Ádám Dér

Exploring the academic libraries' readiness for research data management: cases from Hungary and Estonia

Supervisor: Raivo Ruusalepp

Master Thesis
International Master in Digital Library Learning
2015

DECLARATION

I certify that all material in this dissertation which is not my own work has been identified and that no material is included for which a degree has previously been conferred upon me.


…… (signature of candidate)

ACKNOWLEDGEMENTS

I would like to thank all the people who made these years so special and unforgettable: first and foremost to my supervisor, Raivo Ruusalepp, for his help, his insightful comments, and his optimism throughout my last months in Tallinn; to all our professors at the Oslo and Akershus University, Tallinn University, and the University of Parma, as well as to all the visiting scholars and guest lecturers; to the people at the Loughborough University Library and every other place I had a chance to visit in these two years.

I would also like to thank for the countless, wonderful memories to all my classmates, friends, family, and every other people whom I met during this master programme.

Last, but not least, I would also like to express my gratitude to all the people who have participated in this study for taking their valuable time and sharing their expertise with me.

ABSTRACT

The aim of this thesis was to examine the academic libraries readiness for research data management in countries where this issue is only getting recognised. Therefore, eight academic libraries from Estonia and Hungary has been contacted and asked to participate in this study. The main purpose was to gain understanding about how the different staff members of the participating libraries see their institutions' role in research data management, as well as to gain insight about current practices at the institutions that could serve as a basis for possible future services in research data management. The thesis intended to use a multiple case study approach to examine and to be able to compare the current situation and future priorities of the participating institutions. The study found that while libraries are aware on the raising demands from various stakeholders to introduce RDM services, the issue is so complex that librarians alone certainly cannot handle it. The respondents noted a great level of optimism about the prospects of introducing RDM services, and librarians, as well as library directors were confident that they are able to tackle the issue. However, even if a very limited amount of data from an IT staff member suggests, there might be gaps between how different units within libraries perceive their institutions' possible role in RDM. Limitations of the research were either planned delimitations like the country choices, or unintended, but natural flaws of the data collection technique. The implication of this thesis for relevant communities is that it presents a baseline for comparison of different countries, institutions, and units; moreover, it sheds light on issues that can be studied further by subsequent research.

Keywords: research data management, academic libraries, institutional readiness, Hungary, Estonia

TABLE OF CONTENTS

List of figures

List of tables

List of abbreviations

ANDS: Australian National Data Service

ARL: Association of Research Libraries

BOAI: Budapest Open Access Initiative

CARDIO: Collaborative Assessment of Research Data Infrastructure and Objectives

DAF: Data Asset Framework (formerly Data Audit Framework)

DCC: Digital Curation Centre

DMP: Data Management Plan

EPSRC: Engineering and Physical Science Research Council

JISC: Joint Information Science Committee

LIBER: Ligue des Bibliothèques Européennes de Recherche - Association of European Research Libraries

NIH: National Institutes of Health

NSF: National Science Foundation

OECD: Organisation for Economic Co-operation and Development

PLOS: Public Library of Science

RCUK: Research Councils United Kingdom

RDA: Research Data Alliance

RDM: Research Data Management

RIN: Research Information Network

SIM4RDM: Support Infrastructure Models for Research Data Management

SURF: Samenwerkende Universitaire Reken Faciliteiten - Co-operative University Computing Facilities

“Everything not saved will be lost”

- Nintendo quit screen message

1. INTRODUCTION

In the last decades, technological advancements have enabled researchers to produce, analyse, store, and share digital data on a continuously growing scale. Data sharing has been common practice in certain disciplines for a long time, while other disciplines are only starting to recognize its benefits. Lately, influenced by the open-access movement, governments, funding agencies, and journal publishers have begun to encourage the sharing of research data:

- governments and political forums are issuing recommendations on the availability of publicly funded research data,
- a growing number of funding agencies require researchers to include a data management plan in the grant proposal and to ensure that research data are accessible after the project is finished,
- numerous journal publishers are starting to require data sets as supplementary materials to articles.

This raising level of awareness on data management, preservation, and sharing adds new responsibilities to researchers, as it requires significant effort to manage research data effectively.

Many have argued that academic and research libraries could serve as an important partner in research data management, since many of their traditional practices (organising, preserving, and disseminating information) overlap with the practices necessary to manage research data. However, as digital research data is relatively different in its nature from traditional books and journals, there is also a concern over the involvement level of libraries.

This study focuses on current activities and future plans in providing services for research data management in academic libraries from two European countries, where questions about the management and sharing of research data are only starting to emerge.

1.1 Definitions

Data, plural form of the Latin *datum,* which originally means "something given", stands at the bottom of the widely acclaimed data-information-knowledge-wisdom pyramid. This rather straightforward hierarchy becomes slightly circular once one tries to define the term *data*: the Oxford Dictionary of English describes *datum* as "a piece of information" (Stevenson, 2010b), and *data* as "facts and statistics collected together for reference or analysis" (Stevenson, 2010a).

Just as with information, there are countless definitions of data as well: a study by Zins (2007), for example, documented "130 definitions of *data*, *information*, and *knowledge* formulated by 45 scholars" (p. 479 emphasis in original). The most prevalent of the description of data is that they "are facts, numbers, letters, and symbols that describe an object, idea, condition, situation, or other factors" (National Research Council, 1999, p. 15), which resembles Machlup's (1983) widely used definition which reads as follows: "data are the things given to the analyst, investigator, or problem-solver; they may be numbers, words, sentences, records, assumptions – just anything given, no matter what form and of what origin." (Machlup, 1983, as cited in Zins, 2006, p. 452)

Note that even though the term data is usually defined relatively openly, the act of analysis, problem-solving, investigation are usually present in most of these definitions. Consequently, the definition of *research data* does not particularly differ from the one of *data* itself, as according to OECD, research data are "factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings." (OECD Principles and Guidelines for Access to Research Data from Public Funding, 2007, p. 13) The Blue Ribbon Task Force on Sustainable Economics for a Digital Planet (2010) has complemented this definition, pointing out that research data are not only the primary sources of scholarship, but "the first order results of that research" as well. (2010, p. 56)

OECD provided a definition for *research* as well, dividing it in two broad categories, *basic* and *applied*: both types are work undertaken in order to acquire new knowledge "of the underlying foundation of phenomena and observable facts". But whereas applied research

has practical aims or objectives, basic research is an investigation undertaken "without any particular application or use in view." (OECD, 2002, p. 30)

It might be reasonable to divide this aforementioned definition of data in two main parts: analogue and digital. Advances in technology are changing the ways we tend to think about data: nowadays we are imagining them more like digital signals, bit streams, or binary numbers rather than observable facts no matter what form or origin. As technology is becoming data-intensive, along with the increasing speed of computing, the decreasing costs of digital storage media, the higher performance of networks, and the evolving sophistication of tools for analysis enable scientists to produce and examine digital data on a continuously growing scale. This phenomenon, often called the data-deluge or the fourth paradigm requires significant efforts from researchers to manage the growing amount of (digital) research data effectively.

The activities generally associated with the handling of data are usually *preservation*, *curation*, *management*, and *stewardship*. According to one definition, "preservation entails standards-based, active management practices that guide data throughout the research life cycle, as well as ensure the long-term usability of these digital resources" while "curation involves ways of organizing, displaying, and repurposing preserved data" (Friedlander & Adler, 2006, p. 12). One other states that "preservation is about ensuring that what is handed over to a repository or publisher remains fit for secondary use in the longer term" and "curation connects first use to secondary use" (Whyte & Tedds, 2011, p. 1).

Stewardship "involves both preservation and curation" (Friedlander & Adler, 2006, p. 12), and management "concerns the organisation of data, from its entry to the research cycle through to the dissemination and archiving of valuable results." (Whyte & Tedds, 2011, p. 1)

Cox and Pinfield (2013) defined research data management as follows:

> It consists of a number of different activities and processes associated with the data lifecycle, involving the design and creation of data, storage, security, preservation, retrieval, sharing, and reuse, all taking into account technical capabilities, ethical considerations, legal issues and governance frameworks. Precisely what these are could be radically different in different contexts. (2013, p. 2)

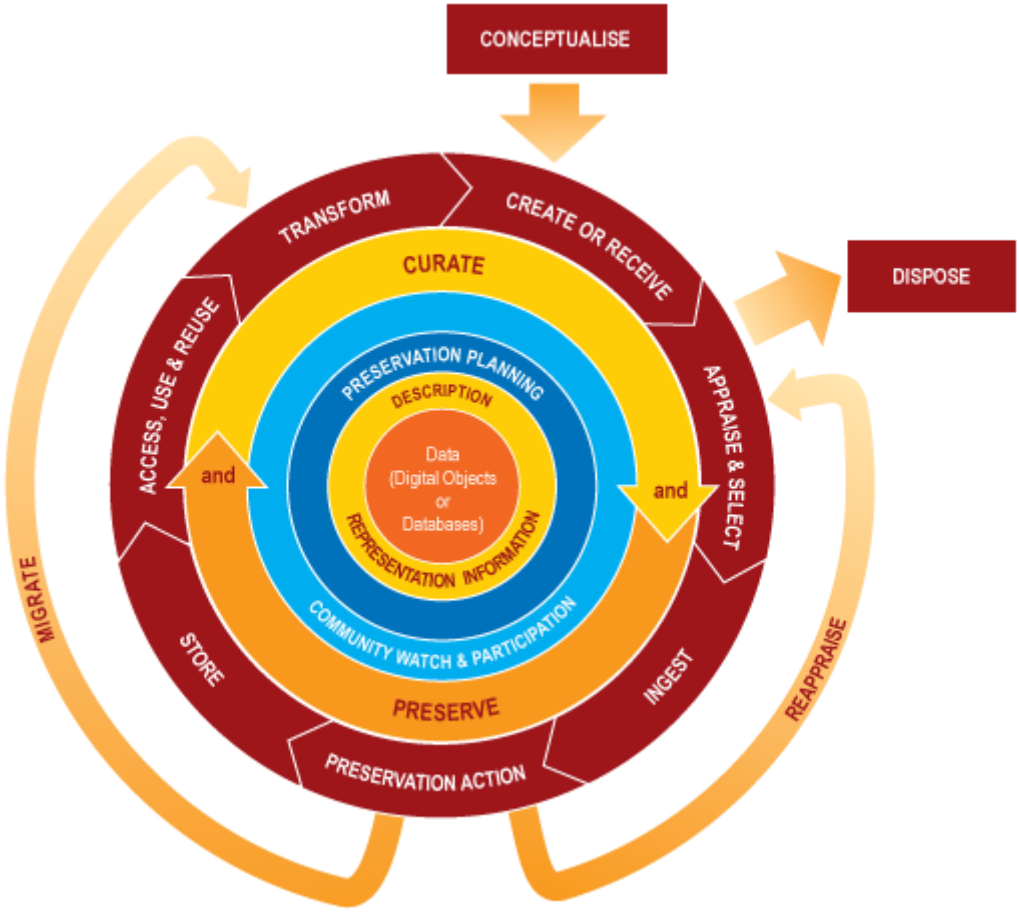


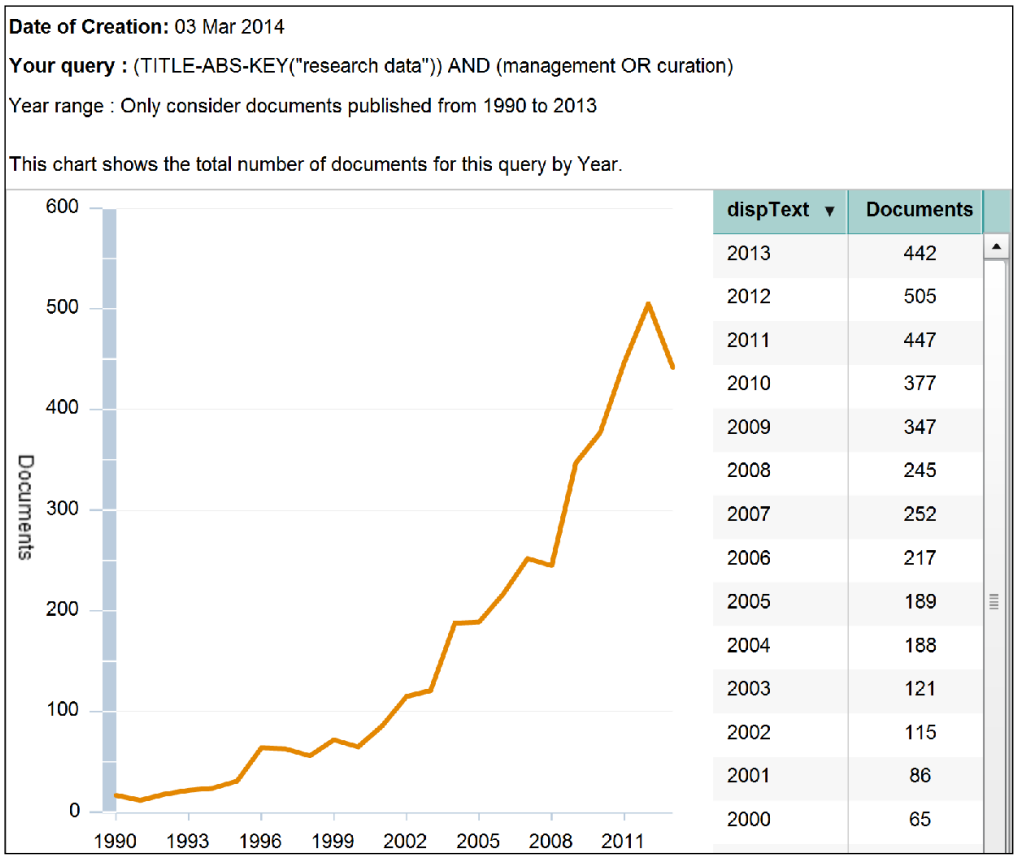
Figure 1: The DCC Curation Lifecycle Model. From http://www.dcc.ac.uk/resources/curation-lifecycle-model

1.2 Scope

Traditionally focusing on supporting the curricula and research at universities, the role of academic libraries in managing research data is regularly emphasised. Moreover, RDM is a 'trending' topic, as figure 2. shows, and as Cox and Pinfield (2013) pointed out: "amongst library and information professionals, there now seems to be a significant amount of positive hype associated with RDM." (2013, p. 15)

But it's not just information professionals: digital preservation of research data is "sliding into the trough" according to the 2014 Gartner Hype Cycle for Education ("Hype Cycle for Education, 2014," n.d.), RDM is described as a "fast trend" in the Library Edition of the New Media Consortium's 2014 Horizon Report (Johnson, Adams Becker, Estrada, &

Freeman, 2014), and even US President Obama "highlighted data management as a critical new job skill for the 21st century." (M. A. Parsons & Fox, 2013, p. WDS33)



dispText	Documents
2013	442
2012	505
2011	447
2010	377
2009	347
2008	245
2007	252
2006	217
2005	189
2004	188
2003	121
2002	115
2001	86
2000	65

Figure 2: Publications in RDM by year. Source: Scopus.

However, as figure 3 suggests, there is be a gap between countries which are engaged in the issue of research data management and which are the subject of this thesis – although this figure can be somewhat distorted because of the journal selection criteria of Scopus.

The aim of my thesis is to discover the capacity of academic libraries and the possible roles of librarians in RDM. In order to achieve this, it will concentrate mainly on communities with no or not very extensive preconceptions about the topic, thus hoping to shed light on traits that could have well stayed hidden by conducting a study within a milieu where policies and practices about the handling of research data are already in place and commonly agreed upon.

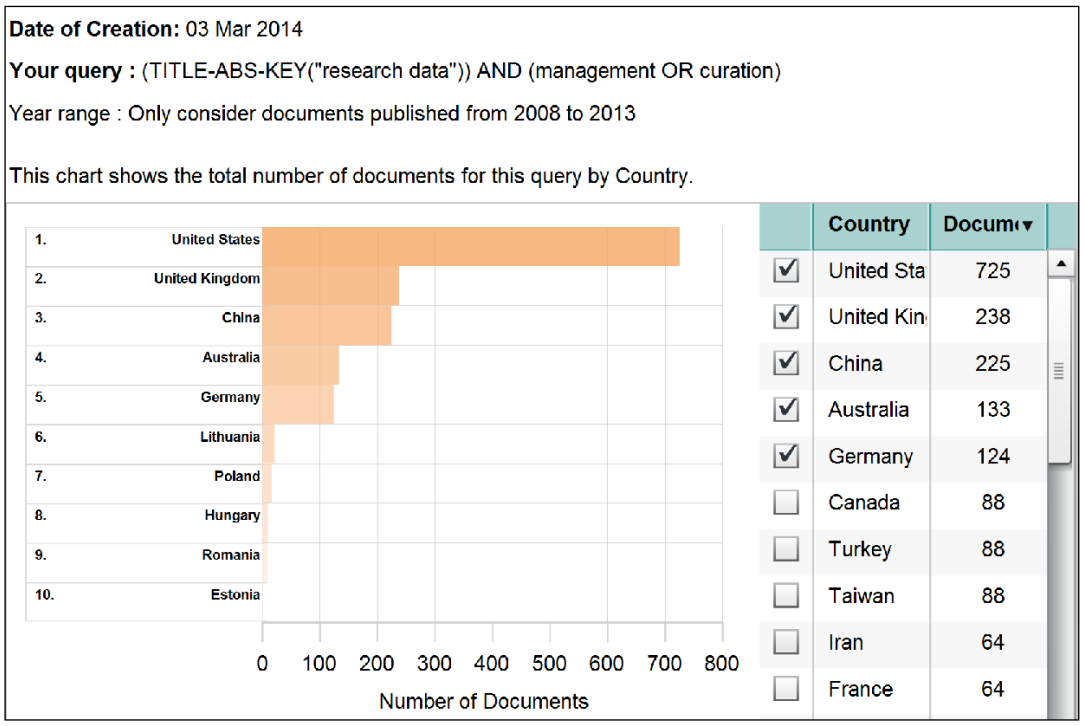


Figure 3: Distribution of publications in RDM between 2008 and 2013 by country. Source: Scopus.

This thesis ultimately aims to answer the following research questions:

- Are academic libraries ready to implement services on research data management?
- How well would research data management fit into the current set of research support services provided by academic libraries?
- In which areas of research data management do librarians see themselves concentrating in the future?
- What are the perceived constrains by academic libraries to introduce RDM services?

1.3 Limitations

As it will be evident from the following chapters, research data management is a complex issue. As there are no homogeneous group of stakeholders, no institutions with the same profiles, no academic libraries with the same community they serve, there are no set of proven solutions to the problem. Thus, the study required a thoroughly developed conceptual framework even to find out what to look for. Moreover, the participants' views, ideas, and thoughts about how their institution can/should tackle the issue of RDM do not automatically mean that their institutions will implement the desired services.

The study was not aiming for representativeness, but rather for an illustrative image on the current practices and future priorities on the issue of research data management in Estonian and Hungarian academic libraries.

1.4 Significance

Sooner, rather than later, effective data management practices will be inevitably required in a growing number of European countries. As research gets more global and more collaborative, localised practices for conducting and funding research will increasingly complemented with international and global engagements, and it is almost certain that the number of funding agencies who require researchers to share their produced data will grow by time. We are currently witnessing this trend in the EU: Horizon2020, the current framework programme for Research and Innovation has a pilot on open research data, and it can be taken for granted that researchers from Estonia and Hungary will participate in these projects.

Moreover, the main research and research funding institutions of these two countries (the Estonian RC, and OTKA and MTA from Hungary) are members of Science Europe, an association of 52 research funding and research performing organisations from 27 countries (see "Science Europe - Member Organisations," n.d.), which recently published its roadmap that includes the following statement: "Science Europe Member Organisations acknowledge that open data should be the standard." (Science Europe, 2013, p. 10)

This thesis adds to the existing literature by providing insights from countries that have not been included in previous studies on the relationship of academic libraries and research data management. This need for broadening the horizon by investigating different communities has been articulated by several previous publications, such as Cox and Pinfield (2013), who called for further work "to capture a sense of the pattern in different countries" in the changing context of research data management and its implications to academic libraries. (2013, p. 16)

2. LITERATURE REVIEW

This chapter provides an overview of the current theoretical, legal, and practical framework surrounding research data management, thereby laying the foundations of my thesis. The chapter consists of two main parts: the first part deals with the complexities of sharing and managing research data, while the second part concentrates on the ways academic libraries can provide research data management services for their user community.

To obtain the most comprehensive overview on the existing literature about research data management, searches have been carried out in various scientific databases. Search terms included "research data", "data management", "data curation", and "academic libraries" without any particular time frame. Once the first set of results have been collected, snowballing, or forward- and backward chaining methods, i.e. following up the references in each article, and finding the citing articles in databases such as Scopus, Web of Science, and Google Scholar have been used to find further relevant articles which were not present in the initial set of results. Additionally, to obtain the most up-to-date information, current-awareness services such as RSS-feeds of related web pages and blogs, secondary sources, and mailing lists of relevant communities have been skimmed frequently.

The scope of the literature review was limited to resources written in English, and to some extent German and Hungarian language, and to sources which were accessible either through the Learning Centre and Library of the Oslo and Akershus University College, the Academic Library of Tallinn University, the National Library of Estonia, or openly available on the web.

2.1 Data intensive science

The "revolutionary changes thanks to digital technology" (S. Carlson, 2006) are transforming the ways of how research is carried out: these changes are so ubiquitous that even terms describing them, such as "data deluge", or "data revolution" became clichés. (Cronin, 2013, p. 435)

The shifting research practices are often referred as the Fourth Paradigm of science (Gray et al., 2005; Hey, Tansley, & Tolle, 2009b), the result of the transition from hands-on,

empirical enquiry, through the use of theoretical models and computational simulations to "e-science", a state where data are actually captured by computers and scientists analyse the resulting products, such as databases and files. (Hey, Tansley, & Tolle, 2009a) It is the "very size and complexity" of the produced data which "enable analysis at unprecedented levels of accuracy and sophistication and provide novel insights through innovative information integration." (National Science Board, 2005, p. 9) Therefore, as Borgman (2007) pointed out, this new type of inquiry is valuable for the whole academic ecosystem: researchers in the sciences are able to "construct better models of physical phenomena", in the social sciences to build "more realistic models of complex social phenomena", and in the humanities to "gain insight into literature and historic texts with new tools to study, explore and compare data". (Borgman, 2007, p. 117)

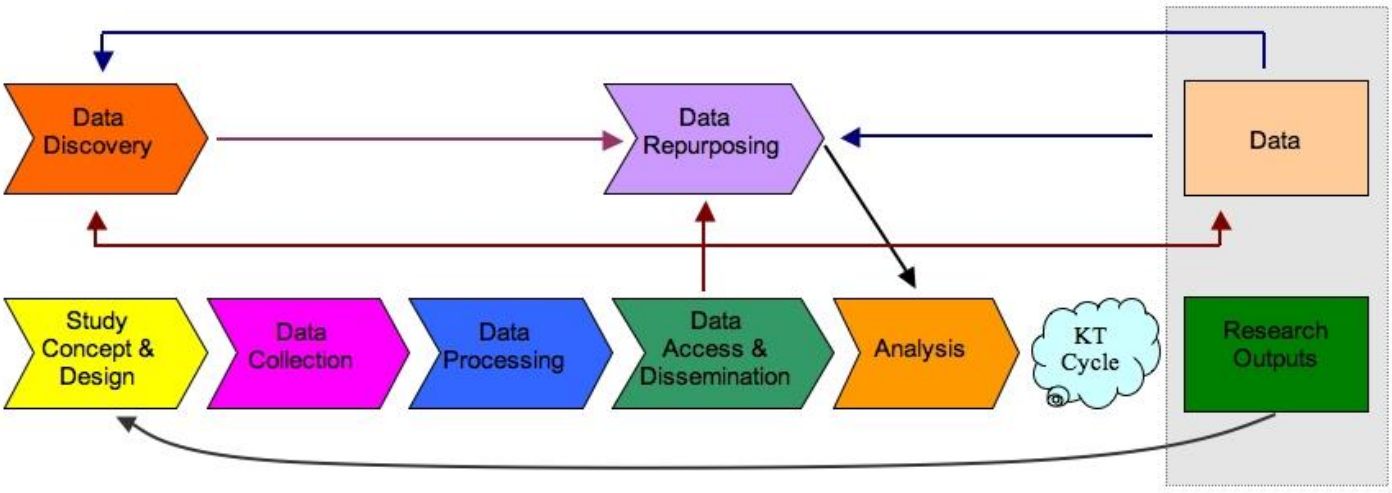


Figure 4: The research life cycle. From Friedlander and Adler (2006, p. 18).

