-date survey by **Cox, Kennan, Lyon and Pinfield (2017)** of 170 academic libraries in Australia, Canada, Germany, Ireland, the Netherlands, New Zealand and the UK, indicates that RDM services in the UK and Australia are the most 'mature' when the "landscape maturity model", reflecting a complete level of service development, is applied. This snapshot study shows that, despite New Zealand libraries' interest in RDM services, they lag behind the UK and Australia, which validates my focus on institutions in those countries.

2.5 Immersion or partnership model

The idea that librarians need to partner with researchers in collaborative RDM, and not just support them, is rapidly gaining traction, and several of the chapters in **Ray (2014)** deal with collaboration, as do the four case studies. **Lyon (2016)** considers an immersive delivery model for RDM services, having explored laboratory-based RDM sessions at the University of Pittsburgh to find out how the model might inform the design of library RDM services. She concludes that best practice RDM services lie on the far end of an evolving service continuum (from transactional through to immersive), because the immersion model provides for the delivery of RDM services at the point of need. **Cox (2016)** offers a thorough review of recent research in the area of library partnerships in digital scholarship, including RDM. He shows that libraries are starting to promote their expertise and facilities as partners in digital scholarship, and adduces as an example the collaboration between the University of Maryland Libraries and the Maryland Institute for Technology in the Humanities in developing a system to sustain and promote digital humanities resources. The partnership model may also get libraries involved in funding proposals and grant applications. Effective initiatives like these have guided the selection of my data.

2.6 Relevance of studies

Taken together, the literature suggests that a domain analytic approach might be applied to humanities RDM (Weber et al., 2012; Roos & Hedlund, 2016), that there are challenges specific to the management of humanities data (Weller & Monroe-Gulick, 2014; Kutay, 2014; Whitmire et al., 2015; Mohr et al., 2015), and that there is

sufficient up-to-date coverage of international projects, policy developments and initiatives to provide the data for a document analysis. Higman and Pinfield (2015) show what could be gained through a content analysis of RDM policies. Institutional RDM models such as those in Pinfield et al. (2014), Lyon (2016) and Cox (2016) could be used as benchmarks to assess activities discovered through my research, while the overview of UK policies in Jones (2012) and literature reviews of Corrall et al. (2013) and Cox and Pinfield (2014) could inform my purposive sampling in order to gain the 'best' data. The data which Brown (2010) collected from New Zealand libraries and Kennan and Markauskaite (2015) from Australian libraries, even if it now lags behind developments, could be useful when compared with the up-to-date data collected by Ohaji (2016) and Cox et al. (2017). The hypothesis tested in Akers and Doty (2013) justifies the humanities as a viable focus for my investigation, while the studies of Burrows (2013), Henry (2014), Trace and Karadkar (2016) and Rath (2016) have demonstrated the variety of initiatives relating to humanities data which need consideration. Finally, the foundational books in the field of RDM, which also offer case studies of humanities RDM for examination (Pryor, 2012a; Pryor et al., 2014; Ray, 2014; Borgman, 2015), are indispensable for my study.

3. Objectives and Research question

As Bawden and Robinson (2013, p. 97-98) point out in their chapter on domain analysis, information practitioners working in a particular domain need to become familiar with the information practices of users in that domain, and with the "forms of documents and content very closely associated" with it, such as archival sources in history or recorded sound in music. One method of doing this is by compiling a data curation profile (DCP). The DCP is drawn up after a reference interview with a researcher to identify their disciplinary needs in managing, sharing or curating their data (Carlson, 2012, p. 8). It describes a researcher's dataset and its lifecycle, indicates how the researcher manages it, and ascertains what the researcher wants to do with the dataset in future (Carlson, 2012, p. 8). Because it is domain-specific, the DCP is a useful tool for information practitioners trying to meet the RDM needs of their constituencies. A Data Curation Profile Toolkit has been developed by Purdue

University Libraries which can be customised to support practitioners to learn more about RDM in a particular domain (Witt, Carlson, Brandt, & Cragin, 2009, p. 102). Analysis of representative DCPs and related forms of documentation should thus provide answers to my main research question: How are research data in the humanities managed by institutions internationally?

My research sub-questions are:

1. What constitutes humanities research data and why do they need curating?
2. What characterises research data use and reuse by humanities scholars?
3. What policies and initiatives are driving institutional humanities RDM?

4. Research design and Methodology

4.1 Research design

A comparative qualitative design informs this investigation, which entails formative evaluation research in its intent to "examine and judge the processes and outcomes aimed at attempted solutions" (Patton, 2002, p. 218), i.e. evaluate the ways in which humanities RDM is engaging with challenges in order to improve current services. Although Patton (2002, p. 220) says that formative evaluation is context-specific, it is likely that what is applicable to a research library in the UK is loosely applicable here, since it is a "parallel institutional environment" (Wallace & Van Fleet, 2012, p. 22). My research has evaluated documentation from a selection of institutional RDM programmes and initiatives in the UK and Australia, looking for emergent themes and patterns of association "without presupposing in advance what the important dimensions will be" (Patton, 2002, p. 56). Purposive sampling was used to discover the 'best' data for my investigation, and so could be considered "intensity sampling" (Patton, 2002, p. 234), i.e. "information-rich cases" (Patton, 2002, p. 46, p. 230), selected because they are intense, but not egregious, examples of their kind. Since Patton (2002, p. 246) advises specifying "minimum samples" and adjusting sample

size to emergent information, the sample size envisaged for this study was 15-20 examples, although this increased to 23 within my flexible research design.

4.2 Theoretical perspective

This investigation adopted a qualitative approach to analysis of international documentation on humanities RDM and relevant projects initiated within the past decade, and applied to the examination Birger Hjørland's domain analysis theory, which Talja (2005, p. 126) explains can benefit scholars' information practices and improve scientific information exchange. The domain analytic paradigm (Hjørland & Albrechtsen, 1995) promotes the study of specific knowledge domains as discourse communities, and uses as its model the UNISIST model of scholarly communication developed by UNESCO in 1971, but later updated and modified to include the humanities and social sciences (Hjørland, Andersen, & Søndergaard, 2005, p. 1). The model reflects the "communication structures inside a scientific domain" which connect knowledge producers with knowledge users, and also with a grid of information practitioners, institutions and technologies (Christensen, 2014, p. 264). In this model, research data counts as an information source alongside informal and formal (both published and unpublished) sources, and the updated UNISIST model (Hjørland et al., 2005, p. 8) includes sources important to the humanities.

Richard Whitley's (1984) influential study of the social and intellectual organisation of scientific fields, republished with a new introduction in 2000, was used to underpin and enrich the domain-specific approach to this investigation. Whitley classes the human sciences as a "fragmented adhocracy" (2000, p. xxxii) because research in this domain is "rather personal, idiosyncratic, and only weakly co-ordinated across research sites" (1984, p. 159), although it has become more team-driven, especially when researchers are practitioners of the digital humanities. This still consolidating field, which Waters (2013, p. 4), following Kathleen Fitzpatrick, has tentatively defined as "the application of digital resources and methods to humanistic inquiry", was included in this study only in relation to research data and e-research methods. Also excluded as a focus of this examination was humanities *information-seeking* behaviour, which is well documented (see e.g. Talja, 2005).

4.3 Data collection

Research data are evidence of a process, have evidential value for research, and thus function as records (Gilliland, 2014, pp. 234-235; Borgman, 2015, p. 24, p. 28). If credible recordkeeping relies on literary warrant – an analysis of regulations, standards, guidelines and best practices codified in the professional literature (Duff, 1998, p. 91) – to endorse recordkeeping procedures, so should responsible RDM. My methodology comprised a qualitative analysis of policies, reports, guidelines and data curation profiles from the UK and Australia – regions identified, together with the US, by the CONZUL Working Group (2015, p. 7) as being ahead of the pack. Although I initially intended to collect samples of RDM policies from the US, it became clear that US universities do not place the same weight that Australian and UK institutions do on institutional RDM policy to guide RDM practices. The US is less systematised and its institutions more entrepreneurial when choosing RDM models and activities (Ohaji, 2016, p. 49). Moreover, given the limited scope of the project, and because New Zealand is more likely to look to the UK and Australia for guidance in RDM, I included US documentation only where it enriched my argument.

Documents constitute a cost-effective and efficient method of data collection which does not require ethics approval, and the data is available online. My approach followed naturalistic enquiry (Bowen, 2008, p. 138; Wallace & Van Fleet, 2012, p. 27), characterised by purposive sampling, contextual awareness, inductive analysis, trustworthiness, and data saturation (Bowen, 2008, p. 139). Selection of specifically information-rich data from the documentation available proceeded in order to reach saturation point. Data saturation is achieved when data is gathered "to the point of diminishing returns" (Bowen, 2008, p. 140). I also actively looked for inconsistent data which could indicate the extent of variability (Leedy & Ormrod, 2015, p. 280) in good practice documentation. Data credibility required selecting "a wide array of documents providing a preponderance of evidence" (Bowen, 2009, p. 33). Samples included flagship initiatives, like Supporting Data Management Infrastructure for the Humanities (SUDAMIH) (University of Oxford, 2009), pioneering guidelines, like those developed by the University of Sydney (2015) for its Conservatorium of Music (SCM), and influential reports, such as the RIN report (Bulger et al., 2011). The criteria for inclusion were English-language documents and studies produced since

2009 which focus on institutional RDM in the humanities, i.e. networked initiatives like the European Union's Digital Research Infrastructure for the Arts and Humanities (DARIAH) or Australia's HUNI would not be subject to analysis. The 23 documents which comprise my research data are listed in **Appendix 1**, and include relevant RDM policies, RDM project reports, technical plans, and data curation profiles.

It became apparent immediately from my data collection that institutional RDM policies, which have only really been finalised in the past four years, do not yet distinguish between different scholarly fields. This is short-sighted, according to Borgman (2015, p. 38), for "policies that fail to respect the substantial differences in theory, practice, and culture between fields are likely to be implemented poorly, be counterproductive, or be ignored by their constituents". It could be expected, as RDM gains traction, that institutional policies start to make specific recommendations for the various knowledge domains. The University of Sydney has shown direction here in that, in addition to its RDM Policy and Procedures (University of Sydney, n.d.), it is developing a suite of domain-specific RDM provisions to give practical effect to the earlier document. The SCM's 18-page RDM provisions (University of Sydney, 2015) elaborate on requirements relating to the nature of music research data; data management plans (DMPs); ownership, retention and storage of data; preferred formats; sharing of research data and metadata; and disposal of data. In its treatment of the complex nature of research data generated by five very different music sub-disciplines, it sets the standard for one field of humanities RDM.

4.4 Data analysis

The documentation was analysed and evaluated using document analysis (Bowen, 2009; O'Leary, 2014, pp. 250-251), which combines content analysis and thematic analysis. Content analysis organises meaningful passages of text into thematic categories related to the research questions, while thematic analysis recognises patterns within the data which then inform the categories for analysis (Bowen, 2009, p. 32). Document analysis requires interpretive rigour, since documents created for purposes other than my own might have insufficient detail, or be biased through their alignment with organisational policy or agendas (Bowen, 2009, pp. 32-33; O'Leary, 2014, pp. 250-251). Document analysis must therefore consider the purpose of the

documents selected, the context of their production, and their intended audience (Bowen, 2009, p. 38), what O'Leary (2014, p. 251) describes as "the latent content or 'unwitting' evidence". I used the qualitative analysis software NVivo 11 Pro to import, code and analyse the documentation and to record my insights in memos.

Through the coding process and comparison of patterns and similarities, allowing the findings to emerge from the data through inductive analysis (Patton, 2002, p. 56, p. 453), I identified core themes and found connections within categories. Categories for content analysis (NVivo's "nodes"), were based on the 18-page RDM provisions of the Sydney Conservatorium of Music (University of Sydney, 2015). But concepts and themes emerging from the evidence, such as case studies in the creative arts (NVivo's "units of observation"), informed the thematic analysis. These categories and themes are listed in **Appendix 2**, which reflects their prevalence and frequency. To ensure authenticity, I heeded Patton's (2002, p. 437) warning against resolving the gaps and imperfections which invariably occur: "Finding patterns is one result of analysis. Finding vagaries, uncertainties, and ambiguities is another". My focus was on providing evidence of trustworthiness and self-reflexivity so that my findings may be accepted as convincing and worthwhile (Leedy & Ormrod, 2015, p. 106).

