Data Science Competencies In Danish Academic Libraries

I. Introduction

1.1 Delimitations of the study

The main focus for this thesis will be on the data science competencies that Danish research librarians have acquired and the ways in which they are working with these skills. In today's competitive research environment, the need for librarians to be knowledgeable about this field is growing. Data-savvy librarians are able to better assist their patrons with the resources they need for their research, as well as extract useful insights from library data. Denmark's research institutions are noteworthy both in a Northern-European context and on a global scale. Yet how prepared are Danish research librarians to assist their users in matters of data science? And from a institutional standpoint, what challenges does this bring?

To answer these questions, I will have a phenomenological approach, starting with a brief timeline of technological advancements in Danish libraries in order to discover what data skills librarians possessed a couple of decades ago, and how they have evolved up to the present day.

After acquiring some insight into the past, I will move into the present times, looking at the data competencies of librarians in Denmark. In the analysis I will use qualitative methods to interpret the data. Through this, I hope to uncover some common grounds that can define the current skillset of a Danish data librarian, as well as some common challenges that they are facing.

By opening a discussion around data science skills, I will also uncover some of the Danish initiatives related to this subject - I will touch upon and mention training initiatives, research support services and data labs and discuss their importance in a research institution. I will keep the discussion inside the research library space, which means this study will not discuss data skills from a business or economic perspective.

1.2 Research questions

In order to analyze the current data science landscape in Danish libraries, I decided to focus on the data skills and competences that librarians have, the ways they use the skills in a library context, as well as what challenges they perceive to have.

This creates the following main research question:

I. "What are the current data science initiatives in Danish research libraries?"

This question can be explored by answering the following secondary research questions:

1. "What data competencies do Danish librarians possess? What skills do they wish to develop?"

2. "What challenges do librarians face when dealing with data science in their work?"

1.3 Objective of the study

This study presents an overview of the data science competencies of academic librarians in Denmark. The aim is to identify some common patterns and challenges that librarians and library managers face in this field.

This will be achieved by a series of interviews with professionals working with data science in and for Danish libraries. Furthermore, the thesis also presents the results of empirical research in the form of a survey sent out to members of the Danish Union of Librarians' Data Science group.

In this way this thesis aims to follow up on a wider international discussion of data science initiatives in libraries, a topic that has increasingly come to the attention of librarians and library managers in the USA. By focusing the discussion on Danish data science practices in a research library context, I am looking to provide a more clear picture of the librarians' situation in the Nordic space.

By presenting the current data science landscape in Danish research libraries, the thesis will uncover trends which will be useful for the development of further initiatives in the field.

II. Data science in relation to librarianship

2.1 Data science as a field

In this chapter I would like to define data science, as well as explore how this field appeared and the delimitations of data science in the context of my thesis.

Question: What is data science?

A few centuries ago, humankind was reorganizing itself around mechanical inventions that were boosting production and bringing unforeseen profits. The 19th century industrial age was speeding up goods production, which in turn, led to the speeding of commerce and productivity; however, it also created the problem of pollution.

Around the 20th century, we further optimized our technology, creating machines capable of computing math equations and running simulations – in this way, we started using computers to gather and store data about ourselves and our environment with the purpose of understanding our universe. The changes that followed the invention of computers and data storage systems reshaped the way we organize knowledge.

The 20th century brought an informational overload. Since inventing computers, our data has doubled at astonishingly fast rates. In 2020 alone, it is estimated that we will create 40 trillion gigabytes of data (Ozdemir, 2016, p. 2). However the data we create can help us enhance our world.

This is where data science steps in. From a broad perspective, data science is "the art and science of acquiring knowledge through data."(Ozdemir, 2016, p.4)

A mix of methods from sciences such as statistics, computer science and domain-specific knowledge, data science takes proven-to-work methods and blends them in new ways in order to deal with today's Big Data questions.

In order to understand what a "data question" might be, it is important to make a distinction between structured and unstructured data (Blei & Smyth, 2017). Unstructured data is data that needs further processing in order to be useful for a specific query. Structured data is data that can be arranged in tables and rows, which makes it easier to get insights from and further process it.

Because at some point, any data can be considered "unorganized" enough for a specific question, data science is involved in all stages of the process, structuring it using different methods and tools, until the data question is answered.

By applying data science methods to different topics and disciplines, one is able to use data to explore, better understand a certain area of study or to predict outcomes. Due to the large amounts of data that data scientists work with, it is necessary for them to have the capacity to simplify, be critical about, and effectively communicate the results of their data analysis. As a science in which one always has to find an appropriate method for the problem at hand, it requires an overview and understanding of statistical and computer science methods and tools, in order to, -as Blei affirmed- "balance appropriate assumptions with computationally efficient methods". (Blei & Smyth, 2017, p.8690)

While some see it as "the child of statistics and computer science" (Blei & Smyth, 2017, p.8689), in the 1960s, Tukey believed that it will come to have a broader definition than just computational statistics: "It will still be true that there will be aspects of data analysis well called technology, but there will also be the hallmarks of stimulating science: intellectual adventure, demanding calls upon insight, and a need to find out 'how things really are'" (Tukey, 1962, p.63)

It seems Tukey was right: after all, what separates a good data question from a great one, might be who asks it. Technology has evolved to the point where computers can store massive amounts of data and perform truly impressive and precise calculations. However "applying statistical and computational tools to modern scientific questions [still] requires significant human judgement and deep disciplinary knowledge" (Blei & Smyth, 2017, p.8690).

Blei & Smyth propose a triangular approach to studying the field of data science; they argue that a data science problem can be studied from three perspectives: statistical, computational, and human. Upon adoption of these three ways of thinking, the following becomes apparent:

- statistical thinking offers a suite of methods for understanding data;
- computational thinking provides the crucial considerations of how to balance statistical accuracy with limited computational resources;
- human thinking offers a understanding of the context of the data, providing a way to answer a specific question.

While providing examples of ways in which statistical and computational thinking might be applied in a library context, this thesis is based on the assumption that a librarian's ratio of technical and domain-specific skills is good enough to be able to use data science in a library setting. Therefore, this paper is concerned with whether a librarian can act as a data scientist in an academic library, from a human, competency-oriented perspective.

2.2. Data scientist competencies

Here I will discuss what competencies a person has to have in order to be called a data scientist. Refer to educational background but also methods of working.

Questions: What does a data scientist do? What competencies do data scientists have?

In order to understand whether a librarian could be a data scientist, I propose first exploring what makes a good data scientist. What are the skills and competencies necessary to be one? And what is the educational background that these specialists have?

Broadly speaking, Stanton suggests that a data scientist will be most involved in what he calls "the four A's" of data: data architecture, data acquisition, data analysis, and data archiving.(Stanton, 2012)

In order to understand the skills and competencies one might need, we have to first take a look at a data science project's lifecycle. There are five steps that describe what a data scientists' work comprises of: *Obtain, Scrub, Explore, Model* and *Interpret*.(Mason & Wiggins, 2010)

Obtaining the data goes beyond retrieving a few documents from a database. For a data scientist, this process involves seeking a sufficient amount of data from different sources, especially by coding. (Mason & Wiggins, 2010) For this purpose, a data scientist should be familiar with a programming or scripting language through which he can automate

and manage the data obtaining process. One should also be familiar with APIs, as well as different formats of data, which will help when sorting it, in the next step:

Scrubbing, or cleaning the data, is the process through which the data obtained is adjusted to remove incorrect, incomplete or duplicated data. During this process a data scientist will standardize all differently formatted data to a common format which can best serve the project. It is important to have a broad knowledge of different data formats and their capabilities. The data has to be validated, either by rejecting incomplete data, or correcting and enhancing selected records.

The purpose of data scrubbing is to get a dataset that is valid, accurate, consistent and uniform, as "a simple analysis of clean data can be more productive than a complex analysis of noisy and irregular data."(Mason & Wiggins, 2010)

Exploring the data implies looking at the data without attempting to make any predictions or test any hypotheses yet. It may involve visualizing, or "zooming in" or "out" of the data and viewing it from different angles. By visually displaying data, interesting 'threads' can be observed which can be more closely followed in the next steps of the analysis.

Models are built and applied in order to predict and to interpret an outcome. The process of building a model includes wrangling with data, transforming it, and feature selection; further, the model is trained on algorithms, its parameters tuned, then retrained; this step implies model management activities, while maintaining good communication with the rest of the team (programmers and domain experts alike); finally, the model is evaluated using methods such as cross validation, model reporting, or A/B testing.

As data alone has little meaning, the final step for a data scientist is to *interpret* his findings. A good model will suggest through its data which analysis will be interesting to perform next. This step makes a data scientists' 'soft skills' shine through; of course, technical skills will still be necessary to present the findings. However, a degree of curiosity, good communication skills, and wish to innovate by applying the data findings will go a long way in a data science team.

These steps represent a general walkthrough of the process, however as any project is unique, it is good to take into account a certain degree of iteration. As well as that, Mason states that a data scientist is not expected to be proficient in all of the above areas; it is only human to have weak points. A data scientists' curiosity and passion to find insightful data through his work will help him overcome and polish those weaker areas over time; it is far more important to have an understanding of the bigger picture and the tools one has available.

In 2017, the UK Government's Digital, Data and Technology department (DDaT) has outlined a set of essential skills any data scientist should have[1].

[1] They can be found in the DDaT Capability Framework, accessible at https://www.gov.uk/government/collections/digital-data-and-technology-profession-capability-framework

Following the data science project lifecycle summarized above, the skills outlined by the DDaT Framework are:

- "Applied maths, statistics and scientific practice skills;
- Data engineering and manipulation;
- Data science innovation;
- Developing data science capability;
- Domain expertise;
- Programming and build (data science oriented);
- Understanding the analysis across the life cycle."

The above skills are useful to determine *what* a data scientist can do. In order to assess how suited a person might be in a data science role, it is a good idea to also understand *how* they use their skills to successfully solve a problem. Therefore, perhaps a better way to measure data science proficiency might be found by analysing a person's data science competencies.

The aforementioned DDaT report suggests the following competencies are desirable traits for a data scientist to have:

- an openness to change, improve and look for a smarter ways to solve a problem;
- a collaborative, team-player mindset, which encourages building a supportive network of colleagues and people outside the organization;
- delivering timely performance and taking responsibility for project outcomes; dealing with challenges in a responsive, constructive way;
- seeking out and implementing the solutions which offer the best quality and the least waste of resources;
- communicating purpose in an enthusiastic, forward-oriented manner, showing pride and passion in performing the services;
- seeing the big picture; understanding how their role fits the wider picture, and focusing their contribution appropriately.

I suggest also including "data-oriented thinking" to the list above. This comes, as Blei & Smyth mentioned, through understanding the context of the data and knowing what to ask of the dataset and how.

The term "competencies" captures three facets which offer a understanding of how well a person might perform in a project: skills, knowledge, and abilities.(Beckett, 2015) Even though it is easier to point out and describe specific skills and tools that are used, a competency-oriented approach can paint a wider picture of a data scientists' capabilities. It can reveal many of the 'soft skills' that are required to be a good data scientist.

And it makes sense from a hiring perspective, too – as we will see further on, when it comes to data science applied in libraries, a librarian's thought processes are quite similar to a data scientists' – why deny one the opportunity to be involved in data science projects just because he doesn't have formal IT qualification? Passion is more difficult to teach than programming; where there's enough of the former, the latter will follow in no time!

A 2017 study with a sample of 1001 LinkedIn data scientist profiles sketches a common educational profile of a data scientist: 77% of them held a Master's degree or higher [2], in the following fields: Economics and Social Sciences, Statistics and Mathematics, Natural Sciences (Physics, Chemistry, Biology), Data Science and Analysis (which includes Machine Learning), Computer Science (which excludes Machine Learning) and Engineering. The ranking of the university doesn't seem to matter more than the actionable knowledge they have; this is further supported by the fact that over 40% had extracurricular qualifications such as certificates from online courses.