These changes not only transform the research process, which was divided into three basic activities by Bell (2009): "capture, curation, and analysis" (2009, p. XIII), but as a result of the changing practices, they are transforming the dissemination models of the research results as well, since "no journal could conceivably publish data in the same way as before." (Royal Society, 2012, p. 26)

It has been foreseen that the scholarly publishing scene will have to change to keep up with the pace which scientists are producing data, as Hey (2010) predicted, journal papers that discuss an experiment and its findings and just refer to data sets will eventually morph "into a wrapper for the data themselves, which other researchers will be able to access

directly over the internet, probe with their own questions, or even mash into their own data sets in creative ways that yield insights the first researcher might never have dreamed of." (2010, p. 63)

	Discovery	Analysis	Writing	Publication	Outreach	Assessment
Trends	social discovery tools	datadriven & crowdsourced science	collaborative online writing	Open Access & data publication	scholarly social media	article level (alt)metrics
Expectations	growing importance of data discovery	more online analysis tools	more integration with publication & assessment tools	more use of "publish first, judge later"	use of altmetrics for monitoring outreach	more open and post-publication peer review
Uncertainties	support for full-text search and text mining	willingness to share in analysis phase	acceptance of collaborative online writing	effect of journal/publisher status	requirements of funders & institutions	who pays for costly qualitative assessment?
Opportunities	discovery based on aggregated OA full text	open labnotes	semantic tagging while writing/citing	reader-side paper formatting	using repositories for institutional visibility	using author-, publication- and affiliation-IDs
Challenges	real semantic search (concepts & relations)	reproducibility	safety/privacy of online writing	globalization of publishing/access standards	making outreach a two-way discussion	quality of measuring tools
Most important long-term development	multidisciplinary + citation-enhanced databases	collaboration + data-driven	online writing platforms	Open Access	more & better connected researcher profiles	importance of societal relevance + non-publication contributions
Potentially most disruptive development	semantic/concept search + contextual/social recommendations	open science	collaborative writing + integration with publishing	circumventing traditional publishers	public access to research findings, also for agenda setting	moving away from simple quantitative indicators

Figure 5: Most important developments in 6 research workflow phases. (Kramer & Bosman, 2015)

Today, we are the witnesses of these predictions: *Nature Publishing Group* recently launched an online-only data journal ("Welcome, Scientific Data!," 2014), Thompson Reuters opened its Data Citation Index, and online data repositories are proliferating: in August, 2013, the searchable catalogue Databib listed 594 websites (Van Noorden, 2013a, p. 244), while at the time of writing, the same database (after merging with re3data.org) just exceeded the 1000 repositories mark (re3data.org team, 2014). These databases vary from small, subject specific repositories through medium scale, national initiatives to very large, commercial agents such as Dryad, Zenodo, or Figshare.

2.2 Managing research data

It is widely recognised that "digital research data of long term value arising from current and future research should be preserved and remain accessible for current and future generations" (Research Information Network, 2008, p. 3); nevertheless, achieving this

requires a non-trivial, extensive, and comprehensive approach, which Parsons and Fox (2013) identified as follows:

- Established trust (of data, systems, and people).
- Data are discoverable.
- Data are preserved.
- Data are ethically open and readily accessible to humans and machines.
- Data are usable, including some level of understandability.
- Effective, distributed governance of the data system.
- Reasonable credit and accountability for data collection, creation, and curation. (M. A. Parsons & Fox, 2013, p. wds36)

The management, organization, access, and preservation of digital data has been understandably named as a "grand challenge" of the information age (Berman, 2008, p. 50), and it is widely acknowledged that "leaving digitally based information to languish in personal electronic filing drawers amid a jumble of unrelated information and with no plans for its survival guarantees its disappearance." (Ogburn, 2010, p. 242)

One of the many cases of this disappearance is very well illustrated by Uhlir (2010):

> Take the case of NASA, the National Aeronautics and Space Administration in the United States. That agency is generally regarded as a paragon of technological achievement. It successfully sent man to the moon at the dawn of the digital age. It developed robots to explore the heavens and invented all manner of gadgets. So, of course, NASA has preserved the record of all its great achievements. Wrong. It turns out that the agency did not preserve the data from its very first mission, Explorer 1. It does not have many of the original tapes from the human exploration of the moon. It has lost much of the early Landsat data, which otherwise form such a valuable longitudinal record of our planet's environmental changes. The list goes on. (Uhlir, 2010, p. ES4)

Implementing the abovementioned comprehensive approach to research data "involves many actors, supporting technologies and organisation, including coordination of human and financial resources." (Whyte & Allard, 2014, p. 2) Figure 6 shows a possible illustration of these actors, representing various institutions from commercial services to libraries and archives, the different levels of the value of the data that these institutions are dealing with, and other factors of trust, risks, and responsibilities.

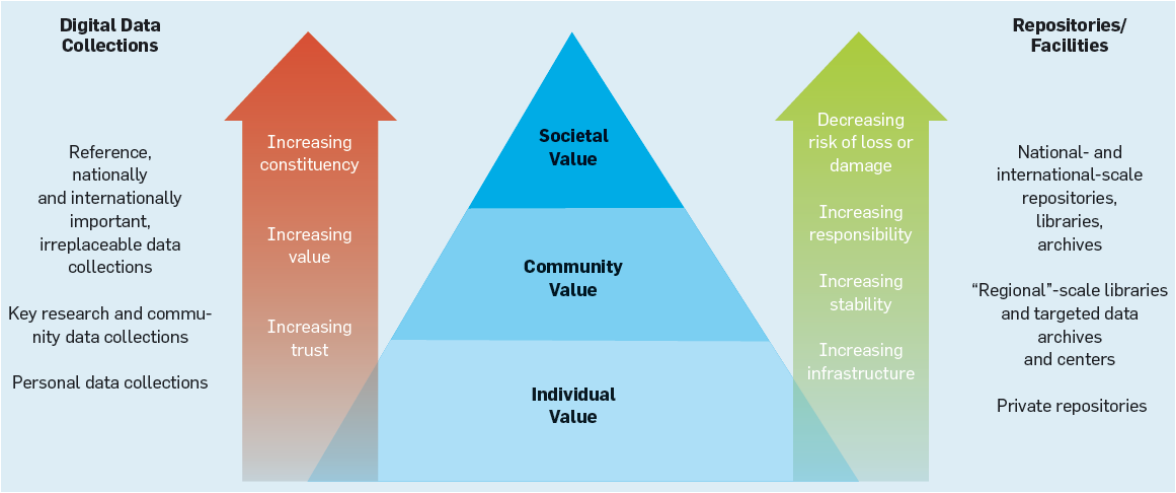


Figure 6: The data pyramid from Berman (2008, p. 53)

However, when it comes to practice, approaches vary greatly: a survey completed by more than 1300 scientists, Tenopir et al. (2011) found hugely mixed approaches to RDM. Although it reported "some satisfaction with tools for data collection and analysis", happiness was lower with "tools for metadata creation and preservation." The study found that "most scientists do not believe their organization is doing a sufficient job in helping them achieve long-term data preservation and many researchers are not currently using international metadata standards." It called the lack of awareness about the importance of metadata "a serious problem as their involvement is quite crucial in dealing with problems regarding data management." (2011, pp. 19–20)

2.3 Sharing research data

The importance of effective data management, especially providing sufficient metadata in standardised format also tends to support a further essential issue, which is not generally present in the ongoing discourse: data mining. As one interviewee pointed out in *Nature*: "a lot of people ask, 'Who reads all these data?' The point is that machines use them – they can search the data" (Peplow, 2014, p. 22).

But not *only* machines: the importance of data sharing has been around for many decades, and some disciplines, like genomics (Smith & Carrano, 1996), or the neuroscience community (Koslow, 2000) have been leading the way to establish standards and practices to share and manage research data effectively; or consider the example of archaeologists: they clearly don't want to dig out an excavation site again once it has been buried.

One of the early spokespersons of the issue was Sieber, who, already at the end of the eighties formulated several questions that are still present in the current discussion about data sharing. (These concerns will be further elaborated in section 2.4.4) Although the ways for sharing research was somewhat different than it is today: the cover of a book on research data (Fienberg, Martin, & Straf, 1985) pictured two hands exchanging a 5¼-inch floppy disc in a position resembling Michelangelo's painting *The Creation of Adam*.

Digital data sets provide "tremendous reuse opportunities" (Lynch, 2008, p. 28), which can take many forms, from depositing them to repositories, through using them as supplementary material to journal articles, to only sharing them upon request. The issue got to the centre of attention due to the previously mentioned technological developments which produced the "data deluge": as Borgman (2012) pointed out: "if the rewards of the data deluge are to be reaped, then researchers who produce those data must share them […]" (Borgman, 2012, p. 1059). However, sharing data is not as straightforward as it first appears, as she continues:

> […] and do so in such a way that the data are interpretable and reusable by others. Underlying this simple statement are thick layers of complexity about the nature of data, research, innovation, and scholarship, incentives and rewards, economics and intellectual property, and public policy. (2012, p. 1059)

Given that the "mere disclosure of data has very little value *per se*" (Royal Society, 2012, p. 14 emphasis in original) Kowalczyk & Shankar (2011) argued that representing context is just as essential as the data themselves. (2011, p. 253) They also formulated three criteria for effective data sharing: "persistence, longevity and sustainability, and quality" (p. 248). A report by the Royal Society (2012) called for a "more intelligent openness" to facilitate the effective communication of data and set out four requirements to achieve this, stating that data to be shared should be:

a. Accessible.
b. Intelligible.
c. Assessable.
d. Usable (Royal Society, 2012, pp. 14-15)

The results of the survey by Tenopir et al. (2011) has shown that scientists are generally open towards the idea of data reuse, but there is a "major gap between desire and current possibility": more than 80% said that they would use other researchers' data sets if they

were easy to reach, but only 36% said that others could access their data easily. (2011, p. 9)

Noteworthy current activities on data sharing are initiatives working towards finding a straightforward solution to the citation of data sets. DataCite ("DataCite Strategy 2013-2016," 2014), Force11, (Data Citation Synthesis Group, n.d.) and CODATA-ICSTI (CODATA-ICSTI Task Group on Data Citation Standards and Practices, 2013) are all trying to formulate standards to help the citation of data.

2.4 Drivers and barriers of managing and sharing research data

The driving forces to effective data management and sharing are coming from various stakeholders, such as funders, research institutions, and journal publishers. The reasons vary, some being theoretical, labelled as "good practice", others purely pragmatic, as the example of International Polar Year shows: it had a policy to make data openly available in the shortest time, because of "the rapidness of change in the poles". Waiting for years to release the data would have meant the Arctic "to be a completely different place." (Nelson, 2009, p. 161) Moreover, actors such as funding agencies are "not calling for data curation, preservation, and sharing because it is abstractly the right thing to do as part of the creation, dissemination, and stewardship of knowledge" (Lynch, 2013, p. 396) Public funders are "under great pressure to show how their funding contributes to broad economic growth, how it addresses the needs of society, and to demonstrate that the requirements that they impose on the work they fund makes discovery ever more rapid, extensive, and cost-effective." (2013, p. 397) This subsection seeks to present some of the arguments for, and the concerns about research data management and sharing.

2.4.1 Openness, the ethos of science

Openness has been traditionally a fundamental norm of science, as it enables scientists to "identify errors, to support, reject or refine theories and to reuse data for further understanding and knowledge." (Royal Society, 2012, p. 7) In an essay originally from 1962, Polanyi used a jigsaw puzzle metaphor to illustrate this phenomenon of openness: he compared the realm of science to a puzzle, on which scientists are continuously working, aiming to solve it. As he pointed out, if one would distribute the pieces of the puzzle equally amongst researchers and then let them work separately on their part, the process of

solving the game "would be totally ineffectual, […] since few of the pieces allocated to one particular assistant would be found to fit together." (Polanyi, 2000, p. 2) To overcome this problem, he described a collaborative system, where people work on the puzzle is sight of others, "by responding to the latest achievements of the others, and the completion of their joint task will be greatly accelerated." (2000, p. 3)

The principle of openness and sharing appears in the *Mertonian* norms of science as well: Merton (1973) listed communism (which later transformed to 'communalism' as euphemism) an integral element of the scientific ethos, and denoted "scientific knowledge as common property". (1973, p. 278)

In the last decade, the scene of scholarly communication took major steps in the direction of greater openness: the open access movement, which started off more than a decade ago with the Budapest Open Access Initiative ("Budapest Open Access Initiative | Read the Budapest Open Access Initiative," 2002) received considerable amount of attention and achieved some remarkable results, just to name one: a report for the European Commission concluded that more than half of the papers published in 2011 were free to read in 2013. (Van Noorden, 2013b)

While there is no room in this work to elaborate in detail the many issues, benefits, concerns, and opportunities surrounding OA publishing, it is worth mentioning how well the past and current state of scholarly communication and the steps taken in the direction of OA are fitting into Polanyi's aforementioned puzzle metaphor. The latest developments towards open access publishing in journals and repositories align with his view of accelerating the research process by enabling researchers to access and respond to the latest achievements of others in an easy, cost-effective, and timely manner. Nevertheless, following the analogy, this current state of publishing would mean that people working on the puzzle could only respond to the announcements of others – while reusing research data would imply that not only are the announcements of latest achievements open to others, but people can see the *actual fitting pieces* of the puzzle as well.

This need for not only making the articles openly available, but to also do the same with the underlying materials has been articulated a year after the BOAI: the Berlin Declaration noted that "open access contributions include original scientific research results, raw data

and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material." ("Max Planck Open Access | Berlin Declaration," 2003)

2.4.2 The benefits of managing and sharing research data

Throughout the years, many advantages of sharing and managing research data have been articulated, and as Borgman (2007) pointed out, most of these arguments "tend to reiterate the principles of open science". (2007, p. 192; see Hitchcock, 2013 for a bibliography about the benefits of open access publishing) These arguments generally are: to boost citation rates (Piwowar, Day, & Fridsma, 2007; Piwowar & Vision, 2013), reduce or prevent fraud (Doorn, Dillo, & Van Horik, 2013), help reproduction (Peng, 2011), make publicly funded research available to the public, enable others to ask new questions of the data, advance the state of science (Royal Society, 2012, pp. 8-11), and even to enhance the research institutes' or the universities' reputation, as well as the researchers' own standing (Hodson & Jones, 2013)

Even though the curation of research data is sometimes viewed as a means to support science rather than the end (Royal Society, 2012, p. 10), some have pointed out that by taking actions to promote the effective management of research data, scientists will become more conscious of their own data handling, thus further advancing their research practices.

According to Lewis (2010), the rewards of managing research data include significant potential benefits for academic research itself:

- The ability to share research data, minimising the need to repeat work in the laboratory, field or library
- Ensuring that research data gathered at considerable cost is not lost or inadvertently destroyed
- The retrieval, comparison and co-analysis of data from multiple sources can lead to powerful new insights
- The ability to check or repeat experiments and verify findings, particularly important amid growing national and international concern about research integrity
- New research themes – and in particular cross-disciplinary themes – can emerge from re-analysis of existing data or comparisons with new data: increasingly data may become the starting point for new research as well as representing an output from current research. (Lewis, 2010, p. 148)

2.4.3 Mandates and recommendations from funding agencies, political bodies, and journal publishers

In 2008, Ruusalepp analysed research data policies of OECD countries, mainly focusing on the UK, US, Australia, Canada, and Germany, but included other several other countries such as Denmark, Finland, France, Greece, Japan, Netherlands, Norway, Spain, and Sweden. The study found "no evidence of either a universal model or agreement on what a data sharing policy should include" (2008, p. 84), "no national level policies or strategic documents that explicitly mandate the sharing of research data" (2008, p. 85), that "institutional policies still remain ad hoc and do not appear to be well coordinated" (2008, p. 86), and that institutions will require significant help and guidance "to develop uniform data sharing policies and put them into practice". (2008, p. 86)

Since the publication of the report, numerous initiatives emerged to promote data curation and sharing, most notably in the Anglo-Saxon world. While there is no space in this subsection to review the full spectrum of these initiatives thoroughly, the following part provides an overview of the current state, and the notable milestones of the international policies about RDM.

Albeit the many advantages pointed out in the previous section, the most vital initiatives of increasing the access to research data almost certainly came in the form of mandates and requirements of various stakeholders, such as funding agencies, journal publishers, learned societies, governments, and other political organisations.

In the US, for example, the National Science Foundation (NSF) has been described as the "epicenter of influence" (Gold, 2010, p. 5), and rightly so: it has been actively promoting the access to research data as part of the development of scientific cyberinfrastructure for years, primarily as a result of the seminal 'Atkins report' (Atkins et al., 2003); but the greatest push was the requirement of submitting data management plans to grant proposals, which is active since the beginning of 2011 ("NSF Data Management Plan Requirements," n.d.). A further boost was the White House's Office of Science and Technology Policy memorandum, which stressed that "citizens deserve easy access to the results of scientific research their tax dollars have paid for." ("Expanding Public Access to the Results of Federally Funded Research | The White House," 2013)

The *public money – public access* argument is also present in the United Kingdom: RCUK, the partnership of the UK's seven research councils stressed in its common principles that

> Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property. ("RCUK Common Principles on Data Policy - RCUK," n.d.)

Even though it has been proven that mandates and requirements generally improve the access to research data (Vines et al., 2013), one must keep in mind that "policy is the easy bit – resourcing, implementation and enforcement is more difficult" (Ruusalepp, 2008, p. 87).

Nonetheless, the implementation 'bit' can be seriously accelerated with enforcement: engagement (which most of all materialised in designing roadmaps of compliance and implementation) in RDM at universities in the UK started primarily thanks to the Engineering and Physical Research Councils' (EPSRC) policy framework on research data (Engineering and Physical Science Research Council, 2013a), which was further divided into a set of seven core principles (Engineering and Physical Science Research Council, 2013c) and nine expectations (Engineering and Physical Science Research Council, 2013b). What made the policy unique was its response to non-compliance: "a potential threat to future funding, with a statement made to all university heads that institutions found to be seriously failing to comply could find themselves declared ineligible for further EPSRC support." (Pryor, 2013, p. 185)

As a result, the EPSRC framework "provided a raison d'être for the creation of policy and the introduction of support infrastructure" (Pryor, 2013, p. 185): between 2011 and 2013, JISC's Managing Research Data programme ran "a set of 17 projects to pilot research data management services in universities" while the DCC "has also undertaken a series of 21 institutional engagement projects providing tailored support to increase research data management capability." (Hodson & Jones, 2013)

Furthermore, as an addition to policies from funding agencies, journal publishers also began to require supplementing materials, such as research data to be archived and shared without restrictions. The most notable example is PLOS's policy on research data that requires authors who wish to publish their papers in PLOS journals to "make all data

underlying the findings described in their manuscript fully available without restriction, with rare exception", strongly recommending to deposit the data in public repositories. ("PLOS ONE : accelerating the publication of peer-reviewed science," n.d.) This policy from PLOS (see e.g. Ganley, 2014 for further details) triggered some strong opposing reactions in the scholarly community, which will be further discussed in section 2.4.4.

International and trans-national initiatives

As "the ease with which data can move across national boundaries" (Kowalczyk & Shankar, 2011, p. 283) increased, the need emerged to regulate the access to research data on an international level as well. *Science* magazine called for an "international framework to promote access to data" in as early as 2004 (Arzberger et al., 2004), while OECD picked up the issue back in 2007, publishing its Principles and Guidelines on research data. (*OECD Principles and Guidelines for Access to Research Data from Public Funding*, 2007) Even the G8 expressed that they are "committed to openness in scientific research data to speed up the progress of scientific discovery, create innovation, ensure that the results of scientific research are as widely available as practical, enable transparency in science and engage the public in the scientific process." ("G8 Science Ministers Statement - News stories - GOV.UK," 2013)

On the European level, a report for the European Commission, *Riding the wave: How Europe can gain from the rising tide of scientific data* called for the development of an international framework for a collaborative data infrastructure and recommended that the EC should earmark additional funds for scientific e-infrastructure, develop and use new ways to measure data value, reward those who contribute it, train a new generation of data scientists, and broaden public understanding of research data. (High level Expert Group on Scientific Data, 2010)

As a reaction to the Riding the Wave report, *Surfboard for riding the wave* (van der Graaf & Waaijers, 2011) presented an overview of the situation about research data in Denmark, Germany, the Netherlands and the United Kingdom and offered "broad outlines for a possible action programme for the four countries in realising the envisaged collaborative data infrastructure." (2011, p. 4) It identified four main areas of incentives for researchers to share their datasets: "re-use and recognition, principles of science, reflected in rules and codes of conduct, requirements by funding organisations and journal data availability

policies", and stressed the need for new skills, labelling the new types of actors as "data scientists" and "data librarians". (2011, p. 4). The report also pointed out that in the system for reward and recognition, research data does not have an adequate place, as rewards are "still mainly based on publishing in high-quality journals and/or citation metrics of their articles". (2011, p. 10)

The following year saw the European Commission to publish its recommendation on the access and preservation of scientific information, which stated that

> Policies on open access to scientific research results should apply to all research that receives public funds. Such policies are expected to improve conditions for conducting research by reducing duplication of efforts and by minimising the time spent searching for information and accessing it. This will speed up scientific progress and make it easier to cooperate across and beyond the EU. ("Commission Recommendation on access to and preservation of scientific information," 2012, p. 3)

The part on research data expressed the need to define clear policies for dissemination and access, thus ensuring that

- research data that result from publicly funded research become publicly accessible, usable and re-usable through digital e-infrastructures
- datasets are made easily identifiable and can be linked to other datasets and publications through appropriate mechanisms, and additional information is provided to enable their proper evaluation and use
- institutions responsible for managing public research funding and academic institutions that are publicly funded assist in implementing national policy by putting in place mechanisms enabling and rewarding the sharing of research data
- advanced-degree programmes of new professional profiles in the area of data-handling technologies are promoted and/or implemented (2012, pp. 6–7)

These recommendations are clearly visible in the new European Framework Programme for research and innovation: Horizon 2020 has areas that participate in a pilot on open research data (see "Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020," 2013, p. 9). The guidelines lay out that scientific research data should be easily

1. Discoverable
2. Accessible
3. Assessable and intelligible
4. Useable beyond the original purpose for which it was collected
5. Interoperable to specific quality standards ("Guidelines on Data Management in Horizon 2020," 2013, p. 6 Note the similarities with Royal Society, 2012, pp. 14-15)

In order to achieve this, submissions in the participating areas are required to include DMPs, "detailing what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved." Beneficiaries must deposit the data in a repository, "and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate – free of charge for any user" ("Guidelines on Data Management in Horizon 2020," 2013, p. 3)

2.4.4 Barriers of managing and sharing research data

Sharing data is not as easy and straightforward as it first turns out: Borgman called it "conundrum" (Borgman, 2012), an editorial in *Nature Genetics* "misnomer", for "it is usually insufficient simply to make one's results available to the greater community." ("Integrating with integrity," 2010, p. 1) As the broad and countless definitions of data suggests, "their status as fact or evidence is determined by the people who produce, manage, and use these data" (Borgman, 2007, p. 121) by their very nature. Consequently, the methods to share these data are as many and as diverse as the disciplines that use these methods, and are "driven by the myriad ways in which researchers need to subsequently access and exploit the information they contain." (Ruusalepp, 2008, p. 2)

Moreover, "not all data are fit for use by others. Each field's experts recognize and use a range of quality measures and caveats that may be more difficult to adapt for other applications than the data themselves." ("Integrating with integrity," 2010, p. 1) As Sieber (1988) pointed out more than 25 years ago, "to examine and extend the work of others, scientists may need access not only to the raw data on which that work is based, but also to actual samples, and to the original research techniques and procedures including »know-how«, software, and other materials and devices." (1988, p. 200) This problem of the "know-how" has been also discussed by Borgman (2007), who labelled *tacit knowledge* as "the greatest barrier" to effective data sharing, as much of the knowledge which is necessary to interpret data "often remains undocumented" (2007, p. 165), sometimes simply because scientists lack "the skills or resources needed to prepare all their data for public sharing". (Cragin, Palmer, Carlson, & Witt, 2010, p. 4035; see also Fleischer & Jannaschk, 2011, p. 575)

Complaints against data sharing are probably just as old as the idea itself: debates on ownership, whether research data are private property or public good data back earlier than

the 80s (see e.g. Dickson, 1980). As Sieber (1991) pointed out, “in reality, openness and its advantages have been ideals, not norms.” (1991, p. 69) She claimed that the concerns about sharing research data are generally in place because scientists are unconcerned, uninformed, and unfamiliar with the idea of sharing. (1991, p. 78) Apart from uninformed concerns, she mentioned two other types: naïve and informed; in the former category she included the questions of credit, reward, costs of documentation and duplication, control, property rights and ownership, confidentiality, timeliness, and competitive advantage; while the latter category contained concerns about the forms of data, correct citation, documentation, or user friendly form. (Sieber, 1991, pp. 81–83)

Further problems are disciplinary differences. Parry & Mauthner (2004) pointed out that while methods for data archiving and sharing is “relatively unproblematic by the quantitative research community, there has been a mixed reaction to data archiving among qualitative researchers.” (2004, p. 140) Some even claim that the whole fundament of the idea is essentially distorted, as using high-energy particle physics or astronomy for reference and comparison “obscures the complexities of production and communication inherent in small science and across subdisciplines”. (Cragin et al., 2010, p. 4025)

Disregarding disciplinary differences or favouring certain fields over others, and then extending these practices to fields which don’t necessary operate with the same kind of data will unsurprisingly result in general, all-purpose models of service, which “does not generally account for research that functions at a more community level” (Cragin et al., 2010, p. 4026). Without doubt, “there is no single infrastructure that will serve all of the varying needs of the scientific community.” (Kowalczyk & Shankar, 2011, pp. 249–250)

Moving further, common reasons for not sharing research data are either insufficient time and lack of funding (Tenopir et al., 2011), legal issues, the fear of being misinterpreted, misappropriated, or the “disregard of good faith practices” (Cragin et al., 2010, p. 4034; see Royal Society, 2012; and Thaesis & van der Hoeven, 2010 as well), and even losing economic advantages in certain fields like chemistry or pharmacology, where “data have high monetary value much of the research is privately funded”. (Royal Society, 2012, 10)

But data can not only mean financial advantage: some researchers and organisations consider themselves “to be either friendly competitors or rivals for funding, leading them

to view their data as a source of competitive advantage" (Kowalczyk & Shankar, 2011, p. 248). As it has been pointed out earlier, a shared and/or cited dataset is not widely rewarded or recognised in the eyes of committees, employers, funding agencies, research administrations, and sometimes not even in the research community. (McGlynn, 2014; Nelson, 2009, p. 162) This state of the scholarly reward-system which lacks or only starts to implement data-citation metrics, or doesn't offer sufficient rewards for published datasets leads to a situation where sharing datasets with others are commonly rewarded with co-authorship instead of proper citation of the datasets. (Cragin et al., 2010, p. 4031)

Managing data is clearly requires significant amount of time and effort from researchers (Borgman, 2012, p. 1072; Goodman et al., 2014, p. 1; Nelson, 2009, p. 160; Van Noorden, 2013a, p. 244), which is often viewed as a burden, since most of these efforts "spent on documenting data for use by others are resources not spent in data collection, analysis, equipment, publication fees, conference travel, writing papers and proposals, or other research necessities." (Borgman, 2012, p. 1073)

Research data requirements from funding agencies and publishers are not exceptions from the burdens (Lynch, 2013, p. 397): some even called this initiative-rich last few years a "decade of nagging and annoyance" (Fleischer & Jannaschk, 2011, p. 576). And even though Nelson (2009) warned that prematurely forcing a sharing requirement on researchers "would be suicidal" (2009, p. 162), the year 2014 bought up some rather surprising events.