Because document analysis, properly done, is "a systematic procedure for reviewing or evaluating documents" (Bowen, 2009, p. 27), my study could be conceived as a systematic review of the evidence towards best practice presented in the documents. A systematic review, as defined by Gough et al. (2012, p. 2), is "a review of research literature using systematic and explicit, accountable methods". While it is typically used to synthesise findings from research studies, it can also be used to synthesise other forms of data like the results of ethnographic studies and outcome evaluations (Gough et al., 2012, p. 54), for "the method itself is agnostic with regard to what is being synthesised". The integrity of a systematic review, according to O'Leary (2014, p. 258), depends on transparent and reproducible methods, an explicit search strategy for studies that meet clear selection criteria, critical assessment of the studies' findings, and rigorous analysis, synthesis and interpretation of the results. O'Leary (2014, p. 257) notes that systematic reviews, typically used in health research, are increasingly employed to inform policy and practice decisions in that they synthesise the results of studies to find some "truth" to aid decision making.

Best practice is likely to become apparent from a systematic review of information-rich evidence related to institutional humanities RDM within the past decade.

5. Results and Discussion

Borgman (2015, p. 38) has helpfully distinguished between "data policy writ large" (by governments), "data policy writ small" (by researchers), and "data policy in the middle range" (by universities). The documentation for my data was selected mainly from "data policy in the middle range", i.e. decisions made by research institutions about preserving and disseminating their research data. Because institutional decisions depend on larger policies about governance and funding formulated by governments and funding agencies, some documents were selected from "data policy writ large". And because institutional decisions also need to take account of the choices made by researchers about what they consider data, what they store, and what they share, some documents were selected from "data policy writ small". The documentation listed in **Appendix 1** comprises four institutional RDM policies, eight RDM project reports, four related documents, two data management plan (DMP) specifications, and five data curation profiles (DCPs), all with application to the arts and humanities. The selection and analysis of this documentation were guided by my three research sub-questions:

1. What constitutes humanities research data and why do they need curating?
2. What characterises research data use and reuse by humanities scholars?
3. What policies and initiatives are driving institutional humanities RDM?

5.1 Need for curation of humanities research data

Governments are promoting effective management of primary materials in research institutions, and requiring of researchers that their research data be made accessible with the necessary controls. Managing primary materials in the humanities requires sensitivity to common perceptions and data practices of researchers so that they may embrace the need for data curation from conceptualisation to disposal.

5.1.1 Domain-sensitive institutional policies

The UK and Australian governments, and their national funding agencies, now have requirements around RDM that research institutions must meet. Institutional RDM policies and procedures are viewed by Research Councils UK (RCUK, 2015) and the Australian National Data Service (ANDS, 2017) as fundamental to responsible RDM. The current ANDS Guide, *Institutional policies and procedures*, does not deal with domain-specific practices or guidelines, but points out the need, when considering the retention of research data, to adhere to legal provisions relating to "materials of heritage value, and disciplinary requirements" (ANDS, 2017, p. 3). However, an earlier document, *Outline of a research data management policy for Australian universities/institutions*, makes provision for "Discipline-specific practices or codes" to be addressed in an RDM policy (ANDS, 2010, p. 4). This document is likely to have informed the University of Sydney's suite of domain-specific RDM provisions for different faculties (University of Sydney, n.d.). The RDM policies and procedures of all Australian institutions are listed in the ANDS Project Registry (ANDS, n.d.), which was used to select for analysis the policies of the Sydney Conservatorium of Music (**University of Sydney, 2015**) and the **University of Adelaide (2015)**.

All Australian universities are bound by the Code for the Responsible Conduct of Research, Section 2 of which states that "Policies are required that address the ownership of research materials and data, their storage, their retention beyond the end of the project, and appropriate access to them by the research community" (Australian Government, 2007). The Code – which is under review for release in late 2017 (Australian Government, 2016) – recognises that "there is no simple, single way to define research for all disciplines" (Australian Government, 2007, p. 1). The Code requires any RDM policy to be consistent with disciplinary practices, including conventions for the retention of research data, and relevant legislation (Australian Government, 2007, Section 2.1). The University of Adelaide's RDM policy closely follows the Code in distinguishing between "research data" and "primary materials" (i.e. data may be derived from objects collected), but also recognises that, in some disciplines, primary materials may constitute research data. Research data may comprise "creative outputs, for example drafts of original literary and musical works,

and musical performances", or function as "evidence to support the formulation of the hypotheses and findings in areas such as text-based research" (University of Adelaide, 2015). Good RDM policies acknowledge the variety of research data and primary materials, even if few yet address discipline-specific requirements.

The RCUK's Common Principles on Data Policy, published in 2011 and fleshed out in *Guidance on best practice in the management of research data* (RCUK, 2015), aimed to regularise (1) open access to research data; (2) data management policies and plans; (3) metadata creation; (4) legal, ethical and commercial constraints on research data; (5) access terms and conditions; (6) privileged use of research data; and (7) the use of public funds for research within UK institutions. The principles have found their way into institutional RDM policies, and a spreadsheet of these policies is updated by the Digital Curation Centre (DCC) (Horton & DCC, 2016). The spreadsheet was used to select for analysis the RDM policies of two creative arts institutions, the University of the Arts, London (**UAL, 2014a, 2014b**), Europe's largest university for art, design, media and the performing arts, and the Glasgow School of Arts (**GSA, 2016**). The RDM policy of the GSA, whose disciplines include fine art, architectural studies, computer science and design, guarantees the management of research materials in any format or media, including "logbooks, computer modelling algorithms, journals, workbooks, sample libraries, sketchbooks, sets of images, video recordings, trials, prototypes, botanical dye recipes, found objects, software, and correspondence" (GSA, 2016, p. 2). The UAL's RDM policy also recognises that, although some analogue materials can be digitised, efforts to manage non-digital data are likely to be required, and it promises reasonable access to "analogue research data resulting from externally funded research" (UAL, 2014a, p. 3).

The University of Edinburgh was one of the first institutions worldwide to institute an RDM policy (in 2011), based on the conclusions of the Data Audit Implementation (DAF) project report (**Ekmekcioglu & Rice, 2009**), which informs my analysis. The DAF report recommended that developing an institutional RDM policy "would provide schools and principal investigators with a basis from which they could develop policies and practices suited to their particular contexts" (Ekmekcioglu & Rice, 2009, p. 16). According to the report (Ekmekcioglu & Rice, 2009, p. 16), a university wanting to safeguard its research data assets needs to

- develop greater data management awareness and guidance on best practice;

- implement an RDM plan and make a DMP template available to researchers;

- facilitate collaboration between the Library and Colleges/Schools to deliver training programmes and supporting materials for specific disciplines;

- encourage Schools to develop discipline-specific guidelines for RDM, procedures for file naming and version control, policies for data storage and backup, and to gather metadata when research is reported;

- provide more networked storage services and an off-line, backed up service.

Although such procedures require substantial technical and human resources, physical space, time and funds, and dealing with ownership and licensing issues (Borgman, 2015, p. 218), applying good practice to research data assets across disciplines, including the humanities, "protects the university's intellectual, financial, human and material investment in research" (Erway, 2013, p. 7). What constitutes humanities research data assets requires consideration if the data curation needs of researchers who generate and share primary materials are to be met.

5.1.2 Common types of humanities research data

The UK's Arts and Humanities Research Council (AHRC) accepts that a precise definition of the humanities is not possible, but, if research is focused on "artistic or creative practices, history, languages, literatures, or on the study of texts or images, then it falls within the domain of the arts and humanities" (AHRC, 2015b, p. 1). These parameters were used to select for analysis five DCPs from three different institutions reflecting a variety of data types and curation needs: Epigraphy (**Eaker, 2012**); Linguistics (**Tancheva, 2012**); Oral Tradition (**Donnelly, 2013**); History (**Forrest, 2014**); and Visual Culture (**Furness, 2014**). Because they provide practical examples, the DCPs will be also be used to illustrate my discussion.

Service requirements for managing humanities data generated by the University of Oxford were investigated in the trail-blazing SUDAMIH project sponsored by the JISC Managing Research Data (JISCMRD) programme from October 2009 to March

2011. The SUDAMIH final report (**Wilson, 2011**), which informs my analysis, confirmed that humanities researchers do not understand data curation terms and avoid using the word "data" unless referring to structured information. Claims by humanities scholars "that they don't actually work with data, an assertion based on the mistaken belief that data is exclusively the stuff of science" are indeed common (Pryor, 2012, p. 2). The SUDAMIH report advised those providing RDM support to emphasise that "data" applies to all the "materials used in the process of reaching conclusions about research questions" (Wilson, 2011, p. 26). Interviews with 40 humanities researchers enabled SUDAMIH investigators to take a broad view of what constitutes their data, including "books, articles, references, photographs, handwritten notes, and electronic files in any format, in addition to datasets as they might traditionally be understood" (Wilson, 2011, pp. 16-17).

SUDAMIH was launched after funding was controversially withdrawn in 2008 from the now defunct UK Arts and Humanities Data Service (AHDS), set up by the AHRC "to enable the discovery, creation and preservation of digital resources across the arts and humanities research, teaching and learning community" (Pryor, 2012b, p. 3). (For an account of its significance for humanities data curation, see Collins, 2012, pp. 165-166.) The AHDS produced a series of Guides to Good Practice available in the disciplines of Archaeology, History, Performing Arts, Literature and Linguistics, and Visual Arts. The Guides were intended to give practical instruction to the arts and humanities research communities "in applying recognised standards and good practice to the creation and use of digital resources" (AHDS, n.d.), and some are still available through the Internet Archive. Although they are a decade old now, the Guides on Visual Arts (**Grout, Purdy, & Rymer, n.d.**) and Electronic Texts (**Morrison, Popham, & Wikander, n.d.**) are used in my discussion to identify good practice data curation in the arts and humanities. While the Guide on Archaeology used the word "data", the other Guides expressly avoided it and instead used "digital resources", "digitised audio materials" or "electronic texts" (see AHDS, n.d.).

The Visual Arts Data Service (VADS), once part of the AHDS but now a research centre of the University for the Creative Arts, London (UCA), led the KAPTUR project from October 2011 to March 2013 to investigate research data in the creative arts (**Gramstadt, 2013**). UCA was one of four institutional partners, the others being the

GSA, the UAL, and Goldsmiths, University of London. Each undertook a case study which fed into the KAPTUR Environmental Assessment report (**Garrett, Gramstadt, Burgess, Murtagh, Nadim, & Spalding, 2012**) and also compiled its own report (Spalding, n.d.; **Burgess, 2013**; **Murtagh, Mahurter, & Hancox, 2013**; **Nadim, Cooke, & Gray, 2013**), which inform my analysis. Questions driving the KAPTUR project were: "What is 'research data' in relation to practice as an artist, performer, film-maker, or musician? What does it consist of? What does it exclude? How can it be identified within the research process? How do you capture it?" (Nadim et al., 2013, p. 2). The term "research data" did not resonate with the four types of contributors to the KAPTUR Environmental Assessment report: artist-researchers in physical media, artist-researchers in digital media, designer-researchers in interior design or architecture, and art historians (Garrett et al., 2012, p. 7). Instead, they experimented with phrases like "documenting the process" and "managing the material" (Gramstadt, 2013, p. 6). Alternative terms discussed were "materialising research; visualising research; making as research" (Garrett et al., 2012, p. 9), which elucidated the evolving nature of their data. The UAL report noted that "research data" is best understood in the context of "specific examples of how the researcher uses research data in their own field", and terminology used should "have a tangible and related application to a researcher's work" (Murtagh et al., 2013, p. 4, p. 5).