As it is a quite young science, data scientists have very diverse backgrounds; their backgrounds are diverse enough that a LIS educated individual could also work with data science, provided that person also has knowledge of data science tools and methods.

When talking about tools, I should mention that according to the same study, the three most popular coding languages are R, Python and SQL. (365DataScience, 2017).

All in all, when talking of data science skills and competencies, a data-oriented thinking seems to be more important (and hard to acquire) than any specific educational degree or tool a data scientist might use. Indeed, analytical backgrounds such as Computer Science, Statistics and Mathematics offer an advantage to someone new to the data science field. Equally important is a data-oriented thinking and applying a sense of curiosity to the information search process.

The latter two sound strikingly in line with what describes most librarians – so going with this theory, a librarian might make a good data scientist if they wish to.

But do they wish to? In the next chapter I will present reasons as to why they might, and the starting point of that argument begins with looking at how libraries and technology developed together through time.

2.3 Libraries' role in the technology world

Questions: How are libraries supporting technological developments? What are ways in which academic libraries can help enhance students' digital skills?

[2] 48% Master's degree; 27% Ph.D.; 2%MBA

Once upon a time, librarians were the "gatekeepers" of collections used by researchers everywhere. In a time of information scarcity, libraries aimed to preserve human knowledge as best as possible and make it as universally accessible as possible. As this role became accepted by society, governments assimilated libraries into a public service, which came with political implications.

Due to the major advancements in information technology over the past few decades, librarians now have control and responsibility not only over physical documents, but also over electronic journals, archives, and databases. In the whirlwind of rapidly expanding knowledge, they remain one of the few entities which can point users towards trustworthy sources of information. And yet, there seems to be a growing public perception that the Internet is making libraries obsolete.(Galluzzi, 2013) It is no wonder, then, that over the past few decades, libraries have tried to reinvent themselves and their role of information mediators.

Libraries have kept up by adapting their collections around technology, and the invention of computers can serve as a good example for this: electronic storage solved many problems that the 19th century library was facing, such as physical storage spaces, and brought welcome developments in areas such as interlibrary loaning. Thanks the invention of integrated library systems, librarians were eased of performing some of the more mundane, time-consuming tasks. Overall, the technological advancements happening in the world improved and advanced library work proportionately over time.

The invention of the Internet brought a huge change to the world, and it affected libraries as well, both in their work processes and in the way the public perceives them. An incorrect generalized assumption that "all good information can be found for free on the Internet" has led the 21st century librarians to lose their place as information experts in the public eye. It created a situation where "libraries are pressed in the middle between the expansive role of the Internet and the shrinking role of national and local authorities in providing public services on a taxation basis."(Galluzzi, 2013, p.4)

This assumption is perpetuated by the ease through which one can use search engines such as Google; why should one go further than the first few relevant search results? It is my opinion that it's important for librarians to hold their place as information mediators in society, as the information landscape nowadays is often political – libraries can shed light in the informational deluge in a neutral way and thus act as a 'diplomat' of democratic data.

In the 90s and 2000s libraries were embracing technology and looking for more ways in which to use it to enhance their services. The opportunity to extend libraries into a digital space has created new roles within the library.(Shahbazi & Hedayati, 2016)

These new roles also came with new responsibilities and considerations: one of these responsibilities was to make sure that society would not be divided between 'the information rich' and 'the information poor', in terms of access to quality information. In this context,

access to information meant not only having the information available for searching, but also *educating* the users on how to find it, using the new technologies. This consequence of technological advancements in libraries brought a new facet to the roles of the modern library: educating the public towards an improved information behaviour.(Carlson & Johnston, 2015)

Yet in the wider discussion on the place of a modern library in society, "an active role of librarians in [...] building [...] digital libraries, as well as in offering services for digital collections, is scarcely perceived by the general public." (Galluzzi, 2013, p.116)

Libraries are fighting for a place in the new information society by extending their services beyond their traditionally-known 'book-lending' role. Librarians are working with library users to find new library tools and methods to help them with their research projects, for example.

In universities, their role is important to novice and experienced researchers alike; however the way in which they serve these two user categories is different. Whereas a novice researcher might be interested in learning how to get the most out of a database or which software to use in their research, what can an experienced researcher get out of this? Librarians are answering this question by attempting to take their information specialist role even further - by becoming involved in the research itself.

2.4 Impact of technology advancements on the librarian profession

Questions: How did technology change librarians' roles? What are some ways in which Danish libraries adapted to the growing need for IT literacy among their public?

The previous chapter focused on the relation between technology and libraries, presenting some of the important changes that have happened in the past few decades. In this chapter, I will take a closer look at how technology has impacted librarians' role. After presenting some general observations, I will talk about the situation in Denmark, and how Danish librarians (and the Danish LIS school) reacted to the rapidly-changing situation.

As pointed out in the previous chapter, technologies and libraries have grown side by side, which caused a numerous amount of changes in the library system over the past few decades. More than half a century ago, librarians were disseminators of knowledge: their duty was to educate the public as to which documents they could read to self-educate on a specific area of study; improvements to the education system combined with the informational explosion meant that as people started to pursue higher education, librarians somewhat lost their role as educators, and best served users by re-emphasizing their role as knowledge custodians.

In the present times, the large amount of information available has led to the creation of numerous tools and software for managing, storing and manipulating it; unlike half a decade ago, people have little difficulty finding the information they are looking for; today's problem is that they find too much of it. And despite there being numerous software that can help filter and shape the data one needs, users are still left confused. This is why librarians have taken back their role as educators – this time around, they are helping users match their data with the appropriate tools and solutions for their needs.

Librarians are educating the users on *how* to reach and get the most out of their preferred information sources, rather than educating them on *which* subjects or authors they should read; in this sense, it is more appropriate to say they are no longer disseminating knowledge, but rather information. And by changing their role from "knowledge disseminators" to "information disseminators", libraries have re-defined the profession of librarians.

The technological advancements that followed means librarians have less control over the information the users access; one aspect on which librarians have an important say is the tools and services the library offers to its users; this puts librarians in a position where they need to follow technological developments and assess which of these would be best suited in a library, considering evolving user needs.

In order to assess this, it is necessary for librarians to learn how to use the new technologies, understand the underlying purpose behind their users' requests, and then figure out a way to guide users towards their specific information need. It is no wonder that this process can be overwhelming for a librarian, seeing as the above steps are repeated again and again: there are so many different user needs, and so many new information technologies appearing each year! (Callahan, 1991; Ogunsola, 2011)

A second area of action for the modern research librarian is partnering with the faculty in the research project life cycle. Ever since around 2008, due to their expertise in research data management, there has been an increasing need for librarians to support faculty during the research process. (Xia & Wang, 2014) Librarians can offer reference support, information literacy support, data stewardship and research data management services. These types of support are useful in that they save time and energy, allowing researchers an increase in productivity. Some suggest that this situation is pushing traditional librarians beyond a reader-oriented approach and into authorship-focused behaviours. (Tsang & Renaud, 2014) Similarly, Borgman maintains that "the processes associated with gathering or producing research data are a form of authorship, whether or not the researchers accept that view…" (Borgman, 2010) While the extent of responsibility of authorship is an interesting area for further investigation, it is beyond the purpose of this thesis.

However, it is important to notice how the discussion around the new roles of librarians goes beyond job responsibilities and into areas such as data stewardship, authorship and information curation.

Recommendations as to how librarians can manage their newly-found roles range from suggesting they need to adapt and become proactive in whatever the future throws at them (Callahan, 1991), to others saying it is the responsibility of the administration to provide professional development opportunities (Farkas, 2016), to recommendations to prepare to take responsibility for even more issues, such as Net Neutrality and copyright issues. (Dougherty, 2010)

Despite big changes in the past few decades, libraries today reside in the eye of the "technology and Big Data" storm; according to Dougherty, technology will steadily advance until 2025, when it will either stagnate or lead way towards the next big technology shift; to survive, libraries will need an actionable strategy plan (Popp, 2012) and librarians will be responsible for developing a library culture that is open towards even more changes.

2.4.1. The situation in Denmark

So how have librarians in the Nordic space reacted to all these changes? Ever since the 50s Nordic libraries have attracted international attention (Torstensson, 1993); the Danish library system has been perceived as enviable (Davies, 1976) due to its integrated ILL practices and high budgets, which allowed for creation of better services.

In order to look at how the libraries in Denmark changed over the past few decades, I propose a 'looking glass' approach, focusing on two decades of particular interest for this thesis: the 1990s and 2007-2017; the former in order to see how the Danish libraries prepared their strategies for the information society; the latter, as it captures what happened after the 2007 structural reform and goes (almost) up to the present times.

2.4.1.1. Library strategies during 1990s

From the 1990s and to the present time, there have been some big changes to library strategy, implemented by the Ministry of Culture and by the Danish government.

The reasons for these changes were legal revisions, adapting to emerging technologies and keeping up to date with the Ministry's new cultural policies. At the same time, the government was attempting to start a society-wide debate on the future of Denmark in the Information Society (this was sparked by a report called '*Info-samfundet år 2000*').

Several official reports highlighted a need to integrate new technologies into the libraries in order to support the democratic process in society. A report by Gitte Larsen (Larsen & Statens bibliotekstjeneste (Denmark), 1996) which takes a look at 50 of the 'larger research libraries' in Denmark presents the situation of Danish librarians in the year 1996: out of a number of 1323 total staff, 408 of those were librarians, and 209 of those were research librarians.

For example, the 1995 'From Vision to Action' report suggests that it is important to make sure the Danes don't get divided between the "information-rich" and the "information poor". They suggest making sure individuals "have access to equipment and [...] have

sufficient skills to use the advanced technology." (Larsen & Statens bibliotekstjeneste (Denmark), 1996, p. 16) This requires that the library turn into a workshop space where the local community has access to technical equipment and training for using it.

Related to this, the Danish IT Political Action Plan, established in 1995, involved an agreement between public library directors of Denmark that before the end of 1996, the main public libraries will have: access to networks, an Internet address, an increase in budget for non-printed (electronic) materials, public computers, and access to the Internet for the users. The action plan also recognizes a need for budgeting for the above areas, as well as for staff training towards multimedia and networks and establishing professional media acquisition advisory services that the library could use to consult before purchasing new media and networks.(Larsen & Statens bibliotekstjeneste (Denmark), 1996, p. 17)

Also, in 1995, the Ministry of Culture was exploring possibilities of expanding the dissemination of culture and knowledge into the digital environment: from this idea was born the "Cultural Network Denmark". The call for culture-related projects from cultural institutions including libraries again highlights the potential that the Danish Ministry of Culture perceived in the technological advancements; institutions such as The Royal Library and the University Library in Århus started bringing their data into the digital environment, and in this way, shaping the information seeking behaviour of the citizens.

Could it be that this early adoption of technology in Danish libraries shifted library users' perception of the library in such a way that even nowadays libraries are perceived as useful community spaces rather than terribly outdated? I do believe that the Ministry of Culture's early integration of technology into digital interactive spaces had something to do with it; through their actions, they not only updated their library systems to more modern ones, but also created more interactive ways for the public to approach the library space. In contrast with this, in many other European countries, the 90s were more of a time to update their integrated library systems than to experiment with new ways their users could benefit from the digital space.

In further support of my above opinion, I can point towards initiatives such as the 1995 "Committee on the Future of Libraries" (UBIS- an abbreviation of the Danish *Udvalget om Bibliotekerne i Informationssamfundet*), the 1996 Danish Board of Technology's initiative to discuss IT development and universal access, and the "Info-Society 2000" Action Plan suggested by the Danish government. (Larsen & Statens bibliotekstjeneste (Denmark), 1996, pp.14-20)

Finally, Denmark's preparations for the information age on an official level are reflected by the 1995 "National Information Policy – with special regard to libraries and documentation centres" and by the 1994 "New Public Libraries Act". In the National Information Policy it is highlighted that the academic (as well as public and special) libraries have a role to provide information for a wide range of purposes, not only pertaining to research but also to politics and other areas that might be of interest to their users.