On the spring of 2014, the National Science Board commissioned a report to the NSF, which stated that "US scientists are spending 42% of their time on bureaucratic chores". ("Time wasted," 2014) Recommendations of the report included that the NSF should „reduce or *eliminate* the requirement for data management plans" (National Science Board, 2014, p. 45 italics added) and "could lead an effort to identify inconsistencies and guideline shortfalls in data sharing across organizations and agencies." (2014, p. 69)

Recent policy changes and recommendations triggered outrage in the scientific community, for example, following the introduction of the new data policy of PLOS (Silva, 2014; Bloom, 2014) even a #PLOSFail hashtag has been created on Twitter, and numerous blog post have been written about how unhappy scientists were with the changes

(see e.g. Strasser, 2014; Crotty, 2014). Some stated that they are considering not publishing any more in PLOS, stressing how much work data archiving requires, saying that "to get these numbers into a downloadable and understandable condition would be, frankly, an annoying pain […]" (McGlynn, 2014).

On the top of this all, sometimes even the applications used for data archiving fail, which does not help to establish the right levels of trust in cloud-based-technologies of third parties. Just as it happened when "Dedoose, a cloud-based application for managing research data, suffered a »devastating« technical failure […] that caused academics across the country to lose large amounts of research work, some of which may be gone for good." (Kolowich, 2014)

2.5 Data management support services in academic libraries

2.5.1 The academic libraries' changing role

Academic libraries, just as any other institutions, are situated in a broader socio-cultural environment, and "are significantly affected by developments in higher education and research economics, organization, communication technology, and the behavior of research communities." (Maceviciute, 2014, p. 297) These developments and changing conditions stimulate libraries to re-envision (Lougee et al., 2007), re-position (Lyon, 2012; Swan & Brown, 2008; Walters & Skinner, 2011; Ward, Freiman, Molloy, Jones, & Snow, 2011), re-profile, re-structure, re-engineer (Lyon, 2012), re-evaluate (Miller, 2012), re-purpose, re-tool (Salo, 2010), re-invent, and even re-boot (Monastersky, 2013) themselves, and librarians to re-examine (Hey & Hey, 2006) their competencies or re-skill (Auckland, 2012; Brewerton, 2012) themselves to keep up with the changing needs of their user community.

The reasons for all these *re-whatnots* are strikingly similar to the changes that took place in the sciences during the 1940s. According to Merton (1973):

> A tower of ivory becomes untenable when its walls are under prolonged assault. After a long period of relative security, during which the pursuit and diffusion of knowledge had risen to a leading place if indeed not the first rank in the scale of cultural values, scientists are compelled to vindicate the ways of science to man. (Merton, 1973, p. 268)

Just as "a frontal assault on the autonomy of science was required to convert this sanguine isolationism into realistic participation in the revolutionary conflict of cultures" (p. 268), so are *librarians compelled to vindicate the ways of librarianship to man*, with efforts to demonstrate their values by developing library infrastructures, changing the practices of acquisitions and the management of collections, delivering access and services anywhere and anytime, fostering open access, maintaining institutional repositories, reaching out for users, assessing user needs, evaluating services (Maceviciute, 2014), and by constantly adding new roles and responsibilities to this list.

Services supporting research data management and sharing seem to fit here perfectly. Based on the circumstances presented in the previous sections, it is apparent that researchers need support to be able to manage research data effectively, and librarians have been called to "step forward to define, categorize, and archive the voluminous and detailed streams of data generated in experiments." (S. Carlson, 2006) Indeed, support provided by academic libraries seems to have a positive effect on data sharing: as a study by Sayogo and Prado (2013) found, skills development and organisational support for research data management are the single most important factor for motivating researchers to publish their research data.

In the recent years, many have argued that academic libraries are in a good position to provide support services for research data management: data curation has been included in the 2012 list of top ten trends in academic and research libraries, where it has been stated that "librarians and information workers have a vital role to play in helping their research communities design and implement a plan for data description, efficient storage, management, and reuse." (ACRL Research Planning and Review Committee, 2012, p. 312) The issue has been included in the recent ACRL top ten trends of 2014 as well, which focused on the unifying theme of deeper collaboration: it still highlighted the libraries' "unique position" in the discovery, reuse, and curation of small and large datasets, but somewhat less enthusiastically acknowledged the fact that "funding organizations, academic institutions, researchers, and librarians continue to struggle towards a shared vocabulary with commonly understood definitions and to develop strategies to support these new initiatives." (ACRL Research Planning and Review Committee, 2014, p. 294)

This current hype around research data management amongst information professionals (Cox & Pinfield, 2013, p. 15) aligns with the ways academic libraries are trying to *re-something* themselves. However, Maceviciute argued that regarding library strategies for development and their influence on the environment, "it is not easy to distinguish between reactions to change and the pro-active behavior of libraries." (2014, p. 295)

Indeed: for example, the US National Institute of Health's National Library of Medicine was one of the earliest implementers of data management, "which in 1988 established the National Center for Biotechnology Information to manage its own collection of molecular biology databases, including the GenBank repository." (Nelson, 2009, p. 162) Moreover, the US National Science Foundation's mandate on submitting data management plans for grant proposals, which is active since 2011, has been recommended to the NSF by a report resulting from an ARL workshop held in 2006. (see Friedlander & Adler, 2006, p. 45)

Although it was proposed by the ARL, this NSF requirement was the trigger that actually "mobilized many research libraries to develop and offer resources and services more specifically dedicated to guiding faculty and students to meet this new condition" (Hswe & Holt, 2011, p. 11), and even the "fellowships in data curation, with the goal of expanding a workforce of expertise in this area, have also emerged since the NSF went public with its requirement" (Hswe, 2012, p. 116). However, this requirement of DMPs shifted the focus of the developments from a technical perspective to a rather theoretical, advisory role: as a needs-assessment study at the University of Houston found, "rather than physical storage capacity, researchers need assistance with funding agencies' data management requirements, the grant proposal process, finding campus data-related services, publication support, and targeted research assistance attendant to data management." (Peters & Dryden, 2011, p. 387)

Similarly to the NSF mandate, library engagement in data management in the UK has been a response to a great extent to the EPSRC principles (Engineering and Physical Science Research Council, 2013c) and expectations (Engineering and Physical Science Research Council, 2013b) on research data, as it has been discussed earlier in section 2.4.3.

2.5.2 Possible roles of libraries in research data management

Most people would agree that "ensuring that the complex output of the research enterprise is collected and is reusable by others is central to the ongoing mission of research libraries." (Gabridge, 2009, p. 20) As Heidorn (2011) stated, "almost all academic libraries have a mission statement that guides the library in acquiring and disseminating information to meet the goals of the institution, including education and research." (2011, p. 663)

But as it has been already argued in section 2.1, "the majority of the knowledge output is neither text nor paper, nor books nor journals." (Heidorn, 2011, p. 662) Nevertheless, most of the studies focusing on the academic libraries' roles in research data management point out the libraries' expertise in handling traditional documents and managing physical collections, in other words "organizing, disseminating, and preserving diverse sets of materials" (J. R. Carlson, 2013, p. 17). As the *Science as an Open Enterprise* report pointed out:

> The traditional role of the library has been as a repository of data, information and knowledge and a source of expertise in helping scholars access them. That role remains, but in a digital age, the processes and the skills that are required to fulfil the same function are fundamentally different. They should be those for a world in which science literature is online, all the data is online, where the two interoperate, and where scholars and researchers are supported to work efficiently in it. (Royal Society, 2012, p. 63)

However, there is a broad spectrum of opinions on whether this know-how of handling traditional documents would make libraries successful in managing digital datasets: from "an awkward fit" (Gold, 2007a) with "significant mismatches" that create a situation which is "neither straightforward nor simple; in some cases, it may even prove impossible" (Salo, 2010), or a "curse" (Heidorn, 2011, p. 663), through "problematic" (Nielsen & Hjørland, 2014) to "essential" (Gabridge, 2009), an "avenue open" (Maceviciute, 2014), calling the library a "key player" (Erway, 2013) having a "unique" (ACRL Research Planning and Review Committee, 2014), or a "particularly well" (Monastersky, 2013) position by this long-established expertise.

Some even warned that the current discourse on the librarians' possible role in data curation strongly resembles previous fads, such as knowledge management, and "there is a risk that a role in RDM will be more written about than be realised in practice" (Cox, Verbaan, & Sen, 2012).

So is managing research data a job for academic libraries? Lewis (2010) gave the most precise answer to this question:

> Yes and no. Yes, in the sense that data from academic research projects represents an integral part of the global research knowledge base, and so managing it should be a natural extension of the university library's current role in providing access to the published part of that knowledge base. No, because the scale of the challenge in terms of infrastructure, skills and culture change requires concerted action by a range of stakeholders, and not just university libraries. (Lewis, 2010, p. 145)

Gold (2010) presented six long-term roles for research libraries in data curation, which are:

1. Supporting interoperability of metadata;
2. Developing metadata;
3. Consulting with individual researchers and research groups on best practices for data management;
4. Contributing as data scientists to ongoing research teams;
5. Developing data use cases that will inform design goals and principles for planned data curation infrastructure; and
6. Collecting digital data. (Gold, 2010, pp. 13–14)

The single most widely anticipated role for libraries in RDM is assistance with metadata creation and documentation. As Friedlandler and Adler (2006) pointed out, without providing appropriate information of the context, data can not be used. (2006, p. 22) However, as Nielsen and Hjørland (2014) argued, "data always need to be described, and that the most obvious description of data are the scientific documents in which they are first presented". Furthermore, they state that data sets, which are accompanied by the information needed to interpret them "should be considered documents", and as such, "the indexing of »data« is here fully in line with the bibliographic tradition of information science." (2014, p. 228)

Indeed, many have expressed that the handling of data is not really different from the information organisation practices of libraries, some even claim that data sets are in some respects "no different from a page in a medieval manuscript". (Monastersky, 2013, p. 432) While it might be true that theoretically data sets and manuscripts are of the same nature, there still might be some *slight* differences between the *Book of Kells* and a caliometric dataset, and this requires different handling practices as well: as Lewis (2010) pointed out, "knowledge of MARC, AACR and even Dublin Core does not represent a licence to curate research data" (2010, p. 162) as "LIS principles for representation and metadata creation

are applicable but not directly transferrable to complex scientific research data." (Weber, Palmer, & Chao, 2012, p. 313)

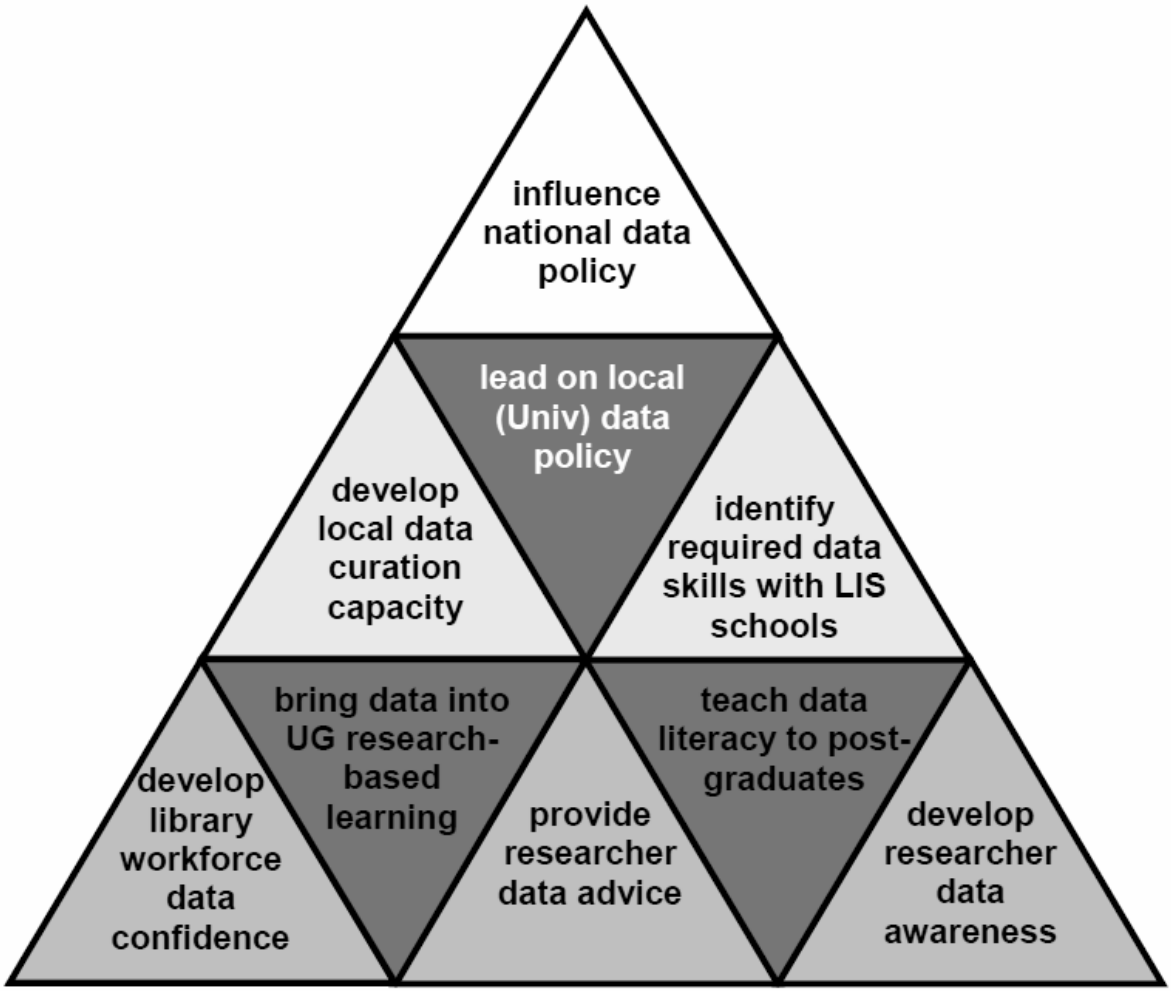


Figure 7: The research data management pyramid for libraries, from Lewis (2010, p. 154)

As some pointed out, "services offered by research libraries are often too generic to be of real value to the research community." (Angevaare, 2009, p. 6) This is probably due to the fact that librarians know "relatively little about current data management practices of scholars" (Heidorn, 2011, p. 665). Therefore, subject-specific knowledge will be important for introducing RDM services: Nielsen and Hjørland (2014) suggested that "information specialists who are also domain specialists could be important partners in providing information about research and data needs in different domains." (2014, p. 237) Some even described curation work as a 'tridge' (three-way bridge), which connects competencies from the domains of "information science and computer science, as well as the natural sciences." (Weber et al., 2012, p. 310)

Lyon specified the knowledge and skills of specialists as follows:

> working knowledge of the research practices and workflows, an understanding of the specific technical standards, metadata schema and vocabularies routinely used in practice, an awareness of the national and international data centres where research data in that domain are deposited, and a good grasp of the data publication requirements of the leading scholarly journals. (2012, p. 132)

Table 1: Librarians' roles in RDM and required competencies mapped against existing roles. (Cox et al., 2012)

Role	**Alignment with existing roles**	**Competencies required**
Policy and advocacy		
Lead on institutional data policy	Advocacy role e.g. in the area of open access	Strategic understanding and influencing skills
Support and training		
Bring data into undergraduate research-based learning, promoting data information literacy	Information literacy training	Understanding of RDM best practices as they apply to relevant disciplines; pedagogic skills
Teach data literacy to postgraduate students		
Develop researcher data awareness		
Provide an advice service to researchers (and research administrators)	Reference and enquiry roles; producing print and Web-based guides; copyright advice.	Reference interview, knowledge of RDM principles
Provide advice as above through a Web portal	Library Web site	Knowledge of institutional and extra-institutional resources
Signpost who in the institution should be consulted in relation to a particular question	Role of library as point of enquiry and the reference interview	Knowledge of institution
Promote data reuse by making known what is available internally and externally; explaining data citation	Marketing of library resources	Knowledge of researchers' needs, knowledge of available material
Auditing and repository management		
Audit to identify data sets for archiving, create a catalogue of materials or to identify RDM needs	Metadata skills	
Develop and manage access to data collections	Collection development, digital library management and metadata management	Audit interviews, knowledge of RDM principles, metadata, licensing
Develop local data curation capacity	Open access role. Preservation role.	Knowledge of RDM principles, relevant technologies and processes, metadata

However, it's evident that "no one individual will have all of the required skills." (Heidorn, 2011, p. 667) Therefore, it has been strongly recommended that data curators collaborate closely with researchers, other institutions, as well as other units of their institutions.

(Block et al., 2010) One notable example of this type of collaboration is Cornell University's Research Data Management Service Group, which "brings together not only librarians and specialists in IT (such as security) but also people with experience in project management, copyright and intellectual property rights issues, high performance computing, and data management system design." (Hswe, 2012, p. 122),

Sometimes the situation at universities is not so ideal, there are occasions when "multiple units are providing varying degrees of data management support to faculty on campus, but no one really knows who is offering what service and to whom." (Peters & Dryden, 2011, p. 397) It's also apparent that with so many players, such as "commercial publishers, information-storage companies and discipline-specific data repositories" (Monastersky, 2013, p. 431) working on the issue at the same time, "libraries will perhaps not always be the first choice" of service on research data management (Nielsen & Hjørland, 2014, p. 237), in some cases simply because "faculty do not often see librarians as being equipped to help them solve their data problems" (Gabridge, 2009):

> for working scientists, who can now browse scientific literature online without leaving their desks, much of this activity goes unseen. For many, libraries seem to be relics that no longer serve their needs. (Monastersky, 2013, p. 431)

Therefore, libraries have to assess carefully "what needs to be done locally, and what might best be done nationally or internationally" (Lewis, 2010, p. 153), discover the needs of their user community, and provide support in ways that serve them the most: "through liaison teams, workshops, one-to-one training or even embedded roles" (Cox et al., 2012), in order to avoid that "scientists themselves […] apply the information management techniques of the new science to their own activities inappropriately". (Joint, 2007, p. 453)

To overcome this problem, providing advice and training seem to be straightforward solutions. As Lewis argued, "many libraries already provide advice on open access and other aspects of scholarly communication, and data management should be seen as a natural extension of this role." (Lewis, 2010, p. 155)

Moreover, as librarians have well-established expertise in teaching information literacy to undergraduate and postgraduate students, it has been suggested that the "best possible

point at which to intervene with guidance and training is very early on in a researcher's career." (Ward et al., 2011, p. 268)

As "top-down, policy-driven, or centralised solutions are unlikely to prove as effective as clear, appropriate and practical support delivered to researchers in a timely manner" (Ward et al., 2011, p. 266), it is difficult to predict the necessary level of approaching the researchers, as most of them rarely have the time to read through long documents and sit through lectures, and are not easy to reach "via face-to-face training sessions." (T. Parsons, 2013, p. 152) Just as it happened in Australia, when a study, which investigated the environment of research data management at universities asked for comments from scientists on their training need and preferences, and, amongst others, received the following answers:

> *What you think I might be interested in, as above, makes me feel sick, frankly! A question -- maybe what you are asking isn't really relevant for humanities disciplines, or at least some of them?*
>
> *No, I do not want "training", I want somebody to do the work for me.*
>
> *NO MORE TRAINING! NO MORE USELESS UNIVERSITY BUREAUCRATS! PLEASE JUST LET ME DO THE RESEARCH.* (Henty, Weaver, Bradbury, & Porter, 2008, p. 18)

Although to be fair, apart from the statements above, the study found an "overwhelming demand" (Henty et al., 2008, p. 1) for training in research data management: "three-quarters of respondents wanted training related to data management planning, either creating a research data management plan at the beginning of a project (52.0%) or after a project has finished (22.4%)." (2008, p. 17)

The question of when is it the best to provide support for researchers during the research process was already raised by the ARL report, *To Stand the Test of Time: Long-term Stewardship of Digital Data Sets in Science and Engineering*, which highlighted that expanding the libraries' portfolio to "include activities related to storage, preservation, and curation […] requires evaluating where in the research process chain curation and preservation activities should take place". (Friedlander & Adler, 2006, p. 42)

Based on these recommendations, Gold (2007b) proposed activities dividing the research cycle into two broad categories: upstream and downstream. She listed the initial activities in the scientific process in the "upstream" category, while the "downstream" side comprises post-production and post-publication services. (Gold, 2007b) As she pointed out later, although libraries contribute to both of these two categories, and the "downstream" part (collecting and disseminating information) is well established, library activities in the "upstream" category (the likes of advising and teaching) are less visible but just as crucial as the former. (Gold, 2010, p. 2)

Figure 8: An illustration of the research cycle. From CIBER (2010, p. 8)

This division of the research cycle appears in a guide by the DCC as well, which categorised the roles and responsibilities for support into three main groups, as follows:

- at pre-award: assistance with the preparation of data management plans, including guidance on costing data management activities and the expert use of online tools;
- throughout the project: advice on data documentation, formats and standards to enable reuse; guidance on storing, managing and analysing data to achieve regulatory compliance and best practice; advice and/or provision of research data storage facilities that meet the needs of a wide range of data types, platforms and access needs;
- post-project: advice on selecting data of long-term value; support to make research data visible and/or available to defined audiences; help for researchers in deciding how to archive data at the end of a project (or at any other appropriate point). (Jones, Pryor, & Whyte, 2013, p. 2)

Salo (2010) expressed need for action throughout the whole research process, rather than as "an add-on at the end": she indicated that "systems designed to steward only final, unchanging materials can only fail faced with real-world datasets and data-use practices" (Salo, 2010), which was also emphasised by many others as well. (Cragin et al., 2010; Fleischer & Jannaschk, 2011; Research Information Network, 2008) Cragin et al. (2010) added that "working with scientists early in the research cycle to identify appropriate metadata standards and to support application facilitates deposition before the end of a project and potentially reduces cost at the point of ingest." (2010, p. 4035)

