Research Libraries and Research Data Management within the Humanities and Social Sciences

Åhlfeldt, Johan; Johnsson, Maria

2015

Research libraries and research data management within the humanities and social sciences

MARIA JOHNSSON - JOHAN ÅHLFELDT | UNIVERSITY LIBRARY | LUND UNIVERSITY

PROJECT REPORT

Research Libraries and Research Data Management within the Humanities and Social Sciences

Project report

Maria Johnsson and Johan Åhlfeldt
University Library, Lund University

LUND
UNIVERSITY

Lund University Library, Department of Research and Study Services
This report is available as PDF, http://lup.lub.lu.se/record/5050462

Printed in Sweden by Media-Tryck, Lund University
Lund 2015

Contents

List of figures

Executive Summary

Research Data Management (RDM) is a process that is designed to deliver high quality datasets, which comply with scholarly, legal and ethical requirements. There are two outputs of the RDM process:

1. Long term preservation of datasets through archiving
2. Sharing and reuse of datasets for further research and other purposes in society at large.

This proposal outlines the creation of a coherent Research Data Management organization at Lund University that utilizes existing resources both within and outside the university and establishes new organizational units and information systems, specific to this new task.

We propose the establishment of a new unit for *Research Data Management and Coordination* at the University Library whose responsibility would be to coordinate the network of existing agents who support research activities such as faculty libraries and ethical, legal, archival and data management experts.

We further propose the creation of a new information system, the *Lund University Dataset Directory*, which will facilitate management of datasets and information retrieval throughout the data lifecycle. We expect that research datasets could be deposited for sharing at national or disciplinary repositories and eventually archived when a solution is in place at the University Archive. Advanced RDM - like semantic web technologies - will require online data services not currently provided by national agents. We therefor propose a *Data Laboratory* within the RDM network at Lund University.

Finally, it's important to recognize that Research Data Management is a new way of organizing information with its own set of tasks for the library organization. Our efforts in RDM will require us to invest significant effort in learning new systems, ways of working and collaboration.

Foreword

We want to thank Riksbankens Jubileumsfond[1] who has supported this project with funding. We are most grateful to their engagement and interest in this research area.

We also want to thank all generous persons who have participated in the many interesting discussions and meetings that we have had in this project. There is a lot of interest and engagement into the themes of this project – i.e. research data management, open data, data publication.

Reference group

Throughout the project we have had regular meetings (in total two meetings) with a reference group consisting representatives from research fields within humanities, theology, social sciences, from University Library, and from the university archive:

[1] Riksbankens Jubileumsfond – the Swedish foundation for humanities and social sciences, http://www.rj.se/ , accessed 2015-02-03

Samuel Rubenson, Centre for Theology and Religious Studies
Marianne Gullberg, Centre for Languages and Literature/ Humanities Laboratory
Matthias Baier, Sociology of Law
Eva Nylander, University Library
Anne Lamér, Record management and Archives

The reference group has acted as a sounding board, with which we have presented and tested different ideas and concepts in the project. The members of the reference group have also contributed a lot with their knowledge and experiences of handling research data, as well as their expectations of a future system for research data management.

Project meetings

The project has been organized within the Unit for Scholarly Communication, part of the Department of Research and Study Services, Lund University Library. We have had regular project meetings (in total five meetings) with the following persons:

Christel Smith, Head of Research and Study Services
Kristoffer Holmqvist, Research and Study Services, Scholarly Communication
Jörgen Eriksson, Research and Study Services, Scholarly Communication

The project meetings have been more practical in its sense, and during these meetings we have discussed and planned in which way we should perform the study of the activities at University Library.

Individual meetings within Lund University

We have met many people individually and interviewed them about their views, knowledge and experiences on research data management and related issues. We have met with people from the following sections and departments:

University Library, Scholarly Communication
University Library, Library Services
University Library, Acquisitions
University Library, Cataloguing and Metadata
University Library, E-resources
University Library, Manuscripts and Special Collections
University Library, Professional Development and Communication
University Library, IT and Technical Support
Library of faculty of humanities and theology
Library of faculty of social sciences
Library of faculty of business and economics
Record management and archives
Research support services
Research ethics network
Department of Arts and Cultural Sciences, Division of Archives, Libraries and Museums

Individual meetings outside Lund University

We have met with people outside Lund University during meetings, conferences and workshops. We have had several meetings with the staff of Swedish National Data Service (SND). The contacts and the meetings with SND have given us a lot of good ideas and inspiration throughout the project.

Correspondence

For further correspondence contact:

Johan Åhlfeldt, e-mail: johan.ahlfeldt@teol.lu.se

or Maria Johnsson, e-mail: maria.johnsson@bibliotek.lth.se

1. Introduction to the project

In this project we have studied the future role of research libraries within research data management with a focus on humanities and social sciences. The project was commissioned by the University Library, Lund University, who wanted to investigate what role the University Library might take in research data management. The original idea for the project came from the authors of the research project *Early Monasticism and Classical Paideia*, who asked whether the University Library could host their database of ancient Christian texts. Through literature studies, interviews within and outside Lund University, workshops and a researcher survey, we have investigated the potential for and the conditions under which Lund University Libraries could get involved in research data management at the university.

Lund University and its library organization – a short description

Lund University (LU) is a major university in the south of Sweden with about 47,000 students and about 7,000 employees – academics and administrative staff. The university hosts a broad range of subject areas across a decentralized organization in which each faculty is responsible for its own library services. There are 26 library units in the university, including faculty libraries, special libraries and the University Library. [2]

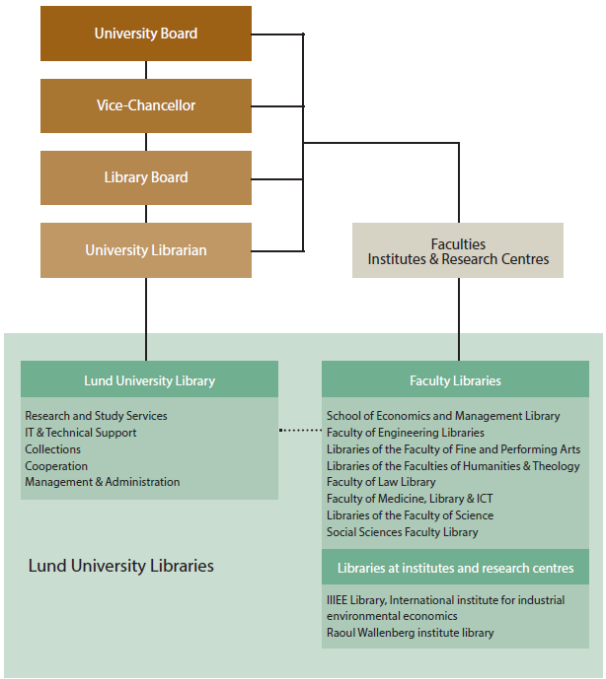

Figure 1

[2] University Library is called Universitetsbiblioteket – UB in Swedish.

The libraries co-operate within a network, with a number of common services. In this report we call this library network "Lund University Libraries (LUL)". Together, LUL manages and maintains a large and broad collection of library and information services, such as databases, e-journals, e-books, print collections, and tools for information management. LUL operates within a common strategy for the libraries at LU, which provides standardized library and information services. The University Library has a role as co-ordinator for the common services in LUL, and it is also a national deposit library. As this project was commissioned by University Library, to investigate its future role, we have been focusing a lot on studying the internal organization of the University Library. But we have also studied the other libraries at LU, in particular the faculty libraries of humanities and social sciences. We have also studied other departments and functions within LU working with research support services and research data management, such as the archive and legal department.

Research methods

This project is highly local in its character in that it seeks to investigate the conditions for the University Library, the faculty libraries and other relevant agents (e.g. archivists, IT-professionals, experts in law and research ethics) at Lund University to engage in research data management (RDM). But this project can also be regarded as a case study for similar research libraries within Swedish universities. Therefore, we have framed our conclusions around how best to organize RDM in the Swedish context in general. We have concentrated on a detailed description of the current organization of LU. Through interviews, meetings and a workshop, we have tried to do a systematic overview of the library departments and its competencies at LU. Using qualitative methods as interviews and meetings, we were able to initiate more in-depth discussions with the library staff about their competencies and their future roles.

In order to capture the current views of the researchers in humanities and social sciences, we chose to perform a survey on their experiences on digital research data. In this case we were able to check the current status and trends among researchers within these major areas at LU. Through the survey we got some interesting quantitative data that we will analyze further.

The conclusion of both the literature review and our participation in external conferences and workshops on the themes of RDM—such as "Digital Humanities at Oxford Summer School 2014, "Bibliotheca Baltica 12th International Symposium", SND national research workshop, and "IFFIS 2014"[3]—is that RDM is a dynamic topic that has benefitted from much new research.

[3] Conference on research infrastructures organized by Royal Library and Swedish Research Council.

2. Background and previous research and investigations

Recently, there has been tremendous development in the areas of open access and digitization. There has been a major development of open access to scientific publications, and this movement is now to include the underlying research data of the scientific output. The technologies for digitizing books, manuscripts, images and archives have improved immensely the last years, and this has resulted in many projects of digitization at libraries, archives and museums. Digitization of cultural heritage is one important example of this. With increasing access to data and information from academia and cultural heritage institutions, we will be better able to *capture* and *re-use* data and information. In this chapter we will focus on open access research data: their conditions for access and the international movement of research data management.

2.1 Research data management

Early in the project we considered the challenges presented by research data management[4] at universities. Our main case study object—the database belonging to the research project *Early Monasticism and Classical Paideia*—will be rich with research data. Thus, the structure in which the data is contained and the means by which it is accessed will affect how future researchers are able to enrich the dataset and conduct further analysis upon it. So we wanted to investigate what kind of data infrastructure is needed to allow researchers to effectively curate and share research findings.

There is no general policy or co-ordination for research data management at Lund University (LU). The university has a policy for open access concerning research publications, but nothing similar on research data. There is a network on e-science called "eSSENCE" at LU, but this network is not widespread at the university. There is to be implemented a CRIS[5] system at LU in 2015 called "LUCRIS", but this will not include research data in its first version.

The following three key concepts will be central in the discourse of this chapter:

- **Research data management – RDM**

  Research data management is a general term covering how to organize, structure, store and care for the information used or generated during a research project (University of Oxford [6]).

[4] We will use the term research data management and its abbreviation RDM alternatively through the report

[5] CRIS means current research information system

[6] http://researchdata.ox.ac.uk/home/introduction-to-rdm/ , accessed 2015-02-03

- **Data curation**

  "Data curation is the active and on-going management of data through its lifecycle of interest and usefulness to scholarship, science, and education; curation activities enable data discovery and retrieval, maintain quality, add value, and provide for re-use over time" (Craigin et al, 2007)

- **Data management plan – DMP**

  "A data management plan is a formal document that outlines what you will do with your data during and after you complete your research. It describes the data that will be created, the standards used to describe the data (metadata), who owns the data, who can access the data, how long the data will be preserved (and/or made accessible), and what facilities and equipment will be necessary to disseminate, share, and/or preserve the data. Several funding agencies require or encourage the development of data management plans for research."

  (NCSU Libraries [7])

Research data management (RDM) is growing in importance, as a result of funders' requirements of data management plans, and as a result of an increased "data sharing culture" among researchers. There is also an ongoing policy shift from requiring open access to scientific publications to requiring open access to research data. Both on the national level and on the international level countries are discussing the importance of making public access to research data from publicly financed research in order to raise public insight.

Both in the political discussion and in the discussion among researchers there appear many arguments for bringing open access to research data, such as:

- Improved quality of results: It becomes easier to build on previous research results
- Greater efficiency: It fosters collaboration and avoid duplication of effort
- Faster to market, faster growth: It accelerate innovation
- Improved transparency of the scientific process: It involves citizens and society

(European Commission, 2013)

Globally there have been different developments of research data management policies during the last years. In the following sections we will give a background of this expansion, which we believe will have an impact on Sweden and ultimately on Lund University.

USA

In USA in the early years of 2000 questions were raised by the American Congress about the inefficiency or duplication of research projects funded by federal agencies, such as the National Science Foundation (NSF)[8]. NSF responded to this critique by holding different hearings and started a major investigation led by Daniel Atkins. The report from this investigation was

[7] http://www.lib.ncsu.edu/guides/datamanagement , accessed 2015-02-03

[8] http://www.nsf.gov/, accessed 2015-02-02

groundbreaking and it is often referred to as the *Atkins Report*. One of the findings of the Atkins Report was an expansion of the "Digital Libraries"[9].

In 2006 The Association of Research Libraries (ARL) followed up the Atkins Report by the report "To Stand the Test of Time: Long-term Stewardship of Digital Data Sets in Science and Engineering", which was the result of a workshop. This report is the first to give a detailed description of the proposed role for research libraries in data management. In the years coming after there were several calls for research project focusing on infrastructure for data management. Both NSF and Institute of Museum and Library Services (IMLS) launched such calls for proposals, which resulted in projects such as "DataOne" and "Data Curation Profiles". In 2010 NSF started to require data management plans with all grant applications beginning in the 2011 grant cycle (Ray, JM, Editor, 2014). To summarize the development in USA there has been a lot of investments in infrastructure supporting and encouraging RDM in which libraries have taken a major role in building up support services.