That the correct use of data curation terminology is important and should be drawn from research communities themselves is evident in two other reports which inform my analysis, Incremental (**Freiman, Ward, Jones, Molloy, & Snow, 2010**) and Curating Artistic Research Output (CAiRO) (**Gray, 2010**). Incremental surveyed researchers at the Universities of Glasgow (in 2009) and Cambridge (in 2010) to improve data curation practices and develop resources of value to other institutions (Freiman et al., 2010, p. 2). The Incremental Scoping report noted that researchers tend to be suspicious of 'policies' but "are receptive to 'procedures' or 'advice' which may be essentially the same thing, but convey a sense of purpose and assistance" (Freiman et al., 2010, p. 18). CAiRO was a project led by the University of Bristol from August 2010 to July 2011 under the JISCMRD programme to create RDM training materials for performing arts researchers who were "producing complex yet formally assessed research data" (Gray, 2010, p. 2). The survey informing the CAiRO User Needs report showed that 83.4% of practitioner-researchers equated

the concept of "data curation" with good research (Gray, 2010, Fig. 2.0), although unstructured answers queried the applicability of the term to an arts context. Researchers' views of the data they generate are informed by their data practices, which must be understood if appropriate support for humanities RDM is to be given.

5.1.3 General humanities research data practices

The findings of the RIN report on information practices in the humanities point to a variety of research data and practices, and variation in technology awareness and use (**Bulger et al., 2011**). The report therefore advised that "strategies for engaging humanities scholars with new types of information and new ways of working with it will require sensitivity to their existing cultures and practices" (Bulger et al., 2011, p. 74). Humanities researchers tend to work on their own but liaise informally through their networks, so that, "although the act of writing may be a solitary one for humanities scholars, their work environment has always been 'virtually communal'" (Bulger et al., 2011, p. 11). Unlike natural scientists, however, "they have no long tradition of working in formal, close-knit and collaborative research groups" (Bulger et al., 2011, p. 6). They have not had to contribute to research programmes, according to Whitley (2000, p. xxxii), but instead "make diffuse contributions to broad and fluid intellectual goals". Sole authorship of a paper is still the norm in many disciplines (Borgman, 2015, p. 255). The last fifteen years, however, have seen a move away from high task uncertainty and low mutual dependence towards more collaboration (Borgman, 2015, p. 58), a shift resulting "in part from new funding opportunities, but also from the possibilities opened up by new technologies" (Bulger et al., 2011, p. 71). A key challenge is to improve researchers' technical ability to link their research data across archives and repositories, since humanities researchers have fewer scruples about data sharing than natural scientists do (Bulger et al., 2011, p. 8).

The RIN report investigated the information practices of specific groups of historians, musicologists, English and Philosophy scholars, and digital humanists collaborating on a global project. All work with texts and images, customarily through annotation or manipulation, and relevant practices and needs have been tabulated below.

Table 1 *Information practices of different humanities communities*

Historians using Old Bailey online database	Musicologists using DIAMM database	English scholars at Birmingham University	Philosophers at University College, London	Humanists on Digital Republic of Letters
Annotate texts and take notes by copying and pasting passages into a Word file, and using the comment and highlight functionality	Make notes in a Word file on an image of musical notation, and either annotate images on their personal copies or by using DIAMM's online annotation tool	Take photos of newspapers or manuscripts, or access images through digital archives, then transcribe images into text files, and use simple apps to edit images	Make annotations on texts selected, or write notes in a Word file and use electronic filing systems to remind them of ideas, authors, or key articles	Do not annotate text or write marginal notes, but instead copy and paste text into a Word file containing quotes and their notes, and then insert into article drafts
Few scholars take notes by hand, but instead toggle between the online texts and a Word file, making digital notes as they go		Need for greater access to source material in digital image format, and for archival collections to be seamlessly connected	Require more annotation tools for PDF files and e-readers (and training in them), rather than using Word files to write their notes	Together edit letters, develop databases, and map epistolary relationships across space and time using visualisation tools
Maintain a simple database (Word files) on their computers, or lists of trials using Excel or Zotero, or keep a folder of each trial				Have developed an archive, database or analysis tool for their research, and share resources with collaborators

Source: Bulger et al., 2011.

More collaborative practices are evident among the historians contributing to *The Digital Republic of Letters*, an international project of humanities scholars, database programmers and data visualisation specialists developing an online "laboratory" of 17th century correspondence. New technologies and text-mining tools are enabling

digital humanists to undertake advanced kinds of analysis, such as visualisation (Bulger et al., 2011, p. 60). The DCP of an Epigraphy scholar at the University of Tennessee, for example, shows how his team considered critical for their project the visualisation of statue base inscriptions from the Late Antique Roman Forum, and this was facilitated by visualisation tools on the project website (Eaker, 2012, p. 5). Though practices vary across disciplines, humanities research is typically a process of iteration (reading texts, thinking, annotating, and re-reading) (Bulger et al., 2011, p. 44), and analysis is a cumulative process, occurring over years, and sometimes even decades, which has implications for data storage, access, and usability.

The DCPs of three researchers at the University of Edinburgh show that humanities data is commonly stored on researchers' own computers or university servers, where it usually remains after a paper is published. Few scholars are aware of repositories such as Edinburgh DataShare, the pioneering institutional data repository started in 2009 (Rice, 2013/14, p. 39). The DCP of a Visual Culture researcher, who was finding out how visitors to the Victoria & Albert Museum during the 2008 financial crisis were reading the Cold War Modern exhibition, indicates that she wanted her interview data stored "in perpetuity" (Furness, 2014, p. 3) so that it could be used "for further historical analysis" (Furness, 2014, p. 5), but she did not specify placing it in Edinburgh DataShare. The audio files were Web-hosted by the company which transcribed them, and the transcriptions were held on the researcher's USB memory sticks, work computer, laptop and emailed to herself (Furness, 2014, p. 3). The DCP of a postgraduate student in History, who was analysing how the US media helped define the public image of the Vietnam War, indicates that he stored his files on the university's home drive and USB sticks, and evinced no knowledge of centralised storage facilities or Edinburgh DataShare (Forrest, 2014, p. 3). The DCP of an Oral Tradition researcher studying the interplay between Dinka musical forms and language shows that, although the Dinka song collection was stored and backed up in a directory on the Linguistics & English Language server (Donnelly, 2013, pp. 2-3), it was archived in both Edinburgh DataShare and the Max Planck Institute. Interest in the endangered language dataset was anticipated to "increase with the passage of time and therefore the longer it is available, the better" (Donnelly, 2013, p. 3).

Similar localised storage practices are evident from the data audits of the School of Divinity and the Department of Economic and Social History at the University of Edinburgh (Ekmekcioglu & Rice, 2009, pp. 19-22). Interviews were held with key researchers on their data types and practices including back-up and data retention. Data types in the case of Divinity ranged from "text corp[ora], documents, reports and transcripts to images, scans, video tapes and photo collections" (Ekmekcioglu & Rice, 2009, p. 20). Data types in Economic and Social History, comprised "archives in local and national collections, documents or reports, MS Word files, MS Excel sheets and SPSS files . . . slides, photo collections, fieldwork data, and websites" (Ekmekcioglu & Rice, 2009, p. 22). The data were stored on the respective College servers but were not structured or easily discoverable. Unique data from various Divinity projects were also Web-hosted for download (Ekmekcioglu & Rice, 2009, p. 20), despite this not being a sustainable archiving solution. Irreplaceable datasets were stored by Economic and Social History researchers on personal computers, laptops or other devices, and "not managed effectively or made readily-accessible to other researchers" (Ekmekcioglu & Rice, 2009, p. 22). Metadata was hardly present and, when it was, the fields were minimally populated or "generated automatically by the data's host proprietary application (and were hence incomplete)" (Ekmekcioglu & Rice, 2009, p. 15). Both case studies demonstrate the haphazard and risky way in which the variety of research data is stored during and after small research projects.

5.2 Research data use and reuse by humanities scholars

Primary materials in the humanities are complex to manage, and the attitudes of creative arts researchers, in particular, to data curation during and after the research process are diverse. Researchers are generally open to sharing their data, which require careful contextualisation and rights clearance before publication, especially since the value of primary materials tends only to appreciate over time.

5.2.1 Attitudes to research data sharing

Although humanities researchers are less likely on the whole to resist data sharing than their natural science counterparts, attitudes nonetheless vary across disciplines. The Incremental Scoping report found that, while researchers tend to be in favour of

sharing data in principle, "they are almost universally reluctant in practice" (Freiman et al., 2010, p. 4). Gaining a scholarly reputation in most disciplines depends on the publication of research results before anyone else. Data sharing is, however, more common on AHRC-funded projects, and among certain disciplines. Researchers in English Language at Glasgow and Cambridge supported data sharing, "noting wider benefits that ensue and hoping the tendency to ring-fence data for personal use could be overcome" (Freiman et al., 2010, p. 16). English Language researchers at Glasgow pointed out, however, that raw data were useless without their context and structure, including their data mining and analytical tools, "as digital resources take on great value as an integrated whole" (Freiman et al., 2010, p. 15), and the report highlighted the importance of contextualisation for data reuse. Contextualisation is a legitimate concern in high task uncertainty and low mutual dependence fields like the humanities, which have few standardised procedures and interpretative norms to enable results "to be compared and assessed in routine ways across research sites or groups" (Whitley, 1984, p. 131). Interpretation characterises the humanities, in that "each object, and even each representation of an object, can be reinterpreted multiple times" (Borgman, 2015, p. 200), and therefore, if data are to be shared, the circumstances which gave rise to the particular interpretation must be recorded.

Researchers who are hesitant to share data argue that they need to be "carefully prepared, annotated, and contextualised" before they can be made available, a slow process which grant funding rarely takes into account (Freiman et al., 2010, p. 4). The DCP of a Linguistics researcher from Cornell University indicates the time-consuming process required to make a corpus of Cheyenne recordings discoverable and usable: transcribe the audio files, clean up and normalise the data, and apply metadata "so that the textual files can be searched for specific linguistic features and then call up the audio files" (Tancheva, 2012, p. 1). Linguistics metadata is usually specific to each project, although metadata crosswalks for interlinear glosses could be considered if data were transferred to a language repository (Tancheva, 2012, p. 3). The DCP of the Epigraphy scholar also shows the careful process applied to the corpus of statue base inscriptions from the Late Antique Roman Forum: collect the text and photographs of the inscriptions from published sources (or take a photo in the field if necessary); assign GPS coordinates; and encode the inscriptions in a form of XML called Epidoc (Eaker, 2012, pp. 2-3). The scholar saw the benefit in

preserving all the data (i.e. the XML files, the images, and the GPS coordinates): "they are all interrelated and interlinked; therefore, they should be preserved as a group" (Eaker, 2012, p. 2). The images and transcriptions were published on the project website and, placing a high priority on access to the corpus by researchers and the general public alike, he set no conditions on their reuse (Eaker, 2012, p. 5).

Humanities data is reused by multiple audiences for different purposes. Loosely coordinated fields like the humanities ("fragmented adhocracies") have historically been more open than the laboratory sciences to lay audiences and influences, and enjoy "diverse and varied legitimate audiences for their work" (Whitley, 1984, p. 84). The lay public is particularly important to traditional humanities fields (Whitley, 1984, p. 277). The Epigraphy scholar's inscriptions were not subject to copyright (Eaker, 2012, p. 4), and, although the full XML files were not accessible on the website, he was willing to share this data on request (Eaker, 2012, p. 5). Multiple audiences were also envisaged by the History postgraduate, who felt that his data on the US media's depiction of the Vietnam War had "a wide ranging potential audience and capacity for further analysis in fields such as American and Vietnamese history, sociology, politics and the study of modern conflict" (Forrest, 2014, p. 3). The data on Dinka song similarly had capacity for reuse by audiences "in fields such as language, anthropology, sociology and ethno-musicology" (Donnelly, 2013, p. 3). The Oral Tradition researcher admitted that "minority languages are 'a buyer's market'", and he was happy to facilitate access to those researchers interested in making use of the collection (Donnelly, 2013, p. 4). The Cheyenne audio files of the Linguistics researcher were not under copyright restrictions and, once ingested into an open searchable database, would be freely accessible online for any researcher or language learner who registered for a password so that users might be counted (Tancheva, 2012, p. 3). Usage statistics were not important for other researchers, however: the Epigraphy scholar would rather know *how* the data is being reused than about how often it is being accessed or by whom (Eaker, 2012, p. 6).