Libraries have been actively involved in the end of the research cycle as for a long time, recently by adding their share to the advancement of cyberinfrastructure by developing and setting up institutional repositories, "to accelerate changes taking place in scholarship and scholarly communication, both moving beyond their historic relatively passive role of supporting established publishers in modernizing scholarly publishing through the licensing of digital content" (Lynch, 2003, p. 327). Institutional repositories have been described as "a logical consequence of library philosophy that embraces the idea of information for everyone regardless of their wealth, status, and opportunities." (Maceviciute, 2014, p. 292)

Even though repositories were labelled as "something extraordinary" by Lynch (2003, p. 327), he expressed his fear that "at some institutions, repositories will be offered hastily and without much real institutional commitment" (2003, p. 334) and warned that "any institutional repository approach that *requires* deposit of faculty or student works and/or uses the institutional repository as a means of asserting control or ownership over these works will likely fail, and probably deserves to fail." (2003, p. 332, emphasis in original)

These warnings have proved to be accurate in many cases: as Salo (2008) pointed out, the "if you build it they will come" approach "has been decisively proven wrong" within the context of institutional repositories. (2008, p. 98) She argued that "most repositories languished understaffed and poorly-supported, abandoned by library and institutional administrators, scoffed at by publishers, librarians, and open-access ideologues" (2008, p. 99), and pointed out that "what institutional repositories offer is not perceived to be useful, and what is perceived to be useful, institutional repositories do not offer. […] They do not help with grant applications, submissions to publishers, or visibility in existing disciplinary

repositories. They do not help a researcher achieve tenure, promotion, or the next round of grant funding." (2008, p. 103) Although operating institutional repositories "takes more effort than originally was imagined and requires solutions to issues such as advocacy (or rather persuading researchers to comply), infrastructure, sustainability, training of library staff and researchers, documentation, roles and responsibilities, as well as acquiring content" (Maceviciute, 2014, p. 292), "creating repositories for data might build on this work" (Gold, 2007b) and "can learn from publication repositories experiences and their efforts to engage researchers to accept and use these new institutional services". (Macdonald & Martinez-Uribe, 2010, p. 6) Nevertheless, Walton (2010) pointed out that only if "libraries have functioned effectively in delivering an institutional repository then the university may seek to widen this to include data curation. If libraries have not delivered on the institutional repository then the role of observer will come the libraries' way." (2010, p. 3)

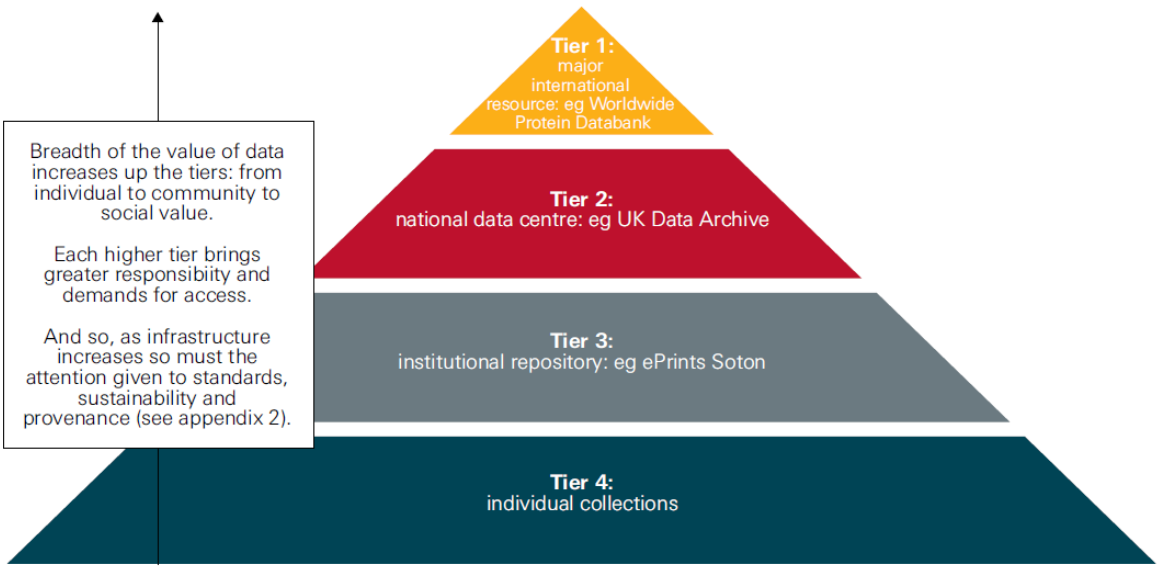


Figure 9: The Data Pyramid – a hierarchy of rising value and permanence (Royal Society, 2012, p. 60)

It has been pointed out that "the scholarly enterprise is sufficiently diverse that disciplinary repositories will never be fully comprehensive" (Lynch, 2003, p. 331), even with the current, relatively high proliferation rate of national or subject-specific repositories for research data. Furthermore, there is a need that "stewardship of digital information should be vested in distributed collections and repositories that recognize the heterogeneity of the

data while ensuring the potential for federation and interoperability." (Friedlander & Adler, 2006, p. 12)

Institutional repositories therefore are seen more "complement and a supplement, rather than a substitute, for traditional scholarly publication venues" (Lynch, 2003, p. 331), with many pointing out university libraries "may need to take responsibility for assisting with the curation and preservation of smaller scale data sets arising from the research of research groups or individual academics" (Hey & Hey, 2006, p. 526), help those researchers "whose discipline does not have any well-established data dissemination system" (Kim, 2013, p. 502), take on a "both/and" approach "as »feeders« of data to discipline repositories" (Gold, 2010, p. 23), or simply "help researchers to find appropriate existing ones." (Monastersky, 2013, p. 432)

Based on Berman's 'data pyramid' (2008, p. 53 see figure 6 in section 2.2), the Royal Society grouped the different types of repositories along with the data that should belong to these as figure 9 shows. As the report claimed, institutional repositories should ideally preserve and curate research data "having potential value, but for which there is no Tier 1 or Tier 2 database available, and which can no longer be maintained by scientists actively using the data" (Royal Society, 2012, p. 63), thus creating facilities "of last resort" (Hamilton, Manuel, Whyte, & Jones, 2014).

2.6 Conclusion

This section provided an overview of the broader context of RDM, covering the international legal context, as well as the cross-country recommendations that could apply to the participating countries. It extensively reviewed the last decade's developments in open science and the various driving forces, benefits, and barriers of research data management. It continued by discussing the academic libraries' changing role and finished with thoroughly reviewing the academic libraries' many possible roles in RDM, providing plenty of examples from the literature, alongside arguments for and against the introduction of these services.

3. METHODOLOGY

The previous section laid out the fundaments of open science and research data management, and the possible roles libraries can take to introduce services in research data management. The following section discusses the methodology used in my thesis in order to investigate if the libraries, which are in the focus of this study, can easily pick up the pace with the developments laid out in the previous section. It presents the choices made during the research, along with detailing the development of the instrument used for data collection, and revealing the paths and dead ends faced during the process. It also includes a discussion about the limitations of this study, as well as the ethical considerations.

Ultimately, the thesis sought answers for the following research questions:

- Are academic libraries in these two countries ready to implement services in research data management?
- How well would research data management fit into the current set of research support services provided by academic libraries?
- In which areas of research data management do librarians see themselves concentrating in the future?
- What are the perceived constrains that could hinder academic libraries to introduce RDM services?

3.1 Research approach and strategy

As it has been seen in the literature review, the question of research data management is a rather complex one, summed up very well by Gold (2010):

> One of the challenges of talking about "data curation" is that the activities of curation are highly interconnected within a system of systems, including institutional, national, scientific, cultural, and social practices as well as economic and technological systems. (2010, p. 3)

The statement strongly resembles the concept of open systems, a theory of the how the many cultural, societal, institutional factors that have an effect on our actions and attitude. According to Pawson (2006):

> A ceaselessly changing complexity is the norm in social life, and this is the open system predicament. […] Such a complex and messy social reality appears to render extremely unlikely the opportunity for experimentally isolating and manipulating all the contributory explanatory elements. However one looks at it, the implication is that we can never exercise control over all the historical and contemporaneous, macro- and micro-conditions that have influenced the situation we wish to explain." (Pawson, 2006, p. 18)

Or, as Popper (2002) pointed out, by analysing certain aspects of social life, it might be possible to find and understand causes and effects, however, "we may nevertheless find that we are unable to formulate *general laws* which would serve as a description, in general terms, of such causal links." (Popper, 2002 emphasis in original)

Without general laws, it becomes hard even to express the problems which we are aiming to solve, and indeed: even a thesis by a former DILL student found "no proven solutions for the management of research data" (Melnarowicz, 2012, p. II). These lacks of clear solutions make RDM a difficult, or if you like, a wicked problem: as opposed to 'tame' problems, which normally occur in the natural sciences, and which have clear true or false answers (at least according to our current understanding), solutions for 'wicked' problems, as described in *Dilemmas in a General Theory of Planning* by Rittel and Webber (1973), depend heavily on the problem-solvers' "group or personal interests, their special value-sets, and their ideological predilections." (1973, p. 163) As they point out, "to find the problem is thus the same thing as finding the solution; the problem can't be defined until the solution has been found." (1973, p. 161) This notion that research data management resembles the characteristics of wicked problems has also been pointed out by Cox, Pinfield, and Smith (2014), who stated that RDM is "significantly more complex problem than those usually encountered by libraries." (2014, p. 2)

Talking specifically about the current study, in order to address the research questions, first, a systematic review was required to formulate a conceptual model of the myriad actions libraries can take to support research data management – which is presented in the literature review. This review of the literature has also been taken into consideration when the tool for this study has been prepared, or, to follow the previous quote by Rittel and Webber (1973), it provided a basis in which the possible solutions were present, in order to address the problems of RDM.

3.1.1 The research method

This thesis ultimately aimed for a deeper understanding of librarians' views and attitudes on research data management, and to assess whether academic libraries should expand their portfolio by providing services supporting RDM. As it has already been laid out sections 1 and 2, such services are getting common in several countries. However, in order to avoid exact replications of previous studies, and to obtain answers which are not

affected by preconceptions of what is required by policy-makers, the chosen institutions for this research were from Estonia and Hungary, from a milieu where policies and mandates on research data are either just emerging, or not quite in place.

The first and probably the most obvious choice to obtain answers for these questions from the broadest range of respondents (both geographically and numerically) would be to implement the survey method. However, the study was not aiming for representativeness, and, as Pickard (2007) points out: while "surveys are designed to produce generalisation within the population, case studies are concerned with individual perceptions, beliefs and emotions." (Pickard, 2007, p. 101) Therefore, instead of a representative survey, an illustrative, multiple case study method has been implemented, as it enables comparison between the sites, shows if certain principles are only typical at a single site or present at a more broad range, gives a rich picture on relationships, and eventually, "multiple cases typically yield more robust, generalizable, and testable theory than single-case research." (Eisenhardt & Graebner, 2007, p. 27)

Implementing multiple case studies instead of a survey allowed me to avoid probability sampling and low response rates throughout the process. Alternatively, a purposive, or theoretical sampling was used, which was fitting for this study, as it enables the researcher to select case study sites in order to "fill conceptual categories, replicate previous findings, or extend emergent theory." (Santos & Eisenhardt, 2004)

The emergent theory this study aimed to extend was the level of academic libraries' involvement in research data management, and to do so, the case study sites have been selected to fill the conceptual categories of academic libraries being situated at the nations' leading, preferably research-intensive universities. In the end, eight libraries have been selected and contacted: the libraries of the Tallinn University, the Tallinn Technical University, and the Tartu University from Estonia, and the libraries of the Budapest University of Technology and Economics, the Debrecen University, the Eötvös Loránd University, the University of Pécs, and the University of Szeged from Hungary.

3.2 Designing the instrument for the inquiry

To assess institutional readiness for specific activities, the usage of capability maturity models are widespread in the fields of industry, business, and management, as they enable

investigators “to determine whether, how easily, and how well a given organization or community would be able, in theory and in practice, to accomplish a given task.” (Lyon, Ball, Duke, & Day, 2012, p. 10)

Maturity models for research data management are rooted in the in seminal *Five Organizational Stages of Digital Preservation* proposed by Kenney and McGovern (2003), which later has been usually referred as the ‘Cornell Maturity Model’. Kenney and McGovern (2003) stressed that “librarians and archivists must understand their own institutional requirements and capabilities before they can begin to identify which combination of policies, strategies, and tactics are likely to be most effective in meeting their needs” (2003), and laid out five stages of organizational stances to digital preservation:

- Acknowledge: Understanding that digital preservation is a local concern;
- Act: Initiating digital preservation projects;
- Consolidate: Seguing from projects to programs;
- Institutionalize: Incorporating the larger environment; and
- Externalize: Embracing inter-institutional collaboration and dependency. (Kenney & McGovern, 2003)

One other model for assessing institutional readiness for research data management is the ANDS RDM Capability Maturity Guide (Australian National Data Service, 2011). In the guide for the tool, the developers acknowledged that “it is up to each organisation to decide where on the model it wants to be: not all organisations will seek to attain Level 5 (Optimised).” (2011, p. 1)

As it can be seen on table 2., activities of research data management have been divided into four categories: Institutional policies & procedures; IT infrastructure; Support services; and Metadata management, and processes into a slightly modified version of the Cornell Maturity Model: Initial (disorganised and ad hoc); Development (under development); Defined (standardised and communicated); Managed (managed and measured); Optimised (focus on continuous improvement).

Table 2: The ANDS Research Data Management Framework: Capability Maturity Guide (Australian National Data Service, 2011)

	Level 1 Initial	Level 2 Development	Level 3 Defined	Level 4 Managed	Level 5 Optimised
	Process is disorganised & *ad hoc*	Process is under development	Process is standardised, communicated	Process is managed, measured	Focus is on continuous improvement
Institutional policies & procedures	Policies & procedures may be undeveloped, not up to date, and/or inconsistent.	Policies & procedures are developed & harmonised.	Policies & procedures are promulgated & absorbed into behaviours.	Policies & procedures accepted as part of the culture & subject to audit.	Policies & procedures are subject to review & improvement.
IT infrastructure	IT infrastructure provision is patchy, disorganised & poorly publicised.	Funds are invested in technology & skills. Responsibilities are defined. Processes are established, defined & documented.	Management shows active support. Facilities are well-defined & communicated, standardised &integrated.	Funding adapted to need. Management actively engaged. Documentation kept up to date.	Concerted efforts to maintain, update & publicise infrastructure. Metrics & feedback used to optimise services.
Support services	Training is ad hoc, curation & preservation services are disorganised, data management planning is unsupported & other services inconsistent & poorly publicised	Investment in skills. Services identified & staffed. Responsibilities are defined. Documentation & training developed.	Active participation in training & widespread availability of services.	Widespread take up of services. Curation & preservation acknowledged as critical to the institutional mission.	Customer feedback used extensively to update & improve services.
Managing metadata	Metadata management is chaotic & understood by only a few.	Responsibilities are defined & skills developed. Processes are established, defined & documented. Metadata applied to key datasets & shared externally.	Processes are standardised & integrated. Metadata applied to new datasets & shared externally.	Metadata quality metrics collected. All datasets described & metadata shared.	Continuous improvement applied to processes & capabilities.

At the initial stage of planning the methodology for this thesis, the implementation of capability maturity models has been strongly considered, as they seem to fit perfectly for the purposes of this thesis. However, during the process, a European landscape study on research data management, which also used maturity models to assess a wide range of stakeholders, including the Library and Information Centre of the Hungarian Academy of Sciences, published its initial results, and estimated the library's overall maturity for research data management "below 1 on a scale of 8". ("Final Short Report | SIM4RDM," 2014, p. 6) As they reported:

> RDM is hardly a priority for the Hungarian scientific budget. […] In Central and Eastern Europe financial resources are needed for digitization in the first place, and questions of data management, data exposure, data organization can only be addressed afterwards. ("Final Short Report | SIM4RDM," 2014, pp. 6–7)

It was believed that assigning maturity levels of zeros and perhaps ones to the participating libraries will not necessarily deepen the understanding of how these libraries see their role in supporting researchers at their universities, therefore the implementation of maturity models for this study has been rejected. However, the review of different maturity models for research data management still helped in designing the methodology and data collection technique of this thesis. Eventually, instead of applying a specific tool, a more general questionnaire has been used for data collection.

Questionnaires enabled me to reach academic libraries in a "geographically dispersed community at relatively low cost" (Pickard, 2007, p. 183), but also brought up some difficulties, as it was slightly difficult to construct a thorough model of *a priori* categories of the libraries' service priorities in RDM.

For the distribution of the questionnaire, the Surveymonkey online system was used. The questionnaire was open from the 29th of May until the 29th of June, 2014. Directors of the eight selected academic libraries have been contacted, and asked to fill and distribute the questionnaire to other staff members who could be associated with RDM: these included the heads of IT services, heads of research support/reference/liaison teams, staff members responsible for the institutional repository, and any other staff members who could in any other way be associated with research data management. The reason of this selection of different units was to have a detailed view on the staff members' attitude, and on the overall picture of the libraries, as it was believed that workers of different units would have different views on the issue of RDM, as they all take a different role in supporting their user community.

3.2.1 Trustworthiness of the enquiry

As it has mentioned above, the main difficulty of designing a questionnaire is to construct *a priori* categories of the phenomenon to be examined. It is even more difficult to do that when we are dealing with wicked problems – a concept which has been discussed earlier in this section. In order to avoid a greatly eclectic approach, but to formulate a

comprehensive, holistic framework, and to possibly cover most of the possible ways libraries can take to implement RDM services, the tool developed for this study builds on previous similar research, and incorporates a wide range of recommendations and best practices from the international discourse on RDM. Therefore, the statements about the importance of research data and the possible aspects of library service priorities in RDM that are included in this thesis are not arbitrary categories that came out of nowhere – they are carefully selected from the literature, as the coming tables show.

More precisely: the statements at the beginning of the questionnaire were borrowed from the RCUK principles ("RCUK Common Principles on Data Policy - RCUK," n.d.), the PLOS guidelines ("PLOS ONE : accelerating the publication of peer-reviewed science," n.d.), Carlson (2013), Kruse & Thestrup (2014), Maceviciute (2014), the Sim4rdm landscape report ("Final Short Report | SIM4RDM," 2014), and a survey conducted by the colleagues of the Loughborough University Library (Hamilton et al., 2014).

The content of the questionnaire covered a broad range of current practices which could serve as a basis for introducing services for RDM, such as the level of development of the institutional repositories, and the training services offered by libraries, for example in research methods, information literacy, and about scholarly communications and open access. It also aimed to assess the respondents' familiarity with the researchers' data handling practices and the broader legal context, such as policies of the universities or funding agencies on open access and research data.

The central part of the study also builds on former, similar research about RDM services in academic and research libraries, which mainly used surveys to map the existing landscape in different countries, such as the US and Canada (Soehner, Steeves, & Ward, 2010; Tenopir, Birch, & Allard, 2012; Tenopir, Sandusky, Allard, & Birch, 2013) Australia, Ireland, New Zealand, and the UK (Corrall, Kennan, & Afzal, 2013), and Denmark (Kruse & Thestrup, 2014). The questions about specific library activities in the current questionnaire aimed to synthetize the aspects of the aforementioned studies, as it can be seen in table 3.

Table 3: Questions on research data management in the current study compared with previous works on the same subject.

Aspects in current questionnaire	Aspects covered in previous surveys			
	Soehner et al., 2010	**Tenopir et al., 2012**	**Corrall et al., 2013**	**Cox & Pinfield, 2013**
Current or future services				
Offer consultancy on data management plans	✔	✔	✔	✔
Consult with researchers on metadata creation and standards		✔	✔	✔
Create metadata for research data		✔	✔	✔
Train undergraduate students on data literacy	✔		✔	✔
Train postgraduate/doctoral students on data management	✔		✔	✔
Train researchers on data management	✔	✔	✔	✔
Create web pages with information and resources about research data management	✔	✔		✔
Embed librarians in research teams				
Offer advice on intellectual property rights surrounding data				✔
Carry out, or assist in an institutional data audit	✔	✔		✔
Provide storage space and preservation for research data	✔	✔	✔	✔
Extend the institutional repository's scope to include data sets			✔	
Assist researchers to submit their data sets to third-party, general or disciplinary data centres	✔		✔	✔
Assist researchers in the discovery of data sets	✔	✔	✔	✔
Assist researchers in the citation of data sets		✔		✔
Assist researchers in the reuse of data sets	✔		✔	
Develop, or assist in the development of an institutional data policy	✔		✔	✔
Assist in the development of a national data policy				
Collaborate with other institutions in the country to provide research data management services	✔	✔		
Collaborate with international institutions in research data management				
Create a permanent 'data librarian' (or similar) position	✔	✔		
Create a fixed term 'data librarian' (or similar) position	✔	✔		
Relocate existing staff to deal with research data management	✔	✔		
Ensure the professional development of current workforce in data management	✔	✔		
Constrains (as open-ended questions)	✔		✔	

Furthermore, as an addition to the activity level, the possible roles which libraries can take to tackle the problem of RDM builds on the former recommendations by Jones et al. (2013), LIBER (van den Berg et al., 2012), and Lyon (2012), as table 4 shows.

Table 4: Questions on library roles in research data management compared with recommendations of previous publications on the same subject.

Aspects in current questionnaire	Recommendations from previous publications		
	Lyon, 2012	**S. Jones et al., 2013**	**van den Berg et al., 2014**
Library roles in RDM			
Discover and address the needs of the researchers	✔	✔	✔
Engage high level university management in the issue			✔
Develop or assist in the development of policies on research data		✔	✔
Raise awareness for data management and sharing in the community	✔	✔	✔
Advocacy and guidance on data management planning	✔	✔	✔
Provide technical infrastructure for storage and preservation	✔	✔	✔
Assist with tools for data analysis	✔	✔	
Assist with documentation and metadata standards	✔	✔	✔
Provide guidance on data discovery and citation	✔	✔	✔
Provide training for students on data literacy	✔	✔	✔
Provide training for researchers on data management	✔	✔	✔
Provide advice on intellectual property rights and licensing	✔		✔

3.3 Ethical Considerations

An informed consent sheet was presented to participants at the first page of the questionnaire form, which explained the purpose of the study and stated what was required from them. Respondents were informed that their personal data will be kept confidential, and that no data will be presented in the study that could help to identify them or their institution. Respondents were made aware that their participation in the study is voluntarily and that they are able to withdraw at any time.

The questions were designed to take about half an hour, and apart from the name of the respondents' institution and their role within the institution, none of them was mandatory. To be able to raise further questions or to comment on any matter if needed, my contact address has been given to the respondents, and an open-ended question have been provided at the end of the questionnaire for any other remarks the respondents wished to share.

3.4 Limitations of the research

Despite of the carefully developed framework, the study has evident limitations as well, some of them deliberate and some unintentional, but incidental to the research design. The most obvious delimitation is naturally the geographical one – but as it has been presented earlier, the whole point of this thesis was to assess the capacity of libraries in countries, which are only at the beginning of implementing services on RDM. A further delimitation is a natural effect of purposive sampling: libraries of research centres could have been incorporated in this study as well; however, as they have some slight differences from academic libraries, it would have taken a more substantial effort to, on the top of the presented framework, cover their possible roles in RDM.

Although the overall writing of this study took more time than originally expected, and therefore the lack of time cannot really be brought up as a limitation, it still affected the data collection process, for the questionnaire could not been excessively piloted on academic libraries, and only personal consultations have been made before distributing it to the study participants. Although these consultations produced a more effective and systematic questionnaire, some of its weaknesses were only apparent during the phase of data collection.

The main limitation of using questionnaires for data collection is the problem of self-reporting, that the researcher has no or limited control over the respondents, and could only assume if the participants share the same vocabulary, understand the concepts in the same way as the researcher does. Moreover, the researcher can only hope that the respondents do what it is asked from them: this got apparent once the responses for this study started coming in, as it got obvious that it will not be possible to reach the desired number of individual responses from the units of the participating institutions.

Conducting focus group interviews might have been fitting better for the purposes of this study for various reasons, like having more control on events, meeting participants in person, and having a bigger control over who participates in the study, and to avoid the excessive efforts of developing *a priori* categories to include in the questionnaire, but rather create a more reactive, ongoing research project. Nevertheless, conducting focus groups would have consequently caused a reduced sample size, as it would have been

virtually impossible to effectively plan and carry out interviews in eight different institutions, and also, focus group interviews would have needed a much substantial effort both from my and the participants' side.

Moreover, studying possible future services might introduce social desirability bias, as respondents might want to stand out, or express wishful thinking about their organisations' future. Social desirability bias might be reduced by making respondents aware of their anonymity, but I had neither any control, nor any clue whether the responses represent the actual situation of the participating institutions. Moreover, even if participants expressed their opinion about the future services honestly, these answers are by no means facts, and do not necessarily mean that the services will materialise in the future.