Europe

In Europe the discussion and development of research data management have also been focused on infrastructure, i.e. how to build good e-infrastructures for research data and how to trigger mechanisms among researchers for sharing data. The pilot study "Riding the Wave: How Europe can gain from the rising tide of scientific data" focused on possible scenarios for the future European researcher. The study lists objectives and actions that must be taken to establish e-infrastructures for research in Europe, and it led to the recommendation by the European Commission regarding access to and preservation of scientific information (European Commission, 2012). This recommendation includes clear and detailed objectives for RDM in the member states, bringing aspects such as open access to research data, preservation and re-use of research data and e-infrastructures for research data. In the EU framework program for research and innovation Horizon 2020 the "Open Research Data Pilot" was introduced which should improve and maximize access to and re-use of research data. A number of research areas in Horizon 2020 are included in "Open Research Data Pilot". Projects within these areas will have the obligation to provide a data management plan.

There are initiatives taken in research data management in different co-operations between universities in Europe, such as LERU (League of European Research Universities)[10], of which Lund University is a member. LERU has published "LERU Roadmap for Research Data", which gives very clear recommendations and instructions on RDM to the LERU member universities. This roadmap is worth describing more in detail, as it could be of great help to universities building up an organization for RDM. As countries are deploying national policies for research data, universities will need to develop their local policies and infrastructures. "LERU Roadmap for Research Data" is an excellent source for help and guidance, not only for the LERU members. It describes RDM from following six different aspects:

[9] "Digital Libraries" was a common project created by Defense Advanced Research Projects Agency (DARPA), the NSF and the National Library of Medicine (NLM).

[10] http://www.leru.org, accessed 2015-02-02

- Policy and leadership
- Advocacy
- Selection and Collection, Curation, Description, Citation, Legal Issues
- Research data infrastructure
- Costs
- Roles, Responsibilities and Skills

It states clearly that research data management needs involvement from all kinds of stakeholders at the university, from the management, the research community, the research support services, the libraries, the IT services. It gives a number of recommendations to the member universities, among the following: forming a steering group for research data management, developing an institutional road map for research data, create a data management service, introduce specific job profiles for data preparation and quality assurance.

United Kingdom

In UK the Government published the report "Innovation and Research Strategy for Growth" in 2011 which included a commitment to ensuring that publicly-funded research should be accessible free of charge; and that it would work with partners, including the publishing industry, to achieve that goal. The organization Research Councils UK (RCUK)[11] has adopted common principles for data policy, which are a framework for individual research councils to set-up their data policies (RCUK). The fact that research councils have launched research data policies, has also meant that many British universities have established policies and have set-up local infrastructures for research data management. The establishment of Digital Curation Centre (DCC)[12] has certainly meant a lot to development at British universities. DCC was started in 2004 and is an important service provider when it comes to support and standards in RDM. They also provide a lot of training and seminars in RDM.

Sweden

In Sweden the Government is preparing for national solutions to fulfil the recommendations in the "European Commission recommendation of 17.7.2012 on access to and preservation of scientific information", and in 2014 it assigned to the Swedish Research Council[13] [14]to prepare national guidelines to open access to research information, including research data (Regleringsbrev, VR, 2013).

The preliminary proposition for national guidelines on open access was publicly presented in October 2014, and at the same time the Swedish Research Council made it possible for the public to give comments on the proposition on their website. Many representatives from the research community and from other communities gave their opinions on the proposed guidelines which declare that research data and scientific publications which emanate from

[11] http://www.rcuk.ac.uk, accessed 2015-02-02

[12] http://www.dcc.ac.uk, accessed 2015-02-02

[13] Swedish Research Council is called "Vetenskapsrådet" in Swedish

[14] http://www.vr.se, accessed 2015-02-02

public funding should be made open access within the legal framework. The guidelines are proposed to come into force in 2025 and will be implemented gradually with proposed pilot projects releasing data for an initial 5-year-period.

In order to engage the Swedish research community in this question the Swedish Research Council arranged several meetings and workshops with different stakeholders during 2014. Among the stakeholders were representatives from Swedish research funders, research infrastructures, libraries, archives and IT staff.

The Swedish Research Council has raised several positive consequences of open access on research data such as: public and democratic access to research, new research opportunities, wider possibilities to innovation outside research community which are the same arguments as in the EU recommendations. The Swedish Research Council has also pointed out that it will be a laborious and time consuming process to fully implement the national guidelines. This process is also dependent on the development of research infrastructures internationally.

The Swedish Research Council clearly states the need for building up supporting infrastructures for open access to research data. They also urge the Government to appoint an actor to be a national hub for the development of open access to research data, and to allocate special funds for infrastructures of research data. Swedish universities, according to Swedish Research Council, should take an active part in the development of archiving and long-term preservation of research data. The universities should co-operate in planning for technical solutions and guidelines for research data. The Swedish Research Council is urging the universities to investigate and engage themselves in the questions around research data management – both the technical aspects and the organizational aspects.

By the end of 2014 the proposed national guidelines are to be handed over by the Swedish Research Council to the Government, i.e. the Department of Education.

Research infrastructures in Sweden

The Swedish Research Council is responsible for supporting national and international research infrastructures, and it acts both as a co-ordinator and a finance sponsor for different research infrastructures in Sweden. The Swedish Research Council publishes guidelines for establishing research infrastructures and lists of current research infrastructures in Sweden.

Swedish National Data Service (SND)

One example of a research infrastructure is the Swedish National Data Service (SND)[15], which is a service provider for research in humanities, social sciences and health sciences. It helps researchers to care for research data management, both by providing a research data-set directory and by providing services for depositing data. With extensive experience supporting the social sciences, SND has cultivated an extensive network with similar organizations outside Sweden, provides training and seminars in research data management and via its website offers dataset search and access functions.

Other examples of research infrastructures in Sweden are Environment Climate Data Sweden (ECDS)[16], and Bioinformatics Infrastructure for Life Science (BILS)[17]. These organizations are

[15] http://snd.gu.se, accessed 2015-02-02

[16] http://www.smhi.se/ecds , accessed 2015-02-02

[17] https://www.bils.se , accessed 2015-02-02

subject-specific service providers from which researchers can get help on data publishing and storage.

RDM at Swedish universities

Several investigations into research data management at Swedish universities have been undertaken, some of which were initiated by the libraries and the archives.

As early as 2007, the libraries and archives of University of Gothenburg, Lund University and the Swedish University of Agricultural Sciences undertook a joint pilot project, financed by the Royal Library in Sweden, that considered research data in open archives and university archives. The study included a survey on researchers' attitudes towards publishing research data and looked specifically at the future roles of open archives and university archives (Björklund & Eriksson, 2007).

SND performed a major study in 2008-2009 on practices of open access to and reuse of research data. This study comprised two surveys targeting professors and PhD students at Swedish universities. The surveys included questions about researchers' experiences of sharing and re-using digital research data, the results of which highlighted several barriers to sharing research data: uncertainty regarding legal and ethical aspects, and lack of resources to prepare research data to be accessible. It concluded that more training in research methods, digital research databases and information about accessible e-tools would be necessary, and that there should be the possibility to apply for funds for preparing research data to be shared and archived. The results of this study have been important in the continued work of SND – its strategies and its activities (Carlhed & Alfredsson, 2009).

About a year later, SND participated in a similar project together with the university libraries at University of Gothenburg, Lund University, Linköping University, Malmö University. This project was financed by the Royal Library and it focused on researchers within humanities and arts. The participating libraries performed a number of interviews with researchers regarding their attitudes of publishing of research data. The project revealed a positive attitude among researchers regarding publishing of research data, and it concluded that researchers required good systems and professional support (Andersson et al, 2011).

In 2012 a workshop by Luleå University of Technology was arranged with support from Riksbankens Jubileumsfond on the themes of research data management and open access of research data. The workshop was attended by representatives from universities and research infrastructures, and it lead to many discussions. It concluded that a common infrastructure and national guidelines for RDM was needed, and that researchers need good tools and systems for the entire research process. As a result of the workshop, a network called ORDER (Open Research Data in E-archives for Reuse) was established (Runardotter & Lindbäck, 2012).

The development of RDM at Swedish universities is slightly expanding now, with different solutions and initiatives. At the Swedish University of Agricultural Sciences, they are quite far ahead in development, and they have recently decided to implement a central system for e-archiving which will include research data management. At two universities, Uppsala University[18] and Umeå University, we have noticed that it is possible to register research datasets

[18] https://mp.uu.se/web/info/forska/open-access/forskningsdata , accessed 2015-01-08

within the institutional archives for research publications. In both cases it is the publishing system DIVA that is used as an institutional archive. DIVA is a publishing system for research publications used by 35 universities and research institutes[19]. We have not further studied the possibilities and policies on research datasets in DIVA.

2.2 Research management in practice – processes and tools

Many processes and tools for research data management on the practical level have been developed. We will first describe and give examples of specific processes or models for RDM, then outline the tools which support data management plans.

Processes and models for research data management

There exist many processes and models for RDM. These processes and models have grown out of the international and national policies, and are designed for researchers and those who support them. Most often they are designed as processes combined with guidelines and checklists. Another common trait of these models is that they embrace the early phases in research, such as the research proposal phase or the project start-up phase. In the book "Research Data Management: Practical Strategies for Information Professionals" (Ray, JM, editor, 2014) they are comparing these processes to *life cycle analysis*, a process in science to assess a product's impact on environment through its different stages. A data life cycle model is then a process to describe the different stages and transformations that data will undergo from its creation to its final sharing and preservation. Using a data life cycle model can provide a useful framework to present and communicate the different stages of data, in order to deliver support for RDM in an organization. The process of research data management is often complex and it involves co-ordination between people, agencies and resources. Breaking down the data process into different component stages helps to identify and assign roles, responsibilities, milestones in the services of research data management. Data life cycle models could be seen as another subset of research life cycle models, as they represent another aspect of the whole research process – focusing on the processing of data in the research. Organizations developing and implementing data life cycle models should take into account their research processes in order to match the researchers' need for support.

[19] http://www.diva-portal.org/smash/aboutdiva.jsf , accessed 2015-01-08

Research data life cycle, University of Virginia

One example of how a data life cycle model could be embedded within the general research cycle model is from University of Virginia Library's Scientific Data Consulting Group [20].

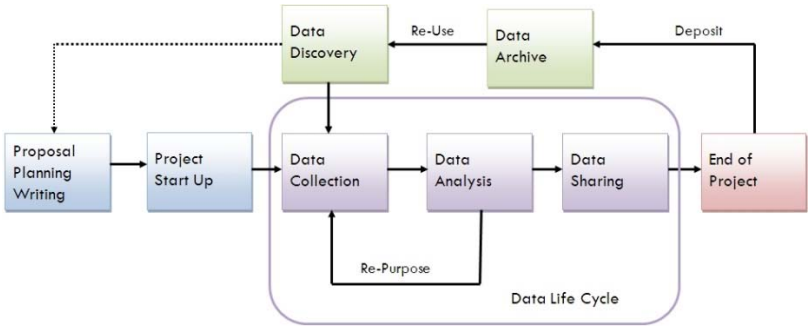

Figure 2.1
Figure reproduced with permission from the University of Virginia Library, Research Data Services. © 2014 by the Rector and Visitors of the University of Virginia. This work is made available under the terms of the Creative Commons. Attribution-ShareAlike 4.0 International license http://creativecommons.org/licenses/by-sa/4.0/

This model has focused the data life cycle components to steps of: *data collection*, *data analysis* and *data sharing*, which are data intensive phases. In its documentation it gives advice and best practice to other research phases too, such as the proposal and project start-up phases. The model was developed by the library of University of Virginia, and it is an example of an academic library where they have taken the initiative to provide a tool for researchers in their RDM processes.

The data life cycle, Inter-university Consortium for Political and Social Research (ICPSR)

The organization "Inter-university Consortium for Political and Social Research (ICPSR)"[21] has developed a data life cycle model which includes all stages in the research process. It describes the different stages or steps on a more generic level, and it is supposed work well for the "typical" researcher in social sciences.