The "New Public Libraries Act" mentions new information technology, stating that libraries "shall endeavour to make available computer programmes and other electronic

materials" and that basic services have to be provided free of charge; (Larsen & Statens bibliotekstjeneste (Denmark), 1996, p.22) there is no specific reference to the Internet, however the Ministry of Culture suggested that access to the Internet *was* included among the basic services, and should therefore be free. Regarding the research libraries, most of them offered Internet access and online searching services (at the end of 1995, 1 in 5 research libraries did not; this number probably rapidly went down in the following years)

The above all point towards devising a plan to evolve traditional librarianship and prepare it for the Information (and soon, Big Data) age. It is my opinion that an early adoption of these technologies positively modified the information behaviour and interactions of Danish library users, in such a way that it created a cycle of libraries being perceived as essential to the public structure, which in turn allowed for higher budgets and explorative initiatives around the new technologies at the beginning of the new millennium.

All of this, however, could not have been achieved without the existence of a good coordination and organisation of library systems nationwide. Initiatives such as the creation of the DanBib system (which created a joint library system for public and academic libraries), installation of integrated library systems ever since the late 1980s, and pooling all bibliographic records to create a nationally shared cataloguing system are examples of strategic moves that the Danish library system made in order to be properly equipped to deal with the upcoming Information age.

2.4.1.2. RSLIS curricula in the 90s

Education-wise, the Royal School of Librarianship (RSLIS) started re-evaluating its strategy in order to better prepare its alumni for the upcoming technological advances. Ever since the 1980s, the curriculum was adapted in order to prepare the students for a broader range of careers in the information field; To keep up to date with the changes, the school created a four-year curriculum which could be "modernized on a regular basis without altering the structure." (Pors, 1994)

Ever since 1985, an information technology-oriented course was part of the LIS undergraduate study curriculum. This course provided students with an opportunity to learn information retrieval, and acquire indexing knowledge through theory and problems; it also included subjects such as "expert systems, artificial intelligence, and database design." (Kajbergand & Ole Pors, 1995, p. 21)

In 1990, the school introduced a Master's programme in Library and Information Science. The 2 year full-time (or 4 year part-time) course allowed information specialists to further enhance their studies in the area, in order to qualify for work at an "advanced professional level in libraries and commercial forms [such as] information manager, head of information services, information analyst, information scientist, information consultant, etc." (Kajbergand & Ole Pors, 1995, p. 22) The Master's programme had courses which aimed to bring the necessary competencies to allow future practitioners to introduce and manage

information technology in an organisation, from an economic but also a computer-based angle.

In 1995, a new postgraduate programme called "Culture and mediation" was introduced, in partnership with the University of Odense. The programme allowed graduates to specialize in cultural mediation and pursue further opportunities in "cultural work, adult education, and mediation tasks in institutions, organizations and companies." (Kajbergand & Ole Pors, 1995, p. 23)

Halfway through the 90s, the Royal School of Librarianship also increased its internationalisation efforts which allowed for further developing of their students' experience by taking part in international seminars and exchange programmes.

Finally, the implementation of a PhD programme at the Royal School of Librarianship was the final step which allowed the school to be fully recognized as a university-based teaching and research institution, in 1998.

A second type of education was available for those who wished to be research librarians. It comprised of eight weeks of courses in librarianship; a base requirement was to prove specialization in a certain subject by having a degree from a university. Their tasks were mainly book selection, classification, and literature search; they were often assigned hours to independently do research or keep up with developments in their fields. This type of subject specialized librarian proved very useful for the universities. Particularly as the "module-based" courses approach became popular in Denmark, research librarians could provide necessary assistance in the absence of a set list of semester books.

Davis Roy points out changes in the educational system are reflected in the way students are using the library. Group work and projects make students focus on a certain area of study at a time. While this might create logistical problems if the library can't accommodate a sudden request for niche books, it also provides a perfect out-of-the-classroom opportunity for further skill-acquisition: the students might go to the library and ask for assistance in the form of lessons on statistics. (Davies, 1976, p. 236) This is an early example of how Danish libraries were used for acquiring data-savvy skills outside the classroom; it is also apparent that research librarians had an important role as educators; besides providing information seeking assistance, they were also teaching students the skills necessary to become more independent researchers.

2.4.2.1. From the 2007 reform up to the present

The current library law in Denmark is from 2000. One of the big changes from the previous library law is the emphasis on a "hybrid" library structure, which combines the traditional library space with a digital library; it "binds the libraries to offer internet and electronic data resources as well as printed books to its users." (Johannsen & Pors, 2010, p. 343) It was at this point that the previous strategic work done during the 90s such as the

Committee on the Future of Libraries and Info-Society 2000 finally materialized into a law that supported the concepts of digital space and technology as an information tool.

In this way, Danish libraries entered into the new millennium with a new updated law, as well a clear vision for the information age ahead.

The next major change which affected libraries was the 2007 structural reform, when 271 municipalities were reduced to 98. Less municipalities meant that some library units merged, and as they grew this made possible the expansion of services, offering better equipment and more specialized staff.

Providing digital services overall had grown in importance starting with the year 2000, and many users were welcoming and expecting a diversification in the services portfolio of the libraries, despite having a rather traditional view on the library as a institution.

The library users' behaviour shifted over time, through an effective interlibrary loan program as well as through modifying the proportion of books-to-other-materials available in libraries. Between 2000 and 2009, it is interesting to watch how collections of music and films grew, and the number of books slowly decreased. A look at loan statistics from that time shows that the number of loaned books doesn't fluctuate much through time (from 54.3 mil. in 2000 to 52.2 mil in 2009), yet there is a slow and steady increase in digital material loans, particularly of films (from 2.5 mil. loans in 2000 to 6.8 in 2009). (Johannsen & Pors, 2010) Furthermore, statistics of library visits show that Danish libraries successfully entered the knowledge age, revitalising their image and continuing to be attractive for their users. (Danish Agency for Libraries and Media, 2011)

The library as a space is intended for *learning* rather than *educating* in a formal sense. This creates alternative and informal lifelong education opportunities, which provides a complementary role to the education one receives in formal institutions. The physical features of the library space encourage self-learning and independent problem solving. Danish libraries have a high potential to support Danish societal development by providing learning opportunities which align with global goals towards an effective knowledge society.

As for the research libraries in particular, they have three interlinked roles which characterize and influence their development after the year 2000: becoming a learning centre, (supporting the educational institution by providing learning materials and tools), a knowledge centre (creating knowledge for the society as part of research groups), and holding an important role in society as a meta-knowledge institution (being responsible for organising and evaluating knowledge). To understand the appearance and development of these roles, it is necessary to look at how the Danish research libraries are situated from a governmental perspective.

Danish research libraries are situated under two ministries: the Ministry of Culture, on one side, which is responsible for the library system nation-wide, and the Ministry of

Science, Technology and Innovation, which is responsible for the universities and for some of the libraries' budgets. As a result of their dual characteristics, research libraries are influenced by both the government's decisions (i.e. the 2007 reform) and by changes in the educational system. This creates additional challenges for research libraries, particularly regarding to their identity and division of roles.

While the necessary skills for research librarians are up to debate and vary between the types of libraries, Danish research librarians' roles can be divided into two groups: there are librarians that focus on their educational role, and serve as a generalized information professional in their collaboration with students; the second group focuses on their role as domain specific knowledge co-creator; in Denmark, this second group has its roots in the 80s, when this type of dual specialization of librarian and subject expert became increasingly common in university libraries, as mentioned in my previous "looking-glass" subchapter.

A 2009 report on the future of Danish research libraries pointed out that technological advancements and the resulting change of user needs are two strong drivers for innovation in today's research libraries. In the near future, larger research libraries which have strong ties with or are part of universities, "may well need to develop into a learning centre." (DEFF, 2009, p. 11) Furthermore, there is a tendency in all developed countries for the research system and research education to expand as it becomes more central to the economy; in the new millennium, there's a strong trust that digital technologies will help us achieve that. (DEFF, 2009, p.10) In order to take full advantage of those technologies, users and librarians alike have to acquire the abilities necessary to work with increasingly vast and complex informational structures.

Starting with 2000, libraries started analysing what could be done to attract non-users (Danish Agency for Libraries and Media, 2011); for research libraries, this meant acknowledging their multi-layered aspect and the different types of users. Typical user groups for a research library might be bachelor students, graduate students, doctoral students, researchers and faculty; technological advances have made it apparent that some persons are more directly involved with the library (physically show up to borrow books or to participate in research activities), whereas other might be indirect users (not aware as to the library being involved when they access digital resources). Making this distinction between different types of users allows librarians to appropriately modify their approach of interacting.

Due to the large amount of data available today, this approach of *teaching* rather than *showing* might end up increasing the efficiency of research libraries and of their users. *In* 2011, the Danish Agency for Libraries and Media reinforced their recommendations regarding the role of libraries in the knowledge society. It is emphasized that libraries are not only a mediator of books and research materials, but they also mediate IT tools and quality data collections that citizens can use to build their own value-creating knowledge products and become independent problem solvers in the society. They stress that "even though the Internet is a rich source of information, [it has become] absolutely essential for an increasing number of the population to be able to use IT to work with information." (Danish Agency for Libraries and Media, 2011)

So is there enough to have just one *data educator* librarian per library team?

An observation of the upcoming roles of research librarians in 2009 revealed four emerging professional identities of research librarians: "archive detective, educational sparring partner, generic knowledge expert, and co-creator." (DEFF, 2009, p. 29) These roles are a mixed consequence of, on one side, research librarians being more qualified to assume an educator role, and on the other side, research libraries adopting a value-creating (rather than knowledge storing) capacity. With the importance that data holds for our society today, it makes sense that all of the roles above will benefit from having awareness and knowledge of good data tools. While some librarians will become specialized in this area, the rest of the team is still involved with data in one way or another – and an understanding of it can do nothing worse than improve their overall librarian competencies.

Due to that, it is only natural for Danish libraries to further evolve their IT instruction services towards areas that are of interest for their users and for society.

2.4.2.2. As IVA turns to INF - Danish LIS education moves towards information

It was Niels Ole Pors's vision that for librarians to successfully adapt to their new roles in the information age, it is the responsibility of the university to build a curriculum focused on the following four areas: analytical capabilities, decision-making skills, technological skills, and communication capabilities. (Pors, 1994, p.19)

I will further summarize the changes in the Danish LIS school system, from 2008 up to 2018.

In the past eighteen years, the Danish librarianship school repositioned itself in accordance with the changes in the library system; of importance to the curriculum was the Libraries Act from 2000; the internationalization efforts started in the 90s continued, and the school became well-known among LIS educational institutions in Europe.

In 2008, the Bachelor and Master's programmes were revised, and introduced problem-based learning and a business orientation, in order to "fulfil society's requirements for new central competences within knowledge, culture and information operations." (IVA, 2010b)

In 2010, the Royal School of Librarianship (in Danish, *Danmarks Biblioteksskole*) changed its name to The Royal School of Library and Information Science (*Det Informationsvidenskabelige Akademi*). The curriculum had been modified in accordance to the "expanded libraries" concept found in the Libraries Act of 2000, but still included "the core competences, knowledge organisation and information retrieval." (IVA, 2010b) In

addition to these core competences, the school was working on providing more international exchange opportunities to its students, and was improving its courses in English.

IVA faculty was working to "prioritise interdisciplinary competence development initiatives with the purpose of knowledge-sharing, learning and cooperation across the organisation, including development of the School's IT competences across the board;" (IVA, 2010a) Other areas of priority were an "increased focus on higher education and employment" and children and culture.

