Master's Thesis in Graduate School of Library, Information and Media Studies

The Combined RDS and RDM Lifecycle Model in Academic Libraries

September 2019
201726109
Asmae ZAIDANE

The Combined RDS and RDM Lifecycle Model in Academic Libraries

Asmae ZAIDANE

Graduate School of Library, Information and Media Studies

University of Tsukuba

September 2019

The Combined RDS and RDM Lifecycle Model in Academic Libraries

Student No.: 201726109
Name: Asmae ZAIDANE

In today's research landscape, the desire to acquire and analyze more data keeps on growing, these changes amongst others has led academic libraries to focus more on reshaping their services to meet the new needs of their research communities in terms of data management. This has required reflecting on the requirements and the challenges to face in developing and dealing with new services, namely research data services (RDS). Several pieces of research have examined and discussed whether academic libraries have the ability to implement and manage RDS, as research data management (RDM) has become a “strategic priority” for universities and academic libraries.

This research explores RDS in academic libraries through investigating models used for RDS and RDM management and the literature discussing the feasibility and requirements of operating RDS in academic libraries. Using theoretical analysis and qualitative content analysis, 18 models were investigated and the content of two websites was analyzed as a lens to examine the current situation of RDS in academic libraries and detect the pattern of change in terms of functions and their importance for the last two decades. Two major issues were concluded through this research: the lack of standardization of models, and the insufficient number of research surveys about the situation of academic librarians working within RDS.

Academic Advisors: Principal: Saori DONKAI
Secondary: Masanori KOIZUMI

CONTENT

List of Figures

List of Tables

Chapter 1

Introduction

The development of research and technology introduced academic libraries to new challenging roles, such as research data management (RDM) development, and new services, namely research data services (RDS).

Several researches have examined and discussed whether academic libraries have the ability to implement and manage RDS (Lyon, 2007; Donnelly, 2008; Lewis, 2010; Corrall, 2012; Tenopir et al., 2014; Tenopir et al., 2017), for instance, Corrall (2012) stated in her article about the roles and responsibilities of librarians and libraries in managing data: "We see three main potential roles for the library: increasing data-awareness amongst researchers; providing archiving and preservation services for data within the institution through institutional repositories; and developing a new professional strand of practice in the form of data librarianship."

As RDM has become a "strategic priority" for universities (Whyte & Tedds, 2011; Cox, Verbaan, & Sen, 2012) and academic libraries are expected to offer such research support to researchers. Furthermore, surveys and case studies were conducted in order to understand the relationship between RDS and academic libraries, and their common activities, for instance: user communities, policy development, awareness, training, data repository development, helpdesk services, and data management plan (DMP) development.

This research was set out to investigate the RDS and RDM models used in academic libraries for the last two decades to examine the situation of RDS in academic libraries and detect the pattern of change in terms of functions and their importance, for a better understanding of the situation of research data services in academic libraries in terms of operation, management and in order to come up with a model encompassing the elements deemed important throughout this study.

Two major issues were concluded through this research: the lack of standardization of models, and the insufficient number of research surveys about the situation of research data services in academic libraries as well as librarians working within RDS.

1.1. Research Background

Over the 21st century, research data has been overwhelmingly produced, costing dramatically high to institutions and organizations (Kellam & Thompson, 2016) and making organizing data challenging; this has resulted in the emergence of different new concepts such as “Data deluge” (Borgman, Wallis, & Enyedy, 2007) or “Data science” as early as 1968 (Wang, 2018). According to studies targeting researchers perception of research data management and data sharing, researchers have no time to deal with all the produced data (Ikeuchi, Harada, Sato, Okabe, & Itsumura, 2017); Often unstructured and inaccessible, data becomes a burden, its loss is inevitable and locating or re-using it is a time and effort consuming task. In order to solve these issues, and generate a return on investment in terms of invested time and money, efficient and effective research data "management" practices were deemed important in order to facilitate the manipulation, preservation, and sharing of the datasets, and that is how the concepts of research data management and research data services were introduced in the 1980s.

As early as 2008, researchers debated the possibility of academic libraries managing research data (Corrall, 2012 ; Lewis, 2010; Lyon, 2007; Donnelly, 2008), for instance in his “Libraries and the management of research data” Lewis (2010) discussed the debate around academic libraries managing research data as he affirmed that in one hand, managing and providing access to research data is a “natural extension” to academic libraries’ current role in the academia and research, however, managing data might not be a job for academic libraries considering the required skills, infrastructure and various stakeholders that concretely or ideally need to be involved in the process of managing data.

Other research studies rounded up this issue from different aspects, for instance: Tenopir (2014), searched academic librarians’ perception of offering research data services and that was one of the few studies dealing with academic librarians’ opinion on engaging in research data services. Cox and Pinfield (2014) investigated the importance of librarians as a key stakeholder in RDS. Federer (2016) debated how librarians and researchers are the most relevant key stakeholder in RDS.

Over the last decade the situation of academic libraries regarding research data services has changed, and new roles are emerging for academic librarians, such as the “data librarian” role, however these changes as well as the involvement of academic librarians in research data services, are yet to be clarified in literature.

Through this research, the situation of RDS in academic libraries will be investigated, using two methods in order to analyze the models used by some academic libraries to provide and manage their research data as well as their research data services. The methodology consists of two parts:

First, a theoretical analysis on the “top down level”, based on deductive reasoning as a research theory,

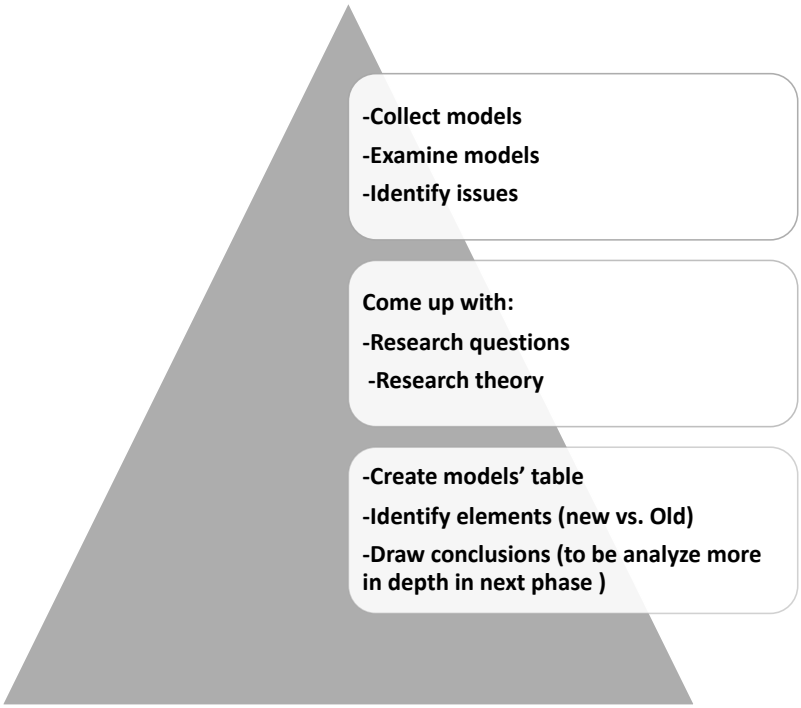


Figure 1 Top-down level (Theoretical analysis)

Second, a qualitative content analysis on the "bottom up level" of this research, which is based on inductive reasoning.

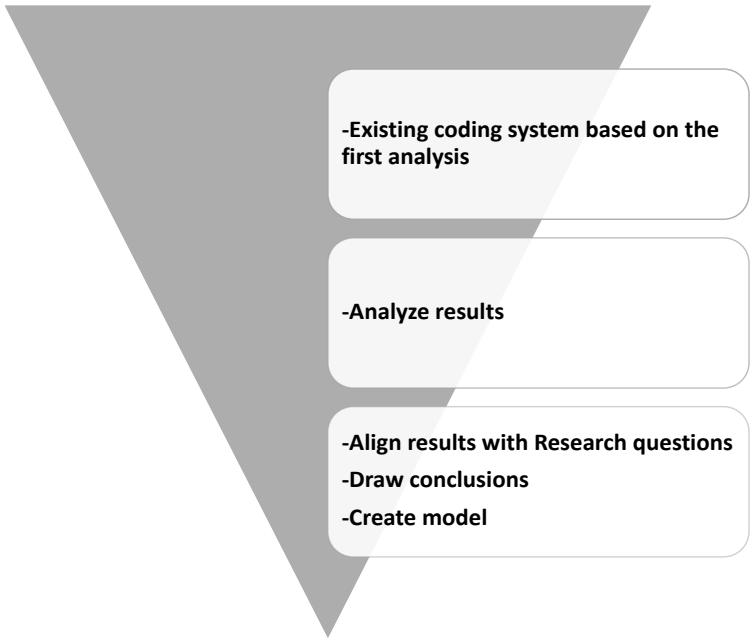


Figure 2 Bottom-up level (Qualitative content analysis)

The objective of conducting the theoretical analysis, is to perceive the bigger picture of the situation of research data services in academic libraries, whilst the content analysis

was chosen as a mean to back up the results from the theoretical analysis and for further investigation of the involvement of RDS in academic libraries.

1.2. Statement of the Problem

The problem this research sought to tackle is that the situation of research data service (RDS) in academic libraries appeared to be uncertain and literature in this field of research disorganized. Most literature and research studies dealt with issues partaking to research data management (RDM) therefore there is not much information about the situation of the service that does the management of research data itself.

Different models are used to manage RDS and RDM, however most studies focus on research data management models instead of models that manage the service in charge of RDM.

These were deemed important gaps in this research field, as the lack of data about research data services, entails a lack of knowledge about the situation and perception of RDS stakeholders, such as: academic librarians. This could be critical considering that RDS' stakeholders are an important factor to sustain and manage the research data service.

1.3. Justification for the Study

For the last two decades, a number of researches have attempted to investigate the practices and services pertaining to research data in academic libraries from different perspectives, For instance: Tenopir et al. conducted a survey study with library directors and academic librarians (Tenopir et al., 2014); Briney et al. (2015) focused on data policy's implementation and impact on research data service within their respective institutions.

Although these previous studies gave insight on the situation of (research) data curation in libraries and can be considered a baseline for future professionals in this research field, they also present some limitations. Firstly, most studies used "self-reporting data" obtained through conducting interviews or surveys, which represents a methodological limitation. Furthermore, studies provide findings resulting of critical and theoretical analysis rather than in-depth studies. Also, the fact that RDS in academic libraries is fairly a novelty in the field of Library and Information Science, every author approaches the topic from their own perspective as there is no theoretical baseline to studying RDS in library science as of today.

This study intends to examine these gaps and contributes to the literature through conducting a theoretical analysis as well as employing a content analysis of web content, in order to understand the best practices of current RDS models used in some academic libraries, such as those investigated in this research.

In sum, this research considers:

- The importance of providing more information about the situation of RDS in academic libraries,

- The contribution to enriching the literature in terms of addressing RDS in academic libraries' issues and Librarians' perception of RDS, as to date, little research has been conducted in some aspects of those matters (Yu, 2017; Koltay, 2017),

- The contribution in enriching the academic librarians' community's knowledge and insight into RDS in the context of academic libraries,

- Providing a perception on the evolution of research data related services in academic libraries for the last 10 years.

1.4. Research Aims

This research examines previous studies and models in order to grasp a better understanding of the status of RDS in academic libraries. Through this research I aspire to investigate, briefly, the engagement of academic libraries in research data services in terms of stakeholders, especially academic librarians, and how does the academic library meet the needs of its academic research community throughout this engagement. Also, this research aims to enrich the academic librarians' community's knowledge and insight into research data services in the context of academic libraries.

1.5. Research Questions

In relation to the aforementioned elements, the main question of this research is: *"What is the current situation of Research data services in academic libraries?".*

However, in order to make the main question manageable, it was broken down into the following research questions:

Research methods	Research Questions
Theoretical analysis (Models)	1-What are the requirements for establishing an effective RDMS?
	2-What kind of models are used to manage and maintain RDS in academic libraries?
	3-Which are the elements emphasized in most of the models? How are the models and the RDS influenced by these elements?
Content analysis (Case study)	1- To what extent are academic libraries engaged in offering RDS to their research community?
	2-Which are the RDMS practices that are most emphasized in this research's case studies?
	3-What is the current situation of RDS in academic libraries?

Figure 3 Research questions according to the research methodology

1.6. Scope of the Study

This research is delimited to the role of academic libraries in hosting and managing research data services. As research data can be managed within research organizations or institutions other than academic libraries, this study will solely focus on the situation of research data services in academic libraries.

1.7. Definition of Key Terms

1.7.1. Data

Data is defined by Pryor (2012) as "the primary building of all information", it can be identified as a group of numbers, characters, images or any other symbols which can become facts, figures and ideas after being contextualized.

1.7.2. Research Data

Is often referred to as data that has been contextualized through research, it is also "...information [which is] relevant to, or of interest to researchers, either as inputs into or outputs from research" (Economic & Social Research Council (ESRC), 2018),

Figure 4 Research data creation

In the field of academic research, data could be defined as “the output from any systematic investigation” (Pryor, 2012), exponentially growing over the last two decades, (Koltay, 2017), the quick growth of research data makes organizing it challenging, which results in unstructured and disorganized datasets and according to previous studies researchers might have no time to deal with all of the produced research data, either from their own researches or others, therefore an efficient and an effective research data "management" was deemed important in order to facilitate the manipulation, preservation, and sharing of research data.

1.7.3. Research Data Management (RDM)

As defined by Steelworthy (2014) research data management are practices applied to research data throughout its Lifecyle, a continuous process in which researchers are supported in collecting, managing, preserving and sharing (if applicable) their newly produced or re-used research data.

Defined also as “**Practices** through the **entire lifecycle of the data**, from **planning** the investigation to conducting it, and from **backing up data as it is created** and **used to long term preservation** of data deliverables after the research investigation has concluded”(Texas A&M UL Research Guides).

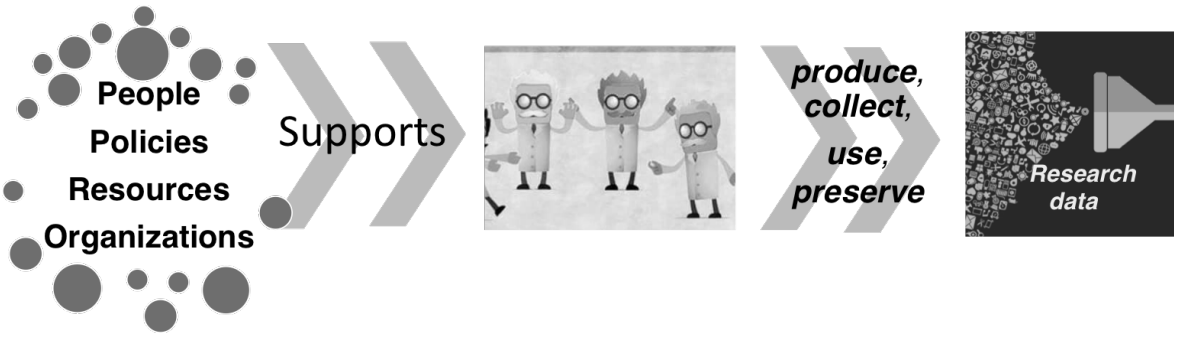


Figure 5 Research data management definition-adapted from M. Steeleworthy (2014)

Practicing RDM helps researchers or other stakeholders decide what to do with their digital and physical research data.

In literature, research data goes through a “life cycle” in which every unit of data is generated (Born) in the initial stages of research and is eventually archived or deleted (dies) making the end of its “useful life”. Through the process of RDM, research data is collected and curated, stored properly and its value appraised to identify whether it should be archived or eliminated, the archived data is to be reused again if it is deemed useful for other researchers, this cycle matches the stages of “the useful life” of research data, and it involves a complex set of tasks at each phase of research. (Fei Yu, Deuble, & Morgan, 2017)

1.7.4. Research Data Services (RDS)

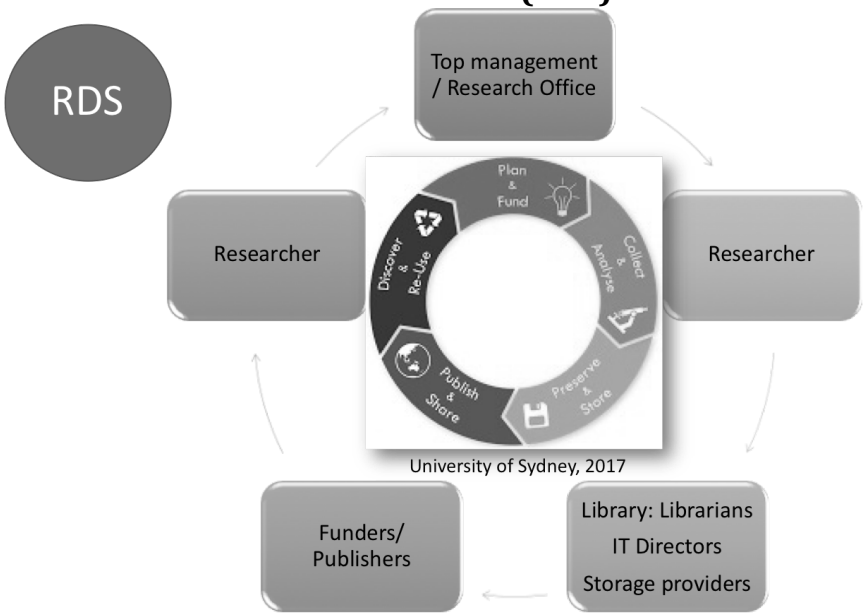


Figure 6 Research data services practices and stakeholders

RDS are services that serve research communities by supporting the researchers' data-related needs (Federer, 2016). It is a as a network of services and various stakeholders, assisting researchers during all phases of the research data lifecycle.

The massive growth of research data, the importance of research data management in handling the generated data and the changing roles of academic librarians in the last decade, have encouraged the creation of research data services as a unit to support researchers in academic libraries.

These services can help the researcher learn how to collect data, organize it, preserve it, submit it to the research data service, understand the requirements of funding agencies and journal publishers in terms of data management plans (DMPs).