[20] http://data.library.virginia.edu/data-management/ , accessed 2015-02-03
[21] https://www.icpsr.umich.edu , accessed 2015-02-02

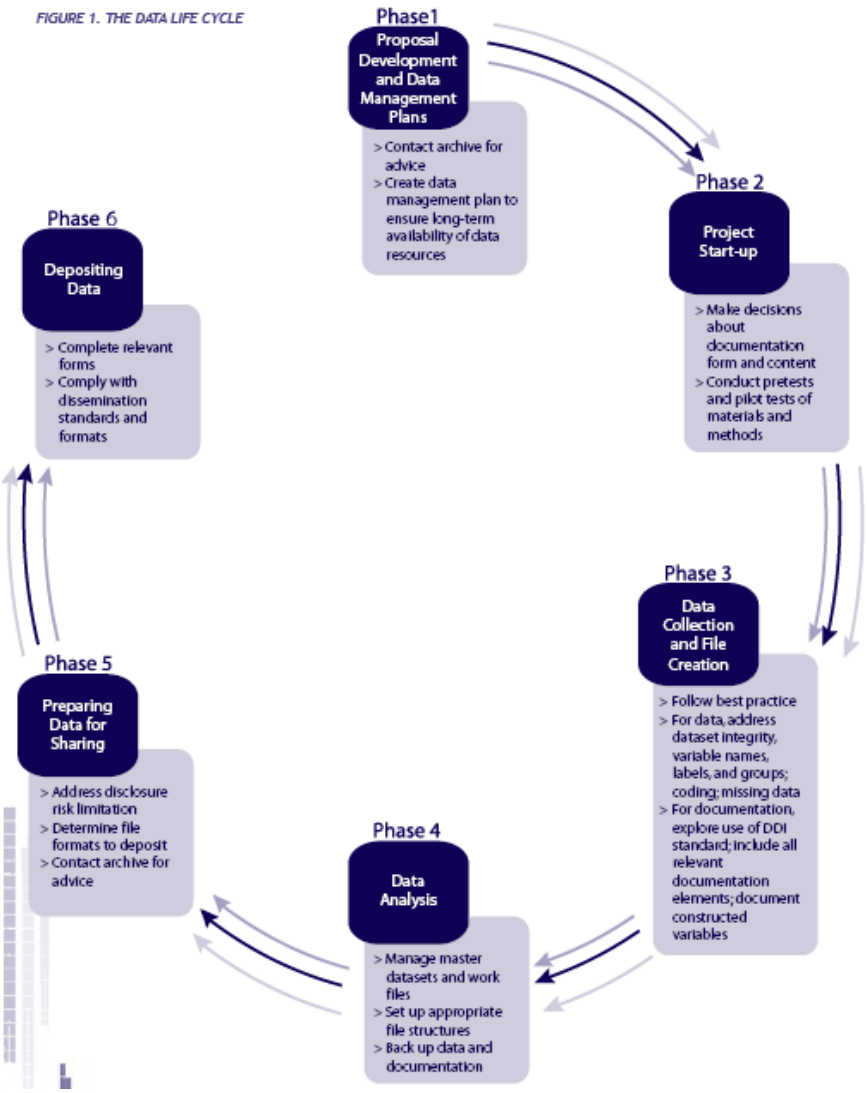

Figure 2.2
(ICPSR, 2012: Data Life Cycle)

ICPSR is a service provider for researchers in social sciences. They offer services for depositing research data and they provide training and other support to researchers in RDM. For their purposes it is essential to demonstrate a process of RDM which clarifies and visualizes the different phases.

Tools for the support of data management plans

Many researchers today are affected by the requirement of producing a data management plan, in particular when they are searching for funding. A data management plan is a planning document which is supposed to describe how research data should be: *collected, stored during research phase, analysed, archived on long-term basis, shared to the public*. Documenting how they will handle their data before they have collected it is quite a new thing to many researchers, who often say they feel they don't have the time, resources or sufficient skills.

Often there exists no infrastructure for archiving and sharing data. The researchers feel this is a new administrative burden put on them in the research funding process. There exist many

different templates and form data management plans. Funders often design their own specific forms for data management plans. At DCC in UK they have developed a web based tool called DMP Online as a support for researchers creating a data management plan. In DMP Online the policies and requirements from all the major funders are included. When using DMP Online the researcher will gradually get help on his/her data management plan, depending which funder it is and depending on which subject it concerns. In the USA they have developed a similar tool called DMPTool designed for the American research community. It is important to stress that while these tools are a support for the creation of a data management plan, it is always up to the researcher himself to carefully read and interpret the instructions of a specific data management plan. At DCC in UK they have created a checklist for a data management plan, which quickly brings the researcher into the important parts of a data management plan (DCC, 2013).

2.3 Research data management and libraries

With the overview of RDM and some scenarios for future researchers, we will now look at what role research libraries may take in this. Research libraries have for a long time given support in the scholarly and communication process for researchers, such as publishing advice, preparing for access to researchers' publications, helping researchers measuring their scientific output etc. (Lyon, 2012). Now the research landscape is changing, and researchers are changing how they work with and document their data. Researchers will certainly need support on the practices of RDM and all the requirements that come along with it.

A UK report by the consultants firm "Key Perspectives" (Swan & Brown, 2008) which gives an overview of future needs for data scientists and curators, states that the libraries have clear role in the data science arena, in particular concerning awareness of data issues and the importance of good data science and curation. Furthermore the report lists three relevant ways in which the libraries might influence development in research data management:

- Training researchers to be data-aware
- Adopting a data archiving and preservation role
- The training and supply of data librarians

There are many examples of academic libraries which have engaged in RDM at their universities. In the following section, we give examples of libraries providing RDM services; Purdue University Libraries at Purdue University and Bodleian Libraries at University of Oxford.

Purdue University Libraries, Purdue University

Purdue University Libraries [22] was one of the pioneers engaging in RDM, starting in the early 2000s. Today it has a full-fledged service for RDM. Early on, it took the approach of engaging the library as a research unit, engaging itself in inter-disciplinary projects at the university. This lead to useful contacts within the university and to extended knowledge of researchers and their processes. Through funding from IMLS, Purdue University Libraries developed Data Curation Profiles Toolkit, that was a set of templates helping researchers to structure the data workflow in research. It also started an internal re-organization of the libraries to be more like an academic

[22] https://www.lib.purdue.edu , accessed 2015-02-03

unit. It started collaborations and partnerships with other units in the university, such as the IT services. D. Scott Brandt, Purdue Libraries, emphasizes the *long-term* work of libraries engaging in research data management: "Once librarians understand what could or should be done with researcher data - and are able to demonstrate and articulate it, then the more likely opportunities will turn up" (Brandt, 2014, p. 342).

Bodleian Libraries, University of Oxford

At Bodleian Libraries [23], University of Oxford, they started off a JISC-funded project called DaMaRO in 2011, which was supposed to create a university research data management policy and the infrastructure for researchers comply with it. The project was supposed to take the result of various research data management projects in the university and develop into better-integrated tools that will support researchers through the data life cycle, from planning to re-use. The project that was ended in 2013 resulted in a number of systems and processes and the university research data management policy.

Poster of the DaMaRO project at University of Oxford

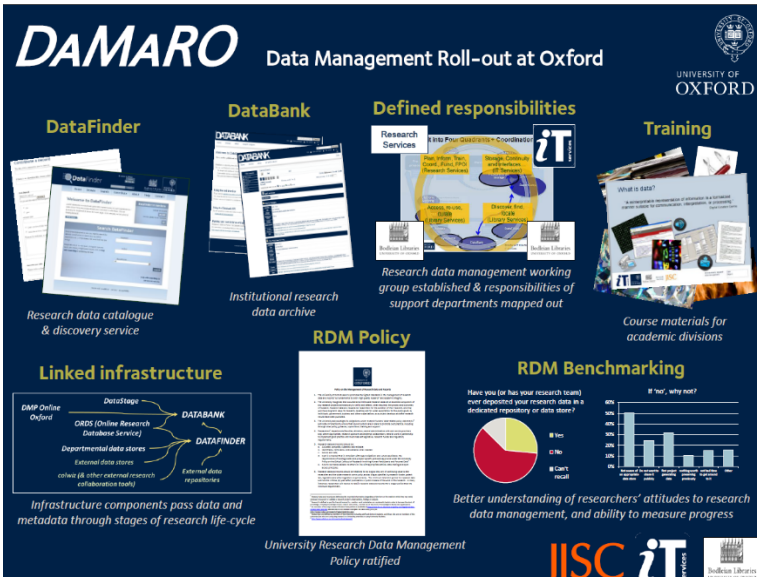

Figure 2.3
(DaMaRO, 2013)

Within the LERU network there are a number of universities which have some kind of official service for RDM, namely 13 out of 21 universities, according to our studies of webpages. It's worth noting that Lund University is not one of 13 out of 21 universities having an RDM policy. The solutions vary a lot between the universities, but it is clear that many have now brought up a discussion on RDM internally and are working on different ways for implementing it. Libraries are often involved in RDM, in 9 cases out of 13 the LERU universities having RDM, the libraries are engaged. This shows that libraries have been entrusted with the task of developing RDM, and are regarded as having major competence on this matter.

[23] http://www.bodleian.ox.ac.uk/bodley , accessed 2015-02-03

Which are the key competences of libraries that could be valuable in the RDM processes of the universities? Based on the literature and studies of various webpages, we may do the following list of crucial skills and knowledge that libraries possess:

- resource description and documentation
- knowledge organization and subject structures
- making data accessible
- the research process
- user training
- guidance and reference work
- process oriented workflow

In planning for an organization for RDM universities may consider the skills and competences of the libraries to see what could be valuable. Competences from other parts and functions within the university are also important to consider in the planning of RDM, but here we focus on the role of the libraries.

Roles of the libraries within the data life cycles

We have provided examples of data life cycles to aid and visualize the process of research data management. Discussing the involvement of libraries in RDM, we could have a look on how libraries may fit into a data life cycle. In what stages and in what ways can libraries support researchers in this process?

As an attempt to illustrate this we have used the different stages in the data life cycle by ICPSR, which is illustrated figure 2.2, and coupled these stages to roles and actions of the researcher and to roles and actions of the library staff (ICPSR, 2012). Please note that this is not a complete list of roles and actions of the researcher and of the library staff. The actions of the library staff may expand as their experiences and competences in RDM are developing.

Phase 1: Proposal development and data management plans		
Actions for	Researcher	✓ Write research proposal ✓ Create data management plan for proposal
	Library staff	✓ Support and advice on data management plan, on long-term storage ✓ Guide to further departments
PHASE 2: PROJECT START-UP		
Actions for	Researcher	✓ Plan for data collection, documentation, storage etc.
	Library staff	✓ Support and advice on data documentation, storage, data description (meta-data) etc.

PHASE 3: DATA COLLECTION AND FILE CREATION		
Actions for	Researcher	✓ Collect data during research ✓ Ensure data is stored securely and effectively ✓ Ensure data is documented
	Library staff	✓ Support and advice on practical data management, such as file structures, file naming, file versions, data descriptions (meta-data)
PHASE 4: DATA ANALYSIS		
Actions for	Researcher	✓ Manage master datasets ✓ Set-up file structures ✓ Back-up data and documentation
	Library staff	✓ Support and advice on file structures, data documentation (meta-data)
PHASE 5: PREPARING DATA FOR SHARING		
Actions for	Researcher	✓ Determine file formats for depositing data ✓ Select data for deposition ✓ Ensure integrity of data such as personal data ✓ Decide which data should be publicly accessible
	Library staff	✓ Support and advice on file formats, methods for depositing data, ways of data discovery in the public
PHASE 6: DEPOSITING DATA		
Actions for	Researcher	✓ Decide where to deposit data, finally prepare data for long-term preservation, finalize documentation of data
	Library staff	✓ Support and advice on ways of depositing data, data preparation for long-term preservation, data documentation (meta-data)

In chapter 5.3 we will continue and describe how these current skills and competences of libraries may be transferred and further developed in the processes of research data management. It is clear that there is needed competence development in certain areas of RDM for the library staff.

2.4 Legal framework

In this section we will examine the legal framework[24] in which Swedish universities and university colleges operate, with focus on research. Public universities are government agencies. Documents kept by agencies are, according to the Constitution, official documents and in most cases they are also public documents. The general public has a right to view such documents if

[24] Statutes and ordinances are referenced by year of issue and number in the Swedish Code of Statutes (Svensk författningssamling, SFS), and can be found in the database RIXLEX (the parliamentary public database) in Swedish only. http://www.riksdagen.se/sv/Dokument-Lagar/Lagar/Svenskforfattningssamling/, accessed 2015-01-07

access is not restricted by law.[25] Exemptions from publicity are found in the Public Access to Information and Secrecy Act (SFS 2009:400). Official documents much be registered, be accounted for and archived. Certain documents can be sorted out but others must be kept forever. Regulations about archiving are found in the Archives Act (SFS 1990:782).

Research material or research data within Swedish universities are regarded as official documents if they have been submitted to the agency or have been created by the agency. Documents in this respect are not restricted to text documents, but also applies to photos, video and audio recording, blood and tissue samples, test results, protocol from measurements, X-ray pictures etc. The format of the document, digital or paper, is not important. This means that documents become official at different stages of the research process, not when the research project has ended. For example, a questionnaire form becomes official when it has been sent out to the participants. A questionnaire response becomes official when it has arrived at the university. A database becomes official when its structure has been determined and is ready to receive data. Every subsequent data entry in a database becomes an official document when it is registered.

Most documents created within research projects are official documents and should be archived. Research material consists of administrative documents, primary data, working material and research results. We are most concerned with research data.

Archives should be kept to fulfill several aims such as the long term right of the general public to view official documents, the need for information within administration and justice, and the need for future research. Agencies are allowed to sort out and destroy official documents only to the extent that is permitted by law and by regulations issued by the National Archives (Riksarkivet). The following research material must always be kept for the future: Documents concerning extensive primary datasets, research registers and databases with high coverage, documents reflecting the evolution of a scholarly discipline, documents that reflect the significance of the cultural context of the universities, documents that reflect the activities of distinguished persons, and finally, documents concerning research that have been debated in public. The perspective of archiving should be part of the entire lifecycle of the research project. This fact makes archiving part of the research data management process described above.