A significant barrier to data sharing in the humanities is copyright of digital materials. Rights management issues are integral to creating and reusing digital resources because any translation into a digital format of an object that already exists, whether a photograph, text, or artwork, effectively makes a copy of it and thus may breach

copyright (Grout et al., n.d.). The Incremental Scoping report found that English scholars at both the Universities of Glasgow and Cambridge encountered copyright problems when digitising materials from other institutions (Freiman et al., 2010, p. 15). The DCP of the postgraduate student in History shows that, although he was fully in favour of sharing his research data, access to it was likely to be restricted unless permission were obtained from copyright holders of material on the Vietnam War in newspaper and government archives (Forrest, 2014, p. 3). His data comprised .pdf files from databases (e.g. newspaper archives) and .jpeg images of print-only documents (e.g. US Government documents), as well as his notes on the primary sources in .doc format (Forrest, 2014, p. 1). The DCP of the Oral Tradition researcher indicates that the project team did not want to restrict access to the Dinka song dataset (185 audio files in .wav format) but publish it as widely as possible (Donnelly, 2013, p. 3). The recordings were archived and shared on a 'not for profit' basis through Edinburgh DataShare, and only in one case, where the song was not the singer's own, was permission to share withheld (Donnelly, 2013, p. 2).

The reuse of creative arts data is even more complicated, because of rights to their work held by practitioner-researchers themselves. Performing arts data is reused most often "as the basis for new works", according to the CAiRO User Needs report, but also "as exemplars for teaching, as the basis for assignment, for illustrative purposes and for reasons of personal interest" (Gray, 2010, p. 1, p. 18). The report advised retaining as much as possible of an art work created as research, especially "developmental materials, documentation of time-based events and any completed versions" (Gray, 2010, p. 3). Such data might be made accessible for scholarly use, but, because of copyright, practitioner-researchers did not want the work itself to be reused through re-staging or mash-ups, even if this was attributed (Gray, 2010, p. 3). They also showed resistance to the idea that long-term preservation technologies should be used throughout the research, perceiving them to impede the process of creativity (Gray, 2010, Fig 4.3). Although 57.1% agreed with the use of technologies (and they were probably assenting to open formats), 42.9% were against them:

One respondent stated "The creative process must not be compromised at all, in any way, by the requirement for data collection, storage or archiving. It is for

the means of archiving etc. to become flexible enough to match creative work".
(Gray, 2010, p. 15)

While some visual arts researchers similarly balk at documenting the creative process to make the research "intelligible for 'outsiders'" (Garrett et al., 2012, p. 23), others undertake the storing and processing of data routinely as part of the research. However, the KAPTUR Environmental Assessment report noted that "their methods could be described as situated rather than normative" (Garrett et al., 2012, p. 4), in that they were context-dependent and "idiosyncratic not only to a discipline but to a specific group of people within a discipline" (Garrett et al., 2012, p. 23).

5.2.2 Curation needs of humanities research data

The complex curation needs of visual and performance arts research data is related to the idiosyncratic nature of artistic research. The domain is characterised by low strategic interdependence and high task uncertainty about "the nature, meaning, and significance of research results" (Whitley, 2000, p. xxxii). Artistic research depends on the *practice* of the arts, and a critical issue is discerning whether an artwork was made as research or not. (For a discussion of this in the context of the production of knowledge in artistic research, see Borgdorff, 2010.) The GSA case study report noted that arts-based or practice-based research data is "complex and complicated and it does not always fit into the natural scheme of data management", requiring policymakers for creative arts RDM to think "outside the box" to craft policies that are "more iterative and open to interpretation" (Burgess, 2013, p. 11) – appropriately reflecting the interpretive practice so characteristic of humanities data scholarship. The GSA report determined that most of the data which their RDM policy would have to deal with would be visual, and speculated it would be more complex to record than forms of visual data in science-based disciplines (Burgess, 2013, p. 2).

Artistic research arguably occupies a distinctive place within academic research because it attempts to convey "content that is enclosed in aesthetic experiences, enacted in creative practices and embodied in artistic products" (Borgdorff, 2010, p. 45). An art historian in the KAPTUR Environmental Assessment report put it thus:

I don't think we have this kind of method-process-output, we have a slippery and sliding scale of things and outputs appear as drawings, photographs, publications, exhibitions, but they are all in a way outputs and they are also all research because they contribute to a process of thinking, exploring, testing, investigating and sharing. (Garrett et al., 2012, p. 16)

The report referenced an essay on research and the self (Griffiths, 2010), noting that a common theme in literature about the creative process is the researcher's notion of self as always changing, which renders visual arts data essentially "unfinished" (Garrett et al., 2012, p. 9). Research expressing the self (or reflective practice) is an ongoing process, for the self is "always in a state of *becoming*, always unfinished" (Griffiths, 2010, p. 168, p. 177). If arts-based research is an infinite process, since "there is no obvious endpoint to the research as both [the artist] and his context change" (Griffiths, 2010, p. 177), the practitioner-researcher is likely to resist RDM when it is construed as the curation of finite, evidential processes. Those providing RDM support "must recognise the concerns among some artists that shoehorning artistic production into the principles of scientific research will threaten processes of inspiration, creativity and idiosyncrasy" (Guy, Donnelly, & Molloy, 2013, p. 102). One outcome of the KAPTUR project was sober recognition of the "enormous" effort required "to engage visual arts researchers with both appreciating the nature of research data and how to manage it appropriately" (Gramstadt, 2013, p. 13).

The nature of performing arts data also presents difficulties. The CAiRO User Needs report surmised that, despite obvious differences in the data management needs of disciplines as various as choreography, narrative cinema and musical composition, "some commonality exists" (Gray, 2010, p. 4). An example is the prevalence of analogue research data and the considerable disinclination to manage it. The report found that, although 71.4% of practitioner-researchers were in favour of working with a specialist to create a plan to manage their creative outputs (Gray, 2010, Fig 3.2), 28.6% resisted the idea, and the report considered the implications for RDM training:

One respondent stated "I suspect [RDM is] not much of a priority for most, especially those working in analogue/non-digital practices" . . . there are few other academic disciplines which would consider an 'analogue' methodology to

be acceptable when producing research output. Even if an activity were essentially non-digital, we might now expect documentation around that activity (for reasons of dissemination, preservation and interoperability) to be in a digital form. If representative, this impacts on the precepts of CAiRO teaching and learning materials and suggests a brief introductory step may be required to present the benefits of digital over analogue data. (Gray, 2010, pp. 9-10)

The RDM provisions of the Sydney Conservatorium of Music, however, recognise that "it is not always possible or appropriate to digitise creative works and collections of physical research data and materials" (University of Sydney, 2015, p. 4), and the RDM policy of the UAL likewise acknowledges that it may not be necessary to create digital versions of non-digital data (UAL, 2014a, p. 4), so the issue is not clear cut.

Humanities scholars also use both analogue and digital resources for their research, which include traditional print materials and manuscripts, and digital materials such as images, audio and 3-D scans of artefacts (Bulger et al., 2011, p. 10). The RIN report found that research data in the humanities are overwhelmingly drawn from text and image resources held in archives and libraries (Bulger et al., 2011, p. 6), and structured again in a researcher's own database. Such resources, according to the VADS Guide on Visual Arts, commonly comprise digital images "stored in some sort of database or other structure and accompanied by textual information to identify and contextualise them" (Grout et al., n.d.). The visual arts have the edge over other humanities disciplines here, for they can draw on the Getty Research Institute's Vocabulary Program to enhance the findability and usability of digitised works of art. The importance of *collections* in humanities research data is emphasised in Collins' (2012, pp. 165-166) discussion of the data curated by the now-defunct AHDS, and it might be said that today's research data is tomorrow's special collections:

> the defining originality and value of research datasets in the humanities tend to lie in the way they have been curated as a collection, rather than in the fact that they are newly generated by researchers. In support of this interpretation, many of the AHDS holdings were digitized versions of existing physical objects, manuscripts or artworks for example, or collections of material such as linguistics corpora that had been produced as the by-products of a research

process. In such a context it is significant that the AHDS referred to its "resources" rather than "datasets", unlike most other data centres.

The diversity of humanities data resources – and the expertise required to interpret them – is recognised by Borgman (2015, p. 166) as characteristic of humanities data scholarship. The wide distribution of resources and multiplicity of access protocols across repositories is both an asset and a liability (Borgman, 2015, p. 170).

Similarly double-edged is the potential of humanities research data to retain their scholarly worth over time. The long-term intellectual value of humanities datasets is emphasised many times in the SUDAMIH final report, and this has implications for preservation and access controls. A database of Roman cities could be of as much use to researchers in 2060 as it is today, provided it is still accessible (Wilson, 2011, p. 16). Indeed, the Epigraphy scholar studying inscriptions on statue bases from the Late Antique Roman forum felt that the data should be preserved indefinitely since their content is already two millennia old: "Since people have been preserving it this long, we might as well try to keep it just as long" (Eaker, 2012, p. 6). Moreover, humanities scholarship "often aggregates to a 'life's work' body of research" (Wilson, 2011, p. 16), and it is common for researchers to return to their old notes, sources, collections or corpora to answer new research questions. This has implications for data sharing which those providing RDM support should recognise, for, "when research depends on evidence accumulated over long periods of time, incentives to release [data] are few" (Borgman, 2015, p. 237). The 'life's work' nature of so much humanities research requires "data to be managed and kept accessible over very long time periods" (Wilson, 2011, p. 29). However, "the long life-span of humanities data" poses real problems, since "the integrity of academic research depends upon the continued availability of sources remaining as those sources were when cited" (Wilson, 2011, p. 31). Ensuring data authenticity, reliability and integrity, as well as usability, throughout the data lifecycle is therefore necessary for humanities RDM.

5.3 Drivers of institutional humanities data management

Institutions intending to give humanities data the management it requires to preserve the scholarly record need to address intellectual property concerns and long-term

storage requirements, consider the benefits of a centralised data infrastructure as well as early intervention and partnership in the research process, and recognise the importance of the data management plan (DMP) for sustainable data curation.

5.3.1 Ownership rights and copyright

Ownership rights require particular attention in humanities data curation. The CAiRO project accepted that their postgraduate RDM training materials "would benefit from guidance around intellectual property rights and especially copyright" (Gray, 2010, p. 18). Though usually willing to share research data (most often curated collections), humanities researchers "need to be reassured that they will not 'lose control' of their data in the process of making it reusable" (Wilson, 2011, p. 32), and so "any tools or repository intended to help manage research data will require high levels of security" (Wilson, 2011, p. 31). Most institutional and many data repositories use open access licences, and ensure that copyright vests with the principal investigator (PI), even as intellectual property rights are held by the researcher's institution. The SCM provisions make it clear that depositing datasets in the institutional repository, Sydney eScholarship, to support publications (and providing a persistent link to the publisher) allows the PI and the University of Sydney to retain their respective rights to the data (University of Sydney, 2015, p. 5). According to Erway (2013, p. 13), many universities claim ownership of research data generated by their employees, as do some funding agencies, but among researchers there is widespread confusion or even disputation on this issue. The interviewees in the KAPTUR Environmental Assessment report expressed various opinions about data ownership, from wanting to share publicly funded research to requiring negotiation of their ownership rights (Garrett et al., 2012, p. 22). Because research data are often not released unless it is clear who has the rights or responsibility to do it (Borgman, 2015, p. 218), policies on data ownership rights need to be specified and understood by all stakeholders.

There is no issue with material in the public domain or if the resource creator is the rights holder, but it is necessary to clear reproduction rights where copyright is held by a third party. Third-party materials which have been re-worked as part of the artistic research process comprise "a huge, huge messy area", according to one artist-researcher in the KAPTUR Environmental Assessment report (Garrett et al.,

2012, p. 21). Copyright issues are especially relevant to practitioner-researchers in the arts; as one respondent in the CAiRO User Needs report warned, "there may be a royalty/fee component here, if the artist's livelihood depends on them" (Gray, 2010, p. 18). Many arts researchers, however, feel that sharing publicly funded research is an ethical imperative; one cultural historian in receipt of public money indicated in the KAPTUR Environmental Assessment report that by "using things like creative commons licences I will try and licence myself out of copyright" (Garrett et al., 2012, p. 25). Permission to reuse materials owned by a third party is often withheld for commercial reasons, and this can impact on access to the research data.