The researcher chooses whether to publish it to the public or not, and they also learn the guidelines of the sharing policy in which they choose whether to share their produced or raw data with other researchers or opt out of it.

Establishing RDS requires the involvement of multiple stakeholders and the academic library is well positioned to host such services.

In providing support services through RDS, academic libraries or research institutions use models, referred to in literature as research lifecycle models, research data lifecycle models or research data management lifecycle models.

Although they have different names however these models share the same goal, which is to guide the provided RDM services; ideally, the models are used to clarify the necessary steps to follow, the policies and strategy of the service providing the RDM and the demanded requirements to sustain and manage the RDS.

1.7.5. The Difference Between RDS and RDM in This Research

As it is crucial to understand the difference between RDS and RDM as envisioned in this study, the following is proposed:

Table 1 The difference between RDM and RDS in this study

Research Data Service (RDS)	Research Data Management (RDM)
A structure, a system within the academic library, offering data management and other data related services to the research community.	**A service** to the research community.
In charge of Managing RDM	**In charge of Managing research data** through its life cycle as long as it is considered useful to the community (CLIR & University of Illinois, GSLIS)
Needs guidelines, collaboration of stakeholders in order **to be managed and sustained.**	Relays on guidelines, namely Data management plan and RDM policies in order to decide on how **to manage, preserve, secure and share research data**

1.7.6. Lifecyle Models in RDS and RDM

Cox & Tam (2018) suggests that the concept of lifecycle can be ambiguous as it could imply two patterns:

a. The model goes from point A, producing data (**Birth**) to point B archiving or deleting data (**Death**),

b. The model illustrates a cycle which starts from point A producing data (**Birth**) and ends up in point B sharing data and re-using the archived data (**Reproduction**).

Through this research the term "lifecycle" represents the pattern (b) in which the process starts from "Birth" and ends with "Reproduction", this implies that the cycle is repeated endlessly.

1.7.6.1. RDS Oriented Models

RDS is operated with the help of a team of stakeholders, therefore, in describing the RDS models, these contain elements such as: Stakeholders (Librarians, IT, Research institutions, etc.), also it involves Infrastructure, Policies and Strategies to manage and sustain this system.

Such model is for instance, Sykes' the UK Research Data Service Model (2009), which only includes stakeholders and their role within the RDS:

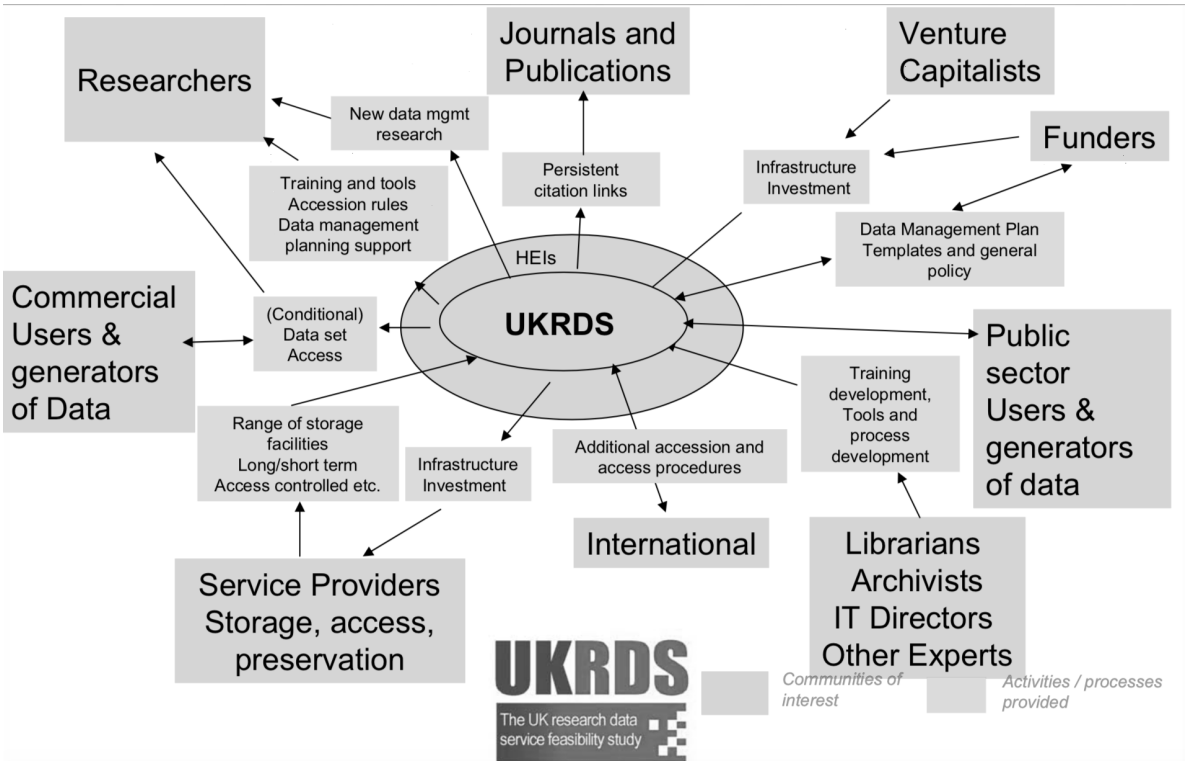


Figure 7 The UK Research Data Service Model (Sykes, 2009)

1.7.6.2. RDM Oriented Models

Ma & Wang (2010) identified six types of RDM lifecycle models:

- **The chain model**: which consists of a succession (chain like) of steps,
- **The matrix:** which uses the same chain model perspective but describes the stages with more details,
- **The circular model:** A model in which the end of the chain becomes the beginning for a new cycle,
- **The spiral model:** which is a circular model that takes into consideration the complexity of each process and how in every repetition of the same cycle the structure of data changes integrated model combines two of the other types as a way to indicate the complexity of processes,
- **The wave:** in which the activity is on-going, but it considers that data loses value through time.

1.8. Outline of the Thesis

This thesis is structured into six chapters, organized under two sections: theory and methodology sections.

In this chapter, I provided the introduction to the research, background information necessary to understand this research, namely, the statement of the problem, justification and research questions, key terms, etc. In Chapter 2 (theory section), I review the core literature relevant for this research, which is also a result of the theoretical analysis that has been conducted throughout this research. In Chapter 3 (methodology section), I explain the research framework and research methodology adopted for this study. Chapter 4 provides the results of this research; Chapter 5 offers the analysis and discussion of the results obtained through chapter 3 and Chapter 6 is a conclusion in which I outline the contributions, implications and limitations of this research, in addition to my future research plans.

Chapter 2

Literature Review

Academic libraries, potentially, working on RDM was discussed earlier in (Lyon, 2007; Lewis, 2010). Later on, the topic started to be approached more professionally through conducting in-depth research in order to understand the implications and involvement of research data services in academic libraries. In this chapter, different aspects relevant to data librarianship, academic librarians' implications in research data services, research data services in academic libraries as well as research data management are to be examined. Firstly, it is important to note that, the reviewed literature was selected following a specific search strategy:

2.1. Literature Search Strategy

In order to identify the core literature relevant to this research and which can provide the required understanding of the field of research intended for this study, a search strategy was necessary.

In this study, the search strategy based on the following question "How are academic libraries involved in offering research data services? And, "Are there models on which are based both RDS and RDM practices?".

The aim of the initial searches was to identify peer-reviewed, relevant literature regarding research data services in academic libraries, therefore three main areas were covered by the searching: RDS in academic libraries, RDM in academic libraries and data librarianship.

The search was run on diverse databases such as ACM Digital Library, Library, Information Science and Technology Abstracts (LISTA), Library and Information Science Abstracts (LISA), Library Literature & Information Science Index, Citeseer, Google Scholar, Library Science, Science Direct.

The keywords used for the searching were "RDS, RDM, research data management, research data services, Academic libraries, data librarians, data curation, academic librarianship, models". Other keywords were added to the list, such as data librarianship, open data, data science, and e-research. The searching for those keywords was limited to results in the English language and to relevance to the area of research. A sources relevance was judged based on the relationship of keywords, for instance: research data

services in academic libraries (rather than in other institutions or research organizations), RDM and academic libraries, RDS and RDM in academic libraries, etc. It was crucial to determine whether the retrieved literature was relevant to the research purpose; in this case, I relied on the title and abstract's relevance to not only research data services, however, specifically in the context of academic libraries.

The search strategy's searching technics were based on stringing the keywords for Boolean or phrasal searching, and backward searching was also taken into consideration as references of retrieved literature were used for this research.

The selected sources were comprised of journal articles, books, eBooks and book chapters, conference papers and proceedings, reports, theses and web documents.

In order to keep the searching as relevant and comprehensive to this research as possible, the search was run repeatedly while progressing in this research, in anticipation of potential new and relevant publications, this was possible through using the same keywords as well as the names of some of the leading authors in this field of research such as Corrall, Cox, Kennan, Lyon, Pinfield, Verbaan, etc.

After retrieving the papers, I ranked them in terms of most cited or most referenced literature by authors working in the field of RDS.

2.2. RDS and Academic Libraries

There has been a significant debate in literature on the roles of librarians and other stakeholders in managing research data and the skills and training needed(Corrall, 2012; Lewis, 2010; Lyon, 2007; Pryor & Donnelly, 2009; Swan & Brown, 2008), that emphasized how the sustainability and management of research data services depends on different stakeholders or communities of interest. Federer (2016) discussed librarians' and researchers' implication and role in the data management process and how both stakeholders are the most important stakeholders in the RDS; A survey conducted by Tenopir et al. (2014) found that librarians and library directors are important factors towards creating an effective RDS, Lewis (2010) focused on his model on Researchers, student researchers and librarians' skills in addition to the government implication in the success of an RDS. In their survey, Cox & Pinfield (2014) investigated the roles that librarians as a key stakeholder could take in RDS and which roles were a priority for the research community. The pyramid of Lewis (2010) implicitly illustrates the stakeholders of and RDS in the context of an academic library, in a model which was modified by Corrall (2012) and has been updated by Ohaji (2016). Corrall (2012) included RDM practices in the Lewis' model (although some of the important elements are not considered in her addition, such as data sharing), and Ohaji (2016) provided a clear idea on the specific stakeholders that intervene in the process of managing data in academic libraries.

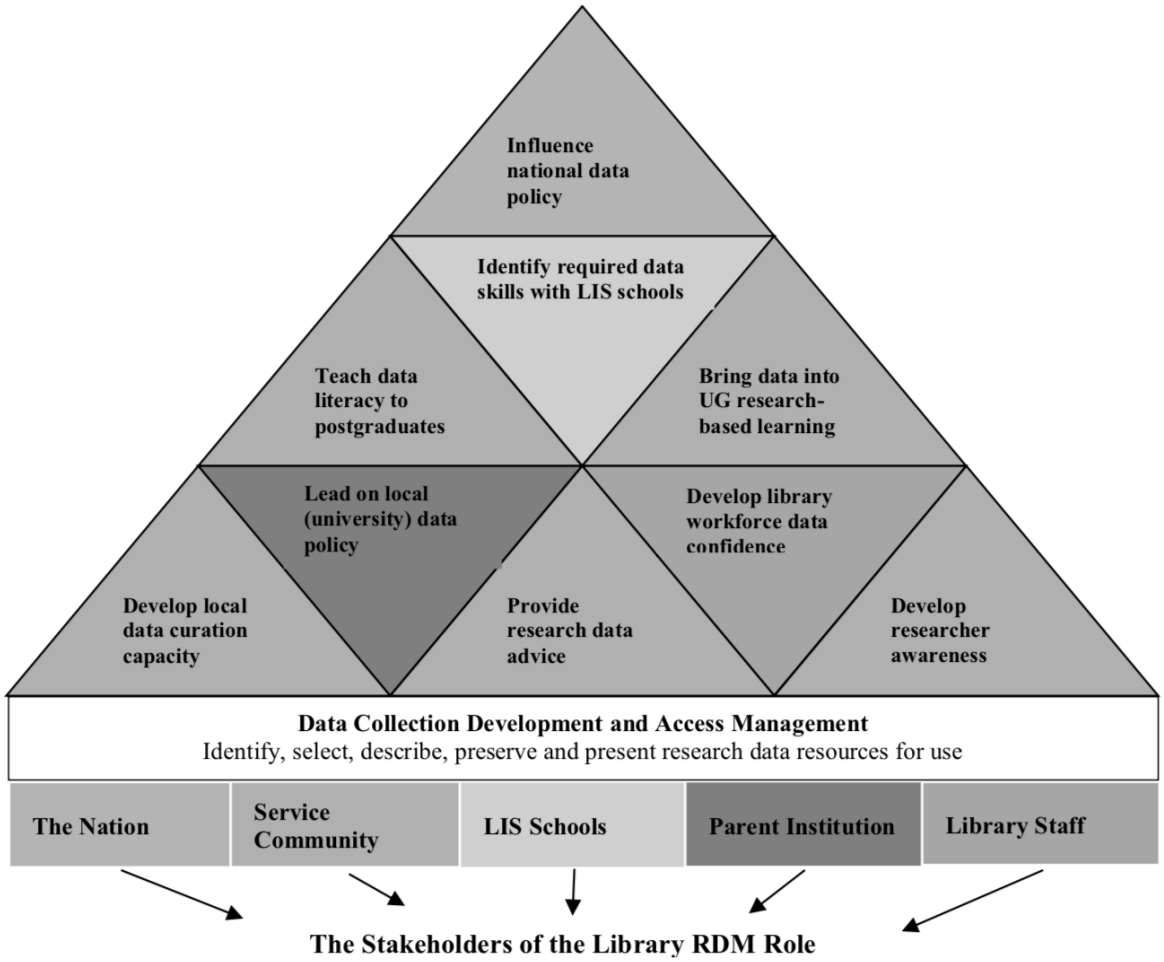


Figure 8 Research data management pyramid for libraries (Ohaji, 2016)

Lewis (2010) suggests the importance of involving Library and Information Science (LIS) schools in identifying the required skills to engage in RDS as librarians or information scientists, and it illustrates the required actions to take on a local, institutional and national level, in order to provide research data service.

According to the vision of Waard' research data management in institutions (Waard, 2014), the success of a research data service within an academic library depends "Ideally" on the collaboration between the following key stakeholders within the institution, namely:

- Researchers or "service community" and "LIS schools" according to Lewis (Ohaji, 2016)
- Librarians (Tenopir, 2014; Faniel & Connaway, 2018)
- Data repositories, IT services (Cox & Pinfield, 2014; Pinfield et al., 2014)
- Offices of research (administration, e.g.: NIMS or AIST in Japan)
- Funding agencies: Government, Journals, Publishers…

However, as reported by Verbaan & Cox (2014) The collaboration is not always achieved amongst the mentioned stakeholders, due to either lack of training, lack of skills and therefore lack of trust.

2.3. RDS and RDM Models

Ma & Wang (2010) discussed the use of lifecycle models in LIS, and how the use of models in Library Science can be critical, as it:

- Ensures providing good service,
- Enables better outreach outcomes in terms of properly explaining the service to the users, the use of models makes the explanation easier,
- Enabling the librarian to understand their precise role within the service.

Carlson (2014) discussed the changing types of lifecycle models depending on the context of data service, and defined three types:

- Individual-based models created for a specific project,
- Organization-based models, describing how every service shall be used depending on the stages of research,
- Community-based, created to describe existing and good practices, it is usually created for a particular academic community, discipline or professionals.

2.4. The Changing Role of Academic Librarians: Data Librarians

The data deluge that is extensively talked about in literature, in addition to the improvement of technology, were an incentive to the change in the type of traditional services that were offered to researchers, the changing information environment is undoubtedly impacting libraries as well, and changes in the way information is handled were necessary, "the librarian cannot sit on the fence of professional spectatorship if he/she is to meet the challenges of the changing environment" (Faniel & Connaway, 2018). As reported in literature, the origins of data librarianship as a new subdiscipline was the result of a need for managing social science data at first and then for bioinformatics data sets' management, different concepts and roles have been created for librarians in order to establish a logic extension between the roles of the traditional academic librarian and research data management's new needs.

2.4.1. The Evolution of Data Librarians in Literature

Librarians that used technology for the sole reason to accomplish service goals in 1997s, and were faced later with new challenges, and various roles requiring "an updated" educational experience that would help librarians adapt to the changes and make them able to offer new services befitting of the changes. The evolution of "data librarians" and

academic librarians in regard to data science can be observed through the following time line.

Severt (2005) mentioned the expression "accidental data librarian" during the IASSIST conference in Edinburgh, referring to librarians who found themselves dealing with research data without any prior training or re-skilling, as they became data librarians in order to integrate new services although they may still lack in terms of skills. Although the emergence of concepts like big data and data science were taking place and changing the needs in terms of research support; the concept of data librarian was far from easy to integrate, in fact, many academic libraries were avoiding getting involved in research data support/services, resulting on librarians who were less informed of the implications of data librarianship in terms of skills, this is also reflected through the models' table, as it noticeable that before 2007, most models were focused on the RDM skills rather than the management of the service that is responsible for RDM, for example, there was less to no interest in the staff training and skills enhancement.

Corrall (2008), brought up the concept of "hybrid information workers", as new roles that have ambiguous status and different titles for the same job, she illustrated the growing interest in data librarians re-skilling and upskilling and talked about data scientists and subject librarians as two different professions that are involved in RDM roles, on the other hand Donnelly & Rusbridge (2008), used the expression "data librarian" in their "core skills for data management model", focusing on the socio-cultural and ethical implication of the data librarian role (e.g.: raising awareness, negotiation skills, etc.), they have omitted some of the commonly traditional roles of the librarian, such as providing help to the research community in order to facilitate access to information, for instance "sharing" and "re-using data", this is apparent through the RDS models' table (the two functions (sharing, re-using) are omitted in eleven out of eighteen analyzed models, mainly those created before 2013).