The Instrument of Government (SFS 1974:152), which is part of the Swedish constitution, states that everyone shall be protected against significant invasions of their personal privacy, if these occur without their consent and involve the surveillance or systematic monitoring of the individual's personal circumstances. This might be at risk even in research, and the constitution does not distinguish between different purposes.

Official documents held by Swedish agencies are in principle public, but the right of the general public to view such documents may be restricted if necessary out of consideration to the individual's personal and financial circumstances. In such cases, there are often varying degrees of secrecy. The strongest secrecy applies to data that has been collected for statistical purposes. Here, secrecy is absolute, which means that the data is in principle not to be disclosed. However, also for this data, there exist certain provisions that override secrecy. For example, the data may be disclosed for research purposes if this can be achieved without risk of damage or harm to the person whom the data concerns. If secret data is disclosed to another agency, such as a university or a university college, for research purposes, secrecy automatically attaches to that data. Data collected for research purposes in a number of specified research fields and conducted by

[25] The Freedom of the Press Act (SFS 1949:105). The Swedish Constitution consists of four fundamental laws and has been translated into English and is published at http://www.riksdagen.se/en/Documents-and-laws/Laws/The-Constitution/, accessed 2015-01-07

universities, have been equated with secrecy for statistical purposes by the Government, where the same provisions apply. This follows from the Public Access to Information and Secrecy Ordinance (SFS 2009:641).

It is the duty of the agency holding the register to consider whether personal data may be disclosed. Where the disclosure is made for research, it is common for the data to be disclosed in anonymized form. This means that researchers do not know which individuals the data concerns. Anonymized data can be coded or de-identified. Coded data means that the researcher receives data with a serial number, that does not carry identity information, while the agency retains a code key that links serial numbers to personal identity numbers and that makes it possible to supplement the data later on. The existence of such a possibility is important, for example in the case of longitudinal studies. When data is de-identified, this possibility does not exist. Even if data is anonymized, there may be a risk of reverse identification, i.e. identifying an individual by compiling a series of data about that person. This is prohibited by the Personal Data Act.

The Personal Data Act (SFS 1998:204, abbreviated PUL) governs the processing of personal data, regardless if the processing is carried out manually or automatically. Such data may be collected only if this is legal and only for specific, explicitly stated and justified purposes. All processing must be authorized by the informed consent of the person concerned. Furthermore, it may not be processed for any purpose that is incompatible with that for which the information is collected. Under PUL, the processing of data for historical, statistical or scientific purposes shall not be regarded as incompatible with the purposes for which the information was collected.

This means, it is possible in research, to make secondary use of data that has been collected for other purposes. The processing of data for purposes other than the original requires either the consent of the persons concerned to this processing or its necessity for a task of public interest. Research might be an example of such a task.

PUL classifies certain personal data as sensitive and prohibited to process. This applies to data disclosing race or ethnic origin, political opinions, religious or philosophical beliefs, membership of a trade union or pertaining to health and sex life.

However, following ethics approval, such data may be processed in research. Ethics approval is granted in accordance with the Ethical Review Act (SFS 2003:460[26]) by an ethical review board consisting of government-appointed representatives of the research community and the general public.

The ethical review board shall also assess whether consent should be obtained or whether the data can be processed without consent. Where utilization of an agency register is extensive, the praxis is not to require consent, as this would presuppose a disproportionate investment of resources.

All major universities and most university colleges have personal data representatives, whose task is to ensure that the controller of personal data, i.e. the higher education institution, complies with existing rules. It is the boards of the higher education institutions that have the ultimate responsibility for ensuring that this activity functions satisfactorily.

---

[26] Ethical Review Act (SFS 2003:460) [Lag om etikprövning av forskning som avser människor] i.e. Act concerning the Ethical Review of Research Involving Humans, see also Ethical Review Board http://www.epn.se/en/start/startpage/, accessed 2015-01-07

2.5 Cultural Heritage Data and Research Data

Research data can be collected directly by researchers or by re-using existing data from the public sector. Both governmental data and cultural heritage data are important sources for research. In the previous section we saw that data from official documents can be disclosed according to the Freedom of the Press Act (TF) if the document is not protected by secrecy or the disclosure violates another law, e.g. the Personal data act. TF formulates an obligation for authorities to disclose documents that are public but does not say in what format (paper format should be assumed). Recently in 2010, two EU directives concerning access to the data from the Public sector have been implemented in Sweden. This regulation concerns the transfer of information from the government agencies to the public. According to the Public Sector Information (PSI) Act[27] and the recommendations issued by the E-Delegation, the main alternative is to make digital information available online, the second alternative is to disclose digital information when requested by potential users. The PSI Act doesn't apply to universities, libraries and archives, although a parliamentary committee has recently proposed for the act to also apply to archives and libraries, including university libraries[28].

Another legal regulation is the Swedish implementation of the EU INSPIRE directive concerning geographical data that may be useful for activities and actions regarding human health and the environment.[29] It aims at establishing an infrastructure of geographical data among authorities and make data available to the public.

Research data can be published in different ways on the internet. This also applies to cultural heritage data and other public sector data. The first way that comes to mind is to provide data that is humanly understandable. But to use and process this data, Internet users will have to organize the data themselves, for instance by entering the data into a spreadsheet. Still, this is by far the most common way to publish research data and other types of publicly funded data online, for instance as an output of data from a database in a table on a HTML web page. Databases are frequently used as backend storage for information that builds webpages dynamically or produce data tables which are part of webpages, but there is often no way to access this information directly, as machine-readable data. Another way is to prepare data files for download and link them to a webpage that presents the data.

Data is commonly understood as machine-readable data, as opposed to information, which require that the data is structured in some way so it can be processed by users' own computers after storage. For example, research data repositories might publish data files as Excel-documents or as Shape-files which is used for geographical data. We can use Berners-Lee's requirements[30] for *Linked Open Data* as a point of departure for a discussion about different ways to publish research data. The term Open Data refers to an open license, e.g. Creative Commons, which makes it possible for the reader to use the data for his own purposes (see previous chapter). This

---

[27] Lag (2010:566) om vidareutnyttjande av handlingar från den offentliga förvaltningen, in English, Act regarding reuse of public sector information. This act is the Swedish implementation of a EU directive.

[28] SOU 2014:10 Ett steg vidare - nya regler och åtgärder för att främja vidareutnyttjande av handlingar, http://www.regeringen.se/sb/d/18055/a/235091, accessed 2015-01-07.

[29] Lag (2010:1767) om geografisk miljöinformation, in English, the Act on spatial information, see https://www.geodata.se/en

[30] Berners-Lee, Tim. 2006, last change 2009. "Linked Data". World Wide Web Consortium (W3C) http://www.w3.org/DesignIssues/LinkedData.html, accessed 2015-01-07.

data model is of course built on the requirements for Linked Open Data but addresses a lot of issues that are important in a wider context of data publishing and the current development in Digital Humanities and Big Data Research. Berners-Lee's first requirement is that data is published online using the internet as a distribution platform and is open, which at least corresponds to what we have called publishing data that is humanly understandable. The list is hierarchical so that all previous requirements must be fulfilled for the next step, see figure 2.4 below.

1. On the web with an Open License, as HTML-documents with or without hyperlinks, in any human-readable format. (unstructured data)
2. Machine-readable format (structured data, e.g. an Excel-document, a SQL database, a statistical dataset, SAS, SPSS, a GIS file)
3. Non-proprietary format (CSV, XML-document, JSON data, MySQL)
4. RDF-standards from W3C (RDF, including RDF/JSON , and SPARQL, a query language for RDF)
5. Linked RDF to provide context

Figure 2.4
Linked Open Data (Five star model).

The second step requires that the dataset is also stored in a machine readable form as structured data that can be openly retrieved online, processed and interpreted by computers as, for instance, a data-matrix with columns and rows (or variables and records). If the dataset has a non-proprietary format it also fulfills the requirements for the third level. Open formats are very important for the data to be long term sustainable and available for future reuse for the case that the data format in question is no longer maintained by the company who owns it and the format specification is secret. The requirements so far indicated by three stars in the model, is usually what is discussed when setting up a local or national research data repository. Such a repository thus constitutes a collection of datasets. This is the kind of service that the Swedish National Data Service (SND) provides concerning data storage and this is also what the Swedish Research Council (VR) has in mind when defining requirements for funding and open access to research information for the future.

A data repository often contains, or should contain, a *dataset directory* where detailed metadata about the content of the dataset, study design, data collection methods and outcome, subject keywords, etc. is stored and is searchable. Some data formats contain information internally about type of data for each variable or column (text, integer, decimal, date and other types), e.g. Excel and SQL, whereas such information must be part of an external documentation for comma separated data files (CSV). There should be metadata about datasets from major directories or catalogs, as can be seen in e.g. data.gov.uk or data.gov (in Sweden SND and the Parliament are good examples, snd.gu.se and data.riksdagen.se respectively). When datasets within a sector are widespread among many contributors and they want to publish data themselves, a dataset directory published online separately from the actual storage of the datasets might be the best solution, e.g. oppnadata.se, which lists open governmental data in Sweden and URL addresses to their storage locations. To our knowledge there are not yet any local data repositories at Swedish universities as part of a general research infrastructure where research data is published, although many university departments, faculties and research groups provide access to data and access to online databases on their websites.[31]

[31] Bohman, Mårten (2013). *Lagring av forskningsdata vid svenska lärosäten* [Storage of research data at Swedish universities and university colleges]. In contrast, there are repositories for open access publications at Lund

We can refer to datasets and their corresponding description and metadata as *Data publications*. There are even scholarly journals specialized in such publications. Dataset described in the directory at SND could also be regarded as a data publication. For the published dataset to be referred to and reused there should be a way to reference them in the scholarly community using stable digital identifiers that allows proper citation. A digital object identifier (DOI) is a character string used to uniquely identify an object such as an electronic document or a dataset. Metadata about the object is stored in association with the DOI name and this metadata may include a location, such as a URL, where the object can be found. The DOI for a document remains fixed over the lifetime of the document, whereas its location and other metadata may change. Organizations that meet the contractual obligations of the DOI system and are willing to pay to become a member of the system can assign DOIs. SND can assign DOI identifiers for datasets within their area of competence. Another system is the URN:NBN protocol which stands for Uniform Resource Name and National Bibliographic Number where the Swedish National Library (KB) assigns digital identifiers for electronic documents in Sweden.[32] The URN:NBN identifiers are used by the Alvin portal, a system for cultural heritage collections in Swedish universities. Both systems operate in a similar manner as they take responsibility for the correct redirection from a central server or resolver to the URL where the document is currently located. These identifiers operate on a "document" level, single "pages" cannot be referred to. There is some confusion about stable identifiers in this context. Any organization with a domain name can make a commitment to publish stable links that are independent of the current technical solution to provide web content and where the URL is likely to change. These stable links or URLs are sometimes called permanent, persistent or canonical links and are, for instance, offered by the National Library of France, the Bavarian State Library and others to reference content in their own digital collections.[33] Below, we will discuss how data items in collections or datasets are addressed in a semantic web context to deliver a RDF and a HTML representation of an object respectively, depending on the request made by the client. The above mentioned DOI and URN protocols have nothing to do with this and are in fact not capable of performing content negotiation to determine whether the server should return RDF/XML or HTML.

Next in the data model by Berners-Lee, i.e. the fourth step, are datasets that make use of Semantic web technologies such as RDF, OWL, SKOS, SPARQL and HTTP URIs. Uniform resource identifiers (URI) are used to give direct and long term access to individual data records within a dataset. URIs are used to name things in the real world, both objects and concepts, and are defined by the resources these identifiers refer to. With linked data the point is that several resources can define the same real world object or concept and link to another URI so they can be discovered. URIs are especially important as authors of scholarly publications will be able to refer to individual items in a dataset, e.g. an ancient or medieval text, a coin, a piece of ceramics, a place or person, etc. which are used as evidence along some line of argumentation. The reader of the publication will in turn be able to evaluate the argument by retrieving the evidence

University and elsewhere which makes it easy to get an overview what publications are produced. Data directories, as discussed here, should play the same role for datasets.

[32] http://www.kb.se/isbn-centralen/urnnbn/ in Swedish, accessed 2014-12-22

[33] Another library is the Herzog August Bibliothek in Wolfenbüttel, which has a guarantee about persistent URLs. The library "guarantees the accessibility of its digital facsimile editions of imprints and manuscripts on the Web on persistent URLs. This permits reliable quoting of entire electronic editions and single pages from them", http://www.hab.de/en/home/library/wolfenbuettel-digital-library/guarantee-declaration.html (accessed 2014-12-22).

instantly, by a single click on the mouse. These technologies make data definition and metadata self-contained so that machines can "understand" the semantic meaning of data. There are currently very few datasets at this level in Sweden but there are reasons to believe that they will grow in importance, also among researchers, if just the infrastructure providing services like this will be developed. The most important example at this point and with an impact also on research, is the Swedish Open Cultural Heritage (SOCH) maintained by the Swedish National Heritage Board.[34] SOCH is a web service used to search and retrieve data from any organization holding information or media relating to Swedish cultural heritage such as museums and archives, and has been operating since 2009. SOCH is the Swedish provider of data to the Europeana collaboration[35]. The Swedish name for SOCH is "K-samsök", which can be translated as "Cultural cross-search" according to their website[36]. SOCH functions as an exchange or aggregator, where data from many different local databases are made searchable and available to the public and to the research community. By creating links between related digital objects from different sources, e.g. objects that originate from the same place using common authority lists for Swedish parishes or municipalities, SOCH enriches the entire body of cultural heritage data as linked data. Unfortunately the Swedish university libraries and archives do not participate in this cooperation, although university museums are.