Planning for this eventuality is necessary from the start of any project. The VADS Guide recommends that "early attention to access issues, such as copyright, user needs and dissemination technology, will pay dividends for projects in the long term" (Grout et al., n.d.). Similarly, the Oxford Text Initiative (OTA) Guide on Electronic Texts advises sorting out the rights at the outset of any digital creation project by obtaining "appropriate permissions from all those who hold any form of rights in the material with which you are hoping to work" (Morrison et al., n.d.). The SCM provisions recognise that music researchers are likely to encounter "multiple and competing ownership and intellectual property rights that may affect their ability to distribute, share and re-use research data" (University of Sydney, 2015, p. 3). The provisions require that PIs or research supervisors plan to make research data, including creative works, available for reuse unless this is prevented by legislation, policy, ethical or contractual obligations (University of Sydney, 2015, p. 4). Data may be shared at any stage of the lifecycle, including uploading performances to the SCM YouTube site if they are free of copyright restrictions and performers have given consent (University of Sydney, 2015, p. 5). Managing humanities research data thus requires early consideration of legal compliance, copyright and licensing issues.

5.3.2 Long-term data sustainability

Humanities data need careful selection, especially when long-term sustainable preservation is required. The GSA case study report warned that the value of research data rests on the views of individual researchers, so "what information is recorded and subsequently kept is the researcher's decision" (Burgess, 2013, p. 12).

The Incremental Scoping report found that "few researchers actively select data for preservation – most hope to keep everything indefinitely", and noted that perceptions that storage is cheap, and concerns about unpredictable requirements for future data reuse, prevent appraisal and selection (Freiman et al., 2010, p. 14). Practitioner-researchers in the CAiRO User Needs report gave unanimous approval to the idea of retaining the whole of a work which has been completed as research, including interactive functionality where applicable (Gray, 2010, Fig. 4.0), although this might not be possible with highly complex works. They also showed unanimous support for keeping detailed documentation of live or performance events conducted as research (Gray, 2010, Fig 4.1), and 83.3% believed step-by-step documentation of a work completed as research should be retained (Gray, 2010, Fig 4.2). Although researchers may be keen to have long-term storage of all their research data, it cannot all be preserved because of the resources required, so a university should plan to invest only in data worth keeping and clarify its policy (Erway, 2013, p. 13).

The aim is a data storage situation regarded as normative, not situated, a distinction drawn by the KAPTUR Environmental Assessment report (Garrett et al., 2012, p. 4). Normative archiving entails "the expected processes and procedures followed by a trained archivist, including selection or disposal of items, cataloguing, and archival quality storage for the longer term", whereas situated archiving depends on the research context (Garrett et al., 2012, p. 27). Following the UK Research Councils' guidance to retain data that underpins findings in publications for at least ten years after publication (RCUK, 2015, p. 4), the RDM policy of the GSA requires that data be retained securely for ten years or destroyed (GSA, 2016, p. 3). For this purpose, the EPrints platform of the GSA's institutional repository, RADAR, was explored "for storing data related to the final research output such as initial thoughts, sketches and numerical data" (Burgess, 2013, p. 8). Information about research data deposited outside of RADAR or the GSA Archives needs to be registered with the GSA to maintain control of data assets (GSA, 2016, p. 3). Rather than use its institutional repository, Goldsmiths Research Online, for data storage, Goldsmiths piloted an open access data archive, Goldsmiths Data Online (Nadim et al., 2013, p. 3).

According to Schedule 3 in the SCM provisions (University of Sydney, 2015, p. 14), the minimum data retention period varies from five years in the case of creative

works to twenty years for research with children. On expiry of this period, research materials are appraised for permanent retention or secure destruction (University of Sydney, 2015, p. 14). There are stringent storage requirements upon completion of a project or leaving the university. Research students must supply "a copy of the research data, research records and any primary materials (retained physical materials)", and create a record of the dataset within the Research Data Registry (University of Sydney, 2015, p. 6). Researchers must store "a copy of the research data on University managed infrastructure", create a record of the dataset "and any associated primary materials" within the Registry (University of Sydney, 2015, p. 6). The University of Adelaide's RDM policy requires "non-digital research data and primary materials" to be stored "in a secure, accessible and identified location", and, along with the digital data, be recorded in the University's metadata store (University of Adelaide, 2015). All research data have to be archived for at least five years after publication, but, "if the work has community or heritage value, research data should be kept permanently, preferably within a national collection" (University of Adelaide, 2015). This shows the need in the humanities, emphasised by the SUDAMIH final report, "for very-long-term data sustainability solutions and cost models designed to deal with effectively permanent storage and access" (Wilson, 2011, p. 32).

5.3.3 Database as a Service (DaaS)

Research data require stable and secure storage from conceptualisation, and the best solution seems to be a federated one. This comprises storage at the discipline or School level, and additional storage and services at the institutional level (Rice & Haywood, 2011, p. 239). SUDAMIH piloted the provision of "Database as a Service" (DaaS) – now known as the Online Research Database Service (ORDS) – to enable the centralised creation, storage and sharing of humanities datasets (Macdonald & Martinez-Uribe, 2010, p. 9). The infrastructure provided by DaaS made it possible for researchers "to quickly and intuitively create, edit, search, and potentially open up relational databases of research data" (Wilson, 2011, p. 3). The SUDAMIH final report suggested that DaaS would be able to serve active in-development databases as well as dormant ones (Wilson, 2011, p. 32). Further development of DaaS would have to meet key user requirements of humanities researchers, among which are an intuitive interface, flexible searching and querying, ability to handle a range of data

types, linkage of records to external sources, support for diacritics and text in non-Roman alphabets, multiple editing via a Web interface, functionality to download and process data on desktops, preservation of data integrity, a stable and secure system, and good user support and training (Wilson, 2011, pp. 18-19). The SUDAMIH project revealed a growing interest in technology-driven research among humanities researchers, as did the RIN report (Bulger et al., 2011, p. 74), but little knowledge about what technologies are available to address specific needs. Assistance might be provided through DaaS, which is better suited to collaborative projects and data dissemination (Wilson, 2011, p. 29), and can reduce the cost to researchers of maintaining digital collections at the end of a research project (Wilson, 2011, p. 24).

However, many humanities researchers are not aware that the institution or other users would be interested in their research data and they do not make full use of what RDM infrastructure, services, or guidance are available (Garrett et al., 2012, p. 31; Gray, 2010, Fig. 3.1). Moreover, they are likely to be reluctant to invest their energy in managing their data using a centralised service, because "anything other than producing research output is unproductive as this is how they are rated" (Ekmekcioglu & Rice, 2009, p. 15). Discussion at the University of Edinburgh about an RDM policy becoming mandatory, for example, was likely to "be resisted by hard-pressed staff who might see it as just another bureaucratic noose around their neck" (Rice & Haywood, 2011, p. 240). Moreover, researchers' willingness to use DaaS depends on institutions clarifying the rights researchers have when depositing their data in a repository or centralised service (Wilson, 2011, p. 31). Because they need to feel 'in control' of their data, when they do relinquish it, they prefer it to be stored by the department or research unit, which also enables active processing and protects against connections going down (Rice & Haywood, 2011, p. 239).

Implementation of a federated model, however, would help to prevent the typical "moderate to catastrophic data loss" experienced by researchers when they store data on a variety of media, including the cloud, without any departmental guidelines or back-up procedures (Freiman et al., 2010, p. 4). Such "ad hoc data storage solutions" and poor file-management conventions are all too common (Freiman et al., 2010, p. 18). Because "a coordinated cyberinfrastructure environment can offer advantages such as economies of scale, integration, and a focused approach to

coordinating technology and expertise" (Erway, 2013, p. 9), existing infrastructure, such as the institutional or data repository, could be developed to support data management and preservation. If a dataset is stored off-site, however, the ingest requirements and retention policy of the repository should be assessed, and a metadata record and link to the dataset included in the institutional or data repository to ensure a full record of the university's research data assets (Erway, 2013, p. 18). Storage of humanities research data, whether onsite in a repository or elsewhere, requires proper naming conventions and documentation, consideration of risks and backup needs, and potential for migration of the standards and formats used.

5.3.4 Intervention upstream in the workflow

Early intervention in the research process is now understood as critical for effective RDM. Academic libraries have traditionally worked "downstream" in the research process, but now work "upstream" through "supporting the use of documentation, best practices, or standards in the production of data as collaborative partners with faculty" (Carlson, 2012, p. 11). Good practice RDM requires that research data be actively curated and migrated throughout their lifecycle, not just stored or backed up, and an academic library's expertise makes it "the most appropriate place on campus for safe, sustained, and trusted stewardship of research data" (Erway, 2013, p. 11). Engaging researchers upstream in the data lifecycle should ensure that important research data "makes it down to the point at which institutional or other long-term curation services can bring their expertise to bear and facilitate the preservation and wider reuse of that data" (Wilson, 2011, p. 30). Upstream engagement may require developing an RDM "toolkit", which could include guidance, a DMP checklist, an audit methodology and tool, and training for researchers (Guy et al., 2013, p. 108). The SUDAMIH final report revealed that humanities researchers want customised training in RDM, and better support for database creation, management and use (Wilson, 2011, p. 29). Although researchers in all disciplines need help with file organisation and technical awareness, if disciplinary needs are not met through targeted training, "issues of data quality are likely to remain" (Wilson, 2011, p. 30). The SUDAMIH report concluded that "subject-based customization and institutional localization is required to maximize training benefits" (Wilson, 2011, p. 30).

The RIN report noted that "humanities scholars are often enthusiastic adopters of appropriate technologies" to enhance research, and welcome the chance to see how these "have been used by their peers to pursue real, current research" (Bulger et al., 2011, pp. 74-75). Training courses for different faculties at the University of Oxford are run separately by IT services (University of Oxford, n.d.b), and those for the Humanities Division include the two-hour course, "Preparing Humanities Research Material for the Future: What You Can Do Now to Avoid Problems Later", and a half-day introduction to "Databases: An Introduction to Oxford's ORDS". Since 2008, Digital Humanities at Oxford has offered week-long summer school workshops, beginning with training in the TEI (Text Encoding Initiative) markup language. TEI is the non-proprietary de facto standard in the digital humanities (Morrison et al., n.d.). Among the workshops offered in July each year of interest to literature researchers, musicologists, linguists and historians is "Humanities Data: A Hands-On Approach" (University of Oxford, n.d.a). It introduces useful tools, methods and concepts for managing digital data, and the datasets of real-world digital humanities projects serve as case studies for use with GitHub, OpenRefine, SQLite, and YesWorkflow. The topics include information organisation, data modelling and workflows.

The Edinburgh DAF report endorsed disciplinary input into "training programmes and supporting materials, targeted at researchers and postgraduate students in specific disciplines" (Ekmekcioglu & Rice, 2009, p. 16). Other reports made the point that training young researchers to improve their RDM skills would have significant benefits down the line. The GSA case study report noted how the GSA Project Officer taught early career researchers and postgraduate students the terminology associated with research data, and determined with them how best to handle specific examples of arts research data, such as art works, paintings, artefacts, written text and numerical information (Burgess, 2013, p. 5). The Incremental Scoping report showed that researchers believed that the best time to intervene with RDM training, guidance or tools was early in their careers, "as this is one point where habits begin forming and young researchers do or do not learn standard practices from their more senior colleagues" (Freiman et al., 2010, p. 18). The resources developed to meet researchers' needs had to employ discipline-specific examples, avoid data curation jargon, and be able to be accessed "at the point of need" (Freiman et al., 2010, p. 2). Choosing to use departmental champions and offer discipline-specific training, and

connecting researchers with support staff for individual advice and partnering were some steps by which the Incremental project proceeded (Freiman et al., 2010, p. 6).