In Auckland (2012) the data librarian profession was still considered a new role for librarians and she focused on subject librarians' re-skilling to meet the new needs of researchers, nonetheless, she insisted on the continuous re-skilling of librarians in order to meet the research support needs in a continually "changing research environment". Cox & Pinfield (2014) illustrate a big shift of interest regarding the re-skilling and upskilling of "academic librarians", without using the appellation "data librarian", the interest is more focused on the development and management of the research data services rather than research data management. This aligns with the changes of the research environment and information professionals' work environment's needs that have been taking place for the last two decades.

Cox (2016), talked about research support librarians, and focused on the skills of information professionals in dealing with the challenges of data science, he was oriented more towards the management of research data services and the research data

management roles that can be taken on in an academic library's environment. Again, this marks a shift in terms of the aspects of interest towards the librarian that engages in data science related work/projects, the interest is not only in technical or RDM oriented issues, but it has become skills and team work management. Koltay (2017) continued stressing on the importance of re-skilling and offering data literacy not only to researchers but also to data management professionals, while Wang (2018) take the discussion to a new level, the author discussed how data librarians play a critical role within the data science "ecosystem" and their roles covers all steps of the (research) data management life cycle, including sharing, re-using, data visualization…, and data literacy training. Wang (2018) stressed more on the necessity of making data librarians "domain specialists" instead of only focusing on introducing data science and data literacy in the library and information science (LIS) curriculum, in order to be able to help researchers efficiently and effectively.

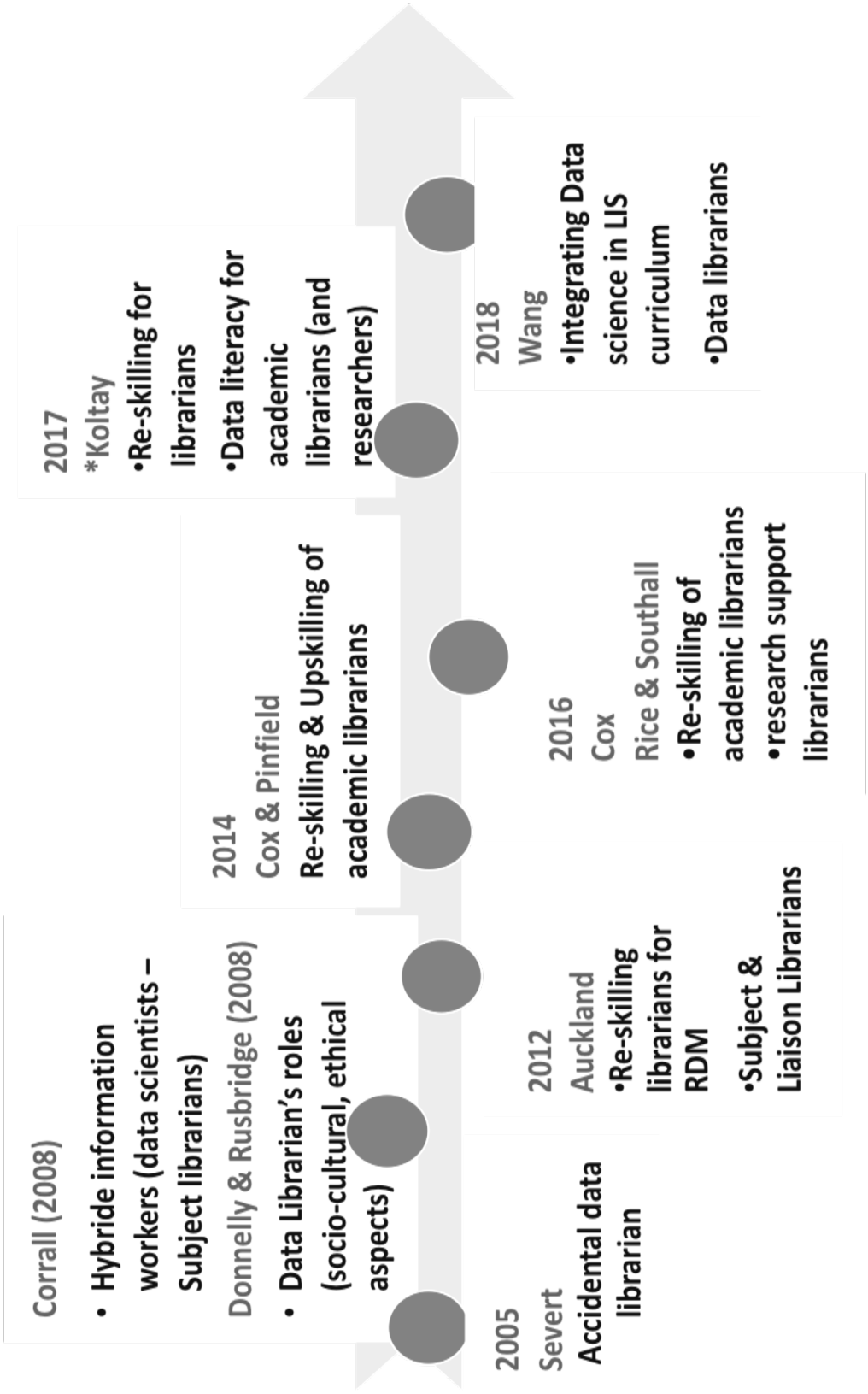


Figure 9 Timeline of data librarians' roles and definitions, as addressed by different authors (Zaidane & Koizumi, 2019)

2.4.2. Factors Influencing the Data Librarian Role

As great information managers and organizers, librarians might have the skills in terms of data curation and communication, but the need to integrate research data services and deal with data science has led many librarians to take on new roles, expand their competencies and enhance their skills in order to meet the expectations of researchers and other information professionals (Osswald & Strathmann, 2012; Stanton, 2012), this manifests in the creation of a new subdiscipline, namely data librarianship and the "data Librarian" profession (Corrall, 2012; Wang, 2018), which can be defined according to Swan & Brown (2008) as "people originating from the library community, trained and specializing in the curation, preservation and archiving of data". Data librarians come from various backgrounds, essentially: IT, statistics, social science and library and information science. Data librarianship can be perceived as a connection between library science and data science, as librarians must gain technical and scholarly communication expertise to perform as a data librarian.

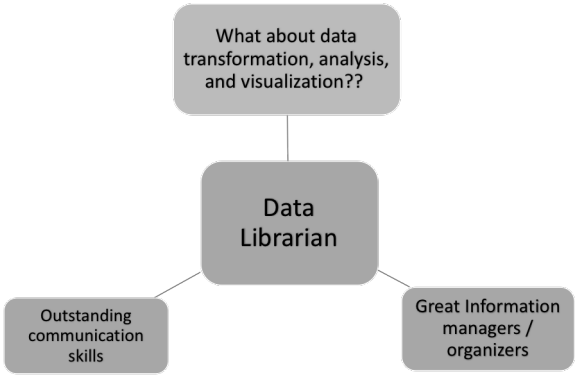


Figure 10 Illustration of data librarians' roles regarding data science & big data problems adapted from "Data science: What's in it for the new librarians?" by J. Stanton, 2012

Previous studies report the involvement of data librarians in the RDM life cycle (Burton & Lyon, 2017; Cox, Kennan, Lyon, & Pinfield, 2017), however, data librarians' role have yet to be defined and specified considering the required skills and knowledge in data science, and the attitude of librarians regarding RDS and RDM (Auckland, 2012; Cox, Kennan, Lyon, & Pinfield, 2017), previous studies have shown that a clear agreement on the definition of the data librarian's role(s) and skills needed to perform these roles needs to be clarified. (Yoon & Schultz, 2017; Tenopir et al., 2014)

Chapter 3

Research Framework and Methodology Design

3.1. The Research Theory in This Study

Theory plays an important role and significance in research inquiries (Reed, 1984; Kumasi, Charbonneau, & Walster, 2013). In library and information science the use of theory was mainly in quantitative studies.

For this research, the adopted theory was the inductive reasoning or the scientific research method (Leedy, 1989:37). This methodology seeks to identify the problem through gathering data and conducting observation as a first step, then a hypothesis is formed and tested empirically. In this research the inductive reasoning was chosen in order to facilitate understanding the context of the study, as well as gaps to be studied through investigating literature. Inductive reasoning allows an easier way to approach a complex research topic such as research data services in academic libraries, as starting from samples such as studying RDS and RDM models, then drawing conclusions about the larger picture involving the situation of RDS in academic libraries is easier to handle, rather than tackling the concept of RDS in academic libraries as a whole. This is represented in the following figure (Figure 11):

Figure 11 The conceptual model for studying the situation of RDS in academic libraries

In order to understand and identify the gaps in literature regarding the situation of RDS in academic libraries, the lifecycle models used to manage research data and research data services were investigated as well as the websites of two universities' academic

libraries which were examined using content analysis. Conclusions related to the situation of RDS in academic libraries were drawn using results of both investigations.

3.2. Overview of the Conceptual and Methodological Basis of the Study

The inductive and deductive approaches from Norton's "Deductively Definable Logics of Induction" (2010) offered a conceptual and methodological basis for the models that guide this research.

The first part of this research which consists of a theoretical analysis, follows a top-down approach as a conceptual basis, and the deductive reasoning methodology as a research framework (this methodology works from the more general to the more specific).

In the deductive reasoning the problem is identified through gathering data, then a hypothesis is developed, and a test is conducted in order to examine the hypothesis. In this research, we gathered data about research data services in academic libraries then research questions were developed after identification of the issues surrounding the research topic.

The second half of this research consists of a qualitative content analysis, following a Bottom-up approach as a conceptual basis in which I start from the conclusion, examine previous results, and I back it up with the Inductive reasoning methodology as a research framework for this part of the research.

The inductive reasoning methodology requires that I use previous results, which in our case are results from the theoretical analysis in order to answers my research questions, instead of starting the investigation from scratch, and through which I can generate a new theory or model from the analyzed data.

3.3. Research Methodology: Theoretical Analysis

As mentioned before this study is set out to examine the situation of research data services in academic libraries, in that regard a methodological approach based on a theoretical analysis of eighteen RDS and RDM models was adopted.

Although the study seeks to identify models and case studies of RDS in academic libraries exclusively, the majority of identified models are used for RDM, but since some of these models can still be applied to RDS, such as the DCC Curation Lifecycle Model (Higgins, 2008) which is relevant to data management or archiving/record management but is used by the UK Research Data Services, I considered including RDM models too.

3.3.1. Case Studies: 18 Models

The following models were those investigated:

- The Scholarly Knowledge Cycle (Lyon, 2003)

- DDI version 3.0 Combined Life Cycle Model (Data Documentation Initiative (DDI), 2004)
- The Lifecyle Model of Research Knowledge Creation (Humphrey, 2006)
- Domain Data Deposit Model (Lyon, 2007)
- Digital Curation Center's Curation Lifecycle Model (Higgins, 2008)
- The Integrated Scientific Life Cycle of Embedded Networked Sensor Research (ISLCENSR) (Pepe et al., 2009)
- The UK Research Data Service Model (Sykes, 2009)
- Research Lifecycle (Rin & Nesta, 2010)
- The Research Data Management Pyramid for Academic Libraries (Lewis, 2010)
- I2S2 Idealized Scientific Research Activity Lifecycle Model (Patel, 2011)
- Revised-RDM Pyramid for Libraries (Corrall, 2012)
- USGS Science Data Lifecycle Model (Faundeen et al., 2013)
- Research Lifecycle Workflow (Grigorov, 2014)
- A library-Oriented Model of Institutional RDM (Cox & Pinfield, 2014)
- Research Data Management Practices Model (University of Sydney Library, 2018)
- Research Data Service at The University of Edinburgh (UED) (The University of Edinburgh, 2018)
- Research Data Management at UCF (University of Central Florida, 2018)
- University of Oxford Research Data Management Chart ("Research Data Oxford," 2018)

3.3.2. The Selection of Models for This Research

In order to accomplish this study, a number of models were extracted from the reviewed literature amongst which eighteen models published for the last two decades. It is important to note that the list of collected models was inevitably uncomprehensive.

The Models were selected based on the following:

Models are in use either in a library or a research institution, to manage either research data or the unit that manages the process of RDM. For instance: The Digital Curation Centre (DCC) "curation lifecycle model" and Lewis's "Research data management pyramid for libraries" model, are highly cited in literature in addition to widely being used in practice.

3.4. Research Methodology: Qualitative Content Analysis

In order to carry out this study, a major method used is content analysis:

An "almost universal" method in qualitative research (Elliott, 2018) used to examine "written, verbal and/or visually communicated information" (Cole, 1988) in order to quantify a phenomenon, describe it and understand it (Elo & Kyngäs, 2008). It is known as "versatile tool" for social science researchers (Ohaji, 2016).

For this part of the study, content analysis was chosen as "Text data are dense data, and it takes a long time to go through them and make sense of them" (Creswell, 2015), as well as for the following reasons as identified by Krippendorff (Krippendorff, 1980):

- Permitting the structuring of the analyzed data in a practical way,
- Being an unobtrusive measure as the analyzed data is not solicitated directly from research subjects,
- Being context sensitive, it offers a certain level of objectivity that leads to less bias results,
- Allows handling large sets of data, as it is simple, economic.

3.4.1. Qualitative Content Analysis for This Study

In this phase, two universities were chosen as case studies, in which the websites of their research data services and other similar services (especially in the case of Harvard University) were analyzed using content analysis. The frequency of updates of the websites was taken into consideration.

We have chosen to analyze their websites, for the following reasons:

- Possibility of getting real-time information that can provide an insight on how the research data service operates,
- Understanding the type of RDS it offers and if an RDS or RDM model is in use or not,
- Build an insight the influence of the RDS activities within the academic library.

It is important to mention that analyzing web content can be challenging, as it is subject to changes and its content can be of mixed types of media. Previous studies offering guidance in dealing with applying content analysis to web content were consulted for this study, such as Kim & Kuljis' (2011) case study on applying qualitative content analysis to blogs.

3.4.2. The Type of Content Analysis Used in This Study

As a method that may be used for analyzing either quantitative or qualitative data, content analysis may be used in:

- A deductive approach, in which there is not enough knowledge from former data about the phenomena,
- An inductive approach, which is used in case the structure of the content analysis could be constructed and operationalized based on previous knowledge, in this case the purpose would be testing a pre-existing theory.

For this study, the content analysis is based on the results and knowledge deducted from phase 1 (theoretical analysis), therefore the use of inductive content analysis. (Kyngas & Vanhanen 1999, Elo & Kyngas, 2008)

The process of conducting the content analysis was based on the "content analysis through nine steps" illustrated by Neuendorf (2002) as follows:

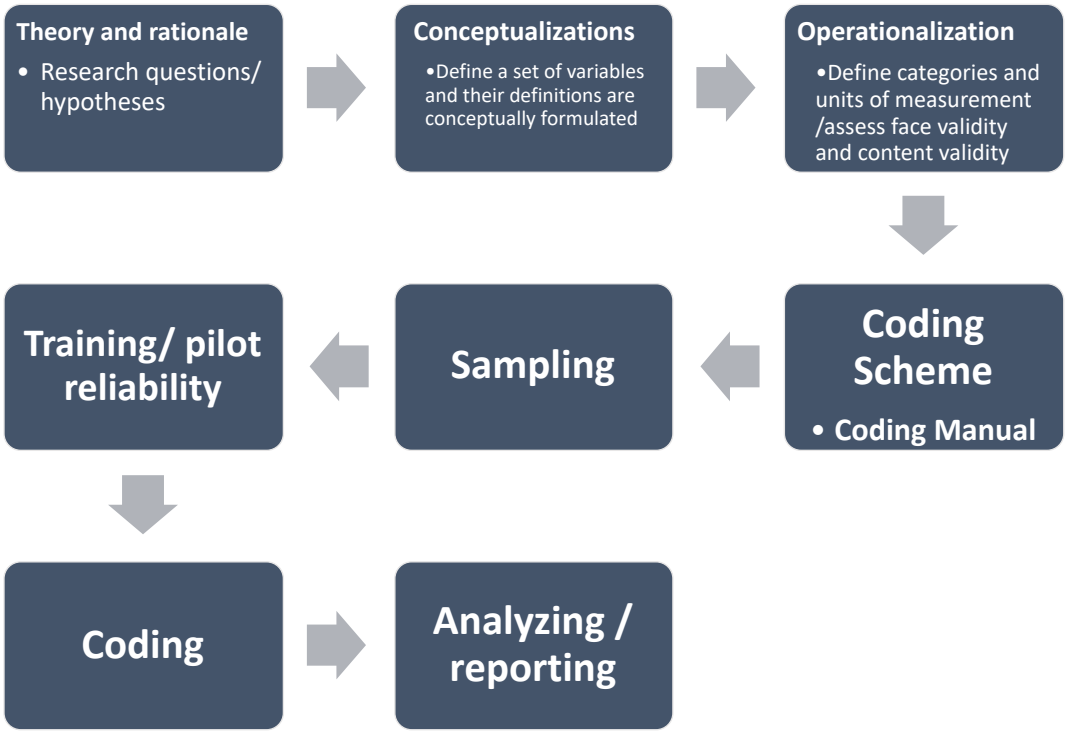


Figure 12 Flowchart of content analysis process, adapted from Neuendorf (2002)

3.4.3. The Software Used for This Study

In order to conduct the content analysis, we used MAXQDA which is one of the most comprehensive programs in the field of qualitative and mixed methods research. Easy to use and offers various data visualization and presentation tools.

3.4.4. Case Studies for the Qualitative Content Analysis

The study cases for this research are two advanced cases in the field of research data services in academic libraries, namely: Harvard University and University of Oxford, as

they are offering research data services to their respective research communities. The choice was based on two reasons:

- The availability of enough data that will help and ensure the continuity of this research, as both Universities are of the oldest in the world.
- The status of these universities as prestigious, ensure that enough funding is available to sustain research data services in their academic libraries, in addition to the high probability of offering quality RDS service.

3.4.4.1. Study Case 1: University of Oxford

As the world's second oldest university in the world, the University of Oxford has been ranking first in *Times Higher Education (THE) World University Rankings* for three years in a row. It is also important to note that the University of Oxford is a research-driven University "having recently ranked number one in the world for medicine (if its Medical Sciences division was a university in its own right, it would be the fourth largest in the UK) and among the top ten universities globally for life sciences, physical sciences, social sciences, and the arts and humanities." (Times Higher Education (THE), 2019).