Returning to the model, this is what is meant by the fifth and last step in the Berners-Lee's model. It represents datasets that are connected to other datasets through the use of common concepts and relations between data items, e.g. RDF-links as identity relations saying that two resources refer to the same real world object which in turn enable the emergence of an interconnected and searchable web of data. As we have already seen, this is becoming a reality for Swedish cultural heritage and library data, which form part of a web of data and is extended internationally through DBpedia[37], Europeana and the Virtual International Authority File (VIAF)[38]. In research, there are also initiatives of building a connected web of data in particular areas. The British Pelagios project[39] creates the infrastructure to enable the use of common identifiers of ancient place among a number of digital resources of the Ancient Greek and Roman civilizations. It builds on the digitization of the Barrington Atlas of the Greek and Roman World[40] and its map-by-map directories of ancient names. There are currently 34,764

---

[34] An equally extensive and important dataset is Libris, the Swedish Union Catalogue, i.e. the joint catalogue of the Swedish academic and research libraries, which has been available as Linked Data since 2008. http://librisbloggen.kb.se/2008/12/03/libris-available-as-linked-data. In 2011 a subset of Libris, the National Bibliography and authority data on persons, organizations and subjects, was released into the public domain as Open Data with a Creative Commons 0 (Zero) license. http://librisbloggen.kb.se/2011/09/21/swedish-national-bibliography-and-authority-data-released-with-open-license. Data is available as Atom feeds and the OAI-PMH protocol.

[35] http://europeana.eu

[36] http://www.ksamsok.se/in-english/ accessed 2014-12-22

[37] DBpedia is a community effort to extract structured information from Wikipedia and make this information available on the Web as RDF triples with extensive linking to other similar datasets. Recently the Wikimedia Foundation has released the Wikidata project, http://wikidata.org, with a similar purpose and to serve as a backend database of facts in localized versions of Wikipedia.

[38] http://www.europeana.eu and http://www.viaf.org.

[39] http://pelagios-project.blogspot.se One of the authors of this report participates in the Pelagios project with the Digital Atlas of the Roman Empire, which is a gazetteer, a map and a backend database, http://imperium.ahlfeldt.se

[40] Talbert, Richard J. A. *Barrington Atlas of the Greek and Roman World*. Princeton University Press, 2000.

numerical identifiers of ancient place in the digital gazetteer. A similar project is underway to enable stable identifiers of ancient Greek and Roman persons.[41]

What about research data in general and its relation to data publishing on the internet? For long it has been up to the researchers themselves to decide. What support and infrastructure are available to researchers at universities for data publishing in any machine-readable form? We will look into these issues in chapter 4.

[41] Standards for Networking Ancient Prosopographies: Data and Relations in Greco-Roman Names. King's College, London, http://snapdrgn.net

3. Overview of the organization of Lund University with focus on research data management

In the previous chapter we gave a background of research data management (RDM) activities outside Lund University. In this chapter we will describe the current organization of Lund University with focus on RDM. Our findings are based on meetings, interviews and seminars that we have had within LU during the project. We will describe the different departments and units, their current tasks, responsibilities and competences. Furthermore we will highlight which of their key competences could be useful in RDM.

3.1 LU libraries - organization and competence

The university has a decentralized library organization, where each faculty is responsible for its library services. There are 26 libraries in the university, i.e. faculty libraries, some special libraries and the University Library (in Swedish Universitetsbiblioteket – UB). The libraries are co-operating within a library network, with a number of common services. The University Library has a role as co-ordinator for the common services and is a national deposit library. Within this project we have been in contact with the departments of the University Library and with the faculty library of humanities and theology as well as the faculty library of social sciences.

University Library

Throughout a series of interviews and meetings we have undertaken a major study of the University Library.

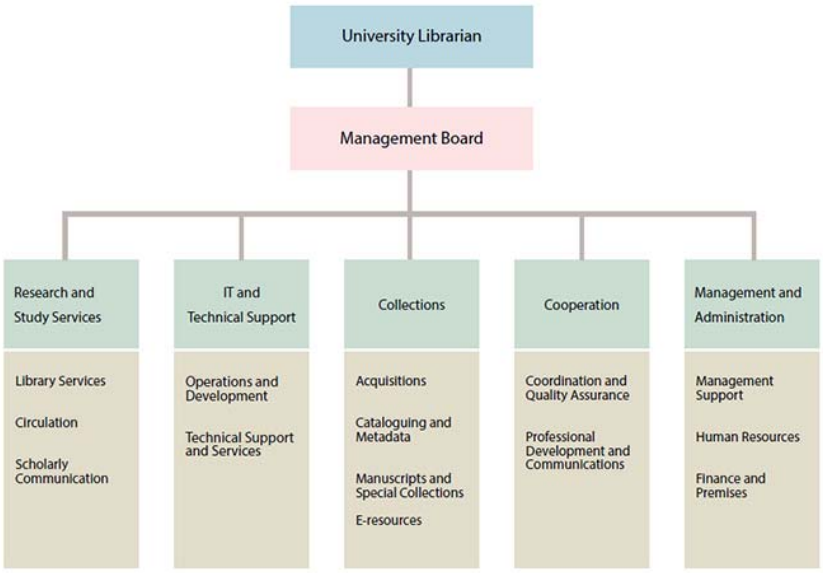

Figure 3.1
Organization chart University Library

The University Library is headed by the University Librarian, and the library is organized through five major departments (green fields in organization chart). The department "Cooperation" co-ordinates a lot of the co-operation with the faculty libraries. The other departments also liaise with the faculty libraries and they have specific areas of responsibility, with their different units (brown fields in organization chart). We have been in contact with the following departments and units: Scholarly Communication, IT and Technical Support, Acquisitions, Cataloguing and Metadata, Manuscripts and Special Collections, E-resources, Professional Development and Communications.

Scholarly communication

The Scholarly Communication unit works with import of publication metadata from external databases, such as Web of Science, into the university institutional archive Lund University Publications – LUP. The unit is also responsible for the data export from LUP into "SwePub", the Swedish research publications database. The unit is also offering a lot of training and seminars on publishing both for librarians and for researchers. It has extensive expertise in open access and bibliometrics. The unit is currently involved in the implementation of a current research information system (CRIS) at Lund University. With its knowledge and experience on publishing and the publishing process this unit could play an important role if the University Library is taking an active role in RDM.

Acquisitions

The University Library is a national deposit library and it receives one copy of every printed item from the Swedish printers. The unit Acquisitions takes care of the delivery of the national deposit items, thus preserving it on long-term basis, and making accessible to the public. The material is very diverse, ranging from newspapers, journals, books to flyers and posters. The

material is an important part of the national cultural heritage, and it has a great potential to be used for research purposes. Both the technical system for preservation and access to the national deposit items, as well as the competence of the Acquisitions unit are of great importance if the University Library is taking an active role in RDM. This unit has knowledge both on how to store the material physically as well as electronically. From 1 January 2015 the new law of e-delivery[42] of national deposit items will be effective, which means that the University Library will get the delivery of national deposit items electronically.

Cataloguing and Metadata

As a national deposit library the University Library has a long and strong tradition of cataloguing and archiving all kinds of published material. The unit Cataloguing and Metadata is cataloguing the collections in Libris, which is the Swedish national bibliography. The cataloguing is mainly consisting of *published material*, i.e. monographs, reports, serial publications etc. The unit is currently working on implementing the new version of Libris called Libris XL which will expose records in the global system WorldCat. This unit holds extensive experience in resource description including field structures and formats for bibliographic records. It is also monitoring the development of metadata standards and new data formats. This unit has a great potential of becoming a unit of expertise in metadata if the University Library is taking an active role in RDM.

Manuscripts and Special Collections

The University Library has been engaged in digitization of special collections for a couple of years and through a number of projects and initiatives, such as the Per Bagge photo collection, the St. Laurentius Digital Manuscript Library and the archive Voices from Ravensbrück. The collections are accessible on the library website through different user interfaces. The library is working on providing a common user interface of the digitized collections, and is currently involved in the major project Alvin/ArkA-D. If the University Library should offer support and services in providing access to research data, the current experience and technical equipment on digitization could be of great value and use.

E-resources

E-resources is responsible for the in-flow of external research information, such as e-journals, e-books and scientific databases into the information system called LUBsearch. This unit is also working with resource description and with metadata. In this matter it concerns the processes of importing metadata from publishers and other providers of e-journals and e-books into LUBsearch. This unit also co-ordinates the subscriptions of the e-journals, e-books and databases for the entire Lund University library network. Both units, i.e. Cataloguing and Metadata and E-resources hold competence which could be of value when it comes to helping researchers describing and enriching their research data in the preparation of making it publicly accessible.

Professional Development and Communications

The Professional Development and Communications unit works with professional development for the library staff of the entire Lund University library network. Every year there is an extensive program of continuing education courses for library staff at Lund University, organized by both internal and external staff. This unit is also responsible for the communication activities at the

[42] This is in Swedish ”Lagen om e-plikt”, SFS 2012:492

University Library, and it co-ordinates the common website for Lund University library. The common library website[43] is an important communication channel for all the libraries, and it is supposed to market the library and information services in a joint and unified way.

There are also more departments and staff within the University Library that could be of importance in the process of engaging the University Library in RDM, but they are not further described here.

The faculty libraries

As we described in the first chapter Lund University has a decentralized library organization where each faculty have the responsibility of their library services. This has resulted in specialized and user-oriented library and information services throughout the whole university. We believe that the faculty libraries could play an important role in the future development of research data management. The faculty libraries are situated in the front to their researchers and they have good extensive knowledge of the subject areas of their researchers. In 2012 there was a common project within the library network which investigated the research support services of the libraries, and which resulted new strategies and ways of working with research support at the libraries. This project did not cover the processes of RDM.

In this project we have mainly been in contact with the faculty library at Humanities and Theology, Social Sciences and Business and Economics. At the faculty library of Humanities and Theology they have been involved in some research data related activities at the faculty, such as the Humanities Laboratory. They have also started to orientate themselves in RDM, and they have provided some information to researchers on their library webpage. At the faculty libraries of Social Sciences and Business and Economics, they are also starting to investigate the field of RDM. All three mentioned faculty libraries are very positive to start working with support in RDM, and to develop their professional skills in this field.

3.2 Other LU departments related to research data management

There are some central departments and units in Lund University which are related to RDM, and which could play an important role in the future. We have identified the following LU central departments: Record management and Archives, Legal Division, Research Support Services, Research Ethics Information. There are even more departments within LU which that could be engaged in RDM in the future, such as the center of GIS[44] and the division of ALM[45]

In the following section we will describe the first mentioned departments and what role they could play in RDM at Lund University.

[43] http://www.lub.lu.se , accessed 2015-01-07
[44] GIS – Geographical Information Systems
[45] ALM – Archives, Libraries and Museums

Record management and Archives[46]

The LU central unit Record management and Archives is responsible for the central registry of official documents at the university, and is managing the central archives. The archives are mainly focused on the administrative documents for the university, but also include long-term archiving of research data. The department is currently engaged in procuring a central system for e-archiving, which is supposed to be implemented during 2015. The competences of this department such as the legal aspects of archiving and the technical and practical aspects of archiving could be of great importance in the development of RDM at the university. The department has good contacts externally, with other universities and with external agencies.

Legal Division[47]

The Legal Division consists of a group of lawyers with expertise within copyright, administrative law, contract law and the secrecy and privacy protection law. The legal aspects of RDM will certainly be very essential in the future. Researchers will need help and advice in questions such as licensing of research data, protection of privacy in research data etc. The competences and expertise of this department will be great importance in this progress.

Research Support Services[48]

Research Support Services is mainly focused on providing support in the grant process and to give advice on funding opportunities. They co-ordinate the EU research applications for the university, and they have extensive knowledge on funders nationally and abroad outside Sweden. This department is also responsible for the implementation of LUCRIS, i.e. the CRIS[49] system for Lund University. With its expertise in grant application processes and in the specifications of funders, the department could play a most important role in the development of RDM. The department also has a good overview of the all the different research areas within the university.

Research Ethics Network[50]

There is a Research Ethics Network within the university that offers training and seminars, as well as advice on matters of research ethics. It could be advice on the management of personal data, on legal aspects of animal research, on bio-bank research. This expertise of research ethics is of great relevance in the development of RDM at the university, both as an educating and advising resources to the researchers and librarians.