Other limitations arose from the broader context in which academic libraries are situated. It is understandable that the "one size fits all" approach does not work in research data management, libraries and universities where these libraries are situated are complex systems *per se*, with their own values, culture, and community, which could differ substantially from institution to institution, resulting in a situation in which a solution might work in one place, but would be ineffectual in an other. Nevertheless, even if these institutions have their own values and culture, they *also* exist in a broader, globalised context, where, indeed, working in silos is what creates ineffectuality. The literature review of this thesis began with the principles of open science, and to conclude this section, the second *Mertonian* norm, universalism comes into the picture: while many institution have the aforesaid unique character; the infrastructure researchers use, the models of publishing, the frameworks for research evaluation, and the rules of research funding agencies (especially in international programs) are not, and probably should not differentiate between scientists just because they work in some place instead of an other. However, as it has also been pointed out earlier, the aspects of RDM services in libraries were taken into consideration in order to cover the broadest possible spectrum, and to include a set of core principles of the international discourse on RDM. Naturally, no one library will be active in every field that is presented in the questionnaire.

Despite the previously noted limitations, the study still managed to formulate relevant conclusions, which could be useful for various communities involved in RDM.

4. DATA ANALYSIS, DISCUSSION

This chapter provides an analysis of the collected data, and a discussion of the findings in the light of previous, similar studies about the academic libraries' role in research data management.

Out of the eight contacted institutions, seven has returned at least one fully filled questionnaire: the Academic Libraries of the Tallinn Technical University and the Tartu University from Estonia, and the Budapest University of Technology and Economics, the Debrecen University, the Eötvös Loránd University, the University of Pécs, and the University of Szeged from Hungary. The profiles of these universities are either general/universal, covering virtually all disciplines, or specialised in technology/natural sciences.

A previously noted limitation of the data collection technique, i.e. that the researcher has a very limited control over the respondents got apparent when the results started to come in. Instead of the initially desired snowball-sampling method, that the key informants would pass the questionnaire to other staff members at their institutions, the respondents took a slightly different approach to fill the questionnaire: at some institutions, only the director provided answers, while at other places only heads of reference services or the IT units. One library took a collaborative approach by first consulting on the questions with different staff members and then returning only one set of answers. Only one library provided multiple answers (two) to the questionnaire.

With only such a limited number of individual responses from each library, the analysis of the data and the presentation of the findings will take a different path than it was planned and laid out in the methodology section: although the study aimed to present different detailed cases about the institutions, the depth of the collected data does not allow the researcher to show the authentic context in which the institutions are situated. Nevertheless, the richness of the data enables me to present the findings as if the study used the survey method, although it must be stressed again that with such a small amount of responses, the findings of this thesis are by no means representative, nor statistically significant. The collected and anonymised data is openly available online, under the following link: http://dx.doi.org/10.6084/m9.figshare.1431881

4.1 Data analysis

Out of the eight set of responses, four has been provided by directors of libraries, one from a deputy director who was primarily responsible for IT services in the library, one from a head of reference services, and two respondents indicated their role as ‘librarian’.

As the only set of answer by a respondent with an IT background showed, there might be a gap between different units within libraries on how they see their libraries’ role in research data management. Therefore, where the data showed traits which seemed to have a connection with the respondents’ role in their institution, the results has been grouped in three categories: ‘director’, ‘librarian’, and ‘IT’. Seeing that the views of heads of reference services and librarians did not differ significantly from each other, these two positions comprised the ‘librarian’ category.

4.1.1 General attitude towards RDM

As figure 10 shows, there was a shared agreement amongst the respondents with the RCUK principles, and basically all agreed that publicly funded research data should be made openly available, that data with long-term value should be preserved, and that effective data reuse requires recording sufficient metadata - with only one respondent being neutral about the “public funding – public good” principle.

Directors and librarians all indicated their confidence that the libraries are the researchers’ neutral collaborators in implementing the system for data management and that librarians are neutral partners in data management given their knowledge and skills in organising, disseminating, and preserving various types of materials. However, even if the amount of data is not sufficient to draw strong conclusions, the respondent with IT background did not agreed with the libraries’ role in implementing the systems and slightly disagreed with the librarians’ role in addressing issues of data management.

Directors generally agreed stronger than others that financial resources are required for digitisation of materials and that issues of data management and sharing can only addressed afterwards.

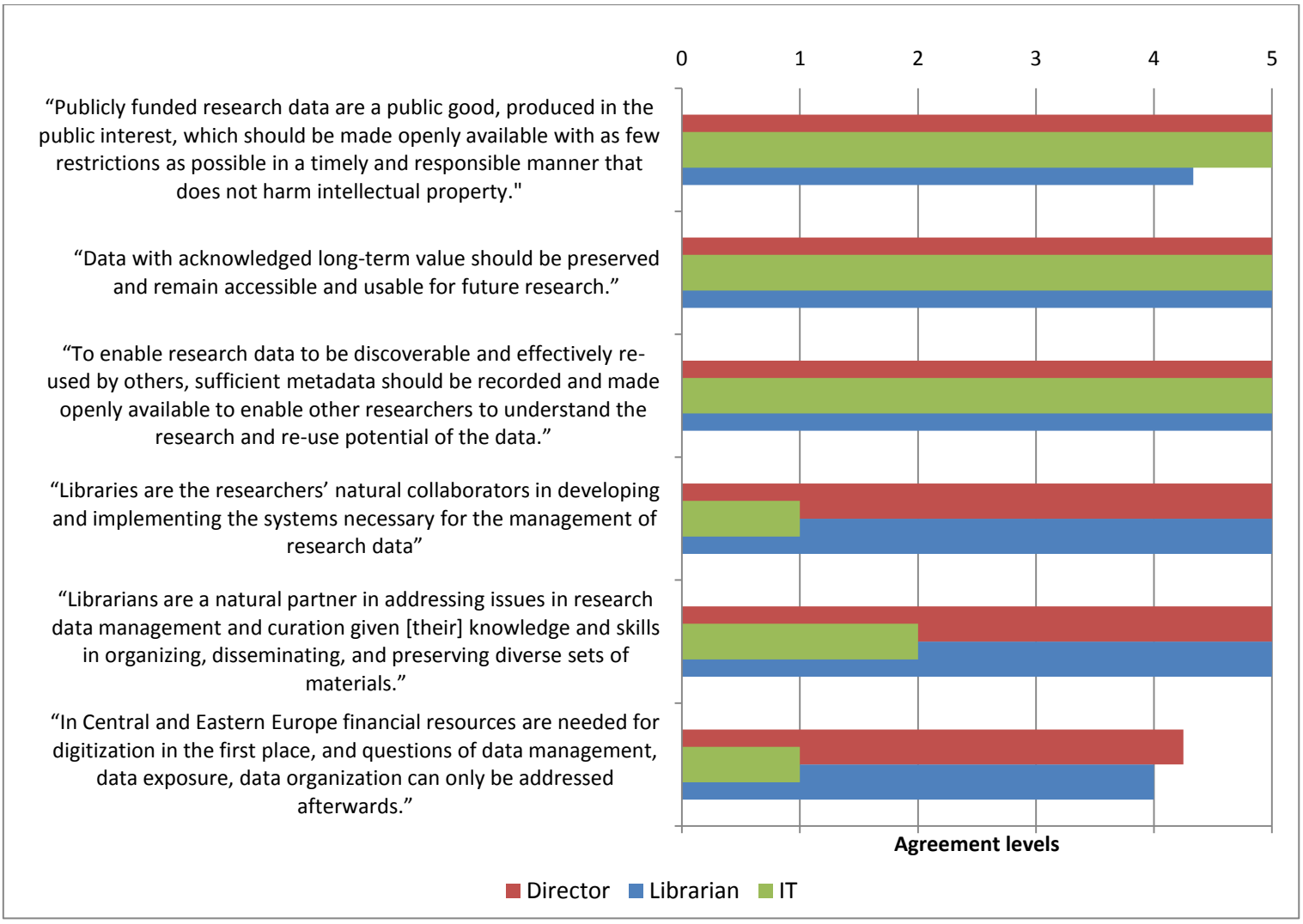


Figure 10: Agreement levels with the opening statements grouped by job titles.

A further question about the attitudes of the respondents sought answers to how they perceive the circumstances concerning smaller institutional, and larger international, subject specific databases for research data. As it can be seen on figure 11, respondents indicated their belief that large databases are more likely to persist, and that smaller, institutional repositories for research data would more likely serve as 'last-resort facilities' when relevant subject-specific databases are not present, but had mixed thoughts about the question of which of these types of databases serve the needs of their users better.

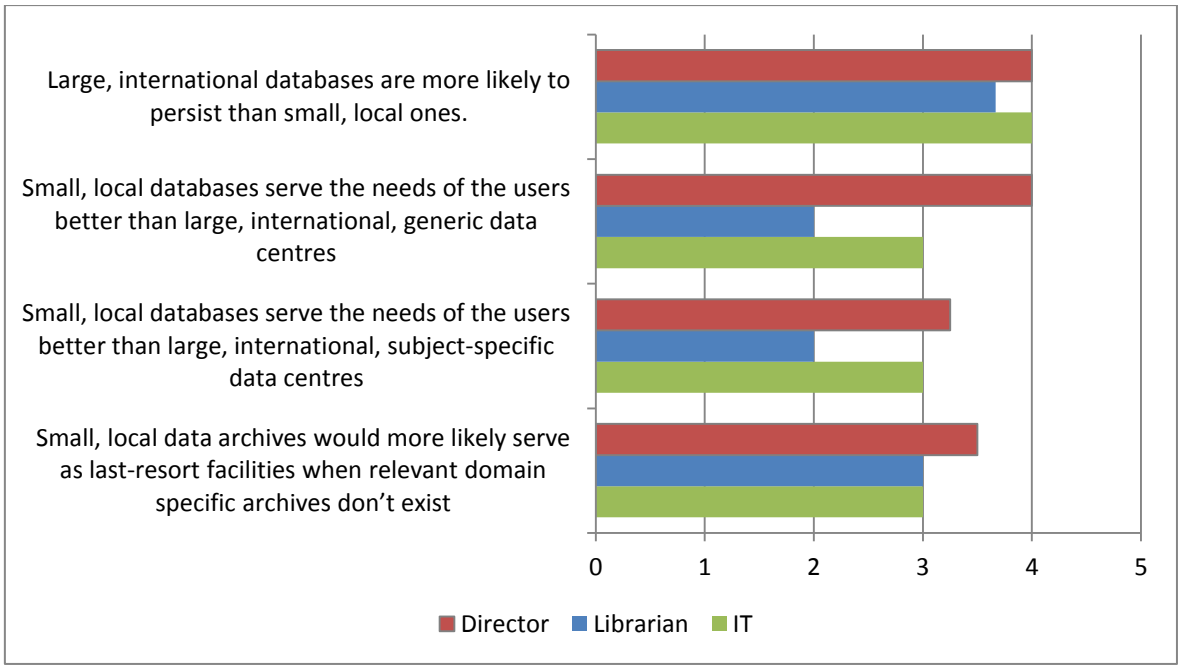


Figure 11: Agreement levels with statements about local and international databases

4.1.2 Current practices

The study aimed to deepen the understanding about how easily libraries could engage with RDM by examining some of the libraries' services, which could either serve as a basis for RDM, or, as institutional repositories, have similar profiles where the experience with these facilities could serve as a valuable lesson to smoothen the introduction of RDM services.

Most of the libraries offered face-to-face consultancy and training on information literacy (6 and 6, respectively), research methods (5 and 4), open access (5 and 4), bibliometrics (4 and 6), and scholarly communications (6 and 4), while web guides are not as widespread, but still prevalent, mostly in the areas of open access (5) and scholarly communications (5).

The primary target groups of these services show that the libraries are actively engaged with the full spectrum of their user community: undergraduate students (mainly in the area of information literacy), postgraduate students (information literacy, scholarly communication, and bibliometrics), researchers (open access), and faculty staff (open access).

The majority of the participating institutions had institutional repositories; only two Hungarian libraries stated that their institutional repository is currently under development. On the possibility of storing data sets alongside documents in the repositories, three libraries stated that it is possible, three said that this feature is under development, and two answered no to this option.

Virtually every respondent agreed (6) or slightly agreed (2) that the setting up of their repository took more effort than originally imagined. The problems encountered were mainly organisational, like the lack of funding, no personnel for management, the difficulties arising from the structure of the university, and some respondents noted that the philosophy of IRs is closer to an archive than to a library system. Other issues came from the lack of clear policies and regulations about the types of works that should be deposited in the repositories, as well as concerns about copyright, synchronizing the different types of data, and that the system in use doesn't allow the needed data input control.

4.1.3 Awareness of the local and global context of RDM

The heterogeneity of policies about open access publications, which were noted as a problem concerning institutional repositories also came up later, when the respondents were asked about their familiarity with the policies on open access publications and research data on the national and institutional level. Concerning open access publications, funding agencies in both countries have either formal or informal recommendations, whereas on the institutional level, the answers were rather mixed: half of the respondents stated that their university have policies regarding open access publications, at one institution certain faculties with their own rules, and the other three institution were working on their open access policy at the time of the study.

However, policies on research data were not this prevalent, as most of the respondents indicated that they didn't know if there are national-level policies, and even those who answered yes or no, had conflicting views: respondents from Hungary simultaneously indicated that national policies on research data are either in place, currently under development, only present in certain disciplines, or missing. Responses about Estonian funding agencies were more consistent (probably because of the fewer answers), but it is still hard to tell based on the responses if Estonian funding agencies have, or currently working on policies on research data. To my current knowledge, both countries have

published strategic documents and roadmaps that recognise the importance of managing research data, but are very much in the middle of the process of developing and implementing formal policies on the issue. On the institutional level, three respondents indicated that their university have policies on research data, the remainder of the respondents either didn't had, or didn't knew about any formal institutional policy on research data.

Three libraries indicated that there were special audits carried out at their institutions in order to gain insight about the data researchers produce and handle: one on an institutional level, and two on a faculty level. At the time of this study, a fourth library was currently at the planning stage of an international data audit together with other libraries from Central Europe within a framework of a Visegrad Fund project (n.b. the Visegrad countries are Czech Republic, Hungary, Poland, and Slovakia).

These audits covered mainly the subject and the restrictions/levels of openness of the data (which were present in every study), followed by the size, type, format, storage media, metadata, creator, manager, and the intellectual property rights about the data (2, respectively). The less frequent aspects of these audits were the purpose of the data (in one study) alongside with the levels of documentation and the hardware/software requirements (in an other study). Questions about the update frequency as well as the potential ways of reuse were not present in any of these audits. About the outcomes of these studies, only one respondent stated that it helped them to become more aware of researchers' practices, the other two indicated that they didn't knew if their university or library used these audits to shape their services according to the findings or used them in any other way.

Out of those institutions that had no any data audits before, one respondent stated that departments and faculties along with liaison librarians know the data handling activities of the researchers in their field, but the other respondents stated that they do not know about anyone who would be aware on the nature of the data that researchers produce.

Respondents provided a mixed set of answers to the question if the management of research data is a priority in their country. Most of them stated that it is getting recognised nowadays, while others pointed out that it is only prevalent in certain disciplines, or when researchers receive funding from certain sponsors. When asked about the situation of

getting funded by international programmes, most respondents indicated that they don't know if researchers are required to submit data management plans in their research proposals or have to provide their produced data when they are asked to do so.

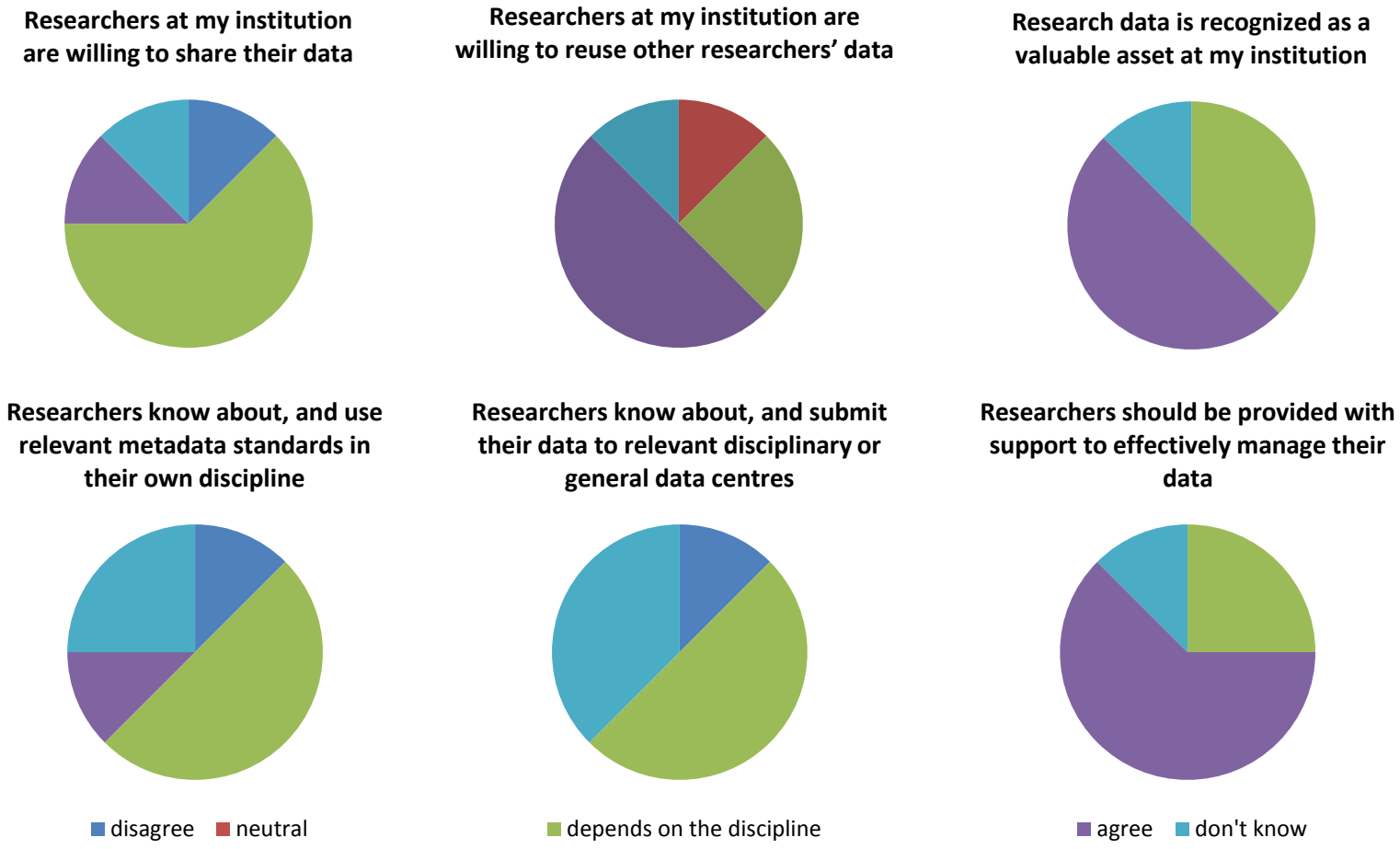


Figure 12: Proportions of answers given to questions about the researchers' practices and needs

These diverse levels of awareness and the perceived difference between disciplines were also apparent in the responses for the questions that were about the researchers' practices and needs. As figure 12 shows, much of the activities connected to research data management are perceived as discipline-dependent, with the only exception of the need of adequate support, which is recognised as a general need.

Respondents indicated that during the research process, the managing of research data is mostly the responsibility of the research team members and principal investigators, but as

it can be seen on figure 13, the level of uncertainty raised when asked about the same responsibility concerning finished projects.

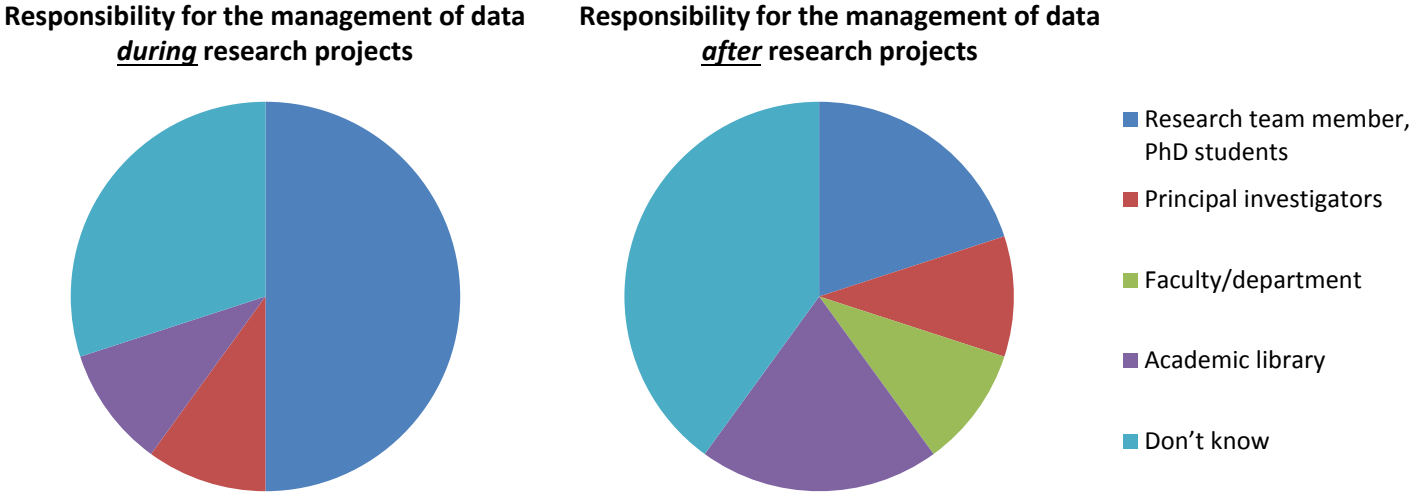


Figure 13: Responsibility for data management during and after the research process

When asked about where the researchers would seek help if they needed assistance on RDM, the vast majority (except one, with the IT background) indicated the academic library. Other answers included that they would figure it out themselves (4) or ask other research team members for help (3). Most of the participating libraries (6) indicated that they encountered questions from the researchers' site about RDM; most of these questions were about archiving, identifying, and searching for data sets.

4.1.4 Future priorities

On the future priorities of implementing RDM services, half of the respondents indicated that it would be the most appropriate to implement these services on an institutional level, one stated that a cross-university partnership would serve better, and three respondents preferred to tackle the issue on a national level. It is worth mentioning that both Estonian respondents opted for the institutional services, while the majority of the Hungarian participants favoured national or cross-institutional services in RDM.

Talking specifically about which units or institutions would be best suited to help researchers with their data management needs, the majority of the respondents logically identified academic libraries (6), but also recognised the importance of collaborating with

other institutions (4). Further mentioned, suitable units for addressing RDM included the local research office (2), the national research council (2), IT services at the universities (1, surprisingly not by the respondent with the IT background), and international data centres (1).

About possible collaborating partners, every respondent stated their willingness of working together with other academic libraries of the country. Other possible partners were the faculties or departments (6), the national research council (6), the local research office (4), IT services, and academic libraries and other relevant organisations from other countries (3, respectively).

Generally, most respondents agreed that libraries should play a both proactive and reactive role in RDM. As figure 14 illustrates, there were differences between the respondents about what they saw as an appropriate stage of the research process to provide help to the researchers. Most respondents expressed the need to get involved during the phase of planning (e.g. to assist with data management plans), and the phase of data collection and analysis (e.g. to assist with tools and documentation), while less than half stated this need during the phase of idea discovery (e.g. to direct researchers to relevant data archives and help them to find and cite datasets), and unexpectedly, even fewer at the stage of disseminating the results (e.g. storage, preservation, publication, and measuring impact).