The University of Oxford counts some of the most acknowledged libraries in the world, having the largest library system in the UK, the University of Oxford tends to its community's research and information needs through 100 libraries (University of Oxford, 2019b), these libraries are the following:

- The University's college libraries, as every college has a library of its own.
- The Bodleian Libraries: thirty libraries across Oxford which include research libraries, College libraries and department, other institutions' libraries, and the Bodleian Data Library. All are of course attached to the main Bodleian library. They hold 13 million printed items, over 80.000 e-journals in total (University of Oxford, 2019).
- The Bodleian main library: the main "research library" of the university and has been a legal deposit library for around 400 years.

a. "Research Data Oxford"

As earlier mentioned, the University of Oxford is highly research oriented and its academic community produces large sets of research data, this called for the creation of a service which can ensure the preservation and sharing of their research data.

Providing a world-wide access to research data in order to enhance the engagement of the research community as well as the public, helping in the maximization of the impact of research in different aspects (social, economic, etc.) are some of the goals of the University of Oxford's research data service.

Managed by the **Bodleian Data Library**, this service strives to help in organizing, structuring, storing and curating the data generated or used for or during a research project, and different stakeholders are involved in this process, participating hand in hand with the Research Data Oxford to fulfill the goals of this service. The main stakeholders are:

- **The Bodleian libraries**: Providing a Data Librarian who can offer guidance on data management and preservation, Subject librarians that could help during the research process and online library guides such as LibGuides,
- **Data protection**: which is a service that provides guidance for researchers on how to protect their research data and introduces them to the data protection policy of the University,
- **IT services**: facilitating secure data sharing, data preservation and backup,
- **Research data support**: which is a service that offers researchers: data tools, services and training,
- **Advanced Research Computing (ARC)** a central resource offering researchers HPC: high performance computing, if needed,
- **Open Access at Oxford (OAO)**
- **ORCID** -or the unique researcher identifier, goes along with the services that help the researchers to easily and globally link to their publications and grants and it is also requested by funders and publishers,
- **Oxford University Research Archive (ORA)** or the institutional repository of the University of Oxford, which offers a "permanent and secure archive" to the research community.
- **Oxford eResearch Centre**
- **Research publications management via symplectic elements**

3.4.4.2. Study Case 2: Harvard University

Established in 1636, it is the oldest university in the United States and the one of the most prestigious universities in the world. Comprising ten graduate schools, a total number of 17.875 teaching (Faculty) and non-teaching staff, including the staff of their allied institutions such as the Harvard Art Museums (Harvard University, 2019).

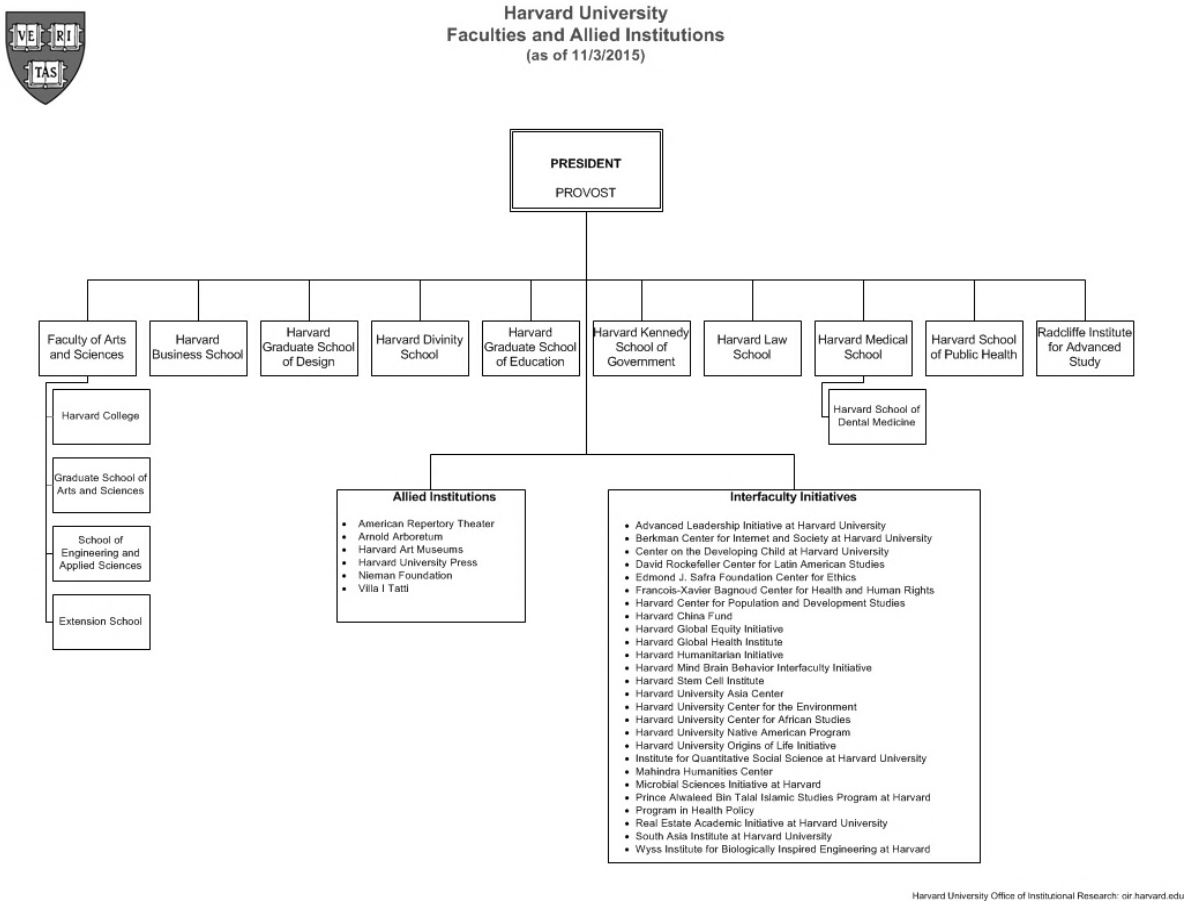


Figure 13 Harvard University's organizational chart - Faculties and allied institutions

Located in Cambridge, Massachusetts and established in 1638, Harvard Library is part of the University's central administration, as shown in the following Figure (Figure 14).

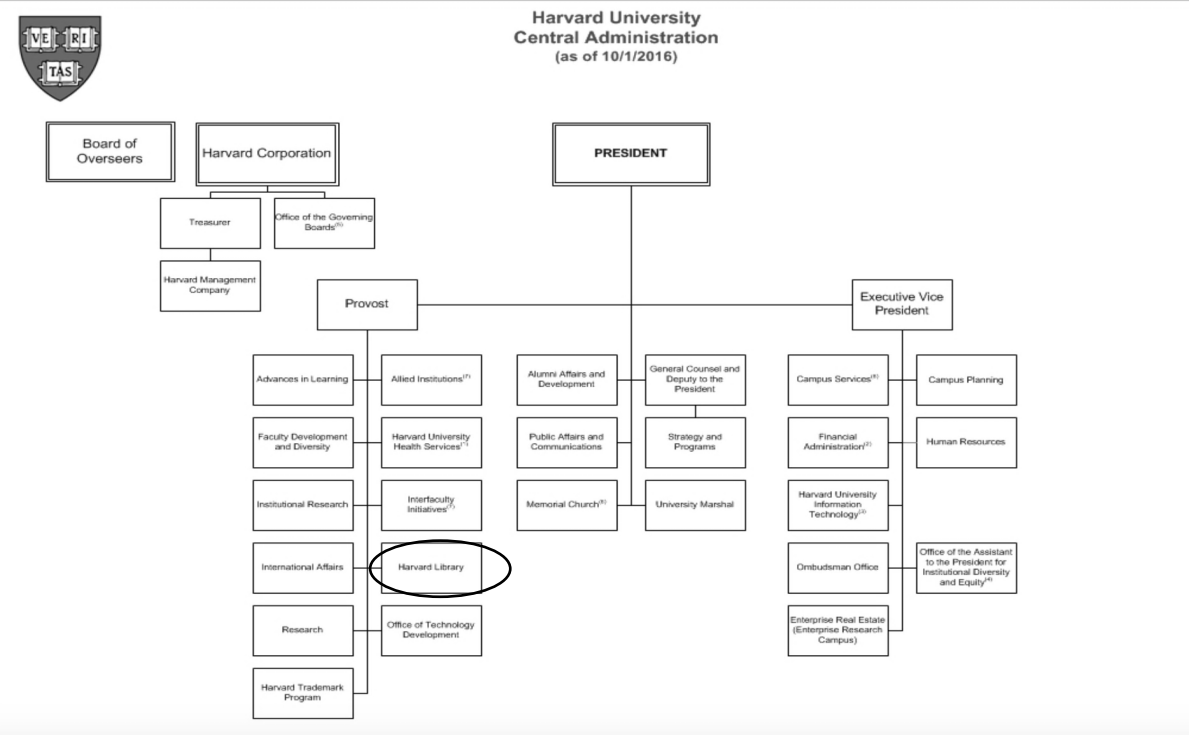


Figure 14 Harvard University central administration chart (Harvard University, 2019)

It is an organized system which includes seventy-six libraries all within Harvard University. It is considered the largest private library system in the world. The organizational structure of Harvard library consists of seven main services:

1. The "**Access Services**" which offers the academic community access to the library's large collection and resources.
2. **Information and Technical services** in charge of acquisition and licensing of library materials.
3. **Program Management** in charge of supervising potential and approved projects and its management.
4. **Finance Supports**, in charge of financial transaction and contributes in decision-making.
5. The **Harvard University Archives** or the institutional archives of the University. In charge of the university's records and monitors its record management activities, the collection of manuscripts, papers or historical materials related to Harvard.
6. **Preservation, Conservation and Digital Managing Services,** in charge of keeping library materials secure, accessible for the long term through various preservation methods, such as digitization, and even providing guidance on how to use the resources through library education and outreach programs.
7. The **Office of Scholarly Communication** offers the research community an open access service to works of scholarship created by the university's community, this office is also in charge of sharing and preserving scholarship.

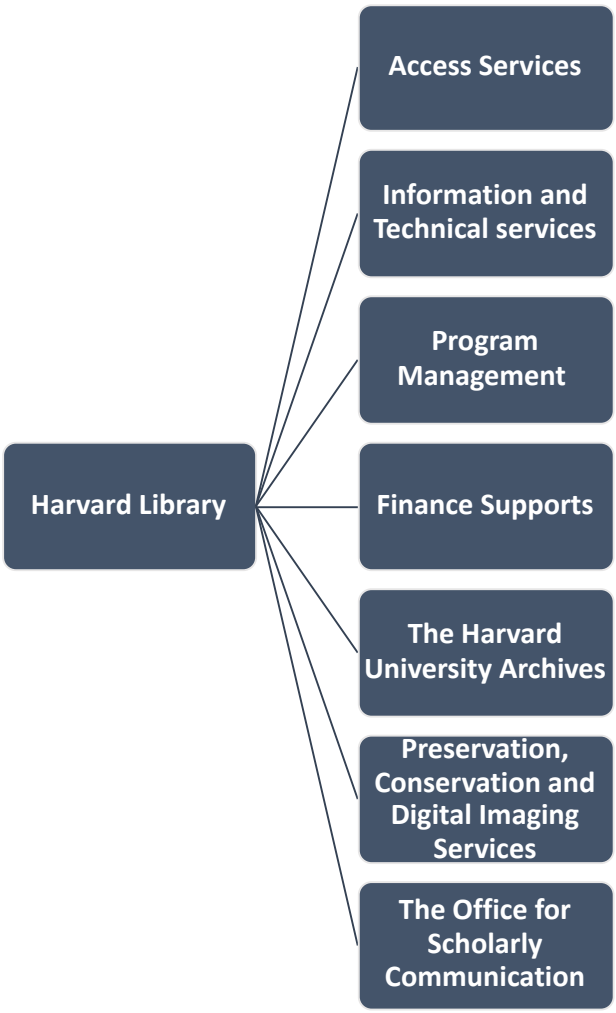


Figure 15 Harvard Library organizational structure

a. Harvard Library's Research Data Services

Harvard University is highly research oriented, investing up to $1.0 billion on research in 2015 as reported by the annual financial report issued by the University (Harvard University, 2015). Its research activities are sponsored either by federal sponsors, such as: The National Institutes of health (USA), NASA, etc., or by non-federal sponsors, for instance: foundations, foreign governments, etc., in 2016 the total amount of funds collected from sponsoring for research purposes was $ 868 million in the academic year 2016-2017. This reflects the importance of research within the University as well as the probable amount of produced and used research data.

Sponsored Research - All

	2012-13	2013-14	2014-15	2015-16	2016-17
Federal	$639.00	$608.00	$578.00	$593.00	$613.00
Non-Federal	$182.00	$204.00	$218.00	$249.00	$255.00
Total	**$821.00**	**$812.00**	**$796.00**	**$842.00**	**$868.00**

Source: Office for Sponsored Programs
Figures are in $M (Millions)

Figure 16 Sponsored research at Harvard University in numbers (Harvard University, 2019)

In order to ensure the accessibility, interoperability and reusability of the research data of Harvard University's research community, six libraries within the Harvard Library System are offering research data services, which are chaperoned by the Harvard Library Research Data Management Program, as shown in the following figure (Figure 17):

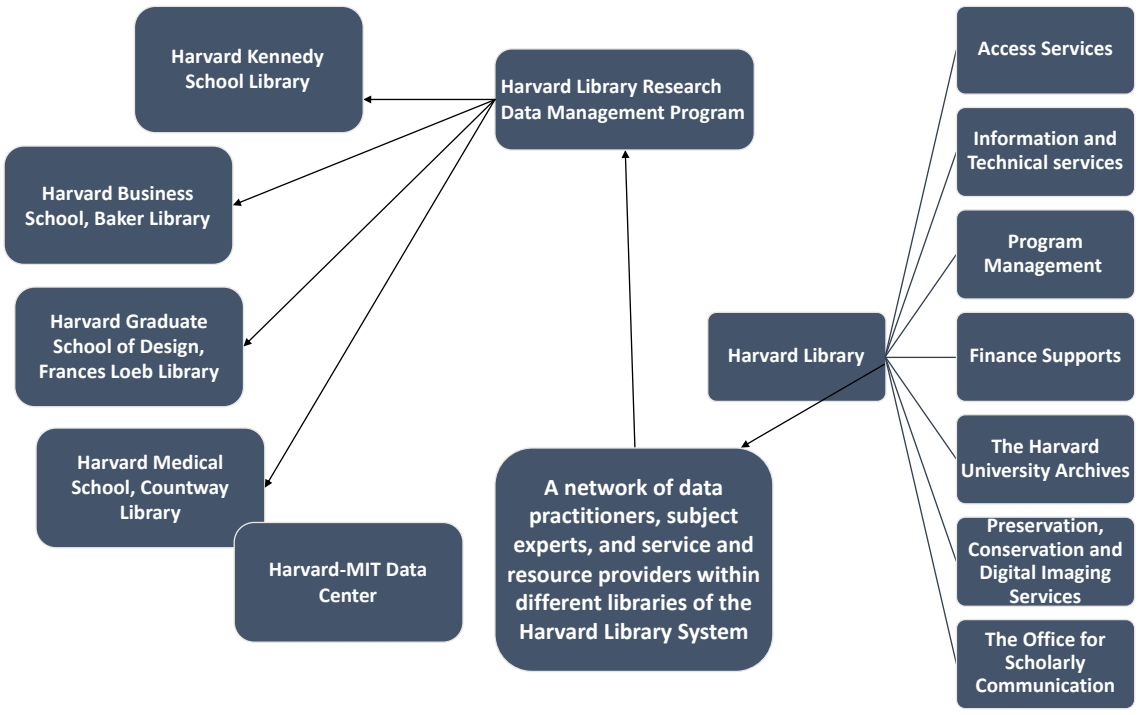


Figure 17 The organizational chart of Harvard University's research data management and services network

3.4.5. The Construction of the Coding System (Defining Categories & Sub-Categories)

Before analyzing the data, a coding system was developed base upon results of Phase 1 of this study in which we sought to examine eighteen RDS and RDM models and identified patterns that exist between models' elements. The coding system was constructed based on the results of phase 1, using the elements used in the RDS and RDM models' table as shown in the figure below (Figure 18):

Research Data services development & management										Research Data collection development & management								
(1)Planning						(2) Training		(3) Funding		(4) Collecting			(6) Preserving / Storing					
(1) RDM policy (Strategy)	(2) Policy / guidlines requirements	Technological Infrastructure / service design (3)	(4) Human infrastructure (RDM team)	(5) Outreach policy	(6) Data Management Plan	(1) Staff training / Skills enhancing	(2) Researchers	(1) Business plan	(2) Sustainability plan (for Preservation & Curation)	(1) Identifying resources	(2) Selecting	(5) Analyzing	Institutional repositories	External repositiories	others	(8) Publishing / Presentation	(9) Sharing	(10) re-use (Repurposing)

Figure 18 Research data management & services models table

The author also defined categories and subcategories of the coding system based on their expertise as an academic librarian.