---

[46] http://lucat.lu.se/LucatWeb/MainServlet?task=view_english_organization_information018010020&username= , accessed 2015-01-08

[47] http://lucat.lu.se/LucatWeb/MainServlet?task=view_english_organization_information018010010&username= , accessed 2015-01-08

[48] http://www5.lu.se/anstaelld/forska/forskningsservice , accessed 2015-01-08

[49] Current Research Information System

[50] http://www.med.lu.se/english/research_ethics_information , accessed 2015-01-8

4. Case Studies

In the following chapter we will describe the results of two case studies in the project. Firstly we will describe the research data resource "Apopthegmata Research Tool" of the Early Monasticism and Classical Paideia (MOPAI) research project[51]. We will analyse in what way the text database of the MOPAI project could be managed and hosted on a regular basis within the library organization or another organization within Lund University. A related question is what competence is needed within any of these organizations. The idea of this project (on research data management) originates from these concerns and the University Library was willing to investigate the question as part of their future orientation towards a library of digital data, both cultural heritage objects and research data publications.

Secondly, follows a description of the results of a researcher survey which we performed at the faculties of humanities and theology, social sciences and economics. Both case studies can serve as snapshots of researcher's data sources and their willingness and aptness to share research data.

4.1 Research project: Early Monasticism and Classical Paideia

Early Monasticism and Classical Paideia (MOPAI) is a research project at the Center for Theology and Religious Studies at Lund University with a 7-year grant from Riksbankens jubileumsfond (RJ). MOPAI investigates how early monasticism can be understood in continuity with the classical paideia of Late Antiquity, i.e. in continuity with elementary education, higher education in philosophy, and literary production of classical Antiquity. MOPAI studies early monastic texts and other source material with a variety of perspectives and methods, with the aim at a substantial contribution to the understanding of the emergence of early monasticism, but also to the emergence of a Christian culture and its relation the Classical Antiquity. At the center of the program stands the various collections of the Apophthegmata Patrum: sayings attributed to the first generations of monks. Apophthegmata are part of a tradition of pagan and Christian wisdom literature and are characterized by sayings and stories about moral matters usually told by a master to his or her disciples. Besides textual studies of these collections, the MOPAI program will also provide new editions of several unedited texts in Greek and Syriac, but also a digital and interactive research database, the Apophthegmata Research Tool (ART), with collections in Greek, Latin, Syriac, Arabic and other languages.

The Apophthegmata Research Tool is an outcome of the MOPAI program that is supposed to stay online as a permanent digital resource to be updated and accessed by a network of scholars within this area, or by anyone interested in using the material. The ART text database constitutes an important part of the digital resources about Late Antiquity (300-640 AD) in the Eastern Mediterranean. This is the reason why the material is published as Linked Open Data so it can be used and reused by other scholars. It also enables the integration of external data from other

[51] http://www.mopai.lu.se

projects. MOPAI will be part of the Pelagios collaboration and referencing places mentioned in our texts using common identifiers. That enables the research team to connect to data from other projects in real time to provide context. Especially important are two other research projects in the same field as MOPAI, Sharing Ancient Wisdoms (SAWS) at King's College, London, and Corpus der arabischen und syrischen Gnomologien (CASG), Luther-Universität Halle-Wittenberg. Just like MOPAI these two research projects work with wisdom collections, from earlier and later historical periods respectively, but also with editing such material in digital databases published as Linked Open Data. An integration using a common vocabulary for encoding concepts is planned with these two projects to enable an even closer comparative perspective and finding a common denominator for all three projects. The SAWS project is already part of Pelagios and their data is published online.

ART is a collaborative editing tool for text as well as an analytical tool for the study of the transmission and transformation of the texts into different cultural settings and over time. It publishes the edited texts and research results in both human and machine readable formats according to the standards of the Text Encoding Initiative (TEI) and other Semantic web standards and technologies like RDF, persistent Uniform Resource Identifiers and automatic content negotiation. ART is also reusing data of other related projects to describe context, textual sources in CASG and SAWS, related textual sources, geo linked data in archeology, epigraphy, papyri, and a historical geographic context by integrating maps of the Late Antique world. In terms of data publishing discussed in chapter 2.5, the ART database is Linked Open data.

The ART online platform is managed by the research program. An expert in Digital Humanities methodology has been with the project for two years. MOPAI runs the platform on a Virtual server currently hosted by the IT-department of the Faculties of Humanities and Theology using Open Access and general purpose software like Linux operating system, Apache web server and PHP server script. To a large extent, MOPAI develops its own interface and software for data sharing, editing, visualization and analysis. External software libraries, both on the server and client side and with an open license, are used and integrated into the platform - the latter requiring both methodological and IT expertise. Considerations are made to enable the maintenance of the platform over a long period of time as a responsibility of general service staff at the university, in collaboration with the IT-department. The IT solutions are only the last step of the methodological development that involves the implementation of international standards for text and data encoding, collaborative text editing and scholarly annotations, visualization, linguistic and statistical analysis.

The MOPAI project has been inquiring what will happen with the ART platform when the funding period of the project is finished at the end of 2015. Who will be responsible for the continued maintenance of the online database and its tools? What requirements can be imposed on an organization for research data management that will be in charge of the long-term accessibility of an online database like ART? As we have seen in previous chapters no one offers a regular service of maintaining live databases as part of a data management service.

Any university IT-department can offer virtual hosting and general support for maintaining a database and a webserver. Transforming data from the database according to international standards, for example as TEI, RDF, or any other textual output, requires knowledge in the scholarly field. It is not a problem that could be solved in the IT-department alone, but a problem that has to be solved in collaboration with the researchers of that particular field. The MOPAI project has competence in the Digital Humanities field. For the successful maintenance of an advanced project like MOPAI and the ART resource, that competence should be part of

a regular service and not just a separate research project, especially if additional projects increasingly are asking for the same infrastructure. Such a service could be part of a faculty or an organization that extends the faculty limits.

An obvious advantage to keep an advanced data management organization within the faculties is that it could specialize in the needs of the researchers. An advantage for the library organization is that it could focus on developing more generalized solutions of data management and data publishing. There would also be a connection to the libraries own digital collections and how to develop those beyond digitization and cataloguing, for instance as resources in education and research. We believe there is a need to establish "data laboratories" at both faculties and libraries to manage the question of linked data, semantic web and other data publishing issues. The development of Open Government and Open Cultural Heritage data will also increase the need for additional competence at the universities in order to utilize these new sources of potential knowledge, both in research and education. The current policy shift towards Open Access to research data is adding to the needs of research organizations to publish their own data.

4.2 Short results from a survey directed to researchers

We wanted to investigate which experiences the researchers currently have regarding research data management, and in particular the management of digital research data. The survey mostly focused on practical experiences and it was divided into the following sections: collection and analysis for digital research data, storing and archiving of digital research data, publication of digital research data, and usage of other researchers' digital data. The survey was distributed to about 1300 researchers at the faculties of humanities and theology, social sciences, and business and economics. It was a web based survey available in two languages, i.e. in Swedish and in English.

136 persons in total answered the survey. 75 persons out of 136 declared that they had experience dealing with digital research data. 11 persons out of 75 declared that they had dealt with data management plans as a requirement from the funder. According to the answers in this survey the awareness is quite low regarding research data management practices. Only a few persons declared that they: have special routines for storing research data during the project, have routines for archiving research data after finished project, have experience of making research data accessible for re-use. It is possible to sense a certain awareness and concern among the researchers for these questions. There are several comments about the time and effort it will take to prepare the research data for sharing. Some have concerns about the context of the research data, the research data must be tightly connected to its context in order to be correctly analyzed and eventually re-used. The answers of the survey indicate both a certain reflection and interest among the researchers for RDM in general. It is clear that there is an ongoing discussion on RDM within the different research communities.

We will analyze the results of the survey further, and hopefully it will lead us to some good contacts with researchers for a further follow-up.

5. Results and discussion

From previous chapters, we have seen that there is currently no coherent data management support available to researchers and that knowledge of the researchers towards current legal and archival requirements vary a lot. This has resulted in research datasets not being properly archived, processing of personal data not reported to the Legal office as required by law. There are some indications that even the concept of personal data is not very well known among researchers. There is also confusion among researchers what kind of secrecy apply to research data. Ethical support offered at the university is underutilized, just to mention a few problems. The outcome is that there are few digital datasets archived for the future and even less data made available for re-use. There is no knowledge within the university regarding the number of research datasets currently held by the university, both datasets containing personal data and other datasets, nor the amount of datasets produced in the past. Our proposals aim at creating a coherent research data management (RDM) service, building on existing resources both within and outside the university. This service should include legal, ethical, archival support as well because all these aspects coincide in the RDM process and should be considered from the beginning of the data life cycle.

From the point of view of the researcher, there are both internal and external demands that coincide in the research data management process as we have seen in previous chapters. Demands from within the research community concerns the availability of data to control and evaluate research results, to develop new fields of research and new methods by reusing existing data, and more efficient use of existing data to avoid expensive data collection. There are also ethical guidelines within each research community that often are prerequisites for publishing in scholarly journals.

Demands from outside the research community concern legal requirements on data processing, archiving, and protection of the research subjects. There are political requirements from the funding agencies as representatives of the Government. With the new directives of Open Access to Research information issued by the European Commission, research data will develop from shared resources within the scholarly community to societal resources available to anyone as far as the legal framework permits. This has great implications for the research community and is to a large extent a new role. The researcher has to take actions to fulfill his new obligations towards external demands. He cannot just say that because data contains personal information and hence are protected by secrecy and data processing restrictions, data cannot be made available to a broader community. Instead he will have to make sure all necessary measurements have been taken in order to make data available for re-use by ways of anonymization, de-identification and other measures to ensure reverse identification of individuals is not possible.

Data management should ensure proper data collection and creation of long term available digital resources as described in chapter 2, taking into account both requirements from the research community as well as requirements from the society. There is no longer a meaningful distinction between resources created for a scholarly community and the society at large.

5.1 Research data management organization

We propose a coherent organization for Research data management (RDM) within Lund University. This organization is described in figure 5.1. Units in green boxes belong to the University Library (UB), Organizations in pink boxes are external collaborators. Brown boxes represent other units within the university. The main idea is that the library organization within LU, Lund University Libraries, will be in charge of Research data management throughout the entire lifecycle of research projects and co-ordination of other parties, both internal and external to Lund University. There will be a division between the responsibilities of the faculty libraries and the University Library (UB). UB will create a new *Unit for Research Data Management and Co-ordination* and will have the responsibility for education and training programs, learning material, information structure and the co-ordination of external agents. This unit at UB will ensure the creation of an organization capable of fully utilizing competences in the network and to learn by experience within the network. The RDM unit at UB will also identify other specialists within LU that can be utilized for various needs, for example, there is a GIS unit already responsible for university wide support in such matters. The Humanities laboratory is an example of another common resource. The faculty libraries will be in charge of the actual RDM efforts towards researchers and teachers. The main responsible agents for RDM activities are of course the researchers themselves. The Library organization will be there to support them and guide them to further resources when needed. For the RDM process to be coherent it is important that all activities related to the research data lifecycle and the parallel RDM lifecycle will be included in one context, and that the support of the archive, legal office and research ethics department are available when needed or required. That will be co-ordinated by the faculty libraries. They will have the responsibility to bring in other specialists when needed. For most research projects the regular RDM service will be sufficient in most cases. They will act as guides to further resources within the university and keep lists of whom to contact, what forms to fill in, and what permission are needed given the circumstances of the research project and data collection. The faculty libraries will also have to make sure the researchers will fulfill all requirements of the university. The research data management plan should have the status of a formal agreement between the researcher and the funding agency on the one hand, between the researchers and the university on the other to provide whatever resources and support are needed to comply with the demands of the funder. The university should have a policy of their own regarding RDM and expectations about data sharing and archiving. That policy should also be expressed in the data management plan.

The research data management organization will be in close relation to the Research and Student services both within the library organization and central offices operating under the vice-chancellor. This concerns the overall issue of research information of the university. We will come back to this in the next section. To co-ordinate activities between different departments of the university in a common research data management process there is a strong need for a steering group that agrees on responsibilities and a division of competences and actions. This is also suggested in the LERU recommendations.

The Swedish National Data service (SND) is a main agent of RDM with a national responsibility for enabling sharing and re-use of research data. SND started in the 1980s as a service for Social Sciences. Since then it has expanded into the research areas of the Humanities and Public Health. SND is part of Gothenburg University and is funded jointly by the university and the Swedish Research Council (VR). They provide RDM services free of charge and

researchers are welcome to deposit datasets on their servers. Their role in the RDM process will be discussed in more detail in the next two sections.