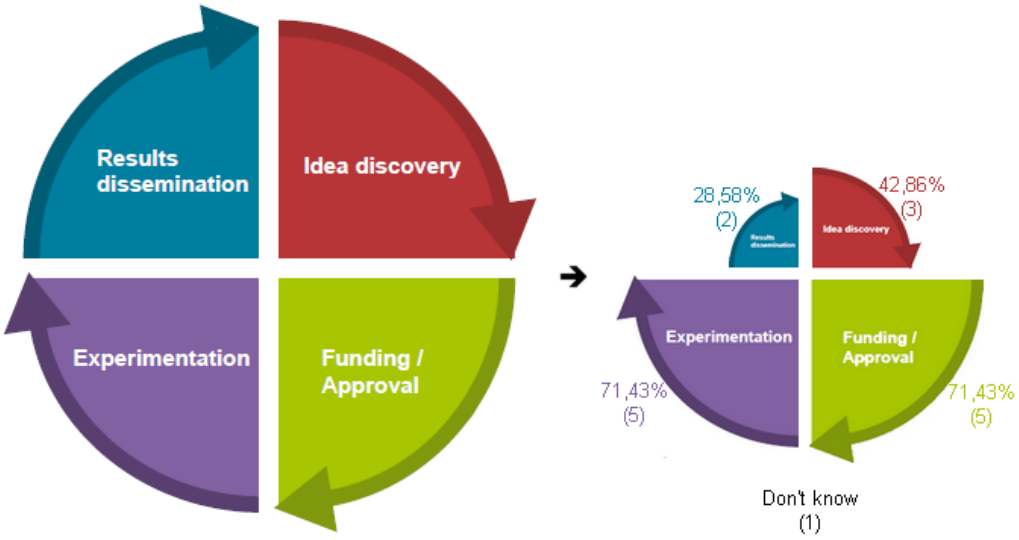


Figure 14: Proportion of answers given on the question about which stage of the research process libraries should provide support for researchers. Original illustration of the research cycle from CIBER (2010)

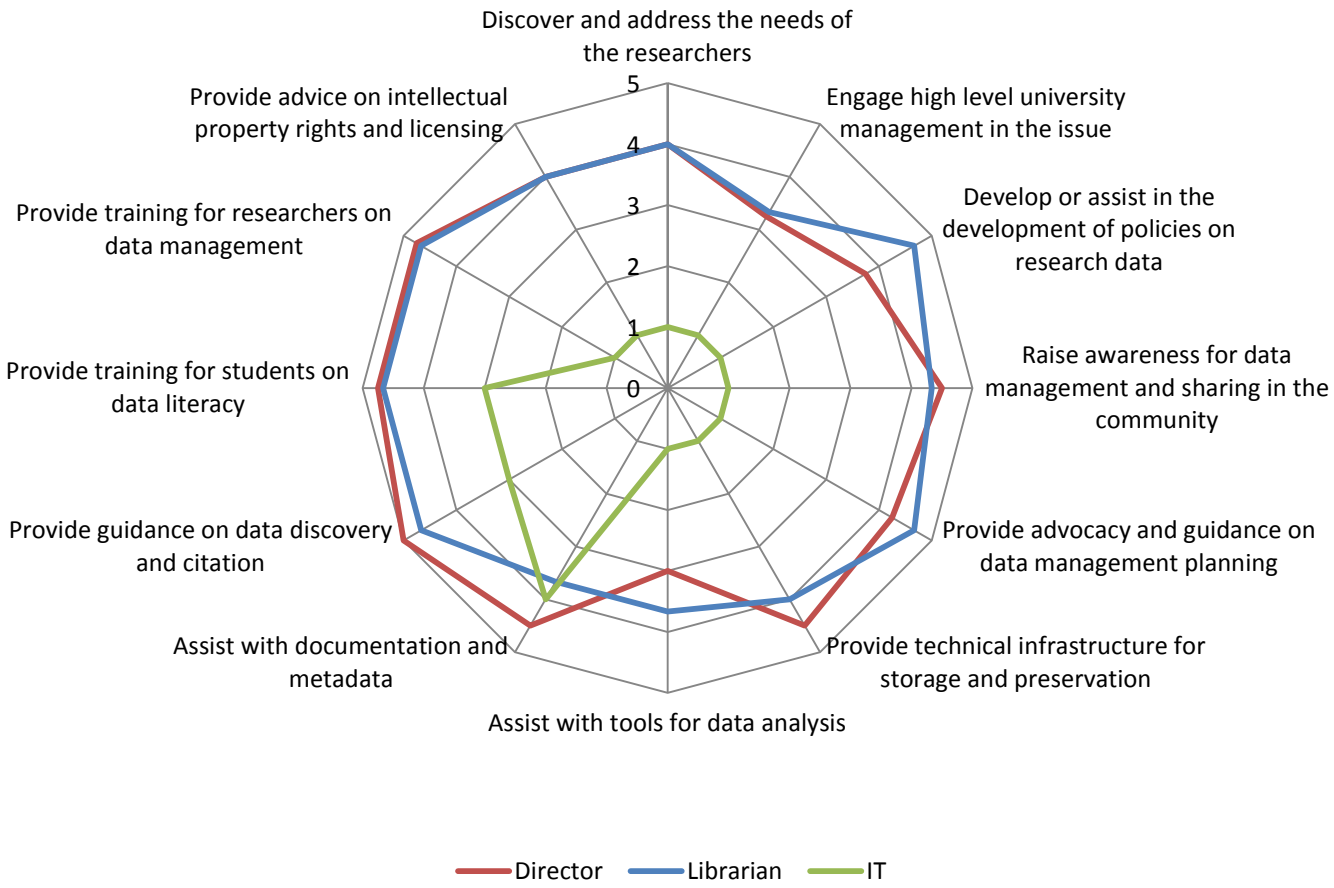


Figure 15: Academic library roles in RDM.

Moving further to the possible roles of libraries in RDM, the difference between the respondent with a technical background and the other participants became strikingly visible at this point. As it is visible on figure 15, while directors and librarians were fully confident that academic libraries should offer services in the means of training, guidance with data discovery and citation, documentation and metadata, assist in the preparations of DMPs, provide the technical infrastructure, and to raise the awareness of their community on the issue of RDM, the respondent with an IT background noted the libraries' role in a more confined spectrum: assistance with metadata, guidance with discovery and citation, and training of undergraduate students.

The study also sought answers from the participants on how they see their institutional services in the future, more precisely, that given the current situation of their institution, on a level from unlikely to likely (with the possible indication that these services are already in place) how they perceive the probability of introducing certain services in RDM in the coming years. Many of the participating libraries were already engaged in many areas at the time of the study, as they offered training and consultancy on data management for

undergraduate students and researchers, created metadata for data sets, had web pages set up with information about RDM, provided storage facilities for research data, and assisted researchers in citing and reusing data. Most of the libraries indicated that it is highly likely that they will create a permanent 'data librarian' position, work together with other institutions, help in the development of an institutional and national policy on research data, and carry out a data audit. The detailed answers are depicted on figures 16 and 17.

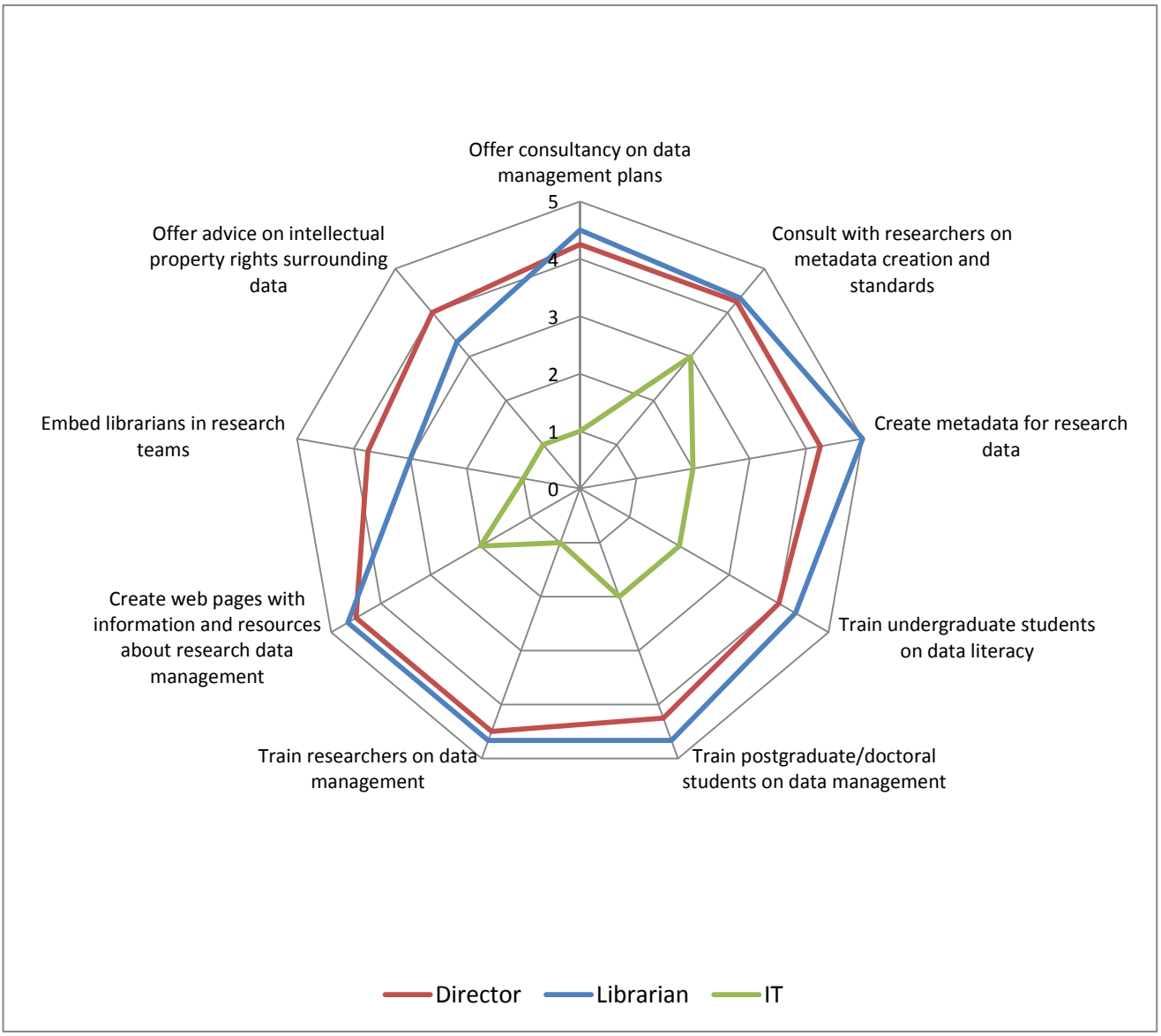
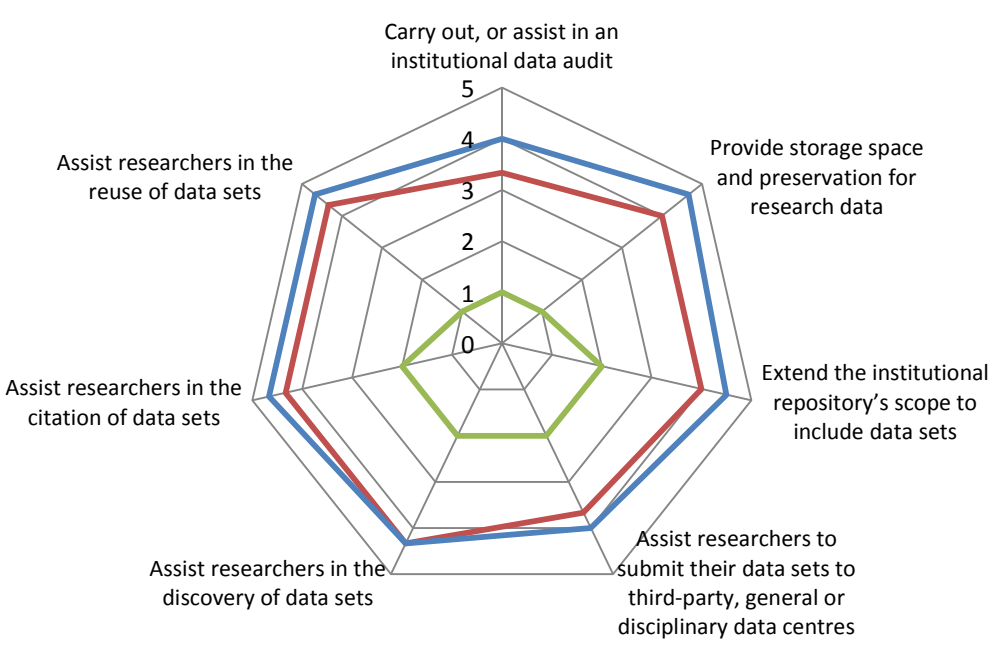


Figure 16: Future service priorities based on the likelihood of their introduction grouped by different job titles 2/1



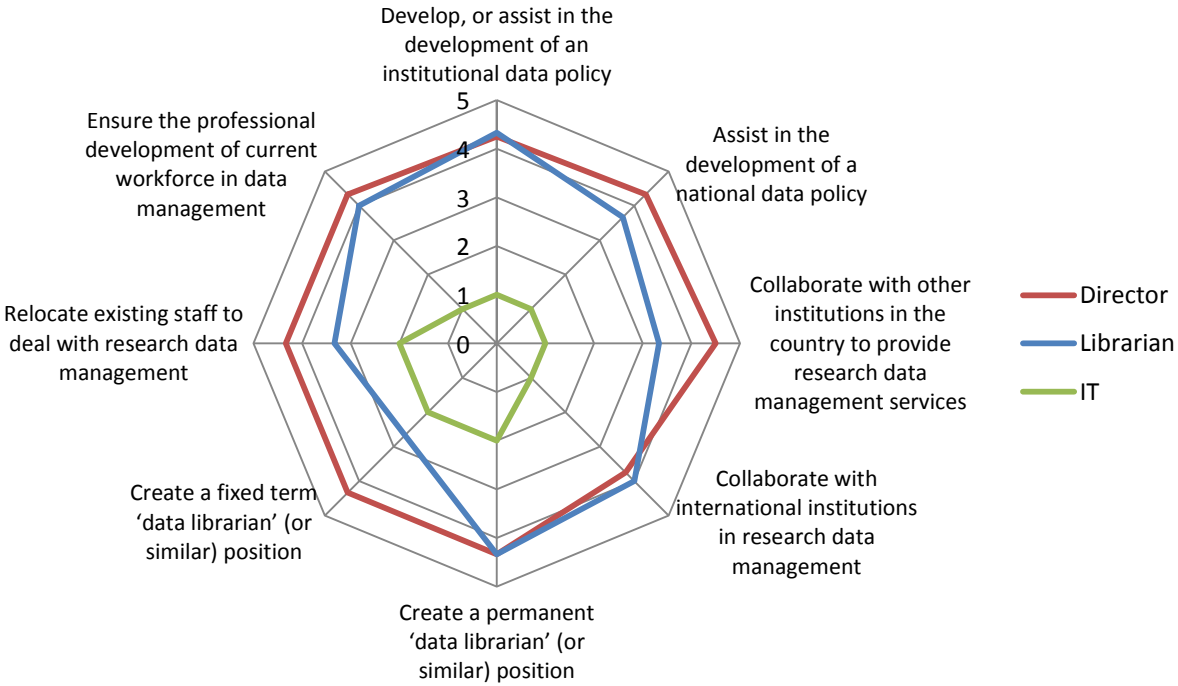


Figure 17: Future service priorities based on the likelihood of their introduction grouped by different job titles 2/2

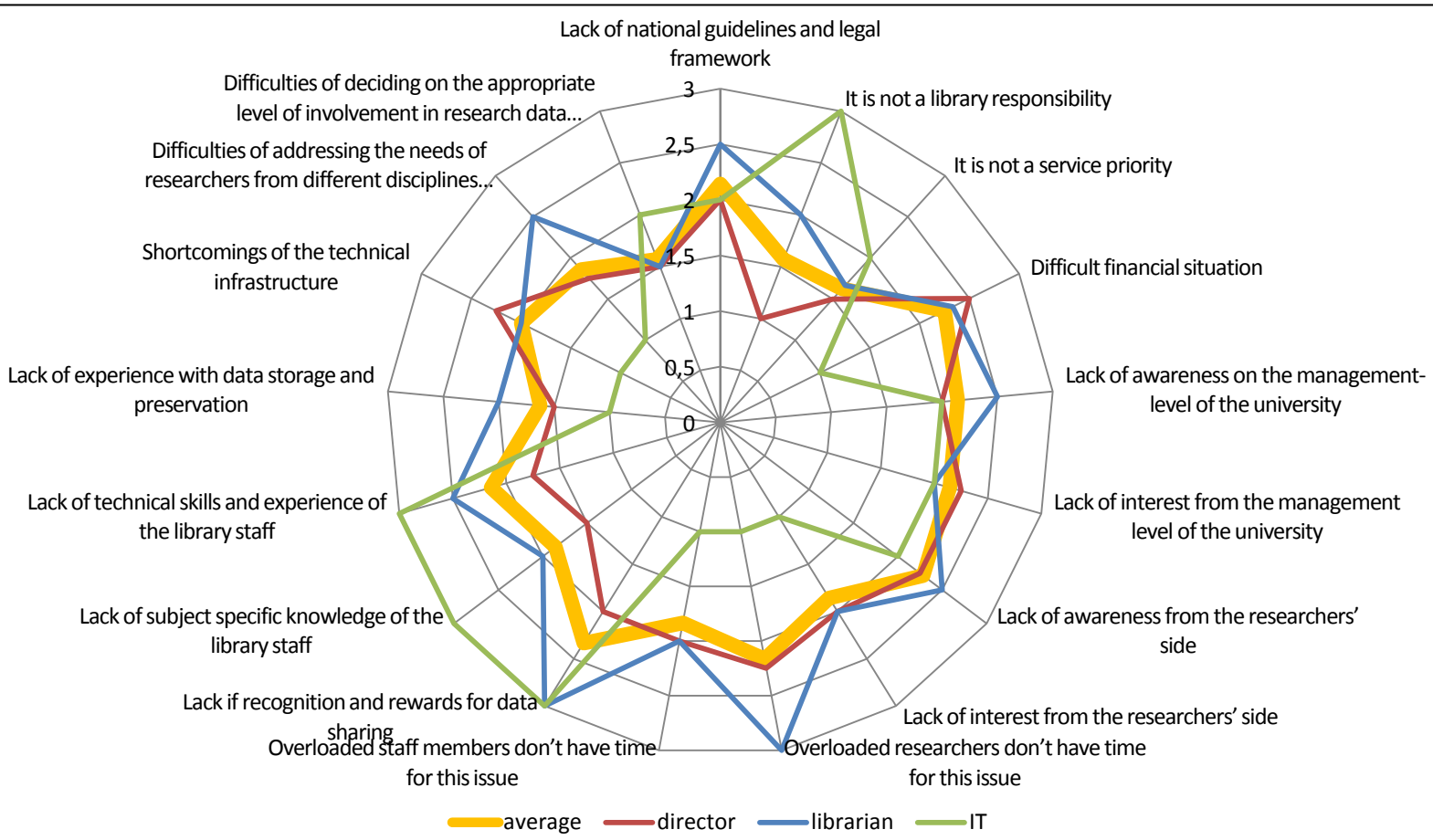


Figure 18: Perceived constraints of introducing RDM services grouped by job titles

Respondents were also asked to (on a level from low to high) name those constraints that could hinder their institutions in the process of introducing or providing RDM services, as it can be seen on figure 17. The top constraints were of an external nature: financial situation, lack of recognition and rewards for data sharing, and lack of awareness from researchers and the university management; while the lowest were internal: appropriate level of services, lack of experience with storage, not a library responsibility, not a service priority. The views of the participants were the most mixed in the area of lack of subject specific knowledge (4 low, 3 high) and overloaded researchers (2 low, 1 medium, 3 high, 1 don't know).

4.2 Discussion

The results of this thesis resonate with much of the findings and suggestions of previous, similar research done in the topic area. As is has been shown, respondents with different job titles gave rather mixed answers to the questions, and had considerably heterogeneous

thoughts about the levels libraries should engage with RDM. This trend between various stakeholders has been similarly pointed out by several studies earlier, for example by Cox et al. (2014), or Tenopir et al. (2014), who discovered "a mismatch between what academic research library directors believe they offer to their librarians and what the librarians themselves perceive to be available to them in the way of RDS". (Tenopir, Sandusky, Allard, & Birch, 2014, p. 6) And not only between directors and librarians: it has also been noted before that staff members with a profile in IT see "RDM as related to an infrastructure for active data storage", whereas librarians are "not concerned with the short-term storage of active data, but with the long-term storage of non-active data and with data sharing." (Verbaan & Cox, 2014, p. 7)

The participants' views on the persistence of large data centres presented in section 4.1.1 contradicts with Berman's (2008) statement, that "to minimize the likelihood of loss or damage and ensure the data will be there for a very long time […] stewardship by a trusted entity (such as libraries, archives, museums, universities, and institutional repositories), whose mission is the public good rather than profit, is generally required." (2008, p. 54) However, it still aligns with the recommendations of the Royal Society presented in section 2.5.2 (and more precisely on figure 9), that the different levels of data archiving solutions should complement each other. (Lynch, 2003, p. 331) The respondents' opinion about institutional repositories being more like 'last-resort facilities' also aligns with the Royal Society's data pyramid model, which stated that institutional repositories should be ideal for data "having potential value, but for which there is no Tier 1 or Tier 2 database available, and which can no longer be maintained by scientists actively using the data" (Royal Society, 2012, p. 63)

Both the Hungarian and Estonian answers which indicated that services on RDM should preferably offered on either a national or on an institutional level reinforce the views expressed in Denmark, where Kruse & Thestrup (2014) found that "storage and preservation facilities for researchers on an institutional and national level in the form of e-infrastructure(s) is the logical answer to the demand for a more open access to research data and research objects" (2014, p. 330) The noted heterogeneity of the participants' awareness on the global and local context of RDM is by no means a surprise: most of the previous surveys found that the regulation concerning research data and the services offered by libraries are rather hectic, for example, a recent survey about "science librarians

at institutions affiliated with the Association of Research Libraries" awareness of repositories and RDM services at their institutions (Antell, Foote, Turner, & Shults, 2014, p. 2) found that "uncertainty" was the strongest theme that emerged from their results. (2014, p. 17)

The respondents' views that libraries should more extensively offer services in RDM at the funding and approval stages of the research cycle than of the dissemination and experimentation stages partly resonate the observations of Peters & Dryden (2011), who found that "rather than physical storage capacity, researchers need assistance with funding agencies' data management requirements, the grant proposal process, finding campus data-related services, publication support, and targeted research assistance attendant to data management." (2011, p. 387) Contrasting this attitude with Gold's (2007b) 'upstream' and 'downstream' model, it is apparent that the libraries are willing to engage in both of the upstream and downstream stages (Gold, 2007b).

The participants' opinion on future service priorities fits in almost perfectly to the international discourse on RDM, where previous surveys found that "more libraries are offering or planning to offer informational/consultative-type services, rather than technical assistance services" (Tenopir et al., 2012, p. 17), that "consultation services such as identifying data sets, providing access to data, and articulating current standards for organization of data in specific subject areas seem to be a natural fit for subject librarians" (Soehner et al., 2010, p. 16) and that "priority apparently assigned to assistance with technology, infrastructure and tools, support for data deposit in an institutional repository, and development of institutional policy to manage data." (Corrall et al., 2013, p. 667)

The perceived constraints of introducing RDM services were mainly in line with the limitations discovered by previous studies: lack of appropriate funding, disciplinary differences, (Cox et al., 2014; Kruse & Thestrup, 2014; Soehner et al., 2010), and the lack of clear information and policies (Cox et al., 2014; Kruse & Thestrup, 2014). However, some of the previous studies noted that the library staff had skill gaps and lacked the confidence needed to successfully implement RDM services (Corrall et al., 2013; Soehner et al., 2010), which were only present in the current study in the opinion of the respondent with an IT background, but librarians and directors were rather confident about their competencies.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The aim of this study was to explore the readiness of libraries for research data management in countries where these libraries are very much at the beginning of implementing these services.

From the collected data, it became apparent that libraries are indeed at the beginning of addressing the issue, as at the time of the investigation, only a small amount of the participating institutions had services in place for supporting research data management at their institutions. Most of the libraries were at the planning or at the beginning of the implementation phase, or, according to the frameworks of capability maturity models, which were discussed in section 3.2, approximately at the first, 'acknowledge', or even lower maturity levels. If the study had assigned readiness-scores to individual institutions, Estonian academic libraries would have received higher numbers than the Hungarian ones.

The detailed conclusions of the thesis will follow the structure of the research questions raised in the first section. Generally speaking, the study found diverse levels of readiness levels amongst the participating institutions.

Are academic libraries aware on the current legal and political framework of research data management that applies to the researchers of their institution?

The answer to this question is not a simple 'no', since most of the institutions or national research funding agencies do not have formal requirements on research data, therefore it is not the libraries' fault if they do not know about the legal and political framework of RDM. Nevertheless, when asked about the current legal situation, respondents from both of the participating countries provided a rather mixed and even contradictory set of answers, which indicates that regardless of the varying awareness levels of the participants, the current national and institutional frameworks on RDM are still very much in a flux, probably because at the time of writing none of the participating institutions had clear nation-wide policies on research data, which, ultimately, causes the confusion amongst practitioners.

Are academic libraries aware of the data practices of researchers?

Yes and no. Virtually half of the participants stated that their institutions either carried out, or were in the middle of the process of carrying out a data audit aimed to discover the data practices and needs of their researchers. From the other half of the institutions one respondent noted that faculties or subject/liaison librarians are familiar with the practices of the researchers of their own department or discipline, but others admitted that they do not know about the researchers' practices, and could not even name any other person at their institution who would possess this knowledge. However, this does not necessary mean that these libraries are unaware of their researchers' practices, even though some gaps were found in the current study as well.

Are academic libraries ready to implement services on research data management?