Until 2013, IVA was an independent institution. After that time, it was merged into the University of Copenhagen, becoming part of the Humanities faculty. In 2017, IVA changed its name to the Department of Information studies (INF). Today, it is a "humanistic, cultural, and social based iSchool." (Department of Information Studies, 2017)

While the focus of the school remains on "systems, platforms, and organization that organize, rank, sort, seek, recommend, and communicate information", the education is approached with digital systems in mind. The INF educational programmes are designed to "provide students with the required skills and knowledge to analyze and design today's digital information and communication technologies, with a special focus on the cultural and social contexts of those technologies." (Department of Information Studies, 2017)

It is clear from these changes that as the field of Danish librarianship shifted towards the knowledge age, the LIS education followed accordingly. Internationalization provided staff and researchers with opportunities to connect with librarians worldwide, and allowed them the exposure necessary to modernize the Danish LIS studies. Today's LIS education at the Department of Information Studies gives future information professionals the knowledge and competences necessary to successfully contribute to the development of the Danish library system.

2.5 Data science in libraries; competencies

Questions: What does data science mean in a library context? How is data science used in a library? What competencies should a librarian have, if they wish to work with data science?

In order to answer the above questions, I'd like to point out two of the new roles of libraries, presented in the previous chapter. As mentioned, libraries have important emerging roles as learning and knowledge centres in the knowledge society. The diversification of a research library's roles comes naturally over time, due to changes in the educational system, the user's needs and expectations, and advancements of society, technology and research.

It is only natural that LIS students want to follow an education that is in line with the current professional standards of practice, and the most attractive librarianship programmes are often those that promise the students will gain applicable job skills by the end of their education. In Denmark, we have the example of the Department of Information Studies. Each revision of the curriculum and strategy of a LIS school reflects "the changes that were the result of the maturing of both the profession and library education." (Stern, 2009, p. 3) The Department of Information Studies is working towards increasing the students' technical competencies in the areas of information and culture; however, it remains an information-oriented school.

Employers have become more and more interested in technically skilled employees, and with each passing year the expected technical level of information schools' graduates has been rising. Because of the rapid technological advances, it is very difficult to understand which technical competencies are the most useful for a librarian to have. This creates a difficult position for library and iSchools when deciding the best educational structure; the students are responsible for exploring their own path, and acquiring transferable skills that they can use in their future careers. (Stern, 2009, p.3)

In other words, while information schools are aware that students might want to acquire more data-savvy skills, and students themselves suppose that such skills will help them in their profession, the only place left to go in order to understand how data competencies are applied in the library, is in the library itself; and in order to understand that, it's necessary to see what data science means in a library context.

2.5.1. What does data science mean when it is placed in a library context?

As established in the previous chapters, technology has changed the way we store and interact with information in the past few decades.

The Internet has made it easy for people to access information from their homes and devices, without needing an information specialist as an intermediary. Specialized software can be used to collect research data, and technological advancements mean that data storage is now cheaper than ever. (Writers, 2012)

Cheap and accessible data storage combined with the use of data as an economic commodity has created a problem. While it is relatively easy to create and store data, this process creates datasets so large that they are impossible to process using regular database software. This type of data is defined as "Big Data." (Gordon-Murnane, 2012) Both public and private sectors are hoping that Big Data will offer them the insights needed to become more efficient and innovative in their work.

Larger amounts of data come with extra challenges when it comes to storing, processing and analysing it:

Among the difficulties that people have when working with large amounts of data are "poorly organized collections; poor search tools and lack of accessibility and findability of internal data sets; lack of awareness of available third-party data sets; and copyright and intellectual property issues." (Gordon-Murnane, 2012, p. 32) The same can be said for

researchers: there are concerns that a large amount of research data is improperly managed, which leads to loss of knowledge; (Gordon-Murnane, 2012, p. 33)

Furthermore, in the past few years there has been an increased capacity for jobs created around Big Data and the concept of data science; (Kim, 2016) the problem is, the definitions around these terms can be quite fuzzy for an employer, and finding the perfect match can be a difficult task. For example, programmers can have very high skills in programming with data, however they miss the scientific approach needed to formulate the right questions; similarly, statisticians have that scientific approach, but they might have weak programming skills and are often missing information science skills. So from an employer's perspective, where should they look to find a specialist who can best manage their data?

With the emerging phenomenon of Big Data, it is clear that there is a gap between skills and specialists in this area. But why should it be considered an opportunity for librarians to get involved in?

While companies have until now tried to look towards individuals such as programmers and other IT specialists to solve their data problems, libraries have been providing similar roles for researchers ever since the 1960s. (Kim, 2016, p. 162)

Data curation services, data stewardship, data management plans, and data repository services are not foreign terms in academic libraries; libraries have been assisting researchers with their data management for some time, and they are in an appropriate place to do so: "Librarians have increasingly become experts in data management because of their combined knowledge of new data sharing standards, information science, and the Semantic Web [...] information literacy has always been a topic of interest to research librarians, and it is natural that their role is expanding to include topics surrounding data curation and access." (Haendel, Vasilevsky, & Wirz, 2012)

The fact that these types of services have been successfully used within libraries to get the most out of their research data shows that libraries as an organization can fulfil this role; As research data grows and becomes more numerous and complex for a research library, data science might be a logical next step when trying to decide which new services to offer.

I believe data science can provide libraries with a new role in today's society. If advertised appropriately, data science as a service can attract a wide variety of persons and companies; if librarians can demonstrate their value in building and interacting with a wide array of data products, it could boost libraries' image as an information expert in our society.

It is also a good idea to introduce data science in a library context in order to boost the librarian's data competences and create lifelong learning opportunities for professionals already established in their fields: "As data science techniques and tools for extracting, manipulating, analyzing, and visualizing data are becoming increasingly important to all fields of scholarship, competency in employing such techniques and tools is needed for librarians and information professionals." (Kim, 2016, p. 163) Introducing data science

competence-building opportunities can provide them with a way to be better prepared for dealing with Big Data, while simultaneously filling the "data workforce gap."

2.5.2. How is data science used in a library?

Considering the above, how can data science be used in a library? Firstly, to sum up the position of librarians in relation to data: librarians have a unique position in that they are professionals whose skills bring solutions to some of the problems we face in today's "Big Data" world. They "facilitate and enable data discovery and retrieval; add value to the data through cataloguing, indexing, and metadata; and [...] "provide for re-use [of data] over time through activities including authentication, archiving, management, preservation, and representation." (Gordon-Murnane, 2012, p. 33) These are skills that librarians have been using to work with printed materials for decades. Furthermore, they do "not only educate the community on data and information literacy, but conduct their own research on how the scientific community can best rise to the data challenge." (Haendel et al., 2012)

Data science can be applied in a library setting in three areas: as an educational tool, towards the students and faculty; as part of the research process, e.g. a service to help the researchers; and to improve the library collections and systems. That is because "data science exists more or less on a spectrum, and depends on an institution's size and mission." (Burton, Lyon, Erdmann, & Tijerina, 2018, p. 7)

As an educational tool, data science could be taught by librarians towards students in a similar way to how information literacy is taught; in "The Accidental Data Scientist" (Affelt, 2015, p. 199), the author analysed data science job postings for information specialists and librarians; among her findings, it is apparent that some US libraries are recognizing the potential usefulness of a librarian that could provide users with aid in this area; for example, one university was seeking a "social sciences data librarian" to be responsible for providing the social sciences department with "data and statistical support, information literacy and research instruction, individual consultations, and collection development."

As part of the research process, librarians might use data science in collaboration with researchers. For example, Affelt mentions the responsibilities that have been assigned to a "Data Services Librarian" at the University of Iowa: "creating, editing, and manipulating metadata, assisting researchers developing data management plans, and initiating outreach and marketing of data services to potential users." (Affelt, 2015, p. 199)

When working with researchers, the role of the librarian is to help researchers transform and manage their data in the shape most suited to their project; although

researchers understand on a certain level what his work comprises of, they will not typically acquire data science skills themselves as part of the interaction. The librarian will trust their domain expertise and will use his skills to find the best way to work with the given data.

When working with students, the role of the librarian is to *help them learn to* transform and manage their data in the shape most suited to the project; this gives the librarian an instructional role. The librarian's capacity to educate the students will be very important in this case; It provides an opportunity to teach students to think and problem-solve independently, in an informal setting. Students should use this opportunity to get familiar with the processes data goes through and improve their data science competencies. They might also request individual consultations, which can be a further learning opportunity; collaboration between professors and librarians can be another way to make sure students acquire data skills outside the classroom; For instance, professors can give assignments which require students to be in close collaboration with the data librarian.

The degree to which the librarian is responsible for students' acquisition of data skills is debatable. (Kim, 2016) Some consider it is not a librarian's place to do that, and that it should be part of the curriculum. Others consider that librarians are in an excellent position to extend this kind of help to students, even if it might be time-consuming. I agree with the latter, as offering students access to instruction from a data-savvy librarian provides them with an excellent opportunity to expand their data skills outside of the classroom.

The third type of use a library might find for data science would be towards improving the library's own collections and systems. As this thesis is focused on librarians using data science rather than on the library as an organization, I will not go too much into detail. It is worth mentioning that there is a lot of potential in applying data science towards improving library systems. Many library systems are already successfully using algorithms to facilitate discovery. (Morawski, Stepan, Dick, & Miller, 2017) Other areas in which data science has proven useful are in improving reference services and helping with citation analysis. (Affelt, 2015, p. 205) A library could also choose to view its collections as data, and may choose to use a business intelligence approach to their library data to facilitate decisions around the "selection, purchasing, circulation, location, development, preservation, and disposal of stock items."(Burton, Lyon, Erdmann, & Tijerina, 2018, p. 13)

The application of data science in a library might enhance the values that it brings to a community, by rebranding the way people perceive it. Libraries have had difficulties in convincing the public of their value in the digital age. Since "data science" has become a trendy term, offering such services inside a library might convince people of the value libraries can add and even attract new types of users. (Burton, Lyon, Erdmann, & Tijerina, 2018, p. 11)

To sum it up, I have identified above two ways in which librarians could use data science in a library. They could take on a role as an educator, helping students acquire the

skills they need to complete their data projects; they could also have a role in helping faculty researchers with their projects, in this way offering a service similar to a data curation or data management service.

2.5.3. Areas in which librarians' skills could be expanded

If applying their skills to Big Data, librarians could extend their roles in a direction that could help optimize the research data lifecycle, or at least, make it more manageable to the researchers. But which skills should librarians expand in order to best demonstrate their value to the community?

I will further review the list of general "data science skills and competencies" and discuss them from a librarian perspective.

When discussing data science, obviously the digital environment and data are important. So to start with, Stanton's aforementioned notion of "four A's" of data are very relevant in a library setting. (Stanton, 2012) Data architecture, data acquisition, data analysis and data archiving are terms that librarians are familiar with due to their LIS education background.

Data architecture in a library setting can be related to the competencies that librarians have when developing data management plans. It is useful for the librarian to have knowledge of a database structure, and experience in how to create and maintain it. (Choi & Rasmussen, 2009, p. 462) This involves an understanding of the dataset, knowledge organization skills, and a certain level of know-how around database software.

Data acquisition and data analysis are part of statistical work competencies that data librarians should have. Librarians' statistics skills are not required to be equally advanced to a statisticians', but as evidence-based librarianship becomes more important to libraries, (Stern, 2009, p. 134) data librarians should have an understanding of and basic skills in statistical methods. Librarians can also apply their information seeking knowledge in order to find relevant materials and reliable sources in the data acquisition stage of a project.