5.2 Research data information system

Our second proposal concerns the creation of an information system to support the data management process. There are several already existing information systems within Lund University which are related to research information and will set the frame for the implementation of any research data system. Basic to all systems are the catalogue of researchers, teachers and administrative staff from which others systems collect information about identity of persons and departments (LUCAT). This system is currently under revision. There is Lund University Publications (LUP) where all research publications are listed and where they are published, including the possibility to deposit a copy of the publication in the repository of the university. All employees of the university have a personal homepage where they can write a presentation and embed a list of their own publications from the LUP database. The faculties of Humanities and Theology maintain a database of research projects, connected to both LUCAT and LUP. There is currently no information system for Research datasets at Lund University. This autumn, the university acquired the PURE research information system from Elsevier as part of the implementation of a Lund University Current Research Information System (LUCRIS). This system is intended to replace the current system as described above and at the same time provide new opportunities in this area. PURE also comes with support for managing research data, but the LUCRIS project decided at an early stage not to implement this from start, pending further investigations into research data management. Before LUCRIS can become operational, it has to undergo a period of testing and evaluation and an adaption to other systems for both input and output. Above all it must await the outcome of the current revision of LUCAT, which is planned to be completed during the first quarter of 2015. For several years, Lund University has been part of collaboration among Swedish universities to develop a system for electronic archiving. Recently they jointly decided not to continue along this route. The archive at LU is now re-investigating other alternatives, which at the moment are purchasing a commercial system or in a few years join the common e-archive of State Service Center. But the main goal still is to implement a system under 2015-2016. All these ongoing changes must be taken into account when planning a future information system for research data management since research information management and archiving of research data cannot be separated from RDM.

Our proposal is based on the distinction between two outcomes of research data management, archiving of research data and publishing of research data. As we have seen earlier, both goals can be derived from the fact that public funded research is by nature a public activity and that research data can be regarded as official documents to which a comprehensive legal regulation apply. The fundamental rule of the Freedom of the Press Act (TF) is that all official documents are available to the public if not stated otherwise by law, e.g. the Secrecy Act. This does not by any means imply that official documents must be published, but recent policy shifts make that perfectly clear, see the previous description of how Open Access to governmental, cultural heritage and research data became part of the recommendations issued by the European Commission and currently being investigated in Sweden by the Research Council (VR) by the request of the Swedish Government concerning Open Access to research information. Another distinction made to solve the problem regarding RDM information systems is to separate the

questions of managing datasets on the one hand from the publishing and archiving of datasets on the other hand. As we will see, this makes it easier to find proper and effective solutions for these related issues.

Fundamental to both publishing and archiving, to managing and storing are of course that the university organization maintains knowledge about the research datasets produced by scholars at the university. For this reason, *The Lund University Dataset Directory* in figure 5.2 is at the center of our proposal. Just like figure 5.1, green boxes are information systems owned or administrated by the University Library (UB), pink boxes represents external and brown boxes other internal information system at Lund University. The Swedish National Data services (SND) currently maintains a Dataset directory, not only of datasets that they are making available for re-use, but also other datasets that owners want to list in this directory. This service is a national responsibility for data in the Humanities, Social Sciences and Medicine and is financed by the Swedish Research Council, free of charge. By our assessment it is the state of the art and complies with international standards and good practice. SND makes a substantial effort in collaboration with the researchers to describe the datasets according to metadata standards and makes sure the datasets are stored using formats that are sustainable for the future. The datasets listed in the directory are also shared within international collaborations. Research data containing personal data are subject to legal regulations according to the Secrecy Act and the Personal Data Act and cannot be stored at the servers of SND. Such data has to be de-identified and additional measurements to prevent individuals from being identified in the dataset, if necessary, has to be taken into account.

In our proposal SND constitutes the primary option to publish data for sharing and reuse. In a pilot project in collaboration with SND, we want to investigate further if we can set up a common dataset directory and thus take advantage of the experience that SND possess and networks in which they participate, for example CESSDA[52], DataCite[53] and the DDI Alliance[54]. SND is in favor of participating in a project like this as it opens an opportunity to explore a new role towards institutional support of the universities, rather than support towards individual researchers. Such a joint database would become a first step towards a *National Dataset Directory* for the research areas within the responsibility of SND. The *LU Dataset Directory* would then constitute a subset of an emerging national dataset directory, making it easier to find and locate other datasets within relevant subjects, regardless where they are produced.

Careful considerations must be made to ensure that a joint information system in collaboration with SND can handle the connections and interoperability with other internal research information systems at Lund University like LUCAT and LUCRIS. From the perspective of the user, that is researchers and librarians, it is very important that input into administrative systems should occur only once. However, as has just been argued, there are strong advantages in relying on SND as a national and public agent for research data, rather than becoming dependent on commercial systems at this early stage when RDM is novel to all parties and much will happen in the future as a result of the new directives from the Research Council (VR). In their preliminary report VR anticipates a development towards a common infrastructure for Research data and a period of Pilot projects within different research areas. The current structure and content of the dataset directory maintained by SND is very rich compared to the corresponding

[52] Consortium of European Social Science Data Archives, http://www.cessda.net

[53] DataCite is a non-commercial organization with the aim of making data available and citeable. Member organizations can assign DOI:s to research datasets, https://www.datacite.org.

[54] Data Documentation Initiative Alliance is developing metadata standards for the description of research data, http://www.ddialliance.org

system that is part of PURE/Elsevier. As a member of DDI Alliance, SND is also involved in the international development of metadata standards for research data.

Currently SND offers sharing of de-identified, if applicable, datasets on file servers together with comprehensive metadata in a dataset directory published online. Direct access to data is not the most frequent alternative. This far, depositing of data at SND is based on a voluntarily agreement between researchers and SND, where the conditions for sharing and re-use are determined by the researcher. We have also looked into the demands of modern research and it is sometimes desirable to offer more elaborate data services, for example, in the *Early Monasticism and Classical Paideia* project, which is our primary case study, researchers would like to share digital editions of source material structured as XML-data and retrievable by permanent identifiers that can be used as references in scholarly literature, and an active API that allows selections and combinations to be made in real-time. This digital source material constitutes a growing and distributed web of data about the Ancient World and is part of collaboration between similar projects, each responsible for an important share of commonly utilized resources. It is also a good example of a born-digital product, that enables new methodological approaches and research questions that could not have been asked using a traditional printed edition. Such digital assets are similar to cultural heritage data organized using modern principles for sharing and reuse of data and for different purposes. Another scenario would be a regular SQL-database that will be shared among researchers at multiple research centers. We also propose that existing solutions for data publication can coexist alongside the main alternative for storing at SND. If for example, there is a research project creating an online database that will be maintained for a very long time, they can handle their own data sharing. Such information systems are labeled *Faculty and Department Resources* in figure 5.2. It has become increasingly common to utilize subject specific resources for data sharing both in Sweden and abroad. All scholarly recognized solutions for online data publishing, or data storage, are acceptable as long as they are published with an open license, the RDM service keeps track of the dataset including the location of its publishing, maintains metadata about the dataset and set up a data management plan.

The archiving of data must be solved at the university in accordance with current Swedish legislation. The entire archiving of data is not an issue of publication of data but long term preservation of data. Data eligible for archiving must be stored in archives without any loss of information. Hence anonymization of personal data might be an option for security, but personal data should never be de-identified. As said above, archiving and publication of data are two different outcomes of research data management and serves different purposes. Archived data can also be subject to use and re-use for other legitimate purposes, but has to be put on trial in each individual case.

Cultural heritage collections are also maintained at the Lund University, in archives, museums and libraries. There is a digitization program running at the University Library that transforms these collections into digital assets. Since data of this kind can also become important assets in education and research they should be managed as well and organized for maximum availability and efficiency in research. Lund University Library is a partner in a national collaboration of university libraries and archives, called Alvin, and aiming at creating a joint web portal and a backend database for cultural heritage collections across libraries. Digitized material will be available through the portal. Cultural heritage data should not, according to our proposal, be part of the *LU Dataset Directory*, but instead be part of an information system of its own. We propose a co-ordinating role for the new *Unit for Research Data Management and Co-ordination*

at the University Library (UB) towards the Digital collections at UB. The Alvin collaboration is similar to the collaboration with SND for research data.

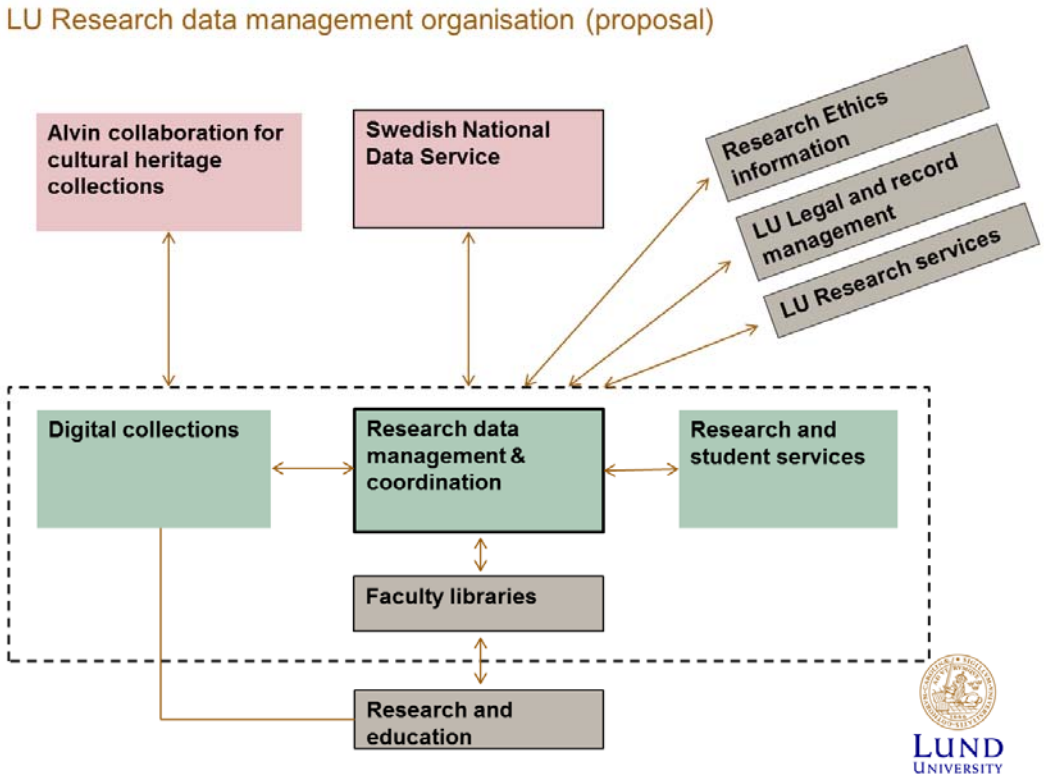

Figure 5.1

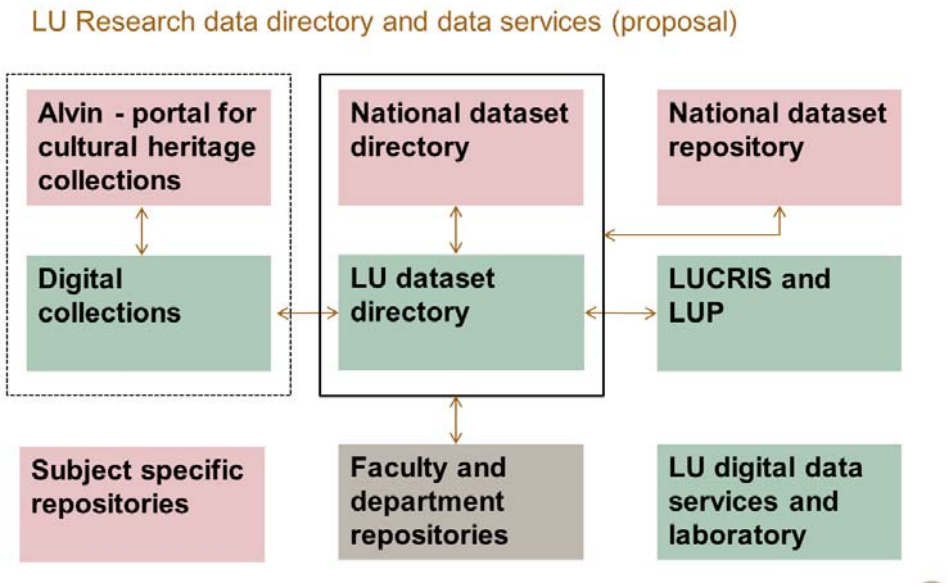

Figure 5.2

5.3 Education and training, development of competence within the library organization

In chapter 2.3 we discussed what competences and skills that libraries and librarians currently have that could be useful in RDM. We will now further describe this and also discuss what is needed in terms of competence development of the library staff. It is not only the library staff that would need competence development, many other groups of staff would need further training too. Much of the training and competence development could be performed in-house, by different university functions and specialists. But the developing of an RDM service must start with a mapping of the current competences within the university and an analysis of the new needs for competences and skills.

In our proposal of an organization for research data management at Lund University we suggest the establishment of the section "Research data management and co-ordination" at the University Library. This new section is supposed to co-ordinate and to promote the activities in RDM towards the faculty libraries and their faculties, university management, and other university research support functions. In chapter 2.3 we described which skills and competences libraries currently possess that could be useful in RDM. We also described possible roles and actions the library staff as supporting in the data life cycle process of the researchers. In order to provide support in the data life cycle process the library staff need competence development into many new areas.