5.3.5 Embedded or partnership model

Early and active partnerships between support staff and researchers are very important to create the trust researchers need for relinquishing control of their data. Partnerships will help allay common fears around privacy, loss of data ownership, or data misuse, and reveal the benefits that accompany devolution of data preservation from the researcher to the institution (Macdonald & Martinez-Uribe, 2010, p. 12). However, arts researchers are likely to reject an embedded approach to RDM. The CAiRO User Needs report found that two-thirds of practitioner-researchers preferred that tools and formats for sustainable archiving of a research work be employed only upon *completion* of the work, even if developmental work created in non-sustainable formats cannot then be retained (Gray, 2010, Fig 4.4). The report commented that "an approach which embedded data management within the creative process would be appropriate for the minority of researchers while others would benefit most from retrospective data 'standardisation'" (Gray, 2010, p. 16), and recommended that multiple options for data curation be offered. Retrospective standardisation, however, is problematic, for standardisation is normally done as part of the data creation process to aid digital preservation; the other preservation strategies of refreshing, migration and emulation are applied to data *after* it has been created.

A possible solution is to target "organisational moments", viewed by the KAPTUR Environmental Assessment report as points in an artistic research continuum at which research data may be *actualised*, such as writing or studio experiments (Garrett et al., 2012, p. 9). Any activities in which the intangible becomes tangible, like writing a grant application or a conference paper, preparing for assessment procedures, institutional duties, learning and teaching events, exhibitions, or even filing information (Garrett et al., 2012, p. 11), might become triggers for partnership with researchers. Supporting the researcher to make some of these "organisational moments" less complicated and stressful contributes to more effective RDM (Garrett et al., 2012, p. 11). The Goldsmiths case study report suggested articulating RDM activities with improvements to existing research support, such as better networked

storage, a DMPonline template customised from the open-source DCC tool, and one-to-one advice (Nadim et al., 2013, p. 4). RDM activities at Goldsmiths are shared between the subject librarians, who underwent RDM workshop training, and the Goldsmiths Research Online repository team, for "the knowledge and experience of subject librarians can make the greatest difference" (Nadim et al., 2013, p. 3). The report noted that researchers appreciate very specific advice, such as "how and where to best store data; what devices to use to record and at what settings; and how best to preserve and describe specific data" (Nadim et al., 2013, p. 4).

Often these recommendations are specified in detail in the institutional RDM policy. Schedule 1 in the SCM provisions indicates the kinds of research data that must be retained (University of Sydney, 2015, pp. 9-11). The records vary across the five sub-disciplines but include physical data (e.g. notebooks and prototypes), digital data (e.g. spreadsheets and recordings), research records (e.g. funding applications and consent forms), and some research outputs (e.g. websites and instrument designs). The preferred file formats for digital data, such as image, audio, video, and text files, as well as qualitative data (e.g. XML and .rtf files), and quantitative data (e.g. CSV or .tab files) are tabulated in Schedule 4 (University of Sydney, 2015, pp. 15-18). The provisions also give information about specialised support available from the Library. The research data manager gives advice on digital file formats (University of Sydney, 2015, p. 4), sharing datasets at any stage of the research (p. 5), and documenting and depositing datasets in the right repository (p. 13). The institutional repository coordinator provides guidance on depositing datasets in Sydney eScholarship (University of Sydney, 2015, p. 5). The manager of copyright services gives advice on copyright and intellectual property in research data, and selecting proper licenses for datasets (University of Sydney, 2015, p. 13). Liaison librarians assist PIs and research students to draw up a DMP (University of Sydney, 2015, p. 2), create a dataset record for researchers (p. 6), and provide RDM training and support (p. 12).

5.3.6 Research data management plans

An appropriate time to initiate partnerships between library staff and researchers is when drafting the DMP. The DMP is a recent requirement, for two of the DCPs do not include preparing a plan (Tancheva, 2012; Forrest, 2014) although the History

postgraduate wished he had done so (Forrest, 2014, p. 1), and the US National Endowment for the Humanities (NEH), despite funding the Epigraphy project, had not yet made it a condition (Eaker, 2012, p. 2). The NEH Office of Digital Humanities has now made the DMP mandatory for its grantees, and updates its specifications annually (**NEH, 2017)**. The DMP is domain-specific in that it "should reflect best practices in the applicant's area of research, and it should be appropriate to the data that the project will generate" (NEH, 2017, p. 1). Aspects to be considered are the rights and obligations of all parties; the types of data (or collections) to be produced and the types to be retained; the period of data retention; the data formats, metadata and provisions to protect confidentiality, security and intellectual property; and data storage and access (NEH, 2017, pp. 2-3). Although the Australian Research Council (ARC) does not require the comprehensive DMPs which the NEH does, it has asked its grantees since February 2014 to indicate how they plan to manage the data from ARC-funded research (ARC, 2015), and Australian universities anyway "have been proactively creating [DMPs] simply because it is good practice" (Erway, 2013, p. 6).

The UK's AHRC prefers to call the DMP a "Technical Plan" (**AHRC, 2015c**), owing to a project's "digital outputs", which are required to be made available as soon as practically possible. The Technical Plan has formed part of a grant application to the AHRC since December 2012 (AHRC, 2015a). It asks researchers to consider contexts for the reuse of the research data, and to give plans for dissemination and impact. It recommends using open standards and formats to facilitate reuse, and procedures and tools to ensure access by the future user base (AHRC, 2015c). It requires information about digital outputs and technologies (Section 1), standards and formats (Section 2.a), hardware and software (Section 2.b), data acquisition, processing and analysis procedures to be followed (Section 2.c), technical support and expertise to be used (Section 3), data preservation (storing ancillary data after funding ends) (Section 4.a) and data sustainability (making important data usable after funding ends) (Section 4.b). Researchers must indicate which digital outputs will be preserved and which sustained, and consider the duration and costs (AHRC, 2015c, Section 4). Information about copyright, intellectual property and ethical issues is required by the Case for Support, another part of the grant application.

Because publicly-funded digital outputs need to be made accessible, metadata and contextual documentation are integral to a DMP, and attention to them early in a project will help to ensure that sustainable formats and metadata are chosen. The RDM planning procedures of the UAL in Annex A (UAL, 2014b, p. 1), which focus on the drafting of the Technical Plan required by the AHRC, require metadata to be created as soon as the funding is awarded. The RDM policy itself specifies that research data is captured in a preferred and open format, and appropriate metadata is created early so that the data is "ready for exposure after the project concludes" (UAL, 2014a, p. 4). The technical aspects of a project are now recognised as part of research and, if addressed early in the planning process, should ensure that project benefits are realised fully (Rans, 2013, p. 4). The SUDAMIH final report warned UK institutions intent on winning research funding to have "a specialist technical advisory service which researchers can consult for assistance with the technical aspect of bids" (Wilson, 2011, p. 31). The DCC has produced a customisable open-source DMPonline tool, and examples of humanities DMPs funded by the NEH and the AHRC (e.g. Rans, 2013) are on the DCC website (DCC, 2017). DMP templates or checklists are provided by many UK and Australian universities on their research webpages, and their focus on the DMP indicates its importance for sustainable RDM.

6. Ethical considerations, Limitations and Assumptions

Because online documentation formed the subject of my examination, there was no need for ethics approval, and data collection began immediately. However, the lone research, short time-span and focus on documents mean that there were limitations. Normally systematic reviews are undertaken by a small team of researchers who can cross-check data collection, coding procedures and interpretation of results in the interests of validity (O'Leary, 2014, p. 258), but my own rigorous methods had to suffice. My choice of documentation for a small research project was of necessity very selective, and there was no opportunity, for triangulation of the data, to hear from librarians themselves about their perspective on the challenges of humanities data curation. However, obtaining that perspective, and extending the scope of the study to include intense examples of European or US documentation could be done

in follow-up studies. This study took for granted the reasonable transferability to the New Zealand context of data collected from RDM documents from culturally similar, English-speaking countries, from which New Zealand has traditionally taken its lead. There would have been no point in undertaking this research if this were not the case (Leedy & Ormrod, 2015, p. 62). This study also assumed a pragmatic bias, that "principles of practice" or "high-quality lessons learned" (Patton, 2002, p. 564), extrapolated from multiple relevant sources, can be applied in similar contexts. A systematic review of information-rich evidence in recent documents from the UK and Australia should help to inform the development in New Zealand academic libraries of institutional services to manage humanities research data effectively.

7. Conclusion

This study adopted Hjørland's domain analytic approach, an evaluation research design, and document analysis methodology to identify policies and initiatives in humanities RDM that may be transferable to the New Zealand context. If policies and strategies are to be effective in promoting scholarly information exchange, however, they must be sensitive to the culture of different research domains (Bulger et al., 2011, p. 7). Owing to their intellectual variety and idiosyncrasy, the human sciences are characterised by high task uncertainty and low mutual dependence. Unlike the natural sciences, "they do not exhibit a stable configuration of specialized tasks or of problem areas, nor do they have strong co-ordinating mechanisms which systematically interrelate results" (Whitley, 1984, p. 168). A focus on multiple reinterpretation, often of the same object in infinitely new contexts, distinguishes humanities scholarship from that of the natural sciences (Borgman, 2015, p. 200):

The ability to interpret or reinterpret—whether artifacts from an archeological excavation, Greek vases, Buddhist texts, the transmission of knowledge in third-century China, or any other aspect of human culture—in light of new perspectives or evidence is the essence of scholarship in the humanities.

Data generated by the human sciences, "ranging from 'traditional' survey data to the elusive process materials that underwrite much of arts-based research" (Nadim et al., 2013, p. 1), and the divergent attitudes expressed towards the systematic management of primary materials, pose many challenges for RDM, which relies on standardised policies and protocols for effective practice.

Nevertheless, the implications of this study are that domain-specific RDM provisions and services developed by UK and Australian universities, and DMP specifications of national funding agencies, could profitably be applied by New Zealand academic libraries to systematise their humanities RDM procedures. Just how the conditions now attached to national agency funding are normalising research outcomes and increasing systematic collaboration is evident in the proliferation of digital humanities projects and team-driven humanities research, and the RDM support services being developed in the UK and Australia to manage the data deluge. The humanities laboratory of the 21st century is increasingly technology-driven, enabling teams of humanities scholars to use resources across international archives and libraries to create new databases and collections, even as the creative arts are continuing to produce non-digital materials. Follow-up studies of European or US institutional RDM procedures, and global humanities infrastructural initiatives, such as DARIAH, are likely to add more definition to the humanities laboratory of today and provide further answers to the research questions which guided my investigation.

7.1 Why do humanities research data need curating?

The documentation has shown that analogue materials are still commonly produced, notably in the visual and performing arts. However, digital audio-visual collections and corpora, using visualisation and other digital tools (such as TEI) are now integral to the intellectual infrastructure of the humanities. Better institutional RDM practices are likely to encourage humanities researchers to reuse existing primary materials, reinterpret extant datasets and corpora, and verify research arguments. Better RDM practices should link primary materials with humanities researchers' publications, make datasets and corpora more discoverable, and advance both scholarly and institutional reputations. Better practices should also safeguard the value of primary materials, allow controlled access to them, and prevent data loss or obsolescence.

7.2 How do humanities scholars use and reuse data?

The documentation has reflected the variety of data and practices in the human sciences, and the range of attitudes expressed towards data preservation and dissemination. The value of humanities data lies in their potential for reuse by the same scholar over a lifetime's work, and by different audiences, scholarly and lay, long into the future. Data are unlikely to depreciate over time; in fact, the opposite is the case, especially regarding endangered languages and irreplaceable artefacts. Humanities researchers seem to be open to sharing their collections and corpora if they are properly documented, although there is likely to be some resistance to institutional RDM policies and initiatives, especially if they are not domain-sensitive.

7.3 What are the drivers of institutional humanities RDM?

The documentation has confirmed the responsibility of library and IT services in a research institution to provide professional expertise, training, and facilities in RDM so that primary materials can be managed to maximise the investment of public funding and the value of the institution's research data assets. Making these assets accessible and usable facilitates scholarly information exchange as envisaged in the UNISIST model, enhances the institution's reputation in a competitive climate, and preserves the scholarly record. Copyright and intellectual property advice, domain-sensitive data stewardship, a federated systems infrastructure, early and embedded intervention in the workflow, and liaison with researchers through the data lifecycle are areas in which support services in New Zealand institutions may contribute to best practice humanities RDM and catch-up investment in e-research infrastructure.