A data librarian can demonstrate strength in this area by applying reference desk good practices and offering more than just a summary of resources: users will feel empowered if given extra materials to go along with their research – such as a bibliographic list of sources, abstracts or full text access to the resources. This is a little effort for the librarian compared to how much value the user will ultimately get out of it. (Hunt & Grossman, 2013, p. 71)

Data archiving practices might be more familiar to information specialists working in areas such as records management, however it is one of the skills which LIS graduates are expected to have; familiarity in this area of information science will mean that a librarian can

transfer this knowledge into understanding how to access and pull differently formatted data from various types of archives, using data science methods and tools.

As can be seen above, there are areas of librarianship that are related to Stanton's "four A's" of data. Most librarians (not only data librarians) are familiar with the above terms and how they relate to a library setting. This constitutes further proof as to why librarians are in a good position to act as data scientists within the library.

While librarians are familiar with the above notions, there is room for improvement if a librarian wishes to get more involved with data science specifically. In the following paragraphs, I will take a look at how the skills and competencies suggested by the DDaT (Digital, Data, and Technology) framework relate to a research librarian's work, and why a librarian might want to acquire some of them (and refresh his knowledge of others). A distinction will be made between three different "levels" of using these skills, which I will later develop into an important reasoning within the thesis.

Applied maths, statistics and scientific practice skills: as discussed above, these skills have useful applications in a library setting. On the most basic level, knowledge of simple statistical concepts can increase a librarian's capacity to select good materials for their users by enhancing their ability to read and review statistically developed results. (Stern, 2009, p. 152) Another level of statistics skills a librarian might want to reach if running a data service would be being knowledgeable of the different methods and tools available up to a point that allows him to provide statistical instruction for his users. If a librarian acquires further skills up to a level that can allow him the ability to create statistical models, this gives them the capacity to create data products for the library itself.

Data engineering and data manipulation: these categories of skills are traditionally used in data science for "architecting distributed systems, creating reliable pipelines, combining data sources, architecting data stores and [...] building the right solutions for [data science teams]." (Dataquest, 2017) It is a database-centric role, and requires having knowledge of systems architecture and how to obtain, scrub and model data in the most efficient way. For a librarian, knowledge of information architecture and an awareness of how to build the right solutions for an informational need are part of their job. This can make acquiring a data engineering-oriented mind-set easier, as librarians already possess information-oriented thinking and the curiosity to find the best solution to the information question at hand. Data librarians can use data engineering and manipulation methods in their data work to clean, transform and prepare the data for their users' data projects. An advanced knowledge of data engineering and data manipulation would allow a data librarian to work with the library systems and collections.

Data science innovation skills: a librarian's sense of curiosity and capacity to identify areas where innovation would bring value to a project can be described as "data science innovation skills." In a practical sense, this could be illustrated as a data librarian's wish to explore the ways in which data science tools and methods might be used in resourceful ways to help users get more out of their data projects.

It can mean looking further than the library-provided resources and into open-source software that can be used for the data projects. A high level of data science innovation skills might describe the skills of a librarian that has had a lot of experience with data science tools and methods, and is confident in his ability to provide creative and valuable extemporaneous suggestions for improving data projects externally (researchers) or internally (library management).

Developing data science capability: at the least, it is a data scientist's –and an aspiring data science librarian's- responsibility to continuously develop both his skills and his colleagues' awareness of data science possibilities. This can lead to a tighter, more data-oriented collaboration between library departments; it might also inspire other library staff to borrow data science methods to use in their own tasks; overall, if a librarian has the interpersonal and instructional skills to make his colleagues more data-savvy, it will inspire an environment of continuous professional development inside the library, which in the context of rapidly evolving technology and Big Data can be considered a good environment to be a part of.

Domain expertise: It is Blei's (Blei & Smyth, 2017, p. 8690) suggestion that a data scientist and a domain expert can be the same person fulfilling different roles, *as long as* he has knowledge of existing theories in the domain, understands the data and the main goals of the discipline, and possesses the necessary computational and statistics knowledge to be able to explore the data in a way that serves the aims of the discipline.

In a library setting, this is an useful skill to have when collaborating with users on their projects; however, exactly how much domain expertise does a librarian need to do data science?

In the ideal world, a research library will have a good mix of domain specialists with different backgrounds that are also knowledgeable at working with data and data science methods. In the real world, limited library resources (staff, time, budgets) mean libraries have a focus in just one or two areas. Another challenge would relate to the librarian's job duties itself: it can take energy to be a good data librarian, up to date with the latest advancements in several domains, while at the same time being occupied with numerous and different user requests. It may be more efficient for a library to teach a generalist librarian how to be data-savvy; in this way, he is able to understand and assist the users, and point them towards more comprehensive data services if needed.

Data science oriented programming: combining domain expertise with programming skills can bring out the best in a data librarian's skills. A librarian with an awareness of different programming languages and software can help guide users towards a method best suited to their project. A data librarian that decides to acquire programming skills himself can prove to be a very valuable both as an instructor and as a developer. A librarian is not supposed to achieve the same level of programming skills that a programmer might have. A good starting point might be to acquire a working knowledge of programming languages and scripting languages. (Choi & Rasmussen, 2009) Luckily, there have been initiatives

developed specifically for librarians interested in data science, and they come in a variety of shapes, from workshops (such as "Library Carpentry", "Data Science and Visualization Institute", "Data Scientist Training for Librarians") to online courses. Although it is not usually part of LIS curriculum, programming is very useful in a library setting, as it allows librarians "to take an active role in the technical aspects of library projects." (Lauersen, 2016)

Understanding of data science analysis lifecycle: acquiring an understanding of the data analysis lifecycle helps librarians get closer to "speaking the researchers' language". Programs such as the ones mentioned above help librarians get first-hand experience with the data analysis lifecycle, and so create an increased understanding of researchers' needs. (Burton et al., 2018) In a practical sense, having a understanding of the data science analysis lifecycle will allow a librarian to easily pick which methods or tools are appropriate throughout the duration of the project.

Since data projects have an iterative character, it is important for a librarian to be aware that certain steps of the project might be repeated before the project is finished; This style of working falls under the Agile (Crowder & Friess, 2015) project management style, which is characterized by short cycles of development followed by getting feedback from the team.

All in all, librarians can be a useful ally in a variety of data projects due to their ability to transfer their research and reference skills to a variety of areas and subjects. In this sense, Affelt (Affelt, 2015, p. 191) claims that just like "lone wolf data scientists" can transfer their skills to a variety of subjects, librarians are also good at working with unfamiliar subjects; their professional background gives them the communication and teamwork skills and the curiosity necessary to approach a variety of data questions.

As can be seen above, there are different levels of data science competencies, according to the different roles librarians have when working with data. For example, the first level of "applied maths, statistics and scientific practice skills" would allow a librarian familiarity around statistics and scientific practices, enough to be able to select the appropriate materials and software for them to work with. In other words, acquiring this level of competencies would "help the librarian help the users help themselves." The second level would allow a librarian to have a working knowledge of data, maths and statistics up to a point where he can be directly involved in instructing the users. The third level would be a librarian whose technical, maths, statistics and scientific practice skills are advanced in a way that allows him to be directly involved in the library and researchers' projects.

This can be compared to, for example, the competences that different types of librarians have around history. Some librarians know enough about this topic to be able to refer users to the right materials. Other librarians are subject specialists in history, and their knowledge and expertise in the field is counted on by researchers and users. And on another higher level, some librarians are directly involved in the library's projects around historical collections.

Because of that, I would like to make the distinction that there are three "levels" of librarianship in connection with data and data science.

The first level - 'data-savvy librarians'; that is, librarians that know *of* data. They might be familiar with the concepts and tools around data science and the data lifecycle, enough so that they can guide users on how to acquire data science capabilities to use in their research projects.

The second level - 'data librarians'; that characterizes librarians whose subject is data; this might mean being involved in a research data management service, or otherwise being 'subject specialists' in data, as well as providing instruction and support for the users in areas related to data science.

The third level - 'data science librarians'. This would characterize highly technically skilled librarians that work with data science within the library, e.g. by making data science products for the library's collections.

It is not to say that any of the levels is "better" than the other, however librarians with the second level have both the competencies of the first level, and then their own set of competencies on top of that. These different levels of competencies mean that data-savvy librarians and data science librarians don't have the same professional interests, and that should be taken into account when assigning them within a library. For example, a data science librarian would not be challenged enough by the same tasks that a data librarian might enjoy, such as providing instruction to students.

It is clear from the above that there are different levels of competencies which librarians might use when working with data in a library. When comparing these competencies with the DDaT's suggested competencies for data scientists, it is quite clear that the librarian profession has a strong argument in favour of introducing data science to libraries.

Firstly, as the term "competency" refers to "having the requisite ability or quality to fulfill a particular need and [...] demonstrate it effectively and consistently" (Stern, 2009, p. 5), this creates several different areas of discussion when applying it to librarians. In other words, required competencies differ from one library department to another, and which competencies are necessary for a specific library job also depends on the person who wrote the job description. Because of that, it is often necessary to restrict the definition of competencies, according to the context of the discussion. In this thesis, I will refer to competencies needed by science librarians. Stern (Stern, 2009, p. 7) highlights several areas of competency that are important for science librarians: communication and collaboration competencies, financial and management competencies, and core library-oriented competencies (he refers to competencies such as knowledge of the publishing industry, knowledge of library processes and preservation and evaluation of collections).

The Special Libraries Association recommends an additional competency, namely the ability to "contribute to the knowledge base of the profession by sharing best practices and experiences, and [continuing] to learn about information products, services and management

practices throughout the life of his/her career." (Stern, 2009, p. 14) It is important for librarians of any kind to be inclined towards lifelong learning, but especially important for research librarians, as this makes them more inclined to be up to date with the latest advancements in their (and their users') fields.

Table A in the appendix unites the DDaT framework (Digital Data and Technology Profession Department, 2018) competencies with suggested competencies for research librarians (Stern, 2009) and with Affelt's competency for a librarian that is interested in working with data science. (Affelt, 2015, pp. 191-193) The first column presents the competency area being discussed; the second column contains competency areas specific to data scientists; the third column contains competency areas specific to research librarian; the third column contains competency areas necessary for a librarian working with data science. I will refer to that type of librarian as a "data science librarian." The purpose of this table is to show that the areas of competency necessary for good performance in each of the three professions are similar.

As can be seen in the table, when discussing the roles of data scientists in relation to research librarians, there are certain areas where the competencies necessary for both professions intersect. And by listing Affelt's suggestions for areas of competency necessary for a data science librarian, one can get an idea of the competencies required for a librarian that works with data science.

For example, for a data scientist, 'innovation competencies' might mean being open to change and seeking constant improvement in one's own methods; this is also true for a research librarian, that has to have a flair for both self-learning along his career, and for making sure the library is up to date with the technologies and tools researchers might need. A data science librarian will need to combine both of the above, which will result in being competent in the areas of professional self-improvement and inspiring a sense of innovation in his library and with his users.

Similar conclusions can be reached by looking at the other areas of the table. While the competencies necessary can be categorized together under similar areas, they are not exactly the same. For example, it is unrealistic to expect that a librarian working with data science would be as good at programming as a full-time data scientist. Also, while a data scientist will be required to collaborate with the data science team, a librarian is responsible for reaching out beyond his immediate library colleagues – often with the purpose of educating them on data science matters.

From that, it is apparent that a librarian's role is compatible with acquiring data scientist competences; he will require more involvement and extra training if he wants to work with data science in the library.

2.5.4. Librarian's approach

But is it really as simple as "data science knowledge + librarian = data science librarian"? Below I will discuss why the role of data science librarian comprises of more than just the skills from both areas, added up.

Apart from looking the skills and competencies necessary, I believe it is necessary to understand the approach that librarians should have, when discussing data science with their users.

The interaction between user (be it researcher or student) and the librarian should be focused on how a librarian can bring value to the user. In this sense, Affelt thinks it is important that the user feels like the product delivered by a librarian goes beyond what he would get from a Google search. (Affelt, 2015, p. 69) In order to do that, the librarian has to have a good understanding of what the user really needs. In this way, he can apply his (data-savvy) knowledge and methods while working with the user's (project-specific) domain knowledge, and collaborate together to create a pleasing final product.