The initial coding system for this study before starting the content analysis was presented as follows:

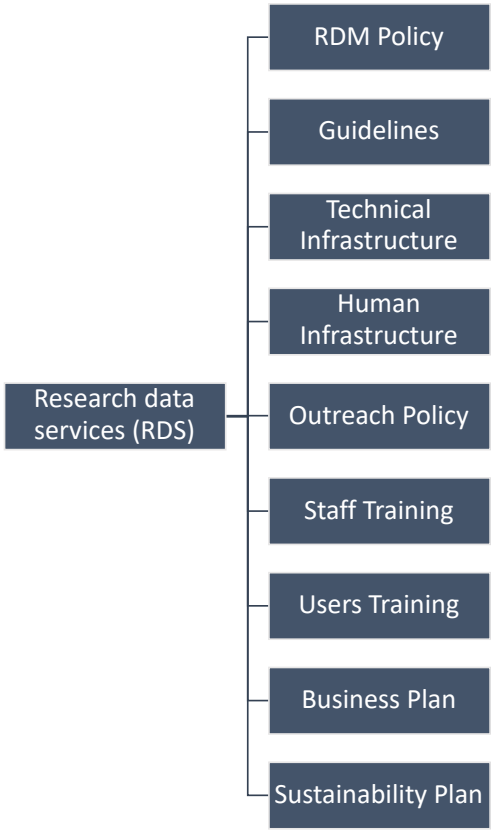


Figure 19 RDS as a main category and its sub-categories

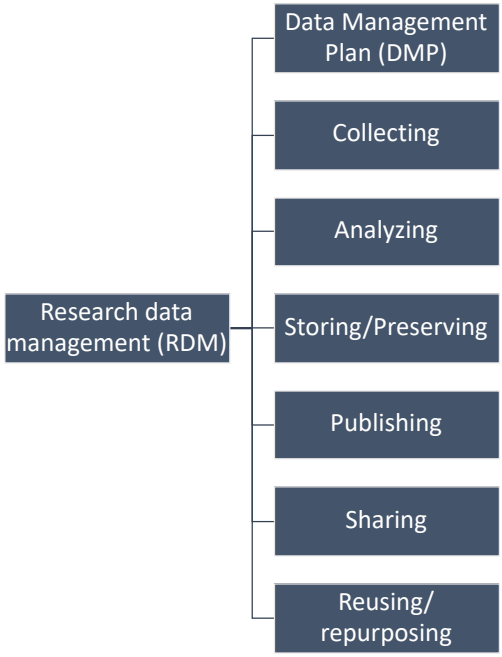


Figure 20 RDM as a main category and its sub-categories

After conducting the content analysis, some changes were introduced to the coding system, in terms of organization of the sub categories, as well as the addition of new elements.

It is important to note that these new elements resulted of the content analysis and that the web-based content analysis influenced these changes, for instance, it was easier to extract new elements from the web-content in comparison to extracting elements from the analyzed models, the reason why goes to the fact that some elements appear in the models but in an abstract or unclear way in some cases (Lyon, 2007). The changes aforementioned are illustrated in the following figure (Figure 21):

It represents the elements and tasks which are needed to maintain, manage and sustain the RDS as a unit/structure within an academic library.

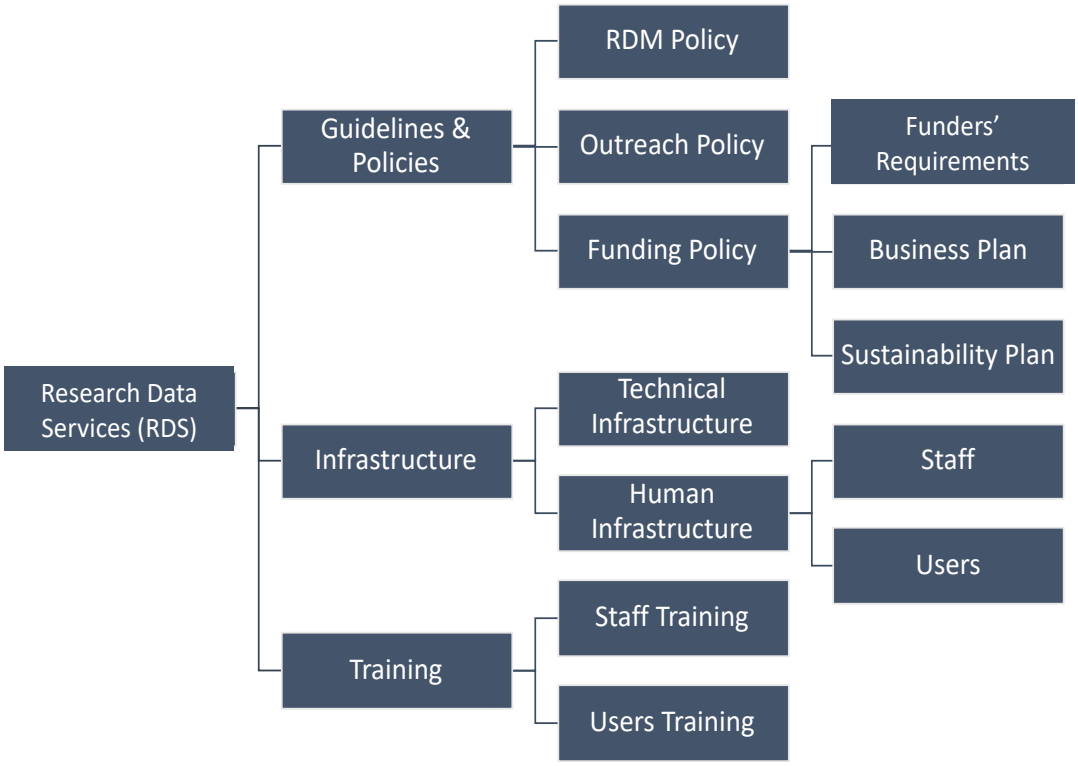


Figure 21 Updated coding system: RDS as a main category and its sub-categories

The following figure (Figure 22) represents the elements and tasks which describe the process of managing research data:

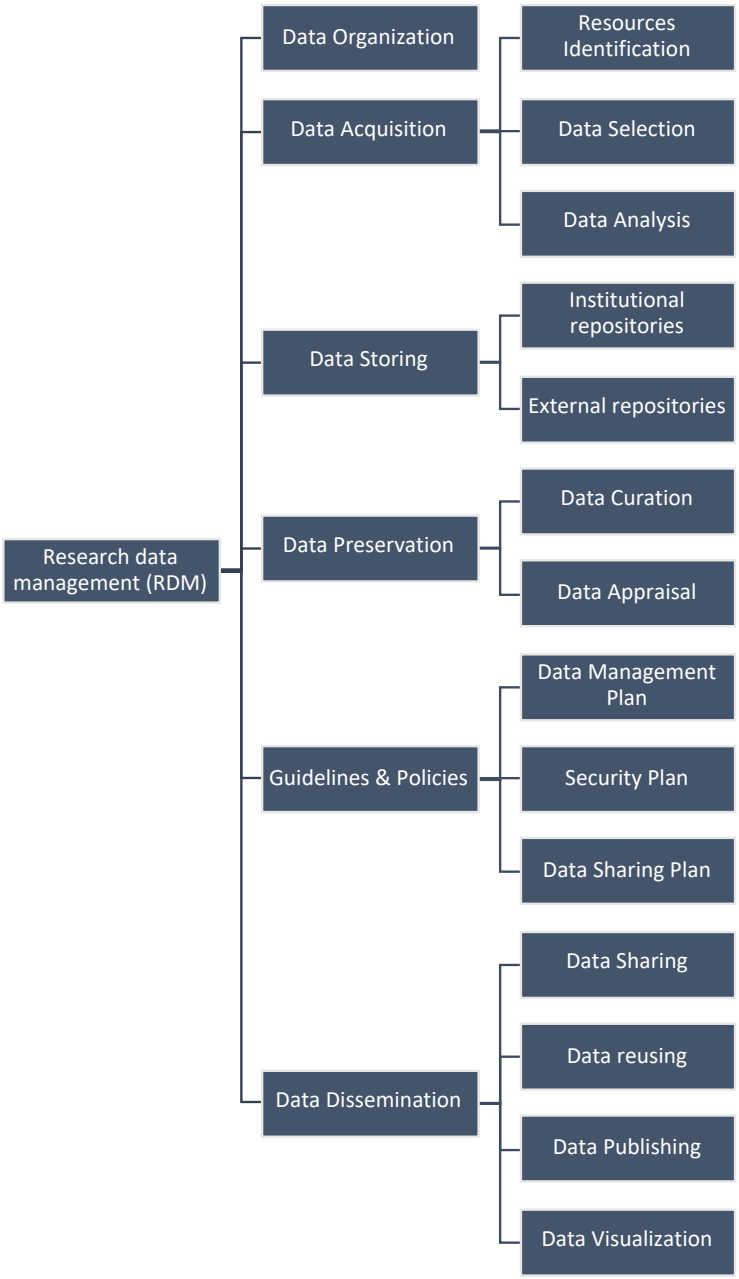


Figure 22 Updated coding system: RDM as a main category and its sub-categories

3.4.6. The Coding Manual

In order to guide the process of examining the analyzed content, a codebook was created, and the categories, generic categories and sub-categories were defined in it with their definitions.

In the coding manual the two main categories are as follows: Research data management and Research data services. In this study, it is important to lay stress on the difference between RDS and RDM as it is the basis of this content analysis.

Table 2 The difference between RDM and RDS in this study

Research Data Service (RDS)	Research Data Management (RDM)
A structure, an institution within the academic library, offering data management and other data related services to the research community.	**A service** to the research community.
In charge of Managing RDM	**In charge of Managing research data** through its life cycle as long as it is considered useful to the community (CLIR & University of Illinois, GSLIS)
Needs guidelines, collaboration of stakeholders in order **to be managed and sustained.**	Relays on guidelines, namely Data management plan and RDM policies in order to decide on how **to manage, preserve, secure and share research data**

The two main categories comprised of nine generic categories under which I have a total of 23 subcategories.

3.4.6.1. RDS Category

Under RDS, I count three generic categories, as illustrated in figure 21:

1- **Guidelines:** which refers to Instructions related to the creation, management and implications of the outreach policy, research data management policy, funders' requirements in terms of research data management,

2- **Training:** training opportunities and RDS skills development for staff and users (as a service),

3- **Infrastructure:** comprises two subcategories, namely:

Technical: related to the all composite hardware, software, network resources and services required to manage and operate the IT environment,

Human: refers to the staff working in the RDS in the academic library, in addition to the other stakeholders involved in the process of managing and sustaining the RDS as well as external stakeholders, as well as researchers, graduate and post-graduate students in this.

In this main category it was important to define RDS as two context or perspectives:

- RDS as a service offered to users (Front office)
- RDS as a system based on teamwork, guidelines, infrastructure, etc. (Back office)

As the inclusion of some elements such as: "staff training" in this category without defining the context can be confusing.

In the following table the categories under RDS are defined (regardless of the hierarchy illustrated in figure 21):

Table 3 Detailed description of the main category RDS

RDS	As a service offered to the research community in order to serve its research related needs	
	Training	• **Users**
	Guidelines & Policies:	• **RDM Policy**: Provided and explained to the users in order to understand policies, data sharing and security policies, etc. • **Outreach Policy** • **Funding Policy**: Explain Funders' requirements to users
	Infrastructure	• Researchers, • Research/Graduate/Post-Graduate students, • Faculty • Platforms, Databases, Materials (Presentations, guidelines, etc.)
	As a structure/environment that needs to be managed and sustained	
	Training	**Staff**, for e.g.: • Trainings offered to library staff • Tools used to offer training • Policies referring to training of library staff
	Guidelines & Policies:	• **Sustainability Plan:** To sustain the RDS • **Business Plan**: To fund the RDS environment
	Infrastructure	• Academic library staff involved in RDS and other stakeholders • Technology & IT environment

3.4.6.2. RDM Category

RDM counts six generic categories, as shown in figure 22:

1- **Guidelines & Policies** refers to instructions about the security policy, data management plan, sharing plan.

The other subcategories illustrate the lifecycle of research data, namely:

2- **Data Organization**

3- **Data Acquisition** includes identifying data, selecting the relevant data,

4- **Data Storing** includes necessarily external and internal (institutional) repositories

5- **Data Preservation** refers to appraisal and curation of research data

6- **Data Dissemination** refers to the different methods to make the research data accessible and visible such as, sharing, re-using, publication and visualization.

3.4.7. Methodological Reliability of the Coding Manual

The creation of the categories relied on the progress of the content analysis as well as were the expertise of the author as an academic librarian, therefore the categories were selected carefully to match the purposes of this research in understanding the situation of RDS in academic libraries.

3.4.8. The Creation of the Manual:

Although the elements of table 1 (from phase 1), are a result of the theoretical analysis, it is important to mention that the elements were chosen according to the research questions as well as the knowledge of the author as an academic librarian, therefore the initially constructed and the newly added categories of the coding manual were chosen with precision according to:

- The specific research questions of this phase
- The research needs of the author as an academic librarian

Also, the main, generic and sub-categories, were defined accurately, again based on the expertise of the author as a professional librarian and a specialist in this field. For instance: as an academic librarian differentiating between RDM and RDS elements was crucial in this research, as understanding the elements related to the research data service such as Librarians' training or the sustainability plan and whether these elements are emphasized in models or RDS websites or not, can be of great importance in revealing the situation of RDS in academic libraries from a librarian's perspective, which in this case is the author.

3.4.9. Pilot Reliability

Before starting the content analysis, a pilot reliability was conducted in order to train the code and also ensure that reliable and relevant results can be obtained using the constructed coding system**.**

3.4.10. Single Coding & Specialist Consultation

As the author was a single coder establishing an inter-rater reliability wasn't achievable at first, therefore two methods were used to ensure the reliability of the analysis:

An intra-rater reliability method was conducted, as described by Mackey & Gass (2013) the author coded the data and after a certain period of time a re-coding of the same data was conducted and there was a slight disparity between both times, therefore no disagreements in terms of codes.

The initial main and some generic categories were chosen as core elements, if during the analysis, no code was found for a certain element, a new category would be created in order to fit the coding (one of the reasons why I created new elements) while always aligning the results with the research questions.

In addition to the above-mentioned measures, a consultation with a specialist, who is an expert in content analysis and a consultant, took place during the "Operationalization" phase of the content analysis, in which I define categories and in sub-categories, as well as the "sampling" and "coding scheme" construction. (Elo et al., 2014)

The discussion revolved around:

- Ensuring the validity of the coding process, through revising the code's structure in terms of categorization,
- Checking the trustworthiness of the analysis.

3.4.11. Sampling

The structure of the web-based content can be challenging in terms of deciding on what content to code Potter (1999). As our research questions for this part of the study require a throughout analysis of the situation of the RDS within their academic libraries.

Before starting the sampling, a contact was established with the manager of Harvard University, who directed us to the specific documents on the website which can be useful to this study.

Therefore the content analysis was conducted on different types of documents, PDFs containing guidelines and other informative data offered by the RDS, Presentations published on the website, and webpages containing text which was converted into a PDF

format in order to facilitate the analysis of the content in MAXQDA (software). Images containing models, charts were also taken into the samples and analyzed.

All pages of both websites were analyzed: 17 pages of the "Research Data Oxford" website, 10 pages and 5 presentations of the "Harvard Library's Research Data Management Program" website. In the case of Harvard University two other pages were added, which are the page on research data in Baker Library's website and the main page of the Harvard Biomedical Data Management within the Countway Library of Medicine, as both libraries propose two models although they are collaboration and supervised by the "Research Data Management Program" of Harvard Library.

The webpages and other documents were collected between 13 and 16 April 2019.

Chapter 4

Results

4.1. Phase 1: Theoretical Analysis

The chosen eighteen models were considered core examples which are the most cited and the most used in research institutions and academic libraries as well. According to the reviewed literature and the collected models, a table to analyze the models was created and it was divided into two sections, as follows:

- Research data services development & management
- Research data collection development & management

The two categories were adopted in order to differentiate the elements or roles that are important to manage research data, namely: research data collection development & management, and those used to manage the research data "service" which were listed under research data services development & management. If an element is explicitly mentioned in a model, it is mark checked.

In some models, the words: re-use or repurposing research data are used. These words could be confusing as I might not be able to differentiate the concepts of sharing and re-using data. Therefore, I put both functions in different categories, as I noticed that re-using research data is mentioned in some models despite the use of sharing (e.g. Lyon, 2007).

No.	Name of the model	Author	Year	Subject matter	Research Data services development & management										Research Data collection development & management								
					(1)Planning						(2) Training		(3) Funding		(4) Collecting		(5) Analyzing	(6) Preserving / Storing			(7) Publishing / Presentation	(8) Sharing	(9) Re-use (Repurposing)
					(1) RDM Policy (Strategy)	(2) Policy / Guidlines Requirements	(3)Technological Infrastructure	(4) Human Infrastructure (RDM Team)	(5) Outreach policy	(6) Data Management Plan	(1) Staff training / Skills enhancing	(2) Researchers	(1) Business plan	(2) Sustainability plan (for Preservation & Curation)	(1) Identifying resources	(2) Selecting		Institutional repositories	External research repositiories	Others			
1	The Scholarly Knowledge Cycle	Lyon	2003	RDM											✓			✓	✓	✓	✓		
2	DDI version 3.0 Combined Life Cycle Model	DDI	2004	RDM											✓	✓	✓	✓	✓	✓			✓
3	The Lifecyle Model of Research Knowledge Creation	Humphrey	2006	RDS /RDM											✓	✓	✓						
4	Domain Data Desposit Model	Lyon	2007	RDS								✓	✓	✓				✓	✓	✓	✓		
5	DCC's Curation Lifecycle Model	Higgins	2007	RDM			✓	✓			✓	✓			✓	✓	✓	✓	✓	✓	✓		✓
6	The Integrated Scientific Life Cycle of Embeded Networked Sensor Research (ISLCENSR)	Pepe [et al.]	2009	RDM/RDS											✓	✓	✓	✓	✓	✓	✓		
7	The UK Research Data Service Model	Skyes	2009	RDS	✓		✓	✓		✓	✓	✓						✓	✓	✓	✓		
8	Research Lifecycle	RIN /NESTA	2010	RDM/RDS			✓								✓		✓				✓		
9	The RDM pyramid for Academic Libraries	Lewis	2010	RDS	✓	✓			✓		✓	✓											
10	I2S2 Idealized Scientific Research Activity Lifecycle Model	Patel/UKOLN	2011	RDM/RDS						✓					✓	✓	✓	✓	✓	✓	✓		✓
11	Revised-RDM Pyramid for Libraries	Corrall	2012	RDS	✓	✓					✓	✓			✓	✓	✓	✓	✓	✓	✓		✓
12	USGS Science Data Lifecycle Model	Faundeen et al.	2013	RDM	✓										✓	✓	✓	✓	✓	✓	✓	✓	
13	Research Lifecycle Workflow	Grigorov [et al.]	2014	RDS/RDM		✓									✓	✓	✓	✓	✓	✓	✓	✓	
14	A Library-Oriented Model of Institutional RDM	Cox & Pinfield	2014	RDM/RDS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓									
15	The RDM Practices Model	Univ. of Sydney	2018	RDM/RDS	✓	✓				✓							✓			✓	✓	✓	✓
16	Researh Data Service at UED	Univ. of Edinburgh	2018	RDS / RDM						✓					✓	✓	✓	✓	✓	✓	✓	✓	
17	RDM at the UCF	Univ. of Central FL	2018	RDM/RDS					✓				✓		✓	✓	✓	✓	✓	✓	✓	✓	
18	University of Oxford RDM Chart	CEOS	2018	RDM/RDS	✓	✓				✓	✓	✓	✓							✓		✓	

Table 4 RDS and RDM models

4.2. Phase 2: Qualitative Content Analysis

Through the conducted content analysis, 17 webpages were analyzed in the case of the University of Oxford's Research Data Service website, and 10 webpages were analyzed in the case of Harvard University's Research Data Management Program website and three pages of Harvard University's Baker Library.