Most of the participating institutions recognised the need of addressing the issue of RDM, and some of them were already providing research data services for their user community. However, the nature of research data makes it a relatively difficult problem to tackle; therefore, academic libraries alone will most certainly not be able to implement effective RDM services. The need of collaborating with other units of the university, as well as other institutions countrywide and internationally was expressed by the respondents, which indicates that they also aware that they are dealing with a complex issue. But as the answers showed, library managers and librarians are confident that they have the skills and competencies to tackle the problem.

How well would research data management fit into the current set of research support services provided by academic libraries?

The majority of the participating libraries offer a wide range of services, which could serve well for RDM services to build upon. They are actively engaged with the full spectrum of their user community through face-to-face consultancy and trainings, and most of the institutions have experience with institutional repositories, and implementing services on RDM could well extend these facilities. Although most of the participants stated that they have encountered unexpected problems while setting up their institutional repositories, these could serve as valuable lessons learned, and libraries could aim for not to make the same mistakes twice.

In which areas of research data management do librarians see themselves concentrating in the future?

Libraries saw themselves as playing both a proactive and reactive role in RDM. Even though there were noticeable gaps between various stakeholder groups, such as directors, librarians, and IT personnel on how they saw their libraries' role in research data management, respondents from basically all of these groups named that the libraries most important role would be in assisting the researchers with documentation and metadata. Other roles that were considered as important library roles included guidance in data discovery and citation, consultancy on data management plans, and training undergraduate students on data literacy.

What are the perceived constrains by academic libraries to introduce RDM services?

Interestingly, the majority of the provided constraints of offering RDM services were perceived as external: lack of clear national guidelines and frameworks, lack of sufficient funding, lack of recognition and rewards for data sharing, and lack of awareness from the university's management, as well as from the researchers side. However, most of the participants were confident in their skills, knowledge, competencies, and experience that are required to implement RDM services effectively. Although the questions about staffing were noted as constraints in previous studies, the respondents seemed confident in this issue, and the directors expressed their willingness to create specific positions or relocate existing staff to work in the area of RDM.

5.2 Recommendations

The study also has several recommendations for different stakeholder groups that could be associated with research data management.

For policy makers

Most of the uncertainties expressed by participants of this thesis originated in the lack of, or the discrepancies of the national and institutional policies about RDM. In order to overcome this confusion, policy makers, such as national funding agencies should implement clear and concise policies and requirements concerning research data management according to international practices in order to avoid inconsistencies.

Most of the participating institutions expressed that one of the highest constraint of introducing services in RDM were the lack of sufficient funding, therefore, governments should provide adequate funding for libraries to help them developing research data management services.

Research assessment policies and frameworks should recognise the sharing of research data more than they do nowadays. As long as scientists do not receive appropriate attention and rewards for a shared cited dataset, they will not engage in the issue.

For academic libraries

Although the following recommendations are not limited only to libraries and librarians, but are more universal, the basis where these originate from makes them a natural choice to include them in the section which is concerned with libraries: Ranganathan's five laws of library science translates very well to laws of research data management.

1. **Research data are for use**

   As it has been extensively discussed in the literature review, the value of research data is increasing with their use. Librarians and other stakeholders should help to ensure that research data are not disappearing after the research projects are finished, but are properly managed, preserved, and possibly reused by others.

2. **Every researcher his/her data**

   Although some of the participating institutions stated that they are aware of their researchers' practices and need, there were many gaps that could be filled, and librarians should initiate, or participate in studies to discover researchers' needs for research data management, thus implementing more diverse and tailored services to their community.

3. **Every data their researcher**

   Most of the previous studies discussed throughout this thesis stressed the fact that research data should be easily discoverable and reusable. Librarians could help this process by helping their research community with guidance on documentation and relevant metadata standards.

4. **Save the time of the researcher**

   Libraries should get more familiar with the level of participation of local researchers in international projects and discover what are required from them in these projects as regards RDM. In addition, libraries should get more familiar with the national and international context of research data management so when researchers seek help from them, they could provide them with answers in a timely manner.

5. **The library is a growing organism**

   Libraries should assess the capacity of their current infrastructure for data storage and preservation and develop these according to the requirements of their user community. In addition, libraries could seek consensus between stakeholders, and cooperate with other institutions to offer adequate services on RDM.

Implications for further research:

This study shed light on some blind spots, which were not anticipated in the planning stages of this thesis, but could serves as a good basis for further research.

One of the most striking of these were the level of how different units of the libraries seen or in fact, did not see their institutions' role in RDM. Further studies could examine this issue on a deeper level, gaining detailed insight from librarians, IT staff, and managers.

This study could also serve as a starting point of a potential follow-up study examining the extent that the library services which were envisioned by the participants of this study were implemented or not, and if not, why?

Most of the respondents were planning to extend the scope of their institutional repositories to handle data sets as well, and it would be worthwhile to assess the use and non-use of these institutional repositories along with subject-specific data centres by researchers.

REFERENCES

ACRL Research Planning and Review Committee. (2012). 2012 top ten trends in academic libraries: A review of the trends and issues affecting academic libraries in higher education. *College & Research Libraries News, 73*(6), 311–320.

ACRL Research Planning and Review Committee. (2014). Top trends in academic libraries: A review of the trends and issues affecting academic libraries in higher education. *College & Research Libraries News, 75*(6), 294–302.

Angevaare, I. (2009). Taking Care of Digital Collections and Data: "Curation" and Organisational Choices for Research Libraries. *LIBER Quarterly, 19*(1), 1–12.

Antell, K., Foote, J. B., Turner, J., & Shults, B. (2014). Dealing with Data: Science Librarians' Participation in Data Management at Association of Research Libraries Institutions. *College & Research Libraries*, Adavnce early publication.

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., … Wouters, P. (2004). An International Framework to Promote Access to Data. *Science, 303*(5665), 1777–1778. http://doi.org/10.1126/science.1095958

Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-molina, H., Klein, M. L., Messerschmitt, D. G., … Wright, M. H. (2003). *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. Retrieved from http://www.nsf.gov/cise/sci/reports/atkins.pdf

Auckland, M. (2012). Re-skilling for Research | RLUK. Retrieved December 12, 2013, from http://www.rluk.ac.uk/content/re-skilling-research

Australian National Data Service. (2011, August). Research data management framework: capability maturity guide. Retrieved May 30, 2014, from http://ands.org.au/guides/dmframework/dmf-capability-maturity-guide.pdf

Bell, G. (2009). Foreword. In T. Hey, S. Tansley, & K. M. Tolle (Eds.), *The fourth paradigm: data-intensive scientific discovery* (pp. XI–XV). Redmond, Wash.: Microsoft Research.

Berman, F. (2008). Got Data?: A Guide to Data Preservation in the Information Age. *Commun. ACM, 51*(12), 50–56. http://doi.org/10.1145/1409360.1409376

Block, W. C., Chen, E., Cordes, J., Dietrich, D., Krafft, D. B., Kramer, S., … Steinhart, G. (2010). *Meeting Funders' Data Policies: Blueprint for a Research Data Management Service Group (RDMSG)* (Working Paper). Retrieved from http://ecommons.library.cornell.edu/handle/1813/28570

Bloom, T. (2014, March 8). PLOS' New Data Policy: Part Two. Retrieved from http://blogs.plos.org/everyone/2014/03/08/plos-new-data-policy-public-access-data/

Blue Ribbon Task Force on Sustainable Digital Preservation and Access. (2010). *Sustainable economics for a digital planet: ensuring long-term access to digital information*. Retrieved from http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf

Borgman, C. L. (2007). *Scholarship in the digital age: Information, Infrastructure, and the Internet*. Cambridge, Mass.: MIT Press.

Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, *63*(6), 1059–1078. http://doi.org/10.1002/asi.22634

Brewerton, A. (2012). Re-Skilling for Research: Investigating the Needs of Researchers and How Library Staff Can Best Support Them. *New Review of Academic Librarianship*, *18*(1), 96–110. http://doi.org/10.1080/13614533.2012.665718

Budapest Open Access Initiative | Read the Budapest Open Access Initiative. (2002). Retrieved March 3, 2014, from http://www.budapestopenaccessinitiative.org/read

Carlson, J. R. (2013). Opportunities and Barriers for Librarians in Exploring Data: Observations from the Data Curation Profile Workshops. *Journal of eScience Librarianship*. http://doi.org/10.7191/jeslib.2013.1042

Carlson, S. (2006). Lost in a Sea of Science Data. *Chronicle of Higher Education*, *52*(42), A35.

CIBER. (2010). *Research Support Services in UK Universities*. London: Research Information Network. Retrieved from http://ciber-research.eu/download/20101116-RSS-report.pdf

CODATA-ICSTI Task Group on Data Citation Standards and Practices. (2013). Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data. *Data Science Journal*, *12*, CIDCR1–CIDCR75.

Commission Recommendation on access to and preservation of scientific information. (2012). Retrieved May 22, 2014, from https://ec.europa.eu/digital-agenda/node/66216

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. http://doi.org/10.1353/lib.2013.0005

Cox, A. M., & Pinfield, S. (2013). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*, Advance online publication. http://doi.org/10.1177/0961000613492542

Cox, A. M., Pinfield, S., & Smith, J. (2014). Moving a brick building: UK libraries coping with research data management as a "wicked" problem. *Journal of Librarianship and Information Science*, Advance online publication. http://doi.org/10.1177/0961000614533717

Cox, A. M., Verbaan, E., & Sen, B. (2012). Upskilling Liaison Librarians for Research Data Management. *Ariadne*, (70). Retrieved from http://www.ariadne.ac.uk/issue70/cox-et-al

Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *368*(1926), 4023–4038. http://doi.org/10.1098/rsta.2010.0165

Cronin, B. (2013). Thinking about data. *Journal of the American Society for Information Science and Technology*, *64*(3), 435–436. http://doi.org/10.1002/asi.22928

Crotty, D. (2014, March 4). PLOS' Bold Data Policy. Retrieved from http://scholarlykitchen.sspnet.org/2014/03/04/plos-bold-data-policy/

Data Citation Synthesis Group. (n.d.). Joint Declaration of Data Citation Principles - FINAL. Retrieved March 26, 2014, from https://www.force11.org/datacitation

DataCite Strategy 2013-2016. (2014, March 3). Retrieved March 10, 2014, from http://www.datacite.org/strategy2013-2016

Dickson, D. (1980). Research data: private property or public good? *Nature*, *284*(5754), 292. http://doi.org/10.1038/284292a0

Doorn, P., Dillo, I., & Van Horik, R. (2013). Lies, Damned Lies and Research Data: Can Data Sharing Prevent Data Fraud? *International Journal of Digital Curation*, *8*(1), 229–243. http://doi.org/10.2218/ijdc.v8i1.256

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory Building from Cases: Opportunities and Challenges. *Academy of Management Journal*, *50*(1), 25–32. http://doi.org/10.5465/AMJ.2007.24160888

Engineering and Physical Science Research Council. (2013a, February 25). EPSRC Policy Framework on Research Data. Retrieved February 17, 2014, from http://www.epsrc.ac.uk/about/standards/researchdata/

Engineering and Physical Science Research Council. (2013b, June 11). Expectations. Retrieved February 17, 2014, from www.epsrc.ac.uk/about/standards/researchdata/expectations/

Engineering and Physical Science Research Council. (2013c, September 5). Principles. Retrieved March 10, 2014, from http://www.epsrc.ac.uk/about/standards/researchdata/principles/

Erway, R. (2013). *Starting the Conversation: University-wide Research Data Management Policy*. Dublin, Ohio: OCLC Research. Retrieved from http://oclc.org/research/publications/library/2013/2013-08r.html

Expanding Public Access to the Results of Federally Funded Research | The White House. (2013, February 22). Retrieved March 14, 2014, from http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research

Fienberg, S. E., Martin, M. E., & Straf, M. L. (Eds.). (1985). *Sharing Research Data*. Washington, D.C.: National Academy Press. Retrieved from http://www.nap.edu/catalog.php?record_id=2033

Final Short Report | SIM4RDM. (2014, March 31). Retrieved May 9, 2014, from http://www.sim4rdm.eu/docs/project-outputs/final-short-report

Fleischer, D., & Jannaschk, K. (2011). A path to filled archives. *Nature Geoscience*, *4*(9), 575–576. http://doi.org/10.1038/ngeo1248

Friedlander, A., & Adler, P. (2006). *To Stand the Test of Time: Long-Term Stewardship of Digital Data Sets in Science and Engineering.* Association of Research Libraries.

G8 Science Ministers Statement - News stories - GOV.UK. (2013, June 13). Retrieved March 25, 2014, from https://www.gov.uk/government/news/g8-science-ministers-statement

Gabridge, T. (2009). The Last Mile: Liaison Roles in Curating Science and Engineering Research Data. *Research Library Issues*, (265), 15–21.

Ganley, E. (2014). PLOS data policy Catalyst for a better research process. *College & Research Libraries News*, *75*(6), 305–308.

Gold, A. (2007a). Cyberinfrastructure, Data, and Libraries, Part 1: A Cyberinfrastructure Primer for Librarians. *D-Lib Magazine*, *13*(9/10). http://doi.org/10.1045/september20september-gold-pt1

Gold, A. (2007b). Cyberinfrastructure, Data, and Libraries, Part 2: Libraries and the Data Challenge: Roles and Actions for Libraries. *D-Lib Magazine*, *13*(9/10). http://doi.org/10.1045/july20september-gold-pt2

Gold, A. (2010). Data Curation and Libraries: Short-Term Developments, Long-Term Prospects. Retrieved from http://works.bepress.com/agold01/9

Goodman, A., Pepe, A., Blocker, A. W., Borgman, C. L., Cranmer, K., Crosas, M., … Slavkovic, A. (2014). Ten Simple Rules for the Care and Feeding of Scientific Data. *PLoS Comput Biol*, *10*(4), e1003542. http://doi.org/10.1371/journal.pcbi.1003542

Gray, J., Liu, D. T., Nieto-Santisteban, M., Szalay, A., DeWitt, D. J., & Heber, G. (2005). Scientific Data Management in the Coming Decade. *SIGMOD Rec.*, *34*(4), 34–41. http://doi.org/10.1145/1107499.1107503

Guidelines on Data Management in Horizon 2020. (2013). Retrieved May 22, 2014, from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. (2013). Retrieved May 22, 2014, from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

Hamilton, M., Manuel, S., Whyte, A., & Jones, S. (2014, February 28). UK Higher Education Research Data Management Survey (Poster for IDCC 2014). Retrieved from http://figshare.com/articles/UK_Higher_Education_Research_Data_Management_Survey_Poster_for_IDCC_2014_/947679

Heidorn, P. B. (2011). The Emerging Role of Libraries in Data Curation and E-science. *Journal of Library Administration*, *51*(7-8), 662–672. http://doi.org/10.1080/01930826.2011.601269

Henty, M., Weaver, B., Bradbury, S., & Porter, S. (2008). Investigating Data Management Practices in Australian Universities. Retrieved from https://digitalcollections.anu.edu.au/handle/1885/47627

Hey, T. (2010). The Next Scientific Revolution. *Harvard Business Review*, *88*(11), 56–63.

Hey, T., & Hey, J. (2006). e-Science and its implications for the library community. *Library Hi Tech*, *24*(4), 515–528. http://doi.org/10.1108/07378830610715383

Hey, T., Tansley, S., & Tolle, K. M. (Eds.). (2009a). Jim Gray on eScience: A Transformed Scientific Method. In *The fourth paradigm: data-intensive scientific discovery* (pp. XVII–XXXI). Redmond, Wash.: Microsoft Research.

Hey, T., Tansley, S., & Tolle, K. M. (Eds.). (2009b). *The fourth paradigm: data-intensive scientific discovery*. Redmond, Wash.: Microsoft Research.

High level Expert Group on Scientific Data. (2010). *Riding the wave: How Europe can gain from the rising tide of scientific data*. European Union. Retrieved from http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

Hitchcock, S. (2013, June 25). The effect of open access and downloads ('hits') on citation impact: a bibliography of studies. Retrieved June 19, 2014, from http://opcit.eprints.org/oacitation-biblio.html

Hodson, S., & Jones, S. (2013, July 16). Seven rules of successful research data management in universities. *The Guardian*. Retrieved from http://www.theguardian.com/higher-education-network/blog/2013/jul/16/research-data-management-top-tips

Hswe, P. (2012). Data Management Services in Libraries. In N. Xiao & L. R. McEwen (Eds.), *Special Issues in Data Management* (Vol. 1110, pp. 115–128). Washington, DC: American Chemical Society. Retrieved from http://pubs.acs.org/doi/book/10.1021/bk-2012-1110

Hswe, P., & Holt, A. (2011). Joining in the enterprise of response in the wake of the NSF data management planning requirement. *Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC*, *274*, 11–17.

Hype Cycle for Education, 2014. (n.d.). Retrieved August 10, 2014, from https://www.gartner.com/doc/2806424/hype-cycle-education-

Integrating with integrity. (2010). *Nature Genetics*, *42*(1), 1–1. http://doi.org/10.1038/ng0110-1

Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. (2014). *NMC Horizon Report: 2014 Library Edition*. Austin, Texas: The New Media Consortium. Retrieved from http://cdn.nmc.org/media/2014-nmc-horizon-report-library-EN.pdf

Joint, N. (2007). Data preservation, the new science and the practitioner librarian. *Library Review*, *56*(6), 451–455. http://doi.org/10.1108/00242530710760337

Jones, S., Pryor, G., & Whyte, A. (2013). *How to Develop Research Data Management Services - a guide for HEIs*. Edinburgh: Digital Curation Centre. Retrieved from http://www.dcc.ac.uk/resources/how-guides

Kenney, A. R., & McGovern, N. Y. (2003). The Five Organizational Stages of Digital Preservation. In *Digital Libraries: A Vision for the 21st Century: A Festschrift in Honor of Wendy Lougee on the Occasion of her Departure from the University of Michigan*. Retrieved from http://quod.lib.umich.edu/cgi/t/text/text-idx?c=spobooks;idno=bbv9812.0001.001;rgn=div1;view=text;cc=spobooks;node=bbv9812.0001.001%3A11

Kim, J. (2013). Data sharing and its implications for academic libraries. *New Library World*, *114*(11/12), 494–506. http://doi.org/10.1108/NLW-06-2013-0051

Kolowich, S. (2014, May 12). Hazards of the Cloud: Data-Storage Service's Crash Sets Back Researchers. Retrieved from http://chronicle.com/blogs/wiredcampus/hazards-of-the-cloud-data-storage-services-crash-sets-back-researchers/52571

Koslow, S. H. (2000). Should the neuroscience community make a paradigm shift to sharing primary data? *Nature Neuroscience*, *3*(9), 863–865. http://doi.org/10.1038/78760

Kowalczyk, S., & Shankar, K. (2011). Data sharing in the sciences. *Annual Review of Information Science and Technology*, *45*(1), 247–294. http://doi.org/10.1002/aris.2011.1440450113

Kramer, B., & Bosman, J. (2015). 101 Innovations in Scholarly Communication - the Changing Research Workflow. Figshare. Retrieved from http://dx.doi.org/10.6084/m9.figshare.1286826

Kruse, F., & Thestrup, J. B. (2014). Research Libraries' New Role in Research Data Management, Current Trends and Visions in Denmark. *LIBER Quarterly*, *23*(4), 310–335.

Lewis, M. J. (2010). Libraries and the management of research data. In S. McKnight (Ed.), *Envisioning Future Academic Library Services* (pp. 145–168). London: Facet Publishing.

Lougee, W., Choudhury, S., Gold, A., Humphrey, C., Humphreys, B., Luce, R., … Young, P. (2007). *Agenda for Developing E-Science in Research Libraries Final Report and Recommendations to the Scholarly Communication Steering Committee, the Public Policies Affecting Research Libraries Steering Committee, and the Research, Teaching, and Learning Steering Committee*. Association of Research Libraries. Retrieved from http://www.arl.org/bm~doc/ARL_EScience_final.pdf

Lynch, C. (2003). Institutional Repositories: Essential Infrastructure For Scholarship In The Digital Age. *Portal: Libraries and the Academy*, *3*(2), 327–336. http://doi.org/10.1353/pla.2003.0039

Lynch, C. (2008). Big data: How do your data grow? *Nature*, *455*(7209), 28–29. http://doi.org/10.1038/455028a

Lynch, C. (2013). "The Next Generation of Challenges in the Curation of Scholarly Data." In J. M. Ray (Ed.), *Research Data Management: Practical Strategies for Information Professionals*. West Lafayette, IL: Purdue University Press. Retrieved from http://www.cni.org/publications/cliffs-pubs/next-generation-challenges-scholarly-data-curation/

Lyon, L. (2012). The Informatics Transform: Re-Engineering Libraries for the Data Decade. *International Journal of Digital Curation*, *7*(1), 126–138. http://doi.org/10.2218/ijdc.v7i1.220

Lyon, L., Ball, A., Duke, M., & Day, M. (2012). Developing a Community Capability Model Framework for data-intensive research. In R. Moore, K. Ashley, & S. Ross (Eds.), *iPres 2012: Proceedings of the 9th International Conference on Preservation of Digital Objects* (pp. 9–16). Toronto: University of Toronto Faculty of Information. Retrieved from https://ipres.ischool.utoronto.ca/sites/ipres.ischool.utoronto.ca/files/iPres%202012%20Conference%20Proceedings%20Final.pdf

Macdonald, S., & Martinez-Uribe, L. (2010). Collaboration to Data Curation: Harnessing Institutional Expertise. *New Review of Academic Librarianship*, *16*(sup1), 4–16. http://doi.org/10.1080/13614533.2010.505823

Maceviciute, E. (2014). Research libraries in a modern environment. *Journal of Documentation*, *70*(2), 282–302. http://doi.org/10.1108/JD-04-2013-0044

Max Planck Open Access | Berlin Declaration. (2003). Retrieved March 3, 2014, from http://openaccess.mpg.de/Berlin-Declaration

McGlynn, T. (2014, March 3). I own my data, until I don't. Retrieved from http://smallpondscience.com/2014/03/03/i-own-my-data-until-i-dont/

Melnarowicz, E. (2012). Because research data matters : a case at the Loughborough University Library [Master thesis]. Retrieved February 25, 2014, from https://oda.hio.no/jspui/handle/10642/1273

Merton, R. K. (1973). *The sociology of science: theoretical and empirical investigations*. Chicago: University of Chicago Press.

Miller, K. (2012). *5 Steps to Research Data Readiness*. Edinburgh: Digital Curation Centre. Retrieved from http://www.dcc.ac.uk/resources/briefing-papers/five-steps-research-data-readiness

Monastersky, R. (2013). Publishing frontiers: The library reboot. *Nature*, *495*(7442), 430–432. http://doi.org/10.1038/495430a

National Research Council. (1999). *A Question of Balance: Private Rights and the Public Interest in Scientific and Technical Databases*. Washington, DC: The National Academies Press. Retrieved from http://www.nap.edu/catalog.php?record_id=9692

National Science Board. (2005). *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century* (No. NSB-05-40). National Science Foundation. Retrieved from http://www.nsf.gov/pubs/2005/nsb0540/start.jsp

National Science Board. (2014). *Reducing Investigators' Administrative Workload For Federally Funded Research* (No. NSB-14-18). Arlington, VA: National Science Foundation. Retrieved from http://nsf.gov/publications/pub_summ.jsp?ods_key=nsb1418

Nelson, B. (2009). Data sharing: Empty archives. *Nature News*, *461*(7261), 160–163. http://doi.org/10.1038/461160a

Nielsen, H. J., & Hjørland, B. (2014). Curating research data: the potential roles of libraries and information professionals. *Journal of Documentation*, *70*(2), 221–240. http://doi.org/10.1108/JD-03-2013-0034

NSF Data Management Plan Requirements. (n.d.). Retrieved March 9, 2014, from http://www.nsf.gov/eng/general/dmp.jsp

OECD. (2002). *Frascati Manual: Proposed Standard Practice for Surveys on Research and Experimental Development*. OECD Publishing. Retrieved from http://www.keepeek.com/Digital-Asset-Management/oecd/science-and-technology/frascati-manual-2002_9789264199040-en#page31

OECD Principles and Guidelines for Access to Research Data from Public Funding. (2007). OECD. Retrieved from http://www.oecd.org/sti/sci-tech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm

Ogburn, J. L. (2010). The Imperative for Data Curation. *Portal: Libraries and the Academy*, *10*(2), 241–246. http://doi.org/10.1353/pla.0.0100

Parry, O., & Mauthner, N. S. (2004). Whose Data are They Anyway? Practical, Legal and Ethical Issues in Archiving Qualitative Research Data. *Sociology*, *38*(1), 139–152. http://doi.org/10.1177/0038038504039366

Parsons, M. A., & Fox, P. A. (2013). Is Data Publication the Right Metaphor? *Data Science Journal*, *12*, WDS32–WDS46.