Other ways in which a librarian can bring value to the user are similar to the traditional roles of a reference librarian: going the extra mile and compiling a list of "best data sources" for a user who feels overwhelmed by all the sources available will undoubtedly be appreciated; by using this approach, the librarian could make his Big Data authority and his information seeking skills work together in favour of the user.

As presented in chapter 2.3, librarians' capability to be educators puts them in a good position to teach their users to work with data science. As previously mentioned, one category of users who would be most interested in picking up data science skills would be students. McKnight supports Kulthau's observation that a librarian should have an awareness of students' information search process. In this way, they can adapt their approach depending on which part of the information seeking process the students are in: "Initiation, Selection, Exploration, Formulation, Collection, [or] Presentation." (McKnight, 2010, p. 27) This informal opportunity for educating students requires the librarian to also have high communication and collaborative skills. (Stern, 2009, p. 21) In addition to that, librarians can take advantage of a library's physical learning space, and turn part of it into a "data science learning space." (Burton et al., 2018) This learning space comes under different names, but universities are increasingly adopting them as part of their libraries.

Data labs and digital labs are two examples of names given to physical spaces where users can increase their skills in various digital tools and technologies and work on their own projects, using the library's equipment and with optional assistance and instruction provided by a librarian. Goodman believes the importance of such as space lies in the fact that they are more than just physical areas with equipment – she believes that "while any community center may provide equipment, a library can add value by providing knowledge and expertise in using the software." (Goodman, 2014) While they don't necessarily focus just on data science, they are great examples of how libraries can provide an opportunity for digital skills acquisition while at the same time empowering the users to turn into content creators.

A librarian working with students and faculty in such a space, in a research library is in a good position to convince the users to start working with data science in their projects; the ability to combine his information seeking skills with technical knowledge, and students and researchers' domain knowledge, together with his instructional capacity give him the leverage needed to increase his users' data skills. In doing so, librarians can empower students and researchers to look beyond their library's available materials and bring their own insights of the world that surrounds them.

2.6 Data librarian as a profession

Questions: How does the term "competencies" relate with "profession"? Is data librarianship a profession? How do data librarians' competencies influence the placement of their profession within the library?

In the previous chapter I have explored the way in which data science can be used in a library, and the skills and competencies necessary for a librarian to have if he decides to work with data science. But what kind of librarian is the most suited to do so? I believe that a librarian working with data science would be called a "data librarian" inside a research library. Yet undoubtedly working with data science will add new roles to their profession. Because of that, I would like to discuss how the profession of data librarian influences the competencies needed. Finally, I will take a look at the placement of a data librarian in a research library, and discuss how his competencies relate to that. After exploring that, it will be clearer how the position of data librarian fits in with a data science approach, in a research library context.

Data is an important part of today's knowledge society. Inside the research world, most of the research data generated nowadays is digital; this makes the processes of storing and sharing data easier than in the case of traditional formats, but it creates the following problem: research data is only as good as the way it is managed (Edwards, 2017, p. 18) Without appropriate description, management and preservation, valuable research data can be lost. This gives librarians an important role in the research data lifecycle, a role that is complementary to the researchers' view of their research data. "Whereas researchers may see data in terms of their specific research objectives, librarians necessarily focus on the wider context of the broader research and scholarly record." (Edwards, 2017, p. 19)

Librarians as a profession have skills that are highly transferable when working with data. They already have the knowledge of how to organise information and make it easy to find; they already teach information literacy to students, so they have a bit of experience as instructors also; they have experience of running publication repositories, which gives them skills that can be transferred into managing data repositories; they have a working collaboration with faculty departments and publishers; their relationship with researchers puts them in the right position to offer data support as well; and lastly, librarians' advocacy

towards open access to publications could be extended to the area of research data, as well. (Ball, 2013, p. 3)

In a research library, these skills make librarians well-positioned to help researchers by providing data support services.

Traditionally, it was subject librarians' duty to assist with data related to their subject; due to more and more data moving into the digital space, it has become more common to have an independent librarian focused on providing research data assistance. (Xia & Wang, 2014)

The information specialist engaged in assisting researchers with their research data is called a data librarian. These specialists have the background knowledge of a librarian, and they run services which can help researchers and students get the most out of their data. While a librarian might reference a researcher (or library user) towards a resource or department (such as IT support) which can help them with their data, data librarians go a step further and become close collaborators and directly involved in researchers' work. Some of the areas in which they are active are data management, metadata management, data analysis and literacy support, and developing data management materials, checklists or training sessions on data analysis tools. (Australian National Data Service, 2013)

Also called data librarianship, this professional area has grown together with the increased complexity of modern research data. (Xia & Wang, 2014, p. 362) Because of that, the job description and requirements have been in a state of constant change since the beginning of this century. Libraries and librarians alike have attempted to experiment with different types of services and approaches (Martin, 2013; Molaro & White, 2015; Svendsen & Lauersen, 2015) in different research libraries in order to find a suitable adaptation of the role of data librarianship within the research environment.

As data librarians are reshaping their role in accordance with the university's data needs, they have moulded themselves into new roles. Ball suggests that the modern data librarian (Ball, 2013, pp. 3–4) needs to be capable of the following: they need to act as a "data detective" to uncover and centralize research data that's been scattered across the researchers' and departments' hard drives; they need to become consultants in order to find areas that researchers need improved; they need to be good negotiators in the relationship with managers and library directors, and be able to prove their contribution to the researchers' work; they need to become advisors and take initiatives in helping create a data policy for the university; they need to adapt their skills to the university and researchers' specific types of data; they need to be able to become instructors and support and train researchers on different data topics.

Using data science within a research library can give librarians the up-to-date tools and methods they need to work more efficiently with researchers' data. A data-science mind-set offers a data librarian additional competencies which can help with several of the above roles. For instance, acquiring data science programming skills might make them better at working with large amounts of differently formatted research data.

As previously mentioned, librarians already have working experience in several of the above areas, which puts them in a good position to specialize themselves as "data librarians",

should they wish to. However, taking into account Ball's suggested roles for data librarians, it also seems like data librarians have to divide themselves into quite a lot of different roles in order to successfully support the faculty.

I propose looking closer into this aspect, by discussing how data librarians' roles relate to their profession. It is my belief that the reason data librarianship is still a blurry area is due to the librarians' wish to mould into an "ideal helper". This is unrealistic due to the diversity of users and their requests towards librarians. What should be a middle way?

Should data librarians restrict and redefine their profession and services in a more rigid way, causing researchers to only come to them with very specific, clearly-defined problems? Wouldn't this cause the "data curation and management" type services to decrease in popularity because of a perceived rigidity on the side of library services?

Or rather, should data librarians continue to indefinitely adapt to the researchers' needs? This might be perceived as useful and it might gain researchers' trust, but wouldn't it create an ultimately unstable environment for data librarianship, filled with confusion around topics such as data authorship, or bring further ambiguity around skills and competencies needed by a data librarian, and so forth? (Tsang & Renaud, 2014)

Like the chicken and egg debate, I believe it is an interesting question to explore. Living in an age where data is becoming increasingly important for decision-making is putting libraries in a position where they chose to redefine their profession with the aim of keeping up with the times and technologies; while they are free to create attractive data services inside their libraries, their required job skills and competencies are very much dependent on their clients.

So while it might be correct to say that "data librarians are innovating and creating their own data services inside their libraries", it is equally fair to say that "the profile of data services inside the libraries is greatly inspired by the "data skills" gap in the outside world."

In order to understand the relationship between areas of data competence in librarians and the data librarian profession itself, I will first clarify the usage of certain terms within the scope of this discussion.

When referring to the "librarian profession" or "data librarian profession", I am using the term "profession" in accordance with McKnight[3] (McKnight, 2010, p. 5),: "An occupation in which a professional knowledge of some subject, field, or science is applied; a vocation or career, especially one that involves prolonged training and a formal qualification."

The above definition is in accordance with Winter's view (Winter, 1983), that librarianship is a profession as it possesses the six following traits: "professional association, formal education, a theoretical and practical body of knowledge, ethics codes for professional practice, service orientation, and community recognition."

[3] Taken by McKnight from the Oxford English dictionary, a version no longer accessible, which I could not verify

Out of those traits, "theoretical and practical body of knowledge", "ethics codes", and "service orientation" are traits that can be discussed from the perspective of librarians' competencies.

When mentioning competencies, I am using the term in accordance with J. Field's definition: "*Competency* refers to having the requisite ability or quality to fulfill a particular need and to be able to demonstrate it effectively and consistently. It can be seen as an expression of expectations but lacks the authority of either certification or accreditation."

What about data librarianship – is it a profession? If taking Winter's model into consideration, we notice that it is harder to precisely point out those six characteristics within data librarianship. For example, while certain LIS schools and iSchools do have a curriculum with an emphasis on data, their graduates are called "librarians" or "information specialists" and not "data librarians" or "data specialists." The validity of considering data librarianship a profession was discussed by Paul H. Bern during the 2005 IASSIST Conference; (Bern, 2005; EDINA National Data Centre & Edinburgh University Data Library, 2005) he points out that due to data librarianship having ethics codes and a service orientation, it certainly qualifies as a profession.

It is clear that data librarians are service oriented, as their duties - for example analysing, managing and archiving data – are often performed for the benefit of the data service's users. He admits that it is harder to prove that librarians' theoretical and practical body of knowledge is specific enough to make them qualify as a profession; however, the placement of librarians within an academic setting makes them uniquely positioned to "set the standard for education and training in the data profession." (Bern, 2005, p. 8) This, he suggests, can be done by "identifying the specific knowledge and skills required to be a data professional." (Bern, 2005, p. 8)

I have given examples of reports and initiatives aimed at defining a set of skills that data scientists, data librarians, and data librarians working with data science should have. (DEFF, 2009; Digital Data and Technology Profession Department, 2018) Through these initiatives, the role of libraries in managing data has gotten clearer, which has made it possible to define required competencies for librarians working with data.

However, advancements in technology and research mean that the competencies required to be a data librarian are rapidly changing; in this way, the terms "competency" and "profession" are changing alongside one another, due to the changes in libraries' structures and the repositioning of the libraries' data service in relation to the researchers.

Consequently it is clear that data competencies and the data librarian profession are dependent on each other while simultaneously being affected by the way researchers and organisations use data.

Because of this, researchers' and data behaviour can be said to have a strong impact on the competencies and strength of the area of data librarianship itself, without researchers being immediately aware of it. In this case, which party should have to adapt their behaviour in order for data librarianship to gain a more clearly defined professional role within the library, university, and research data process?

For my present thesis, I will take the position that both sides need to adapt and reach a middle ground towards clarifying the roles and areas of work for which this profession is responsible. The more clearly defined a data support service is within the library, the better it can serve researchers, and increase the value of their work. The more researchers understand and use the library's data support service, the better defined and understood these services will be. To achieve this balance, librarians should not aim to be "an ideal helper" to a few students, but rather "a reliable helper" to users.

The formats and amount of digital data that researchers are creating and using nowadays as part of their research has influenced librarians to acquire more technically-oriented skills and competencies. Data science oriented programming, for example, is a skill that data librarians are acquiring in order to deal with the increasingly large and complex datasets that researchers are working with.

Furthermore, the various roles that a data librarian has when dealing with researchers and students alike require the librarian to assume different roles (such as being a data detective, a consultant, an educator). In assuming these different roles, a data librarian would benefit from also acquiring more data science-inclined competencies.

While the roles and job requirements of being a data librarian are still changing, initiatives to define this professional area and set some "data science guidelines" for library use will clarify the roles of librarians and give library directors a clearer understanding of this emerging area of librarianship, which will in turn help them set aside the appropriate resources for acquiring talent or training staff towards this area of expertise.