4.2.1. Research Data Services Models and Activities in Harvard University

Although the Harvard Library System counts five different libraries that offer their varied research communities research data management services, the focus was on the Research Data Management Program offered by Harvard Library, as all mentioned libraries are supervised and linked to this program.

But it is important to mention that two of these libraries, namely: Baker Library and Countway Library have their own models in terms of managing research data.

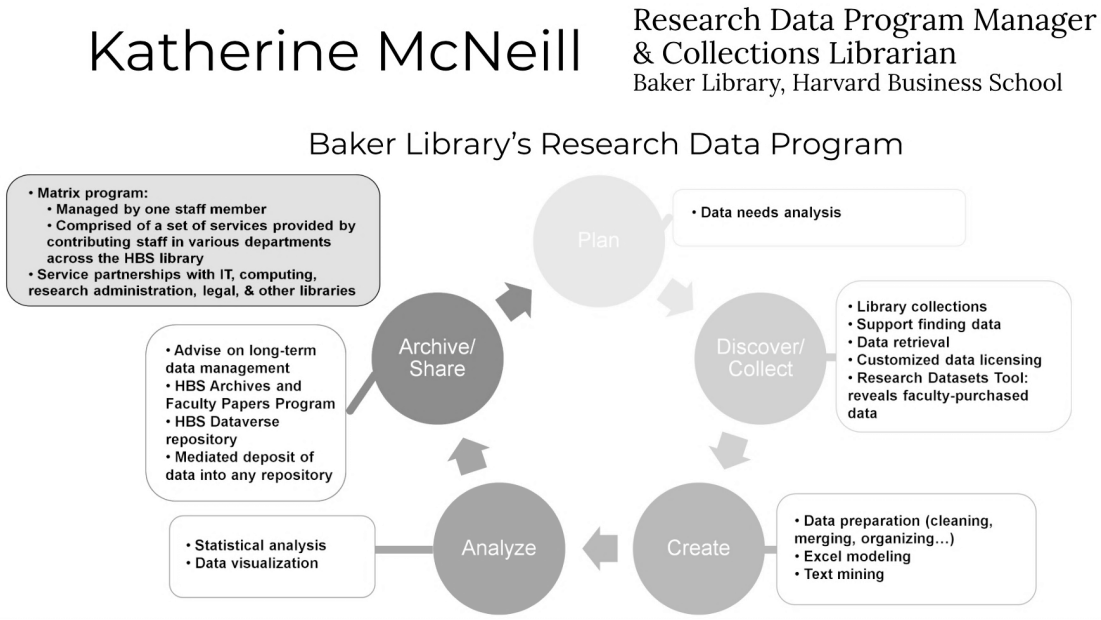


Figure 23 Baker Library's research data program, model (Harvard Library, 2019)

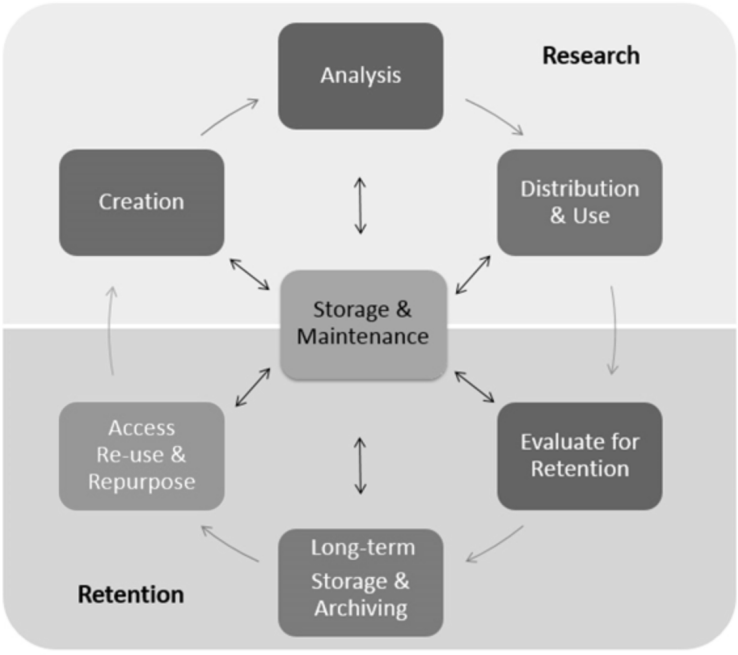


Figure 24 Countway Library- Biomedical Data Management (Harvard Library, 2019)

MAXQDA is the software used to conduct the content analysis for this study (as mentioned previously in the research methodology) this software provides tools for data visualization, through which maps showing the codes use frequencies across the pages, as well as the relationship between codes and the frequency of assigning codes together.

The following map (figure 25) was obtained through selecting the coded pages of the Harvard Library's RDM Program and the codes of both categories RDM and RDS, the retrieved coded segments are analyzed and organized in a map-like display.

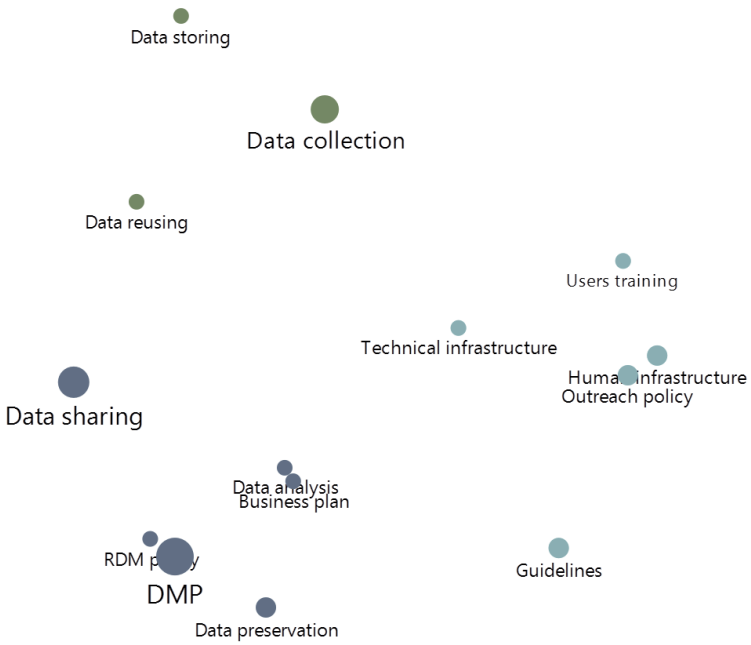


Figure 25 Relationship between codes in a map, for RDM Program, Harvard Library

Through the content analysis, interesting results on the elements that the research data services in Harvard Library focus on, for instance: there is a major focus on information about the creation of the data management plan (DMP) and Data sharing.

4.2.2. Research Data Services Models and Activities at the University of Oxford

Results of content analysis of the web-based content of the Research Data Oxford are as follows:

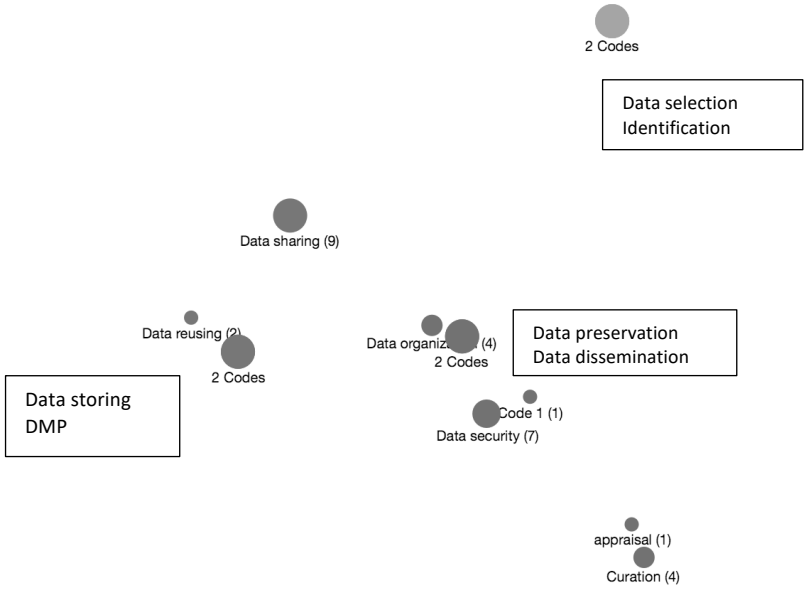


Figure 26 Relationship between codes in a map, for Research Data Oxford

The above map represents the relationship between codes, the more frequently they have been assigned together the closer they will be. Codes which have no relationships were omitted.

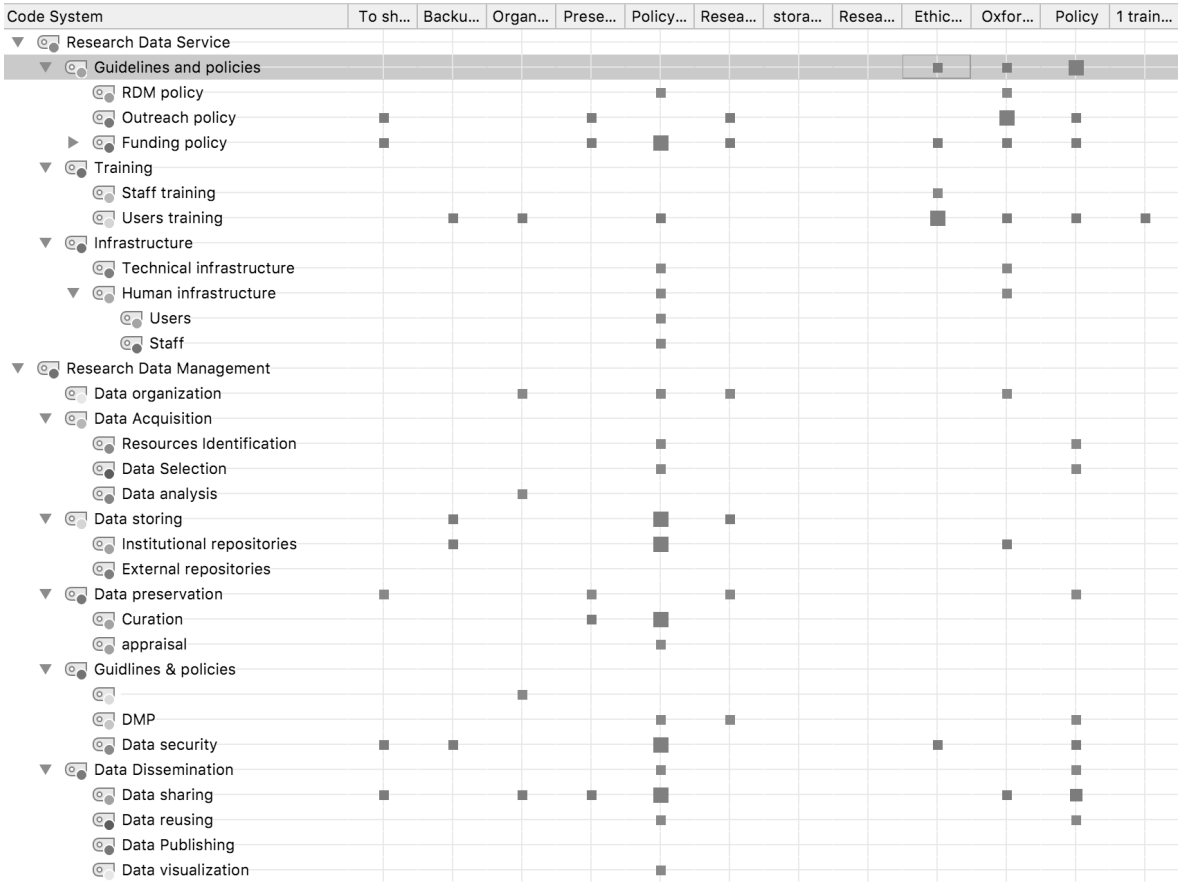


Figure 27 Codes' matrix, Frequency of appearance in some of the analyzed documents

4.3. The Combined RDS and RDM Lifecycle Model in Academic Libraries

The following model is a result of the conducted qualitative content analysis and the theoretical analysis.

It is a simplistic model, comprising the fundamental core elements that could be the baseline for a model that could be used in different contexts, however it englobes the vision of the author in term of elements that could make an efficient model from the perspective of an academic librarian.

As mentioned in earlier stages of this research, RDS is considered a structure or a unit which is in charge of managing research data and delivering RDM to its research community. It was also mentioned that there is a lack in considering RDS as a structure in models and that most models are RDM oriented (Theoretical analysis results). These arguments are a justification for the appropriateness of combining elements of RDS and RDM in one model.

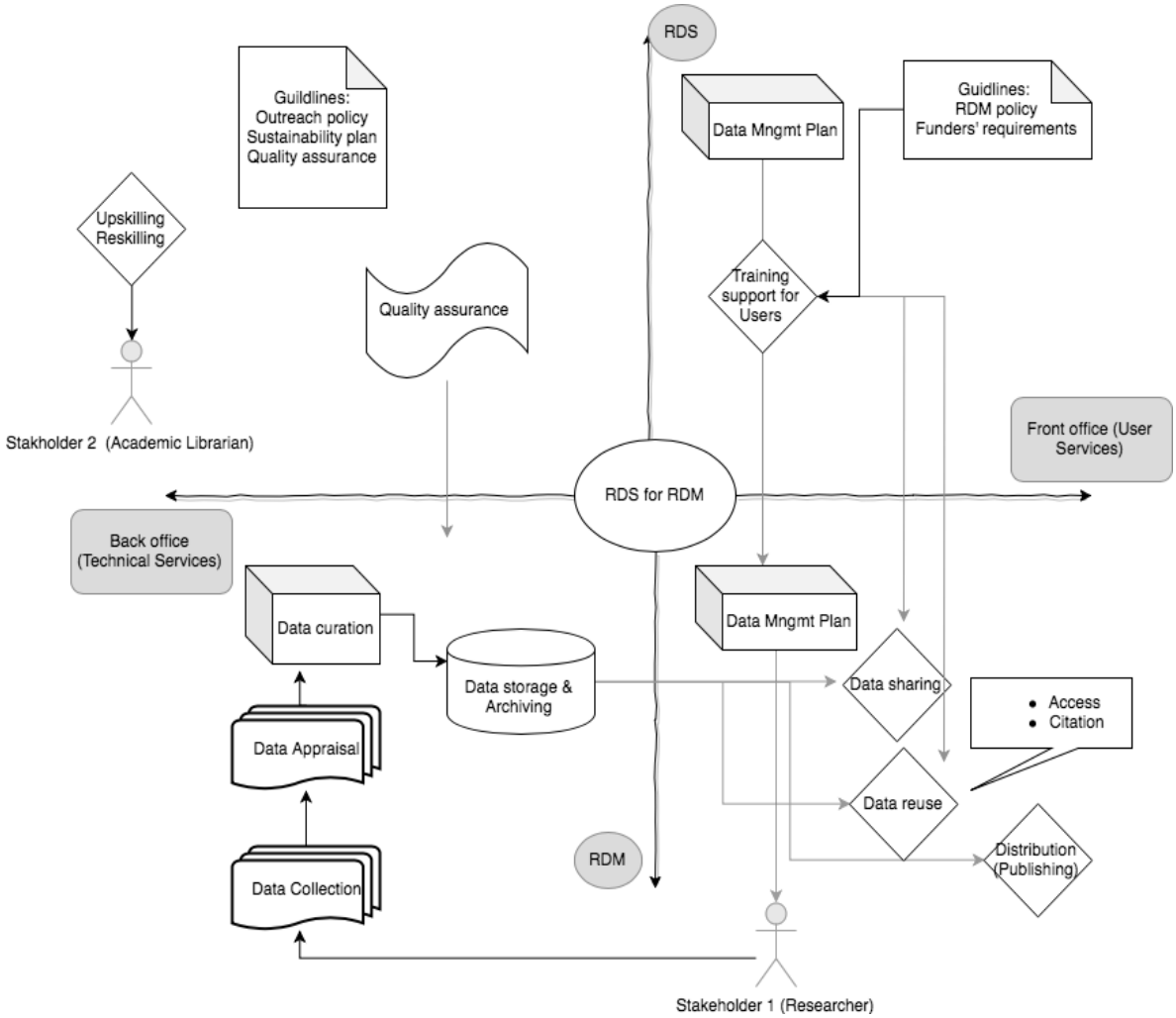


Figure 28 The Combined RDS and RDM Model

Components of the model:

- The vertical lines (lighter colors) are relationships between RDS and RDM
- The horizontal lines (lighter colors) are relationships inter-the same category

The model comprises core elements or tasks of RDS and RDM as well as the two core stakeholders. The elements or tasks are all results of the content analysis. New elements were added following the content analysis, such as:

- Upskilling and reskilling of involved staff, instead of using the word training as it seemed broad in meaning.
- Quality assurance: which the RDS should be in charge of, in order to supervise the quality of offered RDM as well as the quality of the research data service itself.

The process in this model starts from the back office of RDS in which the librarian receives upskilling or reskilling and the RDS's guidelines, strategies and plans are set in order to

have a clear idea about the sustainability (financial, technical, human resources) of the service, its outreach policy and plans to assure quality of the unit as well as the delivered services.