References

Akers, K. G. (2013). Looking out for the little guy: Small data curation. *Bulletin of the American Society for Information Science and Technology*, *39*(3), 58-59. Available at http://onlinelibrary.wiley.com/doi/10.1002/bult.2013.1720390317/epdf

Akers, K. G., & Doty, J. (2013). Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation*, *8*(2), 5-26. doi:10.2218/ijdc.v8i2.263

American Council of Learned Societies. (2006). *Our cultural commonwealth: The report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences*. New York, NY: American Council of Learned Societies. Available at http://www.acls.org/cyberinfrastructure/OurCulturalCommonwealth.pdf

Arts and Humanities Data Service (AHDS). (n.d.). *AHDS guides to good practice*. Retrieved from http://web.archive.org/web/20150315111159/http://www.ahds.ac.uk/history/creating/guides/index.htm

Arts and Humanities Research Council (AHRC). (2015a). *Research funding guide*. Retrieved from https://web.archive.org/web/20150906051137/http://www.ahrc.ac.uk/funding/research/researchfundingguide/

Arts and Humanities Research Council (AHRC). (2015b). *Subject coverage*. Available at http://www.ahrc.ac.uk/documents/guides/ahrc-s-subject-coverage/

Arts and Humanities Research Council (AHRC). (2015c). *Technical plan*. Retrieved from

http://www.ahrc.ac.uk/funding/research/researchfundingguide/attachments/technicalplan/

Australian Government. (2007). *Australian code for the responsible conduct of research*. Available at
http://www.nhmrc.gov.au/_files_nhmrc/file/publications/synopses/r39.pdf

Australian Government. (2016). *Review of the Australian Code for the Responsible Conduct of Research: Status update*. Retrieved from
https://www.nhmrc.gov.au/research/responsible-conduct-research/review-australian-code-responsible-conduct-research

Australian National Data Service (ANDS). (n.d.). *ANDS project registry*. Retrieved from https://projects.ands.org.au/policy.php

Australian National Data Service (ANDS). (2010). *Outline of a research data management policy for Australian universities / institutions*. Available at
http://ands.org.au/__data/assets/pdf_file/0004/382072/datamanagementpolicyoutline.pdf

Australian National Data Service (ANDS). (2017). *Institutional policies and procedures*. Available at
http://www.ands.org.au/__data/assets/pdf_file/0008/738782/Institutional-policies-and-procedures.pdf

Australian Research Council (ARC). (2015). *Research data management*. Retrieved from http://www.arc.gov.au/research-data-management

Bawden, D., & Robinson, L. (2013). Domain analysis. In *Introduction to information science* (Chapter 5, pp. 91-104). Chicago, IL: Neal-Schuman.

Borgdorff, H. (2010). The production of knowledge in artistic research. In M. Biggs & H. Karlsson (Eds.), *The Routledge companion to research in the arts* (Chapter 3, pp. 44-63). Abingdon, UK: Routledge.

Borgman, C. L. (2009). The digital future is now: A call to action for the humanities. *Digital Humanities Quarterly*, *3*(4). Retrieved from http://www.digitalhumanities.org/dhq/vol/3/4/000077/000077.html

Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA: The MIT Press.

Bowen, G. A. (2008). Naturalistic inquiry and the saturation concept: A research note. *Qualitative Research*, *8*(1), 137-152. doi:10.1177/1468794107085301

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27-40. doi:10.3316/QRJ0902027

Brown, E. (2010). "I know what you researched last summer": How academic librarians are supporting researchers in the management of data curation. *The New Zealand Library & Information Management Journal*, *52*(1), 55-69. Available at http://www.lianza.org.nz/sites/default/files/files/nzlimj_vol_52_issue_no_1_oct_2010.pdf#page=55

Bulger, M. E., Meyer, E. T., De la Flor, G., Terras, M., Wyatt, S., Jirotka, M., Eccles, K., & Madsen, C. (2011). *Reinventing research? Information practices in the humanities*. London, UK: The Research Information Network. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1859267

Burgess, R. (2013). *Visual arts research data management: From inception to the production of an RDM policy for the Glasgow School of Art*. Available at https://vads.ac.uk/kaptur/outputs/GSA2013.pdf

Burrows, T. (2013). A data-centred 'virtual laboratory' for the humanities: Designing the Australian Humanities Networked Infrastructure (HuNI) service. *Literary and Linguistic Computing*, *28*(4), 576-581. doi:10.1093/llc/fqt064

Carlson, J. (2012). Demystifying the data interview: Developing a foundation for reference librarians to talk with researchers about their data. *Reference Services Review*, *40*(1), 7-23. doi:10.1108/00907321211203603

Christensen, H. D. (2014). The framing of scientific domains: About UNISIST, domain analysis and art history. *Journal of Documentation*, *70*(2), 261-281. doi:10.1108/JD-03-2013-0038

Collins, E. (2012). The national data centres. In G. Pryor (Ed.), *Managing research data* (Chap 8, pp. 151-172). London, UK: Facet Publishing.

CONZUL Working Group. (2015, February 2). *Research data management framework report*. Wellington, NZ: New Zealand Vice-Chancellors' Committee. Available at http://www.universitiesnz.ac.nz/files/CONZUL-RDM%20Framework%20Report%202015%20FINAL.pdf

Corrall, S., Kennan, M. A., & Afzal, W. (2013). Bibliometrics and research data management services: Emerging trends in library support for research. *Library Trends*, *61*(3), 636-674. doi:10.1353/lib.2013.0005

Cox, A. M., Kennan, M. A., Lyon, L., & Pinfield, S. (2017). Developments in research data management in academic libraries: Towards an understanding of research data service maturity. *Journal of the Association for Information Science and Technology*, 25 March [online]. doi:10.1002/asi.23781

Cox, A. M., & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*, *46*(4), 299-316. doi:10.1177/0961000613492542

Cox, J. (2016). Communicating new library roles to enable digital scholarship: A review article. *New Review of Academic Librarianship*, *22*(2-3), 132-147. doi:10.1080/13614533.2016.1181665

Digital Curation Centre (DCC). (2017). *Example DMPs and guidance*. Retrieved from
http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples

Donnelly, A. (2013). *Data curation profile for linguistics*. Interview with Bert Remijsen,
School of Philosophy, Psychology & Language Sciences, August 20, 2013.
Available at http://datalib.edina.ac.uk/mantra/libtraining/CurationProfiles/DCP-
interview-PPLS-Donnelly.pdf

Duff, W. M. (1998). Harnessing the power of warrant. *The American Archivist*, *61*(1),
88-105. Retrieved from http://www.jstor.org/stable/40294076

Eaker, C. (2012). *Data curation profile – Architectural History / Epigraphy*. Interview,
April 30, 2012. Retrieved from
http://guides.library.ucla.edu/c.php?g=180580&p=1191506

Ekmekcioglu, Ç., & Rice, R. (2009). *Edinburgh Data Audit Implementation Project
final report*. Available at
https://vads.ac.uk/kaptur/outputs/KAPTUR_final_report.pdf

eResearch 2020. (2015, March). *eResearch challenges in New Zealand: Discussion
document*. Available at http://eresearch2020.org.nz/wp-
content/uploads/2015/03/eR2020-Full-Discussion-Document.pdf

Erway, R. (2013). *Starting the conversation: University-wide research data
management policy*. Dublin, OH: OCLC Research. Available at
http://www.oclc.org/content/dam/research/publications/library/2013/2013-08.pdf

Flanders, J., & Muñoz, T. (n.d.). An Introduction to humanities data curation. *DH
Curation Guide: A community resource guide to data curation in the digital
humanities*. Retrieved from https://web-
beta.archive.org/web/20170126012445/https://guide.dhcuration.org/contents/int
ro/

Forrest, M. (2014). *Data curation profile for history*. Interview with Graham Black, School of History, Classics & Archaeology, July 23, 2014. Available at http://datalib.edina.ac.uk/mantra/libtraining/CurationProfiles/DCP-interview-Black-Forrest.pdf

Freiman, L., Ward, C., Jones, S., Molloy, L., & Snow, K. (2010). *Incremental scoping study and implementation plan*. Available at http://www.lib.cam.ac.uk/preservation/incremental/documents/Incremental_Scoping_Report_170910.pdf

Furness, J. (2014). *MANTRA and RDM training: Research data profile project*. Interview with Angela McClanahan, School of Art, February 11, 2014. Available at http://datalib.edina.ac.uk/mantra/libtraining/CurationProfiles/DCP-interview-McClanahan-Furness.pdf

Garrett, L., Gramstadt, M-T., Burgess, R., Murtagh, J., Nadim, T., & Spalding, A. (2012). *Kaptur environmental assessment report*. Retrieved from http://www.research.ucreative.ac.uk/1054/

Gilliland, A. (2014). From custody to stewardship: Digital repositories, preservation, and curation. In *Conceptualizing 21st-century archives* (Chapter 10, pp. 230-254). Chicago, IL: Society of American Archivists.

Glasgow School of Art (GSA). (2016, October 25). *Research data management policy*. Available at http://www.gsa.ac.uk/media/1330796/gsa-research-data-management-policy-2016_10_25.pdf

Gough, D., Oliver, S., & Thomas, J. (2012). *An introduction to systematic reviews*. London, UK: Sage Publications.

Gramstadt, M-T. (2013). *Jisc-KAPTUR final report*. Available at https://vads.ac.uk/kaptur/outputs/KAPTUR_final_report.pdf

Gray, S. (2010). *JISC-CAiRO user needs report*. Retrieved from
   http://web.archive.org/web/20131126030535/http://www.projectcairo.org/node/8

Griffiths, M. (2010). Research and the self. In M. Biggs & H. Karlsson (Eds.), *The Routledge companion to research in the arts* (Chapter 10, pp. 167-185). Abingdon, UK: Routledge.

Grout, C., Purdy, P., & Rymer, J. (n.d.). *Creating digital resources for the visual arts: Standards and good practice*. AHDS/Visual Arts Data Service. Retrieved from
   http://web.archive.org/web/20140827230133/http://vads.ahds.ac.uk/guides/creating_guide/contents.html

Guy, M., Donnelly, M., & Molloy, L. (2013). Pinning it down: Towards a practical definition of "research data" for creative arts institutions. *International Journal of Digital Curation*, *8*(2), 99-110. doi:10.2218/ijdc.v8i2.275

Henry, G. (2014). Data curation for the humanities: Perspectives from Rice University. In J. M. Ray (Ed.), *Research data management: Practical strategies for information professionals* (Chapter 17, pp. 347-374). West Lafayette, IN: Purdue University Press.

Higman, R., & Pinfield, S. (2015). Research data management and openness: The role of data sharing in developing institutional policies and practices. *Program: electronic library and information systems*, *49*(4), 364-381. doi:10.1108/PROG-01-2015-0005

Hjørland, B., & Albrechtsen, H. (1995). Toward a new horizon in information science: Domain-analysis. *Journal of the American Society for Information Science*, *46*(6): 400-425. doi:10.1002/(SICI)1097-4571(199507)46:6

Hjørland, B., Andersen, J., & Søndergaard, T. F. (2005). UNISIST model and knowledge domains. In M. A. Drake (Ed.), *Encyclopedia of Library and Information Science* (2nd ed., pp. 129-135). Taylor & Francis. Available at

http://ccc.ku.dk/employees/?pure=files%2F47044534%2FUNISIST_model_and_knowledge_domains.pdf

Horton, L., & Digital Curation Centre (DCC). (2016). *Overview of UK institution RDM policies.* Version 6 August 2016, Digital Curation Centre. Available at http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies

Jones, S. (2012.). Research data policies: Principles, requirements and trends. In G. Pryor (Ed.), *Managing research data* (Chapter 3, pp. 47-66). London, UK: Facet Publishing.

Kennan, M. A., & Markauskaite, L. (2015). Research data management practices: A snapshot in time. *International Journal of Digital Curation*, *10*(2), 69-95. doi:10.2218/ijdc.v10i2.329

Kutay, S. (2014). Advancing digital repository services for faculty primary research assets: An exploratory study. *The Journal of Academic Librarianship*, *40*(6), 642-649. doi:10.1016/j.acalib.2014.08.006

Leedy, P. D., & Ormrod, J. E. (2015). *Practical research: Planning and design.* Harlow, UK: Pearson Education.