The relationship between the roles and the placement of data librarians within research libraries is also worth looking at when discussing the data librarian profession. The idea behind creating a data support service and assigning data librarians to assist students and researchers is to give librarians a more central and active role around research data. (Australian National Data Service, 2013; Ekstrom, Elbaek, Erdmann, & Grigorov, 2016; Goulding, 2017)

Within the research library, the involvement of a librarian with data presents a certain dualism: in some institutions, a data librarian is mainly responsible for the management and integration of research data into the library's collection (Tsang & Renaud, 2014); in others, data librarians have emphasized roles within services such as research support, where their position within the research team is that of offering assistance with areas such as data management plans, statistics, etc. (Edwards, 2017).

Similar to research support services, initiatives such as the creation of data labs are another way of offering researchers and students alike access to data processing software and equipment, together with the expertise and guidance of a data librarian. As this requires an instructional approach, apart from programming and data specific knowledge, it puts the librarian into a quite dynamic, people-centric position within the library space.

It can be said that the discussion concerning data librarianship as a profession can be spread between several areas of a research library, and looked at from several different perspectives (library directors' perspective, researchers' faculty's, student's, fellow

librarians' and so on). This makes the competencies and skills necessary vary together with the job description, and as data librarianship is still changing its definition and placement within libraries, it is an interesting evolution to follow.

All in all, the librarian profession is well-positioned towards supporting researchers and students with their data. Data librarianship has grown alongside the increasingly complex modern data; it is considered to be a profession due to having ethics codes and being service-oriented. Despite that, it is a profession in which the competencies and skills required are being constantly updated in order to keep up with the technological advancements. Because of this, there is no clearly set path for a data librarian to follow in his career, and some see this as an advantage: it allows persons with different specializations to bring their background skills and knowledge into the library, which can make for a very interesting data librarian profile.

For a data librarian, acquiring data science skills allows him to be up to date with the requirements of working with complex research data, which can make his collaboration with researchers more fruitful. Having a data science oriented mind-set can prepare and inspire data librarians to come up with more innovative data-oriented services, aimed at researchers and students alike.

2.7 Data science initiatives in Danish research libraries

Questions: What are some data-oriented services that Danish research libraries offer? What does the community of librarians working with data science in Denmark look like? What opportunities are there for Danish librarians wishing to enhance their data competencies?

This chapter aims to present the situation of data librarianship and data science initiatives in Danish research libraries.

The situation of the main research libraries will be presented below. This will be achieved by using information found in the official Danish library statistics and on the Danish Union of Librarians website. Information found on the libraries' own websites will be used for the description of data-science oriented services and spaces. For the description of workshops and networking opportunities, the information will be sourced on the Danish Union of Librarians' website and from the workshops' websites. This chapter will be divided into two parts. The first part will show the way research libraries have organized their spaces and services around research data and data science. The second part will focus on the librarians' perspective, presenting the networking and competency development opportunities available for librarians interested in data science.

2.7.1. Data science-oriented services and spaces in Danish research libraries

In Denmark, research libraries are categorized by the Agency for Culture and Palaces, as follows: major research libraries with special obligations, and other research libraries. According to the latest official statistics I have been able to access (from 2015) there are currently 39 major research libraries all around Denmark. In this thesis, I am choosing to include the main research libraries, which are defined by Statistics Denmark as *research libraries with special obligations*, meaning "major research libraries with a staff of at least 3 years, permanent opening hours and headed by a library-trained person." (Statistics Denmark, 2018) I will be leaving out of the discussion the "other research libraries", which are characterized by meeting the requirements of employing "at least ½ library-graduated man-years" and have "a fixed opening time." (Statistics Denmark, 2018)

By exploring the websites of Danish research libraries, I have noticed that the university libraries present the most visible engagement around data science. This might be due to the fact that my investigation of the research libraries' websites did not capture information that is not publically available (such as event pages that are visible only on the intranet). Therefore, below I will discuss the data science-oriented services and spaces at the main university libraries in Denmark. I will mention the data-science oriented hardware and

software that data labs are equipped with. Further on, in the discussion I will discuss this information in relation to questionnaire responses related to software skills of Danish librarians who work with data science.

2.7.1.1. The Royal Danish Library

The Royal Danish Library in Copenhagen acts as the National Library of Denmark. Apart from special tasks such as "preservation of the Danish cultural heritage, dissemination of culture and research, [and being the] main centre for the country's public libraries" (Det Kongelige Bibliotek, 2017b), the Royal Danish Library works with data science as a way to increase the availability of their historical and cultural data. The "Digital Cultural Heritage and Media" department and its Digitization section has some initiatives towards making their materials more readily available for use by researchers; the IT Development and Infrastructure department is looking into improving online access to their collection and optimizing library processes.

In 2017, the State and University Library in Aarhus merged with the Royal Library in Copenhagen. Several branches which are a part of the Royal Danish Library now act as university libraries for the University of Copenhagen, Aarhus University, and Roskilde University. (Det Kongelige Bibliotek, 2017a) Below, I will explore the data science-oriented services in these university libraries.

2.7.1.1.1 University of Copenhagen libraries

The University of Copenhagen (KU) has established several spaces which provide students and researchers with access to equipment and expertise which is aimed at developing their data skills and creating opportunities for inter-disciplinary collaboration. These spaces are positioned either as part of a faculty department, or integrated as part of the faculty libraries.

Of interest to this thesis are data labs situated within libraries; however it is worthwhile to mention that KU has spaces such as the Data Science Laboratory[4] (part of the faculty of Science) or Copenhagen Centre for Social Data Science[5]. These spaces act like interdepartmental centres where researchers and data scientists work together to "enhance the quality of data analysis in research carried out" at the university (Data Science Laboratory, 2017) and to carry out their own "interdisciplinary and disciplinary research involving social and/or digital data." (Copenhagen Centre for Social Data Science, 2016)

The directions in which they offer assistance to the students and researchers are areas such as data acquisition, data handling and data analysis (Copenhagen Centre for Social Data Science, 2016; Data Science Laboratory, 2017); this is achieved through consultations, workshops, courses and collaboration with researchers. Spaces such as these can undoubtedly

[4] https://datalab.science.ku.dk/

[5] https://sodas.ku.dk/

bring value to the research produced by the university, while providing valuable support in the shape of a network of data science savvy peers. These two spaces are annexed to their respective faculties.

Since 2014, KU has come up with an initiative to establish more data labs and place them closer in relation to the libraries. This provides more learning opportunities for students, and "an alternative to the traditional lecture hall situation" (Københavns Universitetsbibliotek, n.d.) for university staff who wishes to provide the students with hands-on learning opportunities.

KUB Datalabs are an initiative that is part of KU's libraries. The Datalabs were created with the purpose of acting as "physical study environments for the students and staff", and are focused on "the use of state-of-the-art information technology and other technology-based features." (Københavns Universitetsbibliotek, 2018)

From 2014 until the present time, three Datalabs have been created. Their activities and aims are divided in two directions: firstly, they serve as spaces where students can work on "media and technology-based projects", and secondly, they are aimed at "staff members interested in integrating technology into their research and teaching" (Københavns Universitetsbibliotek, 2018). I will further present these three labs, mentioning their equipment and how their spaces and services are intended to work. By having an understanding of the services and equipment of a data lab, it is easier to understand which areas of competence they are supporting.

The *Data Lab*[6] (since 2015) at the North Campus is equipped with powerful computers that run software such as "R, Python, ArcGIS, QGIS, Mathematica, Excel, SAS and Endnote." (KUB Nord, n.d.) The computers can be booked in advance for individuals, or for groups of up to four persons who are "interested to learn about digital data and coding." (KUB Nord, n.d.) Through its placement in the university's campus, its purpose is to support the digital data learning process of Natural Sciences students.

The Digital Social Science Lab[7] (since 2016) is "intended as an open platform for education and events focusing on digital methods in social sciences." (Københavns Universitetsbibliotek, n.d.) The lab was founded out of the faculty's need for "incubating data information literacy" in a physical space (Svendsen & Lauersen, 2015); the lab aims to achieve this by bringing together "students, researchers and library staff" and offering them a physical space equipped with a range of tools for data exploration.

The lab is equipped with computers, projectors, a mobile touchscreen, 3D mouse and headphones; the computers have software relevant for multiple stages of the data lifecycle: Browzine and Mediestream for finding data, TCAT and Netvizz for harvesting data, reference management software such as Zotero, Endnote, and Mendeley, OpenRefine for data cleaning, NVivo, STATA, SPSS and Excel for analysing the data, Gephi for data visualization, and Hindenburg for audio editing of interviews. Apart from the above software,

[6] https://kub.kb.dk/datalab
[7] https://kub.kb.dk/DSSL

the lab provides an opportunity for students to learn programming in Python or R. The Digital Social Science Lab comes in support of the students by providing instructional workshops in using the above mentioned software; furthermore, the space is available to KU staff or students who wish to organize workshops and events on its premises. This encourages a dynamic, collaborative and informal teaching environment which creates opportunities for "networking, outreach and external partnerships." (Svendsen & Lauersen, 2015)

HUMlab[8] *(since 2016)* is a learning space for digital humanists at the University of Copenhagen. Apart from creating opportunities for humanists to improve their data literacy, the lab is facilitating inter-disciplinary exploration and collaboration within the humanities campus.

Their software spans a wide range of areas, including "3D scanning software, OCR scanning, GIS software, image, photo and video editing software, recording and sound editing, digital text analysis, web scraping and archiving tools, data cleaning and network analysis." (HUMlab, n.d.) In support of the software, the lab has powerful computers, scanners - including a 3D scanner, and a virtual reality headset. Just like the other two KUB labs, there are regular workshops being held on using their software and hardware; students have the opportunity to experiment with a humanistic approach to programming – for example, using Python for text analysis. The staff at HUMlab are working towards building a community around digital humanities, thus empowering the lab's members to create their own community and therefore encouraging skill sharing practices among students. This is achieved by creating study groups around subjects such as "digital cultural heritage", "digital text analysis", or "network visualization."

2.7.1.1.2. Aarhus University Library

Aarhus University's (AU) library has been recently (2017) integrated into The Royal Danish Library; the library offers courses in areas such as reference management, and online research support in the shape of guides around topics such as H-index, altmetrics, journal metrics, citations and so forth. The library also "provides advice and conducts bibliometric analyses for managers and employees" (AU Library, 2018b), as well as an option to "book a librarian"; since bibliometrics "is the quantitative analysis of science through its products" (Katrenko, 2015), and can involve large amounts of data, it is considered by some to be an area of "Big Data Science." (Moed & Luwel, 2014) They have several employees that can assist students or researchers with reference management questions.

As can be seen on their "Events" page, AU Library's instructional workshops are focused on literature search, academic information seeking, reference management software (EndNote X8) and qualitative data analysis software (NVivo). (AU Library, 2018a)

Aarhus University does have a data lab, and it is situated outside of the library, as part of the Faculty of Arts, serving as a Centre for Digital Social Research. Since it is not positioned within the library, it is outside the scope of this thesis. According to their

[8] https://kub.kb.dk/humlab/

webpage, DATALAB's focus is on "social adaptation of data systems and technologies as well as the social consequences of data applications and uses at small and large scale." It is part of the corporate and research network "Digital Society", and their interests lie in the areas of "data sharing and privacy", "survival of quality content", and "automation and control." (DATALAB, 2018)

2.7.1.1.3. Roskilde University Library

Roskilde University Library does not have a data lab; however, just like AU Library, it offers a service through which students can "book a librarian" for an email or in-person consultation – where librarians with different areas of specialization offer research advice concerning literature and information searches. (Roskilde University, 2018) I have been unable to find information regarding specific data science projects or research data-oriented events on the library's website.