The process resumes with the librarian using his newly acquired or enhanced skills, helps the researcher creating his/her data management plan (DMP), offers training sessions to the researcher, in order to introduce him/her to all they need to know about RDM policies and requirements, these tasks happen in the front office.

In the same front office the researcher learns how to engage in the RDM process, through Summitting research data to the service (RDS) in order to be managed, they also allows or refuse to share their data , and the learn how to reuse data that was managed by the RDS for new research purposes.

This process goes in a circular-like motion in which the cycle starts with the tasks within the back office of the RDS then moves on to enabling the researcher to submit research data then allowing his data to be shared again for another researcher to use (this research will go through the same process).

Chapter 5

Analysis & Discussion

5.1. Phase 1: Theoretical Analysis

5.1.1. RQ1-What are the Requirements for Establishing an Effective RDS?

Theoretically, through the analyzed models' tables, concluding the requirements for establishing an effective research data service might be challenging.

The challenge through this research, was the lack of modelling for the management of RDS, and the lack of explicitly including staff needs and requirements, and the significance disparity in terms of elements in these models, which again are used in the context of academic libraries. This might influence the quality of the offered service in research data management, as well as the perception of researchers of academic libraries offering these services.

5.1.2. RQ2-What Kind of Models are Used to Manage and Maintain RDS in Academic Libraries?

Models used in RDS and RDM

Different models are used for managing research data, such as: the Digital Curation Center (DCC)'s Curation Lifecycle Model (Higgins, 2008) or the USGS Science Data Lifecycle Model (Faundeen, 2013), however few models are explicitly oriented to the management of research data service, therefore through this research, it was assumed according to the type of elements provided in the model, whether it can used for the management of a research data service, thus I have models which are both RDS and RDM oriented, such as the Research Data Management Model at UCF (University of Central Florida, 2018).

It is recognized that academic libraries are evolving environments that need to constantly adapt to the "changing research and learning behaviors" of its community (Dempsey & Malpas, 2018), therefore, the standardization of services can be challenging in such landscape.

However, the lack of standardization of RDM and RDS models can create a disparity in terms of elements, tasks or functions and can be subjectively influenced by the academic library. This can lead to two major issues:

1- Perceiving the situation of RDS in academic libraries can be challenging to researchers investigating the situation of RDS in this field of research.

2- Approaching and dealing with RDS in academic libraries can be challenging for academic librarians.

5.1.3. RQ3-Which are Elements Emphasized in Most Models?

The RDS and RDM table reveals the pattern of change/evolvement that has been taking place, in terms of the importance given to research data management models in contrast to research data services models:

The modeling that has been done between 2003 and 2007, involved the creation of research data lifecycle models or research lifecycle models, which are more focused on how to manage research data (by researchers and other stakeholders), for instance: in the analyzed models, 13 out of 18 models are RDM oriented, in which 10 can be embedded in RDS activities, rather than research data curation only, examples are Lyon (2003), Pepe et al. (2009) and Grigorov et al. (2014), etc.

Prior to 2007, models mostly focused on RDM practices, for instance: Lyon (2007) included more elements which are assumingly associated to the management of RDS as it was not stated that the model is RDS oriented, in addition to describing the traditional RDM process (Lyon, 2007).

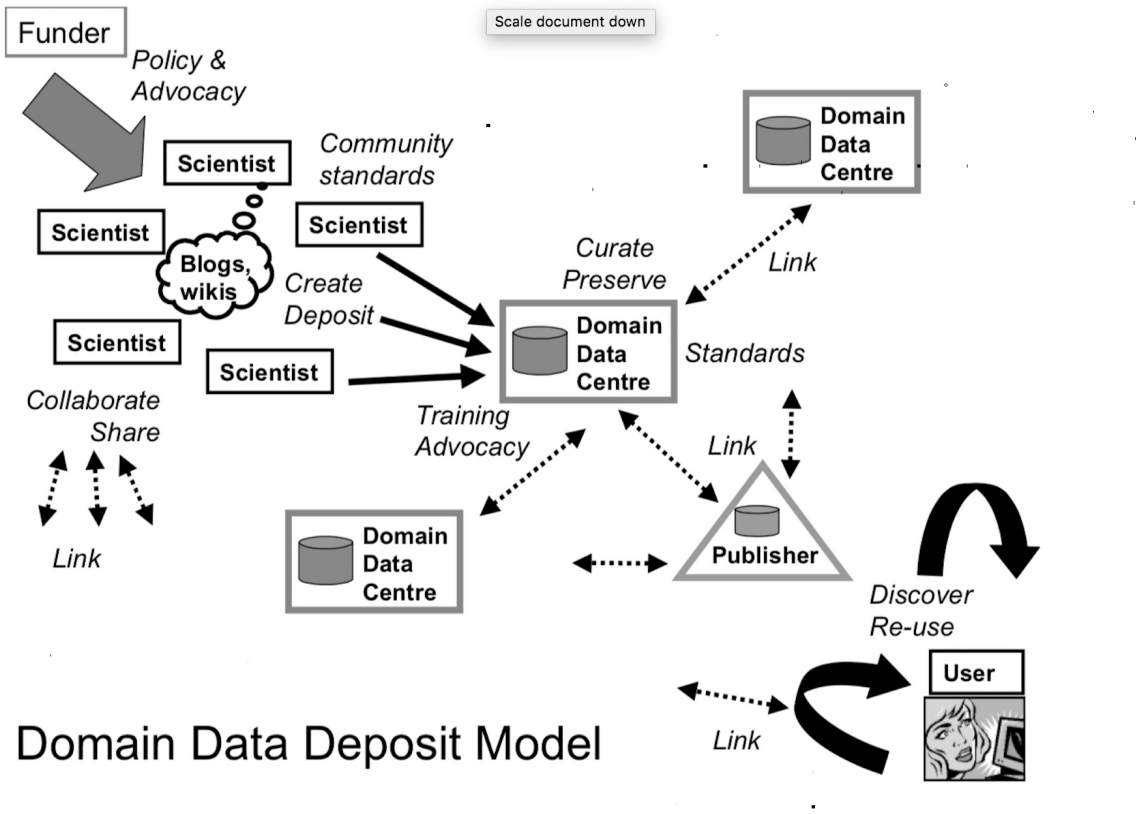


Figure 29 Domain Data Deposit Model (Lyon, 2007)

This slight change in pattern might have been influenced by different factors:

- The growing interest in the management of RDS and the changing perception of traditional academic librarian's roles.
- The maturing of technologies contributed in shifting the interest of different organizations from focusing on technical issues related to preservation and curation of research data to considering new trending elements such as "skills enhancement", "open access" and "sharing data".

After 2007 More components are related to the development and management of research data services as a structure, models like Sykes (2009) introduced a significant change as it showed a clear shift of interest towards this aspect, his modelling, broad and simplistic, focused mainly on the stakeholders and their roles in contributing in the RDS.

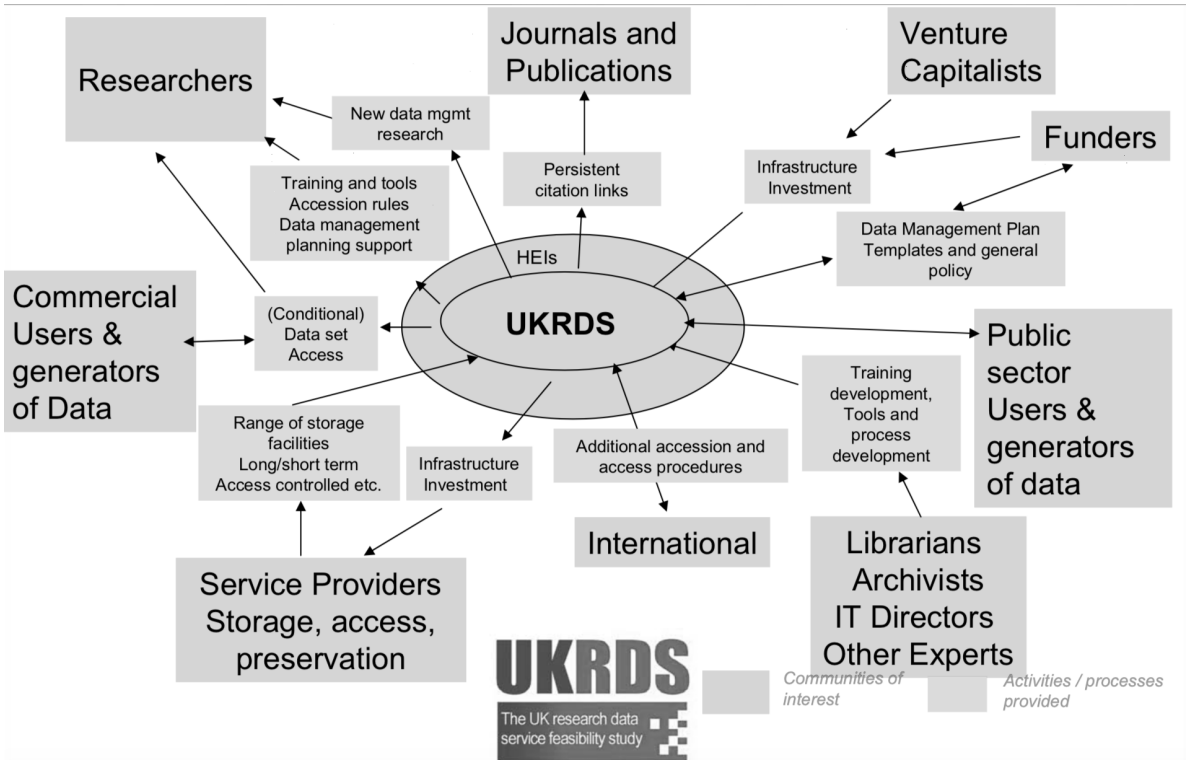


Figure 30 The UK Research Data Service Model (Sykes, 2009)

Throughout our analysis, it is noticeable that the importance has been given to functions such as '**collecting, preserving & storing'** research data, in most of the models. This proves how these functions have been the core functions of RDM.

However, this importance has been shifting towards two elements, namely: **data sharing and re-using** which are both important to researchers and challenging for data librarians.

Although the models are evolving in terms of elements, the researchers and staff's **"skill enhancement trainings"** to use RDS are **not presents in most models**, this aligns with the results of previous surveys (Tenopir et al., 2014; Cox & Pinfield, 2014) in which some academic librarians are unsatisfied with their upskilling or reskilling opportunities and unsure of their abilities in terms of managing RDS and RDM.

5.2. Phase 2: Qualitative Content Analysis

During this phase the content of the two websites of research data services were analyzed, namely the Research Data Oxford of the University of Oxford, and the Harvard Library, Research Data Management Program of Harvard University.

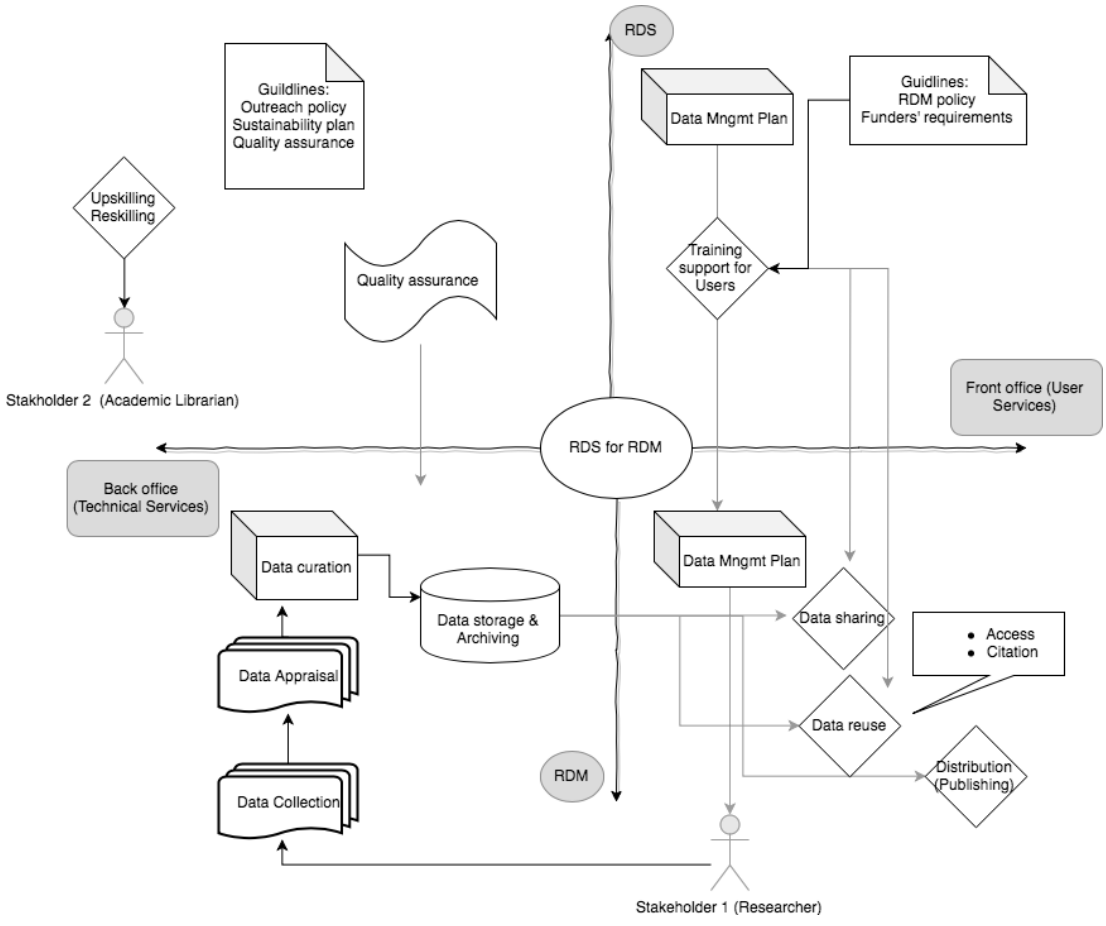
As coding is an interpretive process, there should be a recognition of the fact that coding might be subject to background influence, professional or academic interest and therefore subjectivity. The validity of this study and its conclusions are more likely influenced by my position as an academic librarian and view point as a practitioner in this field.

5.2.1. Newly Found Elements Through the Content Analysis

New elements resulted of the content analysis (as shown the Coding manual part). At the beginning of the analysis, only the elements extracted from the theoretical analysis results were used, however during the content analysis, newly found elements were encountered and added to the coding scheme.

5.3. The Combined RDS and RDM Lifecycle Model in Academic Libraries

As a result of this research, the construction of a model which can ideally combine elements partaking to RDS and RDM, was thought of. It is meant to be a workflow which combines the needs and requirements of both: the service (RDM) and the service (RDS) offering RDM.



(Reshown in Figure 27)

The Combined Lifecycle Model, Why a lifecycle?

The proposed model illustrated the tasks carried out by the research data service in providing research data management to the research community, this process starts with the creation and use of research data and it re-starts through sharing and re-using the same data or the outcome of using that research data.

The same theory applies to the research data "service", which offers the process of managing data, and also follows a process to keep the service itself alive.

This model does not consider RDS (as a structure/ organization) and RDM (as a service) in a process which has a start an end, it is alive and continuous through time, hence the use of "lifecycle".

Why a Combined Model?

The results from the theoretical analysis as well as the content analysis, reveal that:

1- Most models used in academic libraries as far as research data is concerned, are Research data management models, solely used to manage research data. This leaves a gap in terms of models managing the research data service which is responsible of managing the research data itself, in other words, how to ensure that RDM will be delivered efficiently to the research community if I am unsure about the management of the service that provides the RDM?

2- The fact that every library uses their own model and the lack of a standard model, shows that elements in those models can be given importance in different levels, for instance: Research Data Oxford concentrates more on Data Security and Sharing than Harvard Library' RDM Program.

The elements mentioned above can influence directly the performance of the RDS, a concreate example would be the findings of the study conducted by Tenopir et al. (2014), in which academic librarians complained about the lack of training in terms of managing research data, as well as their feeling of inadequacy in terms of skills, which in turn results in researchers distrusting academic librarians with their research data. In this example there are two conclusions:

1- Academic librarians' training is important but might not be emphasized in the context of an RDS, as library directors assume their staff have enough skills. (Tenopir, 2014)

2- The lack of training influences the performance of the service as well as the cooperation of its community.

The proposed model is emphasizing the fundamental elements which can be considered the baseline for creating more complex models which can fit the context of their respective academic library, all the while considering the most important elements that should ensure the efficiency and efficacity of a research data service in an academic library.

Chapter 6

Conclusion

In this research the most cited models used in managing research data in the context of academic libraries or research institutions were investigated and the websites of two prestigious universities' research data services were analyzed, in order to understand the situation of Research Data Services in academic libraries.

Through this research the author concluded that the situation of research data service in academic libraries appeared to be uncertain and literature in this field of research revealed to be disorganized, the conclusion of the author is based on the following reasons: the lack of standardization and research surveys.

6.1. The Combined Model

The model is far from being complex, as it was elaborated in order to present the fundamental elements which should be present in any model that is used within a research data service, as through this research it was revealed that not all models used in academic libraries comprise the same functions. Also, the fact that the function might omitted or not prioritized in the used model could be the reason why it is omitted in the RDS practices of a specific academic library. For instance: according to the results of the content analysis, the Research Data Oxford focuses more on data sharing policy and practices, data Security and user training and outreach., while Harvard Library's RDM Program focuses on data management plan development, data preservation and data sharing.

In this research the combination of RDS and RDM elements is significant as it ensures the equal use of the core elements in one model.

6.1.1. Significance of the Combined Model as an Academic Librarian:

As an academic librarian, standardization is crucial in ensuring the efficiency and efficacity of any provided service in the field of library and information science (Matysek, 2015). Creating a standardized model to manage RDS does not annulate the fact that the model has to be adapted and customized to fit the context of the academic library, for instance: in cataloging Standardization is most important for several reasons such as interoperability, but libraries adapt the rules to their context all while preserving the essential basic elements. The model created in this research has the same perspective.