Parsons, T. (2013). Creating a Research Data Management Service. *International Journal of Digital Curation*, *8*(2), 146–156. http://doi.org/10.2218/ijdc.v8i2.279

Pawson, R. (2006). *Evidence-Based Policy: A Realist Perspective*. London [etc.]: SAGE.

Peng, R. D. (2011). Reproducible Research in Computational. *Science*, *334*(6060), 1226–1227. http://doi.org/10.1126/science.1213847

Peplow, M. (2014). Organic synthesis: The robo-chemist. *Nature*, *512*(7512), 20–22. http://doi.org/10.1038/512020a

Peters, C., & Dryden, A. R. (2011). Assessing the Academic Library's Role in Campus-Wide Research Data Management: A First Step at the University of Houston. *Science & Technology Libraries*, *30*(4), 387–403. http://doi.org/10.1080/0194262X.2011.626340

Pickard, A. J. (2007). *Research methods in information*. London: Facet.

Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE*, *2*(3), e308. http://doi.org/10.1371/journal.pone.0000308

Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, *1*, e175. http://doi.org/10.7717/peerj.175

PLOS ONE : accelerating the publication of peer-reviewed science. (n.d.). Retrieved March 9, 2014, from http://www.plosone.org/static/policies#sharing

Polanyi, M. (2000). The Republic of Science: Its Political and Economic Theory. *Minerva*, *38*(1), 1–21. http://doi.org/10.1023/A:1026591624255

Popper, K. R. (2002). *The poverty of historicism*. London: Routledge.

Pryor, G. (2013). A Maturing Process of Engagement: Raising Data Capabilities in UK Higher Education. *International Journal of Digital Curation*, *8*(2), 181–193. http://doi.org/10.2218/ijdc.v8i2.282

RCUK Common Principles on Data Policy - RCUK. (n.d.). Retrieved February 17, 2014, from http://www.rcuk.ac.uk/research/datapolicy/

re3data.org team. (2014, November 20). Over 1,000 research data repositories indexed in re3data.org | re3data.org. Retrieved from http://www.re3data.org/2014/11/over-1000-research-data-repositories-indexed-in-re3data-org/

Research Information Network. (2008). *Stewardship of digital research data: a framework of principles and guidelines Responsibilities of research institutions and funders, data managers, learned societies and publishers*. Retrieved from http://www.rin.ac.uk/our-work/data-management-and-curation/stewardship-digital-research-data-principles-and-guidelines

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*(2), 155–169. http://doi.org/10.1007/BF01405730

Royal Society. (2012). *Science as an open enterprise*. London: the Royal Society. Retrieved from http://royalsociety.org/policy/projects/science-public-enterprise/report/

Ruusalepp, R. (2008). *Comparative Study of International Approaches to Enabling the Sharing of Research Data*. JISC. Retrieved from http://www.jisc.ac.uk/publications/reports/2008/nationaldatafinalreport.aspx

Salo, D. (2008). Innkeeper at the Roach Motel. *Library Trends*, *57*(2), 98–123. http://doi.org/10.1353/lib.0.0031

Salo, D. (2010). Retooling Libraries for the Data Challenge. *Ariadne*, (64). Retrieved from http://www.ariadne.ac.uk/issue64/salo

Santos, F. M., & Eisenhardt, K. M. (2004). Multiple Case Study. In *Encyclopedia of Social Science Research Methods Encyclopedia of social science research methods*. Thousand Oaks, California: SAGE. Retrieved from http://knowledge.sagepub.com/view/socialscience/n596.xml

Sayogo, D. S., & Pardo, T. A. (2013). Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. *Government Information Quarterly*, *30*, *Supplement 1*, S19–S31. http://doi.org/10.1016/j.giq.2012.06.011

Science Europe. (2013, December). Science Europe Roadmap. Retrieved from http://www.scienceeurope.org/uploads/PublicDocumentsAndSpeeches/ScienceEurope_Roadmap.pdf

Science Europe - Member Organisations. (n.d.). Retrieved May 20, 2014, from http://www.scienceeurope.org/about-us/member-organisations

Sieber, J. E. (1988). Data sharing: Defining problems and seeking solutions. *Law and Human Behavior*, *12*(2), 199–206. http://doi.org/10.1007/BF01073128

Sieber, J. E. (1991). Openness in the Social Sciences: Sharing Data. *Ethics & Behavior*, *1*(2), 69.

Silva, L. (2014, February 24). PLOS' New Data Policy: Public Access to Data. Retrieved from http://blogs.plos.org/everyone/2014/02/24/plos-new-data-policy-public-access-data-2/

Smith, D., & Carrano, A. (1996). International Large-Scale Sequencing Meeting. *Human Genome News*, *7*(6). Retrieved from http://web.ornl.gov/sci/techresources/Human_Genome/publicat/hgn/v7n6/19intern.shtml

Soehner, C., Steeves, C., & Ward, J. (2010). E-Science and Data Support Services: A Study of ARL Member Institutions. *Association of Research Libraries*. Retrieved from http://eric.ed.gov/?id=ED528643

Stevenson, A. (Ed.). (2010a). Data. *Oxford Dictionary of English* (3rd ed.). Oxford: Oxford University Press. Retrieved from http://www.oxfordreference.com/view/10.1093/acref/9780199571123.001.0001/m_en_gb0205520

Stevenson, A. (Ed.). (2010b). Datum. *Oxford Dictionary of English* (3rd ed.). Oxford: Oxford University Press. Retrieved from http://www.oxfordreference.com/10.1093/acref/9780199571123.001.0001/m_en_gb0205860

Strasser, C. (2014, March 13). Lit Review: #PLOSFail and Data Sharing Drama. Retrieved from http://datapub.cdlib.org/2014/03/13/lit-review-plosfail-and-data-sharing-drama/

Swan, A., & Brown, S. (2008). *The skills, role and career structure of data scientists and curators: An assessment of current practice and future needs*. Retrieved from

http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dataskillscareersfinalreport.pdf

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., … Frame, M. (2011). Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE*, *6*(6), e21101. http://doi.org/10.1371/journal.pone.0021101

Tenopir, C., Birch, B., & Allard, S. (2012). *Academic libraries and research data services: Current Practices and Plans for the Future An ACRL White Paper*. Association of College and Research Libraries. Retrieved from http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf

Tenopir, C., Sandusky, R. J., Allard, S., & Birch, B. (2013). Academic librarians and research data services: preparation and attitudes. *IFLA Journal*, *39*(1), 70–78. http://doi.org/10.1177/0340035212473089

Tenopir, C., Sandusky, R. J., Allard, S., & Birch, B. (2014). Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research*, Advance online publication. http://doi.org/10.1016/j.lisr.2013.11.003

Thaesis, & van der Hoeven, J. (2010). *Insight Report: PARSE.Insight: INSIGHT into issues of Permanent Access to the Records of Science in Europe*. Retrieved from http://www.parse-insight.eu/publications.php#d3-6

Time wasted. (2014). *Nature*, *509*(7501), 523–523. http://doi.org/10.1038/nj7501-523a

Uhlir, P. F. (2010). Information Gulags, Intellectual Straightjackets, and Memory Holes: *Data Science Journal*, *9*, ES1–ES5. http://doi.org/10.2481/dsj.Essay-001-Uhlir

Van den Berg, M., Dalsgaard, B. C., Grim, R., Horstmann, W., Jansen, D., Pollard, T., & Roos, A. (2012). Ten recommendations for libraries to get started with research data management: Final report of the LIBER working group on E-Science / Research Data Management. Retrieved March 3, 2014, from http://www.libereurope.eu/news/ten-recommendations-for-libraries-to-get-started-with-research-data-management

Van der Graaf, M., & Waaijers, L. (2011). *A Surfboard for Riding the Wave: Towards a four country action programme on research data.* Knowledge Exchange. Retrieved from http://www.knowledge-exchange.info/surfboard

Van Noorden, R. (2013a). Data-sharing: Everything on display. *Nature*, *500*(7461), 243–245. http://doi.org/10.1038/nj7461-243a

Van Noorden, R. (2013b). Half of 2011 papers now free to read. *Nature*, *500*(7463), 386–387. http://doi.org/10.1038/500386a

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

Vines, T. H., Andrew, R. L., Bock, D. G., Franklin, M. T., Gilbert, K. J., Kane, N. C., … Yeaman, S. (2013). Mandated data archiving greatly improves access to research data. *The FASEB Journal*, *27*(4), 1304–1308. http://doi.org/10.1096/fj.12-218164

Walters, T., & Skinner, K. (2011). *New Roles for New Times: Digital curation for preservation*. Washington, D.C.: Association of Research Libraries. Retrieved from http://www.arl.org/storage/documents/publications/nrnt_digital_curation17mar11.pdf

Walton, G. (2010). Data Curation and the Academic Library. *New Review of Academic Librarianship*, *16*(1), 1–3. http://doi.org/10.1080/13614531003640427

Ward, C., Freiman, L., Molloy, L., Jones, S., & Snow, K. (2011). Making sense: Talking data management with researchers. *International Journal of Digital Curation*, *6*(2). Retrieved from http://eprints.gla.ac.uk/49201/

Weber, N. M., Palmer, C. L., & Chao, T. C. (2012). Current Trends and Future Directions in Data Curation Research and Education. *Journal of Web Librarianship*, *6*(4), 305–320. http://doi.org/10.1080/19322909.2012.730358

Welcome, Scientific Data! (2014). *Nature*, *509*(7502), 534–534. http://doi.org/10.1038/509534a

Whyte, A., & Allard, S. (2014). *How to Discover Requirements for Research Data Management Services*. Edinburgh: Digital Curation Centre. Retrieved from http://www.dcc.ac.uk/how-discover-requirements

Whyte, A., & Tedds, J. (2011). *Making the Case for Research Data Management*. Edinburgh: Digital Curation Centre. Retrieved from http://www.dcc.ac.uk/resources/briefing-papers

Zins, C. (2006). Redefining information science: from "information science" to "knowledge science." *Journal of Documentation*, *62*(4), 447–461. http://doi.org/10.1108/00220410610673846

Zins, C. (2007). Conceptual approaches for defining data, information, and knowledge. *Journal of the American Society for Information Science and Technology*, *58*(4), 479–493. http://doi.org/10.1002/asi.20508

APPENDIX A – QUESTIONNAIRE FORM

In the last decades, technological advancements have enabled researchers to produce, analyze, store, and share digital data on a continuously growing scale. Data sharing has been common practice in certain disciplines for a long time, while other disciplines are only starting to recognize its benefits. Lately, influenced by the open-access movement, governments, funding agencies, and journal publishers have begun to promote the sharing of research data:

• governments and political forums are issuing recommendations on the availability of publicly funded research data[1]

• a growing number of funding agencies require researchers to include a data management plan in the grant proposal and ensure that research data are accessible after the project is finished[2]

• numerous journal publishers are starting to require data sets as supplementary materials to articles[3].

This raising level of awareness on data preservation adds new responsibilities to researchers, as it requires significant effort to manage research data effectively. Many have argued that academic and research libraries could serve as an important partner in research data management, since many of their traditional practices (organizing, preserving, and disseminating information) overlap with the practices necessary to manage research data. However, as digital research data is quite different in its nature from traditional books and journals, there is also a concern over the involvement level of libraries.

My thesis is concentrating mainly on countries where the awareness on research data is only emerging. I am interested in academic librarians' attitudes towards research data management, and my aim is to discover whether academic libraries are well positioned to introduce research data management services to their user community.

I am asking for your participation as I believe that hands-on experiences, as well as knowledge about the broader context where institutions are situated provide library practitioners with a unique perspective. This perspective could be a valuable contribution to the ongoing discussion about the capacity of academic libraries to provide research data management services.

Completing the questionnaire should take about 30 minutes.

Please be assured that your participation is voluntarily, that you can freely withdraw at any time, and that no personal data which can help to identify you or your institution will be used in my thesis or other publications derived from it.

If you have any further questions or concerns, please contact me at deradam42@gmail.com

Sincerely,
Adam Der
MA student

I understood the purpose of the study and agree to take part (tick)

[1] See e.g. "G8 Scienc Ministers Statement - News Stories - GOV.UK," June 13, 2013, https://www.gov.uk/government/news/g8-science-ministers-statement.

[2] See e.g. "Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020," 2013, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

[3] See e.g. "PLOS Editorial and Publishing Policies," accessed May 23, 2014, http://www.plosone.org/static/policies.

1. Name of your institution: (open-ended question)

2. Your role in the institution: (open-ended question)

3. E-mail address (provide only if you are open for possible follow-up questions): (open-ended question)

4. Work experience, years: (open-ended question)

5. Are there any specific disciplines your library is specialising on? Can you please provide these? (open-ended question)

6. How do you agree with the following statements?

	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Don't know
"Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property."[4]						
"Data with acknowledged long-term value should be preserved and remain accessible and usable for future research."[5]						
"To enable research data to be discoverable and effectively re-used by others, sufficient metadata should be recorded and made openly available to enable other researchers to understand the research and re-use potential of the data."[6]						
"Libraries are the researchers' natural collaborators in developing and implementing the systems necessary for the management of research data"[7]						
"Librarians are a natural partner in addressing issues in research data management and curation given [their] knowledge and skills in organizing, disseminating, and preserving diverse sets of materials."[8]						
"In Central and Eastern Europe financial resources are needed for digitization in the first place, and questions of data management, data exposure, data organization can only be addressed afterwards."[9]						

7. Could you please provide a link to, or copy your library's mission statement in the following box? (open)

8.a. Does your library operate an institutional repository? (single choice)

- No (skip to q.9.)
- It's currently under development
- Yes
- Don't know (skip to q.9.)

8.b. If yes, is it possible to store data sets in the repository? (single choice)

- Not
- It's currently under development
- Yes
- Don't know

[4] "RCUK Common Principles on Data Policy - RCUK," accessed February 17, 2014, http://www.rcuk.ac.uk/research/datapolicy/.
[5] Ibid.
[6] Ibid.
[7] Filip Kruse and Jesper Boserup Thestrup, "Research Libraries' New Role in Research Data Management, Current Trends and Visions in Denmark," *LIBER Quarterly* 23, no. 4 (April 15, 2014): 310–35.
[8] Jake Carlson et al., "Developing an Approach for Data Management Education: A Report from the Data Information Literacy Project," *International Journal of Digital Curation* 8, no. 1 (June 20, 2013): 204–17, doi:10.2218/ijdc.v8i1.254.
[9] "Final Short Report | SIM4RDM," March 31, 2014, http://www.sim4rdm.eu/docs/project-outputs/final-short-report.

8.c. How do you agree with the following?

	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Don't know
"Setting up our institutional repository took more effort than originally imagined"[10]	(Skip to q.9.)					(Skip to q.9.)

8.d. Could you please describe the main issues or problems that you encountered when setting up and operating your institutional repository? (open-ended question)

9. How do you agree with the following?

	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Don't know
"Large, international databases are more likely to persist than small, local ones."[11]						
Small, local databases serve the needs of the users better than large, international, generic data centres						
Small, local databases serve the needs of the users better than large, international, subject-specific data centres						
Small, local data archives would more likely serve as last-resort facilities when relevant domain specific archives don't exist						

10. Does your library provide consultancy, training, or websites on the following?

	Consultancy	Training	Web guides	Other	No service	Don't know
Research methods						
Information literacy						
Open access						
Scholarly communication						
Bibliometrics						
Other, please state						

10.b. Which groups are in the primary focus of these services? (leave it blank if you have no such services)

	Undergraduate students	Postgraduate /doctoral students	Researchers	Faculty staff	No specific target group	Don't know
Research methods						
Information literacy						
Open access						
Scholarly communication						
Bibliometrics						
Other (previously stated)						

11.a. Do funding agencies in your country have a policy on open access publications? (single choice)

- No
- No, but it's currently under development
- Only informal recommendations
- Yes, some of the funding agencies
- Yes, but only covers certain disciplines
- Yes
- Don't know

[10] Elena Maceviciute, "Research Libraries in a Modern Environment," *Journal of Documentation* 70, no. 2 (March 4, 2014): 282–302, doi:10.1108/JD-04-2013-0044.
[11] "PLOS Editorial and Publishing Policies."

11.b. Do funding agencies in your country have a policy on research data? (single choice)

No
No, but it's currently under development
Only informal recommendations
Yes, some of the funding agencies
Yes, but only covers certain disciplines
Yes
Don't know

12. Is the managing and sharing of research data a priority in your country? (single choice)

Not
It's not a standard, but it's getting recognized nowadays
Only in certain disciplines
Yes
Other
Don't know

13.a. Does your university have a policy on open access publications? (single choice)

No
No, but it's currently under development
Only informal recommendations
Yes, some faculties/departments have their own
Yes
Don't know

13.b. Does your university have a policy on research data? (single choice)

No
No, but it is currently under development
Only informal recommendations
Yes, some faculties/departments have their own
Yes
Don't know

14.a Have your university/library conducted, or participated in studies or audits which were aimed to discover the produced data, and the data handling practices of researchers? (single choice)

No
No, but it is currently planned
Yes, on a faculty level (skip to q.14.c)
Yes, university-wide audit (skip to q.14.c)
Yes, on a national, cross-institutional level (skip to q.14.c)
Yes, on an international level (skip to q.14.c)
Don't know

14.b. If not, are you or your institution aware of the data practices of the researchers at you university? (multiple choice) (skip to q.15)

No
The departments/faculties are aware of the practices and needs of their researchers
Yes, subject/liaison librarians know the practices and needs of the researchers in their field
Yes, the research office is responsible for this
Other, please state
Don't know

14.c. If yes, could you please provide a link, or send me relevant documents on e-mail, where I can find out more about this study? (open)

14.d. If yes, do you know if the following aspects were covered? (multiple choice)

- Size
- Type
- File formats
- Location/storage media
- Subject
- Purpose
- Level of documentation
- Metadata standards
- Creator(s)
- Management responsibilities
- Usage/update frequency
- Hardware and software requirements
- Potential reuse
- Intellectual property rights and ownership
- Restrictions/level of openness
- Other, please state
- Don't know

14.e. If yes, do you think it was helpful? (single choice)

- No, there was no real interest in participating
- Yes, we got more aware of researchers' practices
- Yes, and we shaped our services according to the articulated practices and needs
- Other, please state
- Don't know

15. If researchers at your institution receive financial support from international funding agencies, participate in international projects, or publish in international journals, are they required to…

	Never	Occasionally	Always	Don't know
Submit data management plans for grant proposals?				
Deposit their data in data archives?				
Add their data sets as supplementary materials to journal articles?				
They are not required to submit the data anywhere, but have to provide it upon request				

16. How do you agree with the following statements?

	Disagree	Neutral	Depends on the discipline	Agree	Don't know
Researchers at my institution are willing to share their data					
Researchers at my institution are willing to reuse other researchers' data					
Research data is recognized as a valuable asset at my institution					
Researchers know about, and use relevant metadata standards in their own discipline					
Researchers know about, and submit their data to relevant disciplinary or general data centres					
Researchers should be provided with support to effectively manage their data					
There is a need from the side of researchers for training and support in data management					

17.a Do you know who is usually responsible for the management of data during the research projects? (multiple choice)

- Nobody
- Research team member, PhD students
- Principal investigators
- Faculty/department
- Local research office
- IT services
- Academic library
- Local data centre
- National research council
- Third party data centres
- Other (please state)
- Don't know

17.b Do you know who is usually responsible for the management of data after the research projects? (multiple choice)

- Nobody
- Research team member, PhD students
- Principal investigators
- Faculty/department
- Local research office
- IT services
- Academic library
- Local data centre
- National research council
- Third party data centres
- Other (please state)
- Don't know

18. In your opinion, where would the researchers seek help, if they needed support with their data, e.g. on preparing data management plans, adding metadata, finding datasets, archiving and preserving datasets, etc.? (multiple choice)

- They would figure out themselves
- Other team members/PhD students
- Principal investigators
- Faculty/department
- Local research office
- IT services
- Academic library
- Local data centre
- National research council
- Third party data centres
- Other (please state)
- Don't know

19.a Have any researchers at your institution approached the library for help with their data, e.g. on preparing data management plans, adding metadata, finding datasets, archiving and preserving datasets, etc.? (single choice)

- Not (skip to q.20.)
- Yes
- Don't know (skip to q.20.)

19.b. If yes, could you please provide what the questions were about? (open question)

20. In your opinion, which of the following would be best suited to help researchers with their data management needs? (multiple choice)

- Local research office
- IT services
- Academic library
- National research council
- A national, cross-university partnership
- International centres specialized in data curation
- Other (please state)
- Don't know

21. Do you think that academic and research libraries should play a role in the management of research data? (single choice)

- Not
- Yes, a rather reactive role, based on the needs of the user community
- Yes, a rather proactive, innovative role
- Yes, both proactive and reactive
- Don't know

22. In your opinion, what would be an appropriate level to provide research data management services? (multiple choice)

- Faculty level
- Institutional level
- Cross-institutional partnership level
- National level
- International level
- Other, please state
- Don't know

23. In your opinion, which of the following would serve as a potential collaborator for your library to offer research data management services? (multiple choice)

- Faculties/departments
- IT services
- Research office
- Other academic libraries in the country
- Other academic libraries internationally
- National library
- National research council(s)
- Relevant international organizations
- Other, please state
- Don't know

24. In your opinion, which stage of the research process would it be the most appropriate to provide advice and support on data management? (multiple choice)

- At the stage of idea discovery, e.g. direct researchers to relevant data archives and help them to find and cite datasets
- At the stage of planning and the grant proposal, e.g. assist with data management plans
- At the stage of data collection and analysis, e.g. assist with tools and documentation
- At the stage of disseminating the results, e.g. storage, preservation, publication, and measuring impact
- Other, please state
- Don't know

25. In your opinion, which of the following should be the libraries' role in research data management?

	Not a library role	Rather not a library role	Neutral	Rather a library role	Definitely a library role	Don't know
Discover and address the needs of the researchers						
Engage high level university management in the issue						
Develop or assist in the development of policies on research data						
Raise awareness for data management and sharing in the community						
Advocacy and guidance on data management planning						
Provide technical infrastructure for storage and preservation						
Assist with tools for data analysis						
Assist with documentation and metadata standards						
Provide guidance on data discovery and citation						
Provide training for students on data literacy						
Provide training for researchers on data management						
Provide advice on intellectual property rights and licensing						
Other, please state						
Other, please state						

26.a Given the current situation of your institution, how likely is it that in the next three years your library will…?

	Unlikely	Rather unlikely	Rather likely	Likely	Already in place	Don't know
Offer consultancy on data management plans						
Consult with researchers on metadata creation and standards						
Create metadata for research data						
Train undergraduate students on data literacy						
Train postgraduate/doctoral students on data management						
Train researchers on data management						
Create web pages with information and resources about research data management						
Embed librarians in research teams						
Offer advice on intellectual property rights surrounding data						
Other, please state						

26.b. Given the current situation of your institution, how likely is it that in the next three years your library will…?

	Unlikely	Rather unlikely	Rather likely	Likely	Already in place	Don't know
Carry out, or assist in an institutional data audit						
Provide storage space and preservation for research data						
Extend the institutional repository's scope to include data sets						
Assist researchers to submit their data sets to third-party, general or disciplinary data centres						
Assist researchers in the discovery of data sets						
Assist researchers in the citation of data sets						
Assist researchers in the reuse of data sets						
Other, please state						

26.c. Given the current situation of your institution, how likely is it that in the next three years your library will…?

	Unlikely	Rather unlikely	Rather likely	Likely	Already in place	Don't know
Develop, or assist in the development of an institutional data policy						
Assist in the development of a national data policy						
Collaborate with other institutions in the country to provide research data management services						
Collaborate with international institutions in research data management						
Create a permanent 'data librarian' (or similar) position						
Create a fixed term 'data librarian' (or similar) position						
Relocate existing staff to deal with research data management						
Ensure the professional development of current workforce in data management						
Other, please state						

27. In your opinion, to what extent would the following be a constraint for your library to introduce data management services?

	High constraint	Medium constraint	Low constraint	Don't know
Lack of national guidelines and legal framework				
It is not a library responsibility				
It is not a service priority				
Difficult financial situation				
Lack of awareness on the management-level of the university				
Lack of interest from the management level of the university				
Lack of awareness from the researchers' side				
Lack of interest from the researchers' side				
Overloaded researchers don't have time for this issue				
Overloaded staff members don't have time for this issue				
Lack of recognition and rewards for data sharing				
Lack of subjects specific knowledge of the library staff				
Lack of technical skills and experience of the library staff				
Lack of experience with data storage and preservation				
Shortcomings of the technical infrastructure				
Difficulties of addressing the needs of researchers from different disciplines within a single service				
Difficulties of deciding on the appropriate level of involvement in research data management				
Other, please state				
Other, please state				

28. Thank you very much for completing the questionnaire. Please use the following box if there is anything else you would like to share (open)