Lyon, L. (2016). Librarians in the lab: Toward radically re-engineering data curation services at the research coalface. *New Review of Academic Librarianship*, *22*(4), 391-409. doi:10.1080/13614533.2016.1159969

Macdonald, S., & Martinez-Uribe, L. (2010). Collaboration to data curation: Harnessing institutional expertise. *New Review of Academic Librarianship*, *16*(S1), 4-16. doi:10.1080/13614533.2010.505823

Mohr, A. H., Bishoff, J., Bishoff, C., Braun, S., Storino, C., & Johnston, L. R. (2015). When data is a dirty word: A survey to understand data management needs across diverse research disciplines. *Bulletin of the Association for Information Science and Technology*, *42*(1), 51-53. doi:10.1002/bul2.2015.1720420114

Morrison, A., Popham, M., & Wikander, K. (n.d.). *Creating and documenting electronic texts: A guide to good practice*. AHDS/Oxford Text Archive. Retrieved from http://web.archive.org/web/20160812142925/http://ota.ox.ac.uk:80/documents/creating/cdet/

Murtagh, J., Mahurter, S., & Hancox, E. (2013). *Research data management at the University of the Arts London*. Available at https://vads.ac.uk/kaptur/outputs/UAL2013.pdf

Nadim, T., Cooke, J., & Gray, A. (2013). *Towards a research data management policy at Goldsmiths*. Available at https://vads.ac.uk/kaptur/outputs/Goldsmiths2013.pdf

National Endowment for the Humanities (NEH). (2017). *Data management plans for NEH Office of Digital Humanities proposals and awards*. Available at https://www.neh.gov/files/grants/data_management_plans_2017.pdf

Ohaji, I. K. (2016). *Data management: An exploration of the data librarian role in New Zealand research organisations* (Doctoral dissertation, Victoria University of Wellington, Wellington, New Zealand). Retrieved from http://researcharchive.vuw.ac.nz/handle/10063/5657

O'Leary, Z. (2014). Secondary data: Documents, data sets and online data. In *The essential guide to doing your research project* (Chapter 12, pp. 243-270). London, UK: Sage Publications.

Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). London, UK: Sage Publications.

Pinfield, S., Cox, A. M., & Smith, J. (2014). Research data management and libraries: Relationships, activities, drivers and influences. *PLoS ONE*, *9*(12): e114734. doi:10.1371/journal.pone.0114734

Pryor, G. (Ed.). (2012a). *Managing research data*. London, UK: Facet Publishing.

Pryor, G. (2012b). Why manage research data? In G. Pryor (Ed.), *Managing research data* (Chapter 1, pp. 1-16). London, UK: Facet Publishing.

Pryor, G., Jones, S., & Whyte, A. (Eds.). (2014). *Delivering research data management services: Fundamentals of good practice*. London, UK: Facet Publishing.

Rans, J. (2013, November). *Planning for the future: Developing and preserving information resources in the Arts and Humanities.* DCC RDM Services case studies. Edinburgh: Digital Curation Centre. Available at http://www.dcc.ac.uk/sites/default/files/documents/publications/case-studies/OU_A_and_H_DMP_case_study.pdf

Rath, L. L. (2016). Low-barrier-to-entry data tools: Creating and sharing humanities data. *Library Hi Tech*, *34*(2), 268-285. doi:10.1108/LHT-07-2015-0073.

Ray, J. M. (Ed.). (2014). *Research data management: Practical strategies for information professionals*. West Lafayette, IN: Purdue University Press.

Research Councils UK (RCUK). (2015, July). *Guidance on best practice in the management of research data*. Available at http://www.rcuk.ac.uk/documents/documents/rcukcommonprinciplesondatapolicy-pdf/

Rice, R. (2013/2014). Edinburgh DataShare: Reflections from a data repository manager. *Bulletin of the Association for Information Science and Technology*, *40*(2), 39-40. Available at http://www.asis.org/Bulletin/Dec-13/DecJan14_RDAP_Rice.pdf

Rice, R., & Haywood, J. (2011). Research data management initiatives at University of Edinburgh. *Journal of Digital Curation*, *2*(6), 232-244. doi:10.2218/ijdc.v6i2.199

Roos, A., & Hedlund, T. (2016). Using the domain analytical approach in the study of information practices in biomedicine. *Journal of Documentation*, *72*(5), 961-986. doi:10.1108/JD-11-2015-0139

Spalding, A. (n.d.). *Raising the profile of research data management: The UCA story*. Available at https://vads.ac.uk/kaptur/outputs/UCA2013.pdf

Talja, S. (2005). The domain analytic approach to scholars' information practices. In K. E. Fisher, S. Erdelez, & L. McKechnie (Eds.), *Theories of information behaviour* (Chapter 18, pp. 123-127). Medford, NJ: American Society for Information Science and Technology.

Tancheva, K. (2012). *Data curation profile – Linguistics*. Interview, March 23, 2012. Retrieved from http://guides.library.ucla.edu/c.php?g=180580&p=1191507

Trace, C. B., & Karadkar, U. P. (2016). Information management in the humanities: Scholarly processes, tools, and the construction of personal collections. *Journal of the Association for Information Science and Technology*, 28 March [online]. doi:10.1002/asi.23678

University of Adelaide. (2015, November 9). *Research data and primary materials policy*. Retrieved from http://www.adelaide.edu.au/policies/4043/

University of Oxford. (n.d.a). *Humanities data: A hands-on approach – Making the most of messy data*. Retrieved from http://digital.humanities.ox.ac.uk/dhoxss/2016/workshops/dhcuration

University of Oxford. (n.d.b). *Research data management – Training courses and events*. Retrieved from http://blogs.it.ox.ac.uk/acit-rs-team/advice/rdmcourses/

University of Oxford. (2009). *Supporting data management infrastructure for the humanities (SUDAMIH): Project outputs*. Retrieved from http://sudamih.oucs.ox.ac.uk/documents.xml

University of Sydney. (n.d.). *Policies: Understand your research data responsibilities*. Retrieved from https://library.sydney.edu.au/research/data-management/policies.html

University of Sydney. (2015, March 13). *Sydney Conservatorium of Music research data management provisions 2015*. Available at http://sydney.edu.au/policies/showdoc.aspx?recnum=PDOC2015/384&RendNum=0

University of the Arts, London (UAL). (2014a). *Research data management policy*. Available at http://www.arts.ac.uk/media/arts/research/documents/UAL-Research-Data-Management-Policy-2014.pdf

University of the Arts, London (UAL). (2014b). *UAL research data management planning procedures (annex a)*. Available at http://www.arts.ac.uk/media/arts/research/documents/UAL-RDM-Policy-2014-Annex-A.pdf

Wallace, D. P., & Van Fleet, C. (2012). Knowing, research, and evaluation. In *Knowledge into action: Research and evaluation in library and information studies* (Chapter 1, pp. 5-38). Santa Barbara, CA: Libraries Unlimited.

Waters, D. J. (2013). An overview of the digital humanities. *Research Library Issues*, *284*, 3-11. Retrieved from http://publications.arl.org/rli284/3

Weber, N. M., Baker, K. S., Thomer, A. K., Chao, T. C., & Palmer, C. L. (2012). Value and context in data use: Domain analysis revisited. *Proceedings of the American Society for Information Science and Technology*, *49*(1), 1-10. doi:10.1002/meet.14504901168

Weller, T., & Monroe-Gulick, A. (2014). Understanding methodological and disciplinary differences in the data practices of academic researchers. *Library Hi Tech*, *32*(3), 467-482. doi:10.1108/LHT-02-2014-0021

Whitley, R. (1984). *The intellectual and social organization of the sciences*. Oxford, UK: Clarendon Press.

Whitley, R. (2000). Introduction. Science transformed? The changing nature of knowledge production at the end of the twentieth century. In *The intellectual and social organization of the sciences* (2nd ed., pp. ix-xliv). Oxford, UK: Oxford University Press.

Whitmire, A. L., Boock, M., & Sutton, S.C. (2015). Variability in academic research data management practices. *Program: electronic library and information systems*, *49*(4), 382-407. doi:10.1108/PROG-02-2015-0017

Wilson, J. A. J. (2011). *JISC-SUDAMIH final report*. Available at http://sudamih.oucs.ox.ac.uk/docs/Sudamih_FinalReport_v1.0.pdf

Witt, M., Carlson, J., Brandt, D. S., & Cragin, M. H. (2009). Constructing data curation profiles. *International Journal of Digital Curation*, *4*(3), 93-103. doi:10.2218/ijdc.v4i3.117

Appendix 1: Documents analysed using NVivo

1. RDM policies/guidance:

University of the Arts, London. (2014a). *Research data management policy*.

University of the Arts, London. (2014b). *UAL research data management planning procedures (annex a)*.

University of Adelaide. (2015). *Research data and primary materials policy*.

University of Sydney. (2015). *Sydney Conservatorium of Music research data management provisions 2015*.

Glasgow School of Art. (2016). *Research data management policy*.

2. RDM project reports:

Freiman, L. et al. / Universities of Glasgow and Cambridge. (2010). *Incremental scoping study and implementation plan*.

Gray, S. / University of Bristol. (2010). *JISC-CAiRO user needs report*.

Wilson, J. A. J. / University of Oxford. (2011). *JISC-SUDAMIH final report*.

Garrett, L. et al. / University for the Creative Arts, London. (2012). *KAPTUR environmental assessment report*.

Burgess, R. / Glasgow School of Art. (2013). *Visual arts research data management*.

Gramstadt, M-T. / University for the Creative Arts, London. (2013). *Jisc-KAPTUR final report*.

Murtagh, J. et al. (2013). *Research data management at the University of the Arts London*.

Nadim, T. et al., (2013). *Towards a research data management policy at Goldsmiths*.

3. Related reports/guides:

Arts and Humanities Data Service (AHDS) / Oxford Text Archive (OTA). (n.d.). *Creating and documenting electronic texts: A guide to good practice*.

AHDS / Visual Arts Data Service (VADS). (n.d.). *Creating digital resources for the visual arts: Standards and good practice*.

Ekmekcioglu, Ç., & Rice, R. (2009). *Edinburgh data audit implementation project final report*.

Bulger, M. E. et al. / Research Information Network (RIN). (2011). *Reinventing research? Information practices in the humanities*.

4. **Data curation profiles (DCPs):**

Eaker, C. (2012). *Data curation profile – Architectural History / Epigraphy*. University of Tennessee, Knoxville

Tancheva, K. (2012). *Data curation profile – Linguistics*. Cornell University

Donnelly, A. (2013). *Data curation profile for linguistics*. University of Edinburgh

Forrest, M. (2014). *Data curation profile for history*. University of Edinburgh

Furness, J. (2014). *MANTRA and RDM training: Research data profile project*. University of Edinburgh

5. **Data management plan (DMP) specifications:**

Arts and Humanities Research Council (AHRC). (2015c). *Technical plan*.

National Endowment for the Humanities (NEH). (2017). *Data management plans for NEH Office of Digital Humanities proposals and awards*.

Appendix 2: Nodes used in NVivo

The NVivo nodes and sub-nodes tabulated below reflect (1) categories drawn from the Sydney Conservatorium of Music provisions imposed on the documents, and (2) themes discovered through reading the documents.

Nodes	Sources (prevalence)	References (frequency)
Best practices	11	44
Context	14	60
Collaboration	6	19
Case studies	10	12
Data as asset	15	25
Data definitions	18	40
Data management plan	16	62
Humanities definitions	4	19
Library role	4	6
Models	7	20
Nature of data	23	79
Research practices	32	306
- Access	19	59
- Annotation	7	21
- Attitudes	10	47
- Storage	29	99
Rights	17	45
Selection of data	5	10
Sharing practices	19	99
- Audiences	7	11
- Methods	3	6
Support	20	77
Technologies	17	61
Training	13	38
Trends	14	50

Types of data	27	199
- Formats	13	77
- Metadata	18	47
- Standards	3	13
- Versions	6	9