In the following areas we see needs for competence development of the library staff:

- ✓ Basic data management
- ✓ Data management plans
- ✓ The data life cycle
- ✓ Archiving
- ✓ Subject classification of research data, i.e. DDI (Data Documentation Initiative)
- ✓ Metadata standards
- ✓ Copyright and licenses, i.e. Creative Commons
- ✓ The research lifecycle
- ✓ Legal prerequisites and Research ethics in managing research data
- ✓ Data collection methods
- ✓ Preparation for sharing, de-identification of personal data
- ✓ Data identifiers such as PURL, URI
- ✓ Linked data, the semantic web and other forms of data publishing
- ✓ Methods for text encoding such as XML and other formats

There is already some knowledge in the library organization on some of the above areas, but there is needed more in-depth competence on these areas, if the library should be an essential partner in research data management at the university. Today we have differvfr5tyent categories of librarians and specialists within the library organization, such as subject librarians (faculty libraries), research support librarians, user-training librarians, cataloguing librarians, e-resources librarians etc. We need to review our personnel categories in order to see where these would be integrated in the RDM process and to see what is needed in terms of competence development for each role. As we stated in our proposal for organization of RDM the faculty libraries would

take the roles as a primary contact point for researchers needing support in RDM. In such a scenario there is needed competence in RDM and data management plans at the faculty libraries. The faculty libraries would also need competence development in the data life cycle process, and they need to consider how they will develop this new way of working together with the researchers. The library staff at the faculty libraries provide highly subject specific support to their researchers, and they have close relationships to their researchers. Different sections at the University Library, such as digitization services, digital collections, scholarly communication, may also play important roles in the RDM process, acting as experts in their specific domains. The library staff at the University Library also needs competence development in RDM, both on the basic level and on specific parts of the RDM process.

Programs for competence development

As we stated in the beginning of this chapter much of the training and competence development could be performed internally, within the university, as a knowledge exchange. But there will be also be a need for external training opportunities, by external training partners. The Swedish National Data Service (SND) would be a suitable partner for training and competence development. As we stated in the proposal for organization of RDM, we suggest that SND are engaged to deliver an *institutional support* to Lund University in RDM – both in terms of technical support and in terms of training support. In such a scenario Lund University and SND would commonly set-up a competence development program for the library staff and other employees at LU.

Competence development on short-term basis and on long-term basis

With the coming national guidelines, RDM will most likely be an important concern for the university in the future. There are obviously needs for competence development on short-term basis and on long-term basis. On short-term basis there would be actions taken such as: introductory workshops on RDM, focus-groups on RDM, production of basic information material on RDM. Seminars and workshops on different parts of RDM would also be included in a short-term training program. In the long-term the university needs to rethink its policy regarding recruiting library staff and the continuing development professional program for the current library staff. A new generation of librarians will grow in the future, with specializations such as data curation or data librarians, and this will form a different recruitment base for the university. The library and information schools will slightly change their education programs for librarians, something we already see happening in USA and UK.

Future role of academic librarians

When planning for an RDM service at the university and planning for the competence development of this service, we can get a lot of inspiration and ideas from other universities and from the literature. There are many discussions in the literature on the new roles and competences of library staff in RDM. Libraries and librarians will encounter many changes with the future development of open access to research and research data, so there is obviously a need

for discussions on the future role and profession of librarians. Many organizations such as ARL[55], RLUK[56], JISC[57], ARIADNE[58], have taken initiatives to studies and investigations, and many researchers have discussed the future professional skills of libraries. In the article "Upskilling Liaison Librarians for Research Data Management" Andrew Cox gives some strong arguments for libraries to play an important role in research data management (Cox, 2012). According to Cox libraries have good networks with institutions built through liaison activities, and they have an understanding of generic information management that be may be applied to data management. He summarizes the potential of the libraries as follows:

"As a well-networked profession, librarians are quick and generous in sharing knowledge, and so can play an important role in rapidly replicating good practice across the whole sector" (Cox, 2012)

[55] ARL – Association of Research Libraries, http://www.arl.org , accessed 2015-02-04
[56] RLUK – Research Libraries UK, http://www.rluk.ac.uk/ , accessed 2015-02-04
[57] JISC, http://www.jisc.ac.uk/ , accessed 2015-02-04
[58] ARIADNE, http://www.ariadne.ac.uk/ , accessed 2015-02-04

6. Conclusion

The main purpose of this project was to investigate the roles and actions the University Library and the faculty libraries at Lund University might take in Research Data Management (RDM) at Lund University. As a case-study we have used an example of research data resource, i.e. the online database of the Early Monasticism and Classical Paideia project, and we have analyzed in what way this resource could be managed and hosted on a regular basis within the library organization. We have performed a detailed study of the library organization at Lund University, and an in-depth analysis of the functions within Lund University that could be considered related to RDM. We conclude that the full range of RDM activities constitute new tasks for the library organization but falls within the general and already familiar competences of a research library to structure, classify and make knowledge available to the public.

We have also undertaken a major literature study of the development of RDM globally and the current trends in Sweden. Having done these studies internally and externally, we outlined a future model of research data management at Lund University. Our model is built upon the transfer of current information systems and competences at the Swedish National Data Service (SND) to build a full range and coherent RDM organization at Lund University involving researchers, librarians, archivists, ethical and legal experts. We have also studied the future needs of professional development at Lund University, as the university will be engaged in RDM. Here we focused mainly on competence development of the library staff.

The position and scope of the library organization within the university makes it a suitable agent to co-ordinate the efforts of researchers and other agents concerned to develop scholarly data resources into resources for the society as a whole according to recent policy shifts underway. We have seen both a willingness and an interest within the library organization to take on this new task because the library is already involved in other digital activities around research information, digital publications as well as information supply.

As authorities, research funders and other parts of the society are expected to put more requirements on researchers to make their research data publicly accessible, universities will have to prepare themselves for these research conditions. One important factor for the universities in this new, coming situation is to invest in professional development in RDM towards the researchers and towards the research support personnel, such as the library staff. Another important factor for the universities is to invest in good technical systems for RDM, which will facilitate and rationalize work of the researchers.

By acquiring the current competences and information system of SND and make it a core component of the local RDM organization at Lund, especially the library organization, we believe this model of RDM can be applied to other Swedish universities as well according to the design and intentions of this study. The outlined co-operation between Lund university and SND, and the changed role of SND - from individual support towards researchers, to institutional support of an entire university - we are taking the first steps towards a national co-ordination concerning RDM at Swedish universities, and a common system for the preservation and dissemination of research data. This co-operation could be seen as an answer to the call

made by the Swedish Research Council in their preparation of national guidelines for Open Access to research information, a period of initiatives from the universities to carry out pilot projects in research data management.

References

Research Data Management: Practical Strategies for Information Professionals, 2014, Purdue University Press, West Lafayette, Indiana.

Guidelines on Data Management in Horizon 2020. Version 1.0, 11 December 2013, 2013, European Commission.

*LERU Roadmap for Research Data, LERU Research Data Working Group*2013, LERU, Leuven.

Innovation and Research Strategy for Growth 2011, Department for Business, Innovation & Skills, UK Government. Available: https://www.gov.uk/government/publications/innovation-and-research-strategy-for-growth--2 [2015, February/03].

Riding the wave: How Europe can gain from the rising tide of scientific data. Final report of the High level Expert Group on Scientific Data. A submission to the European Commission. October 2010, 2010, European Union.

OECD Principles and Guidelines for Access to Research Data from Public Funding, 2007, OECD, Paris.

Andersson, U. et al 2011, *Projektet Forskningsdata inom humaniora och konstnärliga vetenskaper - open access?*, Kungliga biblioteket, Stockholm.

Björklund, C. & Eriksson, J. 2007, *Forskningsdata i öppna arkiv och universitetsarkiv: en förstudie vid Göteborgs universitet, Lunds universitet och Sveriges Lantbruksuniversitet: Projektrapport till Kungl. biblioteket, Programmet för Open.Access.se Augusti 2007*, Kungliga biblioteket, Stockholm.

Bohman, M. 2013, "Lagring av forskningsdata vid svenska lärosäten", *Tillgänglig Forskning: Vetenskaplig Publicering, Evaluering och Open Access,* vol. 1, no. 1.

Brandt, D.S. 2014, "Purdue University Research Repository: Collaborations in Data Management" in *Research Data Management: Practical Strategies for Information Professionals*, ed. J.M. Ray, Purdue University Press, West Lafayette, Indiana, pp. 325-345.

Carlhed, C. & Alfredsson, I. 2009, "Swedish National Data Service's Strategy for Sharing and Mediating Data: Practices of Open Access to and Reuse of Research Data", *IASSIST's 35th annual conference*.

Cox, A. 2012, "Upskilling liaison librarians for research data management", *Ariadne,* no. 6 December.

Craigin, M.H. et al , *An Educational Program on Digital Curation*. Available: https://www.ideals.illinois.edu/bitstream/handle/2142/3493/ALA_STS_poster_2007.pdf?sequence=2 [2014, December/07].

Data Management Consulting Group, *Research Data Life Cycle* [Homepage of University of Virginia Library], [Online]. Available: http://data.library.virginia.edu/data-management/ [2014, December/08].

DCC 2013, *Checklist for a Data Management Plan*, 4th edn, Digital Curation Centre, Edinburgh.

European Commission 2012, *Commission Recommendation of 17.7.2012 on access to and preservation of scientific information*, European Commission, Brussels.

ICPSR 2012, "Figure 1, Data Life Cycle" in *Guide to Social Science Data Preparation and Archiving*, 5th edn, Inter-University Consortium for Political and Social Research, Ann Arbor, Michigan, pp. 8.

Ivarsson, L. & Gullstrand, S. 2014, *Kort omvärldsanalys och kartläggning av hur universitet i Europa hanterar sin forskningsdata och open access: En rapport*, Göteborgs universitetsbibliotek, Göteborg.

Kuula, A. & Borg, S. 2008, *Open Access to and Reuse of Research Data - The State of the Art in Finland*, Finnish Social Science Data Archive.

Lyon, L. 2012, "The Informatics Transform: Re-Engineering Libraries for the Data Decade", *The International Journal of Digital Curation,* vol. 7, no. 1, pp. 126.

NCSU Libraries*, Data Management Planning: What is a Data Management Plan (DMP)?*. Available: http://www.lib.ncsu.edu/guides/datamanagement [2014, December/07].

RCUK*, RCUK Common Principles on Data Policy* [Homepage of Research Councils UK], [Online]. Available: http://www.rcuk.ac.uk/research/datapolicy/ [2014, December/08].

Regeringen, 2014, *Regleringsbrev för budgetåret 2014 avseende Vetenskapsrådet*.

Rumsey, S. & Jefferies, N. 2013, "Challenges in Building an Institutional Research Data Catalogue", *The International Journal of Digital Curation,* vol. 8, no. 2, pp. 205.

Runardotter, M. & Lindbäck, L. 2012, *Workshop: Hantering av forskningsdata och "open access" till forskningsdata: Slutrapport*, Luleå tekniska universitet, Luleå.

Schaffner, J. & Erway, R. 2014, *Does Every Research Library Need a Digital Humanities Center?*, OCLC Research, Dublin, Ohio.

Showers, B. 2012, *Does the library have a role to play in the Digital Humanities?*. Available: http://infteam.jiscinvolve.org/wp/2012/02/23/does-the-library-have-a-role-to-play-in-the-digital-humanities/ [2015, February/03].

Sula, C.A. 2013, "Digital Humanities and Libraries: A Conceptual Model", *Journal Library Administration,* vol. 53, no. 1, pp. 10.

Swan, A. & Brown, S. 2008, *The Skills, Role and Career Structure of Data Scientists and Curators: An Assessment of Current Practice and Future Needs, Report to the JISC*, Key Perspectives Ltd, Truro.

University of Oxford*, About RDM - What is RDM?*. Available: http://researchdata.ox.ac.uk/home/introduction-to-rdm/ [2014, December/07].

University of Oxford*, DaMaRO Data Management Roll-out at Oxford*. Available: http://damaro.oucs.ox.ac.uk/docs/DamaroPoster3.1.pdf [2014, December/13].

Vandegrift, M. 2012, "What is Digital Humanities and What's it Doing in the Library?", *In the Library with the Lead Pipe,* Available: http://www.inthelibrarywiththeleadpipe.org/2012/dhandthelib/ [2015, February/03].

Vetenskapsrådet 2014*, Bakgrundsdokument: Nationella riktlinjer för open access*. Available: http://www.vr.se/download/18.4702e05d148edabbd9c51aa9/1412783218988/Utkast+Bakgrundsdok+Nationella+riktlinjer+OA.pdf [2015, January/4].

Vetenskapsrådet 2014*, Förslag till Nationella riktlinjer för open access*. Available: http://www.vr.se/download/18.4702e05d148edabbd9cdd591/1412862685498/Utkast+Nationella+riktlinjer+OA.pdf [2015, January/4].

Voog, H. et al 2013, *Tillgänglighet, närhet och synlighet – gemensamma utmaningar för LUB-nätverket för att möta forskares behov av stöd vid Lunds universitet*, Lund University, Lund. Available: http://lup.lub.lu.se/record/3404879 [2015, February/3].

Westerberg, B. 2014, *SOU 2014:45 Unik kunskap genom registerforskning. Betänkande av Registerforskningsutredningen,* Stockholm.

Zorich, D.M. 2008, *A Survey of Digital Humanities Centers in the United States*, Council of Library and Information Resources, Washington.

WWW.UB.LU.SE

LUND UNIVERSITY

Box 117
SE-221 00 Lund
Tel +46 46 222 00 00
www.lu.se