A standard model can help in:

- Clarifying the tasks to perform

- Simplifying the concept behind research data services and the roles an academic librarian can have in it,
- Facilitating the engagement in research data services,
- Preserving the needs of academic librarians and other involved staff and stakeholders and taking those into account, for instance: training for staff (reskilling or upskilling),
- The embedding of certain elements in a standard model, therefore it will ensure that the academic library or the research data service will -ideally- abide by the model and consider or the elements crucial to the success of the RDS.

6.1.2. Significance for Researchers:

The model is as well important for researchers, as it will be easier to get an idea of the current situation of RDS in academic libraries through the elements emphasized in the model. As well as understanding the fundamental elements of an RDS as the model is simple and concise.

The clarification of the situation of RDS in academic libraries as well as the clear perception of academic librarians of RDS and their roles in the said service can ensure the improvement of the services offered by an RDS, thus, improve the trust and the relationship between the academic librarians and their research community.

6.2. Limitations of This Research

In conducting the content analysis on web-content, it was difficult to find information about the guidelines and policies used to manage the research data service, as well as trainings of stuff even after getting in touch with the manager of Harvard Library's RDM Program.

Organizing literature about the current situation of RDS was challenging as the topic of research data services in academic libraries is not that old, which is why literature is still disorganized and approached by researchers based on their own perspectives. Models are various and lack in uniformity.

Although it has been more than two decades since the discussion about RDS and RDM in academic libraries has started, but literature in this sense is still to be enriched, specifically in terms of surveys, interviews of the perception of the librarians' community.

Although those limitations may have affected this research's results, but they certainly leave room for further research.

6.2.1. The Lack of Standardization:

Different types of models are used for the management of research data services depending on the academic library's context, amongst which, most are models used to manage the research data lifecycle rather than the RDS, which makes the use of some of

these models questionable as to whether the models might be lacking important elements as it was not conceived to manage an RDS in the first place.

The lack of standardization of the models used to manage RDS makes it difficult for researchers to get a clear idea on how RDS should be operated efficiently. Which was one of the limitations of this research as well.

6.2.2. Lack of Research Surveys:

Currently, academic libraries are starting to engage in offering RDS, however, the situation of these services and their management is unclear in literature as there is few researches conducted to investigate RDS in academic libraries from the perspective of this research.

The issue mentioned above leaves room for questioning the equality and quality in terms of offered services in academic libraries. Therefore, the importance of creating a model that englobes the fundamental elements of RDS and RDM in order to ensure that the academic librarians' and other involved stakeholders' needs as well as the researchers' are taken into consideration, and could be a baseline on which a more complex model could be created to befit the context to which the RDS belongs.

6.3. Further Studies

The author strives to conduct future research on the perception of academic librarians specifically data librarians, of research data services, and further research on the evolution of librarians' roles in academic libraries in light of the emerging concepts such as RDS, open data and open science.

Acknowledgement

I would first like to thank Both my advisors Professor Saori DONKAI Professor of the Faculty of Library, Information and Media Science at the University of Tsukuba and Professor Masanori KOIZUMI Assistant Professor of the Faculty of Library, Information and Media Science at the University of Tsukuba, for their continuous guidance, support and encouragement, without which this research would not have been possible. I am also grateful to all my friends at Koizumi Laboratory for their moral support, constant help and for welcoming me amongst them.

I would also like to thank my classmates and friends who made my stay in Japan more enjoyable and all of my professors for the successful learning experience.

Of course, this acknowledgement would not be complete without thanking my parents, who supported and encouraged my activities, to them go my sincere gratitude for their constant love and support.

References

Auckland, M. (2012). Re-skilling for research: An investigation into the role and skills of subject and liaison librarians required to effectively support the evolving information needs of researchers. Retrieved from: https://www.rluk.ac.uk/portfolio-items/re-skilling-for-research/

Borgman, C. L., Wallis, J. C., & Enyedy, N. (2007). Little science confronts the data deluge: Habitat ecology, embedded sensor networks, and digital libraries. I*nternational Journal on Digital Libraries,* 7(1-2), 17-30. doi:10.1007/s00799-007-0022-9

Briney, K., Goben, A., & Zilinski, L. (2015). Do you have an institutional data policy? A review of the current landscape of library data services and institutional data policies. J*ournal of Librarianship and Scholarly Communication,* 3(2), 1-25. doi:10.7710/2162-3309.1232

Burton, M., & Lyon, L. (2017). Data science in libraries. b*ulletin of the Association for Information Science and Technology*, 43(4), 33-35. doi:10.1002/bul2.2017.1720430409

Carlson, J. (2014). The Use of life cycle models in developing and supporting data services. *Research data management: Practical Strategies for Information Professionals*. West Lafayette: Purdue University Press., 63-86.

Cole, F. L. (1988). Content analysis: Process and application. *Clinical Nurse Specialist*, 2(1), 7-53. Retrieved from: https://journals.lww.com/cns-journal/Abstract/1988/00210/Content_Analysis__Process_and_Application.25.aspx.

Corrall, S. (2008). Research data management: Professional education and training perspectives. (p. 18). Retrieved from: http://www.dcc.ac.uk/sites/default/files/documents/RDMF/RDMF2/07%20Corrall.pdf

Corrall, S. (2012). Roles and responsibilities: Libraries, librarians and data in managing research data. Facet Publishing, London, 105 - 133. Retrieved from: http://jn2xs2wb8u.search.serialssolutions.com.ezproxy.tulips.tsukuba.ac.jp/?url_ver=Z39.88-2004&rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Abook&rft.genre=bookitem&rft.btitle=Managing%20research%20data&rft.atitle=Roles%20and%20responsibilities%3A%20Libraries%2C%20librarians%20and%20data&rft.spage=105&rft.epage=133&rft.date=2012&rft.aulast=Corrall&rft.aufirst=S.&rfr_id=info%3Asid%2Fwiley.com%3AOnlineLibrary

Cox, A. (2016). Research data management and the library and information science professions. Retrieved from: http://conferences.sun.ac.za/public/conferences/31/papers/SULIS-Symp14-RDM-Workshop-AndrewCox.pdf

Cox, A., Kennan, M. A., Lyon, L., & Pinfield, S. (2017). Developments in research data management in academic libraries: Towards an understanding of research data service maturity. J*ournal of the Association for Information Science and Technology*, 68(9), 2182–2200. https://doi.org/10.1002/asi.23781

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

Cox, A., 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

Creswell, J. (2015). Educational Research: Planning, conducting, and evaluating quantitative and qualitative research. (p. 672). New York: Pearson.

Data Documentation Initiative (DDI). (2004). DDI version 3.0 conceptual model (p. 10) [Working draft]. Retrieved from: https://www.ddialliance.org/sites/default/files/Concept-Model-WD.pdf

Dempsey, L., & Malpas, C. (2018). Academic library futures in a diversified university system. *Higher Education in the Era of the Fourth Industrial Revolution* 65–89. https://doi.org/10.1007/978-981-13-0194-0_4

Donnelly, M., Rusbridge, C. (2008). Research data management forum: RDMF2: Core skills diagram. Retrieved from: http://data-forum.blogspot.com/2008/12/rdmf2-core-skills-diagram.html

Economic and Social Research Council (ESRC). (2018). ESRC Research data policy. (p. 9). Retrieved from: http://www.esrc.ac.uk/files/about-us/policies-and-standards/esrc-research-data-policy/

Elliott, V. (2018). Thinking about the coding process in qualitative data analysis. *The Qualitative Report*, 23(11), 2850-2861. Retrieved from: https://nsuworks.nova.edu/tqr/vol23/iss11/14

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). Qualitative content analysis: A focus on trustworthiness. *Sage Open*, 4(1), 5–6. https://doi.org/10.1177/2158244014522633

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107–115. https://doi.org/10.1111/j.1365-2648.2007.04569.x

Faniel, I. M., & Connaway, L. S. (2018). Librarians' perspectives on the factors influencing research data management programs. *College & Research Libraries,* 79(1), 100–119. 10.5860/crl.79.1.100

Faundeen, J. L., Burley, T. E., & Carlino, J. A. (2013). The United States geological survey science data lifecycle model: U.S. geological survey open-file report. (p. 4). http://dx.doi.org/10.3133/ofr20131265.

Federer, L. (2016). Research data management in the age of big data: Roles and opportunities for librarians. I*nformation Services & Use*, 36(1–2), 35–43. https://doi.org/10.3233/ISU-160797

Fei Yu, Deuble, R., & Morgan, H. (2017). Designing research data management services based on the research lifecycle: A consultative leadership approach. J*ournal of the Australian Library & Information Association*, 66(3), 287–208. Retrieved from: https://www.tandfonline.com/doi/full/10.1080/24750158.2017.1364835

Grigorov, I. (2014). Research lifecycle enhanced by an "Open science by default" Workflow. Retrieved from: https://www.rri-tools.eu/-/research-lifecycle-enhanced-by-an-open-science-by-default-workflow

Harvard University, O. of I. R. (2019). Fact book: organizational chart - faculties and allied institutions (2018-19). Retrieved from: https://oir.harvard.edu/fact-book/org_chart_schools

Higgins, S. (2008). The DCC curation lifecycle model. *International Journal of Digital Curation*, 3(1), 134–140. https://doi.org/10.2218/ijdc.v3i1.48

Humphrey, C. (2006). E-Science and the life cycle of research. (p. 5). https://doi.org/10.7939/R3NR4V

Ikeuchi, U., Harada, T., Sato, S., Okabe, Y., & Itsumura, H. (2017). Data literacy perceptions and research data management practices by researchers in Japan. Retrieved from: http://jairo.nii.ac.jp/0025/00043290

Kellam, L., & Thompson, K. A. (2016). Data librarianship: The academic data librarian in theory and Practice. ProQuest. Retrieved from: https://search-proquest-com.ezproxy.tulips.tsukuba.ac.jp/docview/1808433223?pq-origsite=summon

Kim, I., & Kuljis, J. (2011). Applying content analysis to web-based content. *CIT. Journal of Computing and Information Technology*, 18(4), 369–375. https://doi.org/10.2498/cit.1001924

Kyngas, H., & Vanhanen, L. (1999). Content analysis as a research method [Finnish]. Hoitotiede, 11, 3-12.

Koltay, T. (2017). Data literacy for researchers and data librarians. J*ournal of Librarianship and Information Science*, 49(1) 3–14. https://doi.org/10.1177/0961000615616450.

Krippendorff, K. (1980). Content analysis: An introduction to its methodology. (p. 456). California: Sage Publications.

Kumasi, K. D., Charbonneau, D. H., & Walster, D. (2013). Theory talk in the library science scholarly literature: An exploratory analysis. *Library & Information Science Research*, 35(3), 175–180. https://doi.org/10.1016/j.lisr.2013.02.004

Leedy, P. D. (1989). Practical research: Planning and design. (p. 318). London: Macmillan ; Collier Macmillan.

Lewis, M. J. (2010). Libraries and the management of research data. Envisioning Future Academic Library Services (pp. 145–168). Retrieved from: http://www.facetpublishing.co.uk/title.php?id=046916#.VtR2MvmLSUk

Lyon, L. (2003). Ebank UK: Building the links between research data, *Scholarly Communication and Learning. Ariadne*, (36). Retrieved from: http://www.ariadne.ac.uk/issue36/lyon

Lyon, L. (2007). Dealing with data: Report. (p. 65) Retrieved from: http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf

Ma, F., & Wang, J. (2010). A literature review of studies on information lifecycle I: The perspective of value. *Journal of the China Society for Scientific and Technical Information*, 29(5), 939–947.

Mackey, A., & Gass, S. M. (2013). Second language research: Methodology and design. (p. 424). London: Routledge.

Matysek, A. (2015). Standardization in library and information science in selected European countries. AIP Conference Proceedings, 1644(1), 91–96. https://doi.org/10.1063/1.4907822

Neuendorf, K. A. (2002). The Content analysis guidebook. (p. 320). California: Sage.

Norton, J. D. (2010). Deductively definable logics of induction. *Journal of Philosophical Logic,* 39(6), 617–654. https://doi.org/10.1007/s10992-010-9146-2

Ohaji, I. K. (2016). Research data management: An exploration of the data librarian role in New Zealand research organisations (PhD, Victoria University of Wellington). Retrieved from: https://researcharchive.vuw.ac.nz/xmlui/bitstream/handle/10063/5657/thesis.pdf?sequence=1

Omekwu, C. O. (2003). Information technology revolution, libraries and cultural values: Issues, impacts and inevitable challenges for Africa. World Library and Information Congress: 69th IFLA General Conference and Council at Berlin, 23.

Osswald, A., & Strathmann, S. (2012). The role of libraries in curation and preservation of research data in Germany: Findings of a survey. (p. 10). Retrieved from: <https://www.ifla.org/past-wlic/2012/116-osswald-en.pdf>

Patel, M. (2011). I2S2 Idealized scientific research activity lifecycle model. Retrieved from: http://opus.bath.ac.uk/35186/

Pinfield, S., Cox A., & Smith. J. (2014). Research data management and libraries: Relationships, activities, drivers and influences. *Plos One* 9 (12) https://doi.org/10.1371/journal.pone.0114734.

Pepe, A., Mayernik, M., Borgman, C. L., & Van de Sompel, H. (2009). From artifacts to aggregations: Modeling scientific life cycles on the semantic web. J*ournal of the American Society for Information Science and Technology*, n/a-n/a. https://doi.org/10.1002/asi.21263

Potter, R. F. (1999). Measuring the "Bells & Whistles' of a new medium: Using content analysis to describe structural features of cyberspace, 49th Annual Conference of the International Communication Association at San Francisco, 99.

Pryor, G. (2012). Managing research data. (p. 224). London: Facet Publishing. Retrieved from: https://www.amazon.com/Managing-Research-Publications-Titles-Published/dp/1856047563

Pryor, G., & Donnelly, M. (2009). Skilling up to do data: whose role, whose responsibility, whose career? *International Journal of Digital Curation* 4(2), 158-170. https://doi.org/10.2218/ijdc.v4i2.105.

Reed, K. L. (1984). Understanding theory: The first step in learning about research. *American* J*ournal of Occupational Therapy*, 38(10), 677–682. https://doi.org/10.5014/ajot.38.10.677

Research Data Oxford. (2018). Retrieved from: http://researchdata.ox.ac.uk/

RIN, & NESTA. (2010). Open to all? Case studies of openness in research (p. 50). Retrieved from: http://www.rin.ac.uk/system/files/attachments/NESTA-RIN_Open_Science_V01_0.pdf

Severt, C. (2005). Discovering a profession: The accidental data librarian. Presented at the IASSIST Conference, Edinburgh. Retrieved from: https://iassistdata.org/blog/discovering-profession-accidental-data-librarian

Stanton, J. M. (2012). Data science: What's in it for the new librarian? Retrieved from: https://ischool.syr.edu/infospace/2012/07/16/data-science-whats-in-it-for-the-new-librarian/

Steeleworthy, M. (2014). Research data management and the Canadian academic library: an organizational consideration of data management and data stewardship. *The Canadian Journal of Library and Information Practice and Research*, (1), 1–11.

Swan, A., & Brown, S. (2008). The skills, role and career structure of data scientists and curators: An assessment of current practice and future needs [Monograph]. Retrieved from: https://eprints.soton.ac.uk/id/eprint/266675

Sykes, J. (2009). UK research data service feasibility study: A briefing. presented at the group RPEG, London. Retrieved from: http://eprints.lse.ac.uk/25570/1/UKresearchDataServiceFeasibility3_Oct08%28LSERO%29.pdf

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*, 36(2), 84–90. https://doi.org/10.1016/j.lisr.2013.11.003

Tenopir, C., Talja, S., Horstmann, W., Late, E., Hughes, D., Pollock, D., … Allard, S. (2017). Research data services in european academic research libraries. *Liber Quarterly: The Journal of European Research Libraries*, 27(1), 23–44. https://doi.org/10.18352/lq.10180

The University of Edinburgh. (2018). Research data service. Retrieved from: https://www.ed.ac.uk/information-services/research-support/research-data-service

Times Higher Education (THE). (2019). University of Oxford. Retrieved from: https://www.timeshighereducation.com/world-university-rankings/university-oxford

University of Central Florida. (2018). Overview: Research lifecycle. Retrieved from: https://library.ucf.edu/about/departments/scholarly-communication/overview-research-lifecycle/

University of Oxford. (2019). History of the Bodleian. Retrieved from: https://www.bodleian.ox.ac.uk/about/history

University of Oxford. (2019). Libraries | University of Oxford. Retrieved from: https://www.ox.ac.uk/research/libraries?wssl=1

University of Sydney Library. (2018). About research data management, University of Sydney Library. Retrieved from: https://library.sydney.edu.au/research/data-management/research-data-management.html

Verbaan, E., & Cox, A. (2014). Occupational sub-cultures, jurisdictional struggle and third space: Theorizing professional service responses to research data management. *Journal of Academic Librarianship*, 40(3/4), 211–219. https://doi.org/10.1016/j.acalib.2014.02.008

Waard, A. (2014). Research data management at institutions | Part 1: Visions. Retrieved from: https://libraryconnect.elsevier.com/articles/research-data-management-institutions-part-1-visions

Wang, L. (2018). Twinning data science with information science in schools of library and information science. *Journal of Documentation*, 74(6), 1243–1257. https://doi.org/10.1108/JD-02-2018-0036

Whyte, A., & Tedds, J. (2011). Making the case for research data management (p. 8). Retrieved from: http://www.dcc.ac.uk/webfm_ send/487

Yoon, A., & Schultz, T. (2017). Research data management services in academic libraries in the US: A content analysis of libraries' websites. *College & Research Libraries,* 78(7), 920–933. Retrieved from: https://crl.acrl.org/index.php/crl/article/view/16788/18346

Yu, H. H. (2017). The role of academic libraries in research data service (RDS) provision. *Electronic Library*, 35(4), 783–797. https://doi.org/10.1108/EL-10-2016-0233

Zaidane A. & Koizumi M. (2019). Roles of data librarians and research data services in academic libraries. Unpublished manuscript. (Submitted to the conference BOBCATSSS2019)