Much like Aarhus University, Roskilde University's data lab is situated outside the library, and is seen as belonging to the University, which makes it unrelated to my thesis. The Digital Media Lab[9] is currently open by appointment only, and it aims to be a space where faculty and students can get "inspiration, help and consultation on research involving digital data and digital methods." (Roskilde University's Digital Media Lab, 2018) The lab offers digital services focused on data analysis in the form of Twitter data extraction tools and web crawlers; these resources are hosted internally by the lab. They also offer links to external resources such as Instagram scrapers, data cleaning software, data visualization software and other tools for working with data.

The above mentioned university libraries have all been integrated into the Royal Danish Library as of the year 2017. Below, I present several other university libraries and their initiatives around research data and data science.

2.7.1.2. Aalborg University Library

Aalborg University's main research library is located in Aalborg, and has smaller branches near the other university's campuses, in Aalborg, Esbjerg, and Copenhagen.

The library provides support for researchers and students in the form of software and web services which assist them in reference management, and data analysis and collection. Of the latter, they offer access to software for survey analysis (SurveyXact), qualitative data analysis (NVivo), data management plans (DMPoline), and statistics software (SAS, SPSS, STATA). (Aalborg University Library, 2018) Furthermore, they offer an online guide for data management, and a consultation-based data management service.

Aalborg University does have several labs in which students and researchers can develop competences, however none of them are specifically related to data science, and as such will not be mentioned here.

[9] https://digitalmedialab.ruc.dk/

2.7.1.3. Danish Technical University Library

The Danish Technical University's (DTU) Library offers a guide on research data management (DTU Library & Refn, 2018); they hold events on educating the students and researchers on the topic of research data management, such as the Love Data Week. (DTU Library & Lavanchy, 2018) Within their Office for Research and Relations, DTU have a dedicated team working with bibliometrics and ranking.

The DTU Library has supported librarians interested in data science by hosting the first "Data Scientist Training for Librarians" in Europe, ever since 2015. (DTU Bibliotek, 2015) More information about this training will follow further on in the chapter.

2.7.1.4. University Library of Southern Denmark

FabLab[10] is a lab that is part of the University of Southern Denmark's library, located in Odense. It has a unique concept in Denmark: it is based on a "student2student" support concept. (University of Southern Denmark, 2018) This means that students who need support with their projects book the assistance of other students. The lab is equipped with a digital video recording studio, 3D printers and powerful PCs. The PCs come with statistical, simulation and graphics software.

While it is not explicitly a "data lab", it does qualify as providing some degree of support in data science due to the availability of statistical software and of instruction. The concept of "student2student" instruction makes this space an interesting example as to how libraries can create opportunities of collaboration and learning through facilitating peer-to-peer teaching.

2.7.2. Competences development and community-building around Danish data science librarianship

As data science in libraries is a relatively novel concept it is apparent that not many research libraries in Denmark have set up services that involve data science. However, certain librarians have started to show an interest in acquiring competences in data science, and using them within the libraries. Further on, I will discuss what opportunities there are for librarians in Denmark who wish to acquire data science competences, or are interested in joining a community around this area of librarianship.

2.7.2.1. The Data Science Professional Group

In order for data librarianship to increase its presence in Denmark, the Danish Union of Librarians has established a professional working group around the topic of data

[10] https://www.sdu.dk/en/fablab

librarianship. The Data Science Faggruppen[11] ("professional group") is aimed at developing and supporting librarians' competencies in the field of data librarianship, data science and data management.

The group creates and maintains a network of librarians and others with an interest in data librarianship and open science, primarily through organizing workshops, teaching on data science matters, and providing support in the shape of reading groups and knowledge sharing about specific tools and methods.

The professional group was founded in 2017, and so far has brought together around 60 members – librarians, students, and other interested individuals - whose professional interests lie in the area of data librarianship. The group's chairman and contact person is Jeannette Ekstrøm, from the Department of Innovation and Sector Development at the Danish Technical University Library. Jeannette has an interest in teaching information skills, information literacy, and Open Access, publishing and research data; apart from creating the Data Science group, she has been involved in building data librarianship in Denmark, including coordinating the "Data Scientist Training for Librarians", as well as contributing to theoretical knowledge in this field.

The professional group's board is further comprised of five information specialists and researchers: Lorna Wildgaard, Michael Svendsen, Mikael Elbæk, Annette Graae and Hazel Engelsmann (from Copenhagen University's Department of Information Science, The Royal Danish Library, DevoTeam Consulting, and Aarhus University Library) and three information science students, including me.

This Data Science group is the first initiative in Denmark that attempts to build a community around data librarianship and working with data science; the aim is to gather members from both public and research libraries, and teach them the competencies necessary to apply data science tools and how to adapt them to the specific contexts of their libraries.

Several times a year, the group holds meetings in different formats: workshops, meetings and study groups around different topics such as data management, Data Discovery, text and data mining and analysis, APIs, Open Data, and others, as well as hands-on workshops on data science software such as GitHub, Jupiter Notebook, NVivo, Voyant Tool.

Another focus of the group is increasing awareness around data science possibilities in libraries, which requires reaching out to various libraries across the country; most of the group's members reside in Copenhagen; reaching out to communities of librarians across the country through local meetings and presentations of the group's activity are also important for the group's purpose. This includes increasing awareness of the group within the University of Copenhagen's Department for Information Studies, and getting the future information specialists familiar with data science.

Another initiative towards which the group aims to contribute is the theoretical development of data librarianship within Scandinavia: several of the group's board members are involved in the creation of a book presenting theories and methods of data librarianship in a Scandinavian context. This material will be an important addition towards local theoretical

[11] http://bf.dk/ArrangementerOgNetvaerk/FaggrupperSC/Data%20Science%20Faggruppen

knowledge in the domain, as most existing research about data librarianship is written from the perspective of US librarians.

2.7.2.2. Data Scientist Training for Librarians

Most of the workshops I have mentioned up to now have been focused on the way that librarians can instruct researchers and students and help them acquire data science competences. For librarians to have the knowledge necessary to hold such workshops in the first place, they need to be familiarized with the tools and methods used in data science. After acquiring a familiarity around that area, they can decide which areas of competence they would like to strengthen and which skills they should acquire to improve their data librarianship work.

Data Scientist Training for Librarians[12] (DST4L) started out as an experimental course within the Harvard Library, with the purpose of giving research librarians there the data science skills they needed to better assist the researchers. Christopher Erdmann, the founder of DST4L, believes that if librarians acquire programming skills and get more data-savvy, it allows them to "become partners in the data-related work of their organizations."(Erdmann, 2014) In Europe, DST4L has been successfully applied as a data science training workshop in Denmark since 2015. The workshop has been held at the DTU Library.

The activities of DST4L span along three days, in which librarians attend hands-on workshops in in a range of data science areas, such as data collection, metadata handling, data cleaning, web scraping, version control and collaborative software, and data visualization. (DST4L, 2017)

By taking part in this symposium, librarians get more familiar around data science tools and methods; this allows them to understand the ways in which data science can help them in their libraries: this might mean inspiring them to hold instructional data science workshops, or it might spark their interest in learning about methods that they can use to facilitate data discovery in the library's collection.

The purpose of this workshop is not to offer librarians a complete training in data science. By providing librarians with a sense of familiarity around data science, DST4L increases research librarians' awareness of their role in regards to library data, and the initiative tries to empower them to create their own data products within the library. At the least, librarians come back from DST4L with a better understanding of the research data lifecycle, which can put them in a good position to identify areas in which they can support the researchers.

Previous editions of DST4L have registered attendance by research librarians from all over the Nordic countries. The organizers have compiled a "DST4L Toolbox"[13] where anyone interested can access the slides, recordings and workshop guides introduced during the 2016 edition of the event.

[12] http://www.dst4l.info/
[13] http://www.dst4l.info/toolbox.html

III. Method

A mix of methods was used in order to answer the research questions. Employing a triangulation of methods brought the advantage of giving this thesis the ability to "address different aspects of the same research question." (Gorman & Clayton, 2009, p. 12)

Triangulation is defined as "the use of more than one approach to researching a question." (Heale & Forbes, 2013) It can provide a more comprehensive picture of the results than if using just one method and it's useful in avoiding potential bias.

One of the main reasons that triangulation was used in the present thesis was related to the comprehensiveness and credibility of data. The credibility of the data "is enhanced if it can be confirmed from several sources, just as it can be if gathered using several different research methods." (Gorman & Clayton, 2009, p. 129)

As a way of answering the main research question, "What are the current data science initiatives in Danish research libraries?" from different perspectives, two secondary research questions were created.

For the first secondary research question, "What data competencies do Danish librarians possess? What skills do they wish to develop?" empirical evidence was used in the form of questionnaire responses from the Danish Union of Librarians' Data Science group, which was used as explorative material within the discussion.

For answering the second question, "What challenges do librarians face when dealing with data science in their work?" I conducted two interviews with professionals involved in the field, with the purpose of getting two different perspectives on the topic: on one hand, the perspective of an information specialist who is coordinating a data lab, and on the other hand the perspective of the lead data scientist that is developing the data science strategy and products for the Danish Bibliographic Centre.

The following sections explain how the respondents and participants were selected, and provide information on the materials involved, as well as insights into the data collection procedure.

3.1. Questionnaire

Questionnaires are a suitable method for acquiring empirical data with a standardized form from a representative sample of persons within a specific time span. (Kelley, Clark, Brown, & Sitzia, 2003)

For the present thesis, the questionnaire method was chosen due to providing the ability to screen out a relevant professional group of librarians and gather information on

their data science competencies as well as their wishes for further skills acquisition in that area. Using a questionnaire also allowed respondents the opportunity to remain anonymous, in this way encouraging more honesty.

This questionnaire was developed by the Data Science Professional Group that is a part of the Danish Union of Librarians. Its inclusion within the thesis was done with written consent of the group's chairman. My position as student board member inside the group gives me an awareness of the way the questionnaire was developed. The questionnaire was originally sent out with the purpose of uncovering the members' interests and their wishes for future areas of focus around the group's activity.

The questionnaire responses are used to open a discussion around the competences of librarians and their wishes for further developments in the field; the exploration of this data was done with an inductive, pattern-seeking approach.

3.1.1. Participants

The sampling of the questionnaire was focused on the participants' professional background and interest in data science and data librarianship in Denmark. The term 'participants' is used to refer to both the respondents, and to the persons who received the questionnaire but chose not to interact with it. All of the participants who were sent the questionnaire reside in Denmark, and are members of the Data Science Professional Group, and implicitly of the Danish Union of Librarians.

3.1.2. Materials

The questionnaire design was based on a cross-sectional, single-stage approach. (Lavrakas, 2008) The Data Science group sought to understand more about their members, including their existing data science competencies and skills which they wished to develop under the instruction of the group. In order to ensure the validity and coverage of the questions, the questionnaire was validated by getting feedback from experts in the field.

The questionnaire layout was comprised of five sections which contained both open and closed (multiple-choice) questions. The first section contained questions aimed establishing the demographic and professional profile of the respondent, the second section asked participants about their experience working with data science, the third section queried the respondents' wishes as to what skills they wish to further develop, and the fourth question inquired the respondents' preferred form of keeping up with the professional group's updates. Of importance to this thesis are the topics covered by the second and third section.

The questionnaire was created using Google Forms, and the responses recorded in Google Sheets. The questionnaire was in Danish; within the appendix, an English translation is provided. The questionnaire responses are not provided within the appendix, because due to the number of responses, the character of certain questions (workplace, education) could have made the respondents easily identifiable.

3.1.3. Procedure

The link to the online form containing the questionnaire was sent out through email to all the group's members, as an official email from the Data Science group. It was open for responses between March and June 2018. The questionnaire was sent out on the occasion of the group's first anniversary, and to motivate the members to respond, a giveaway was held, through which the randomly selected winner could win a Raspberry Pi computer. The responses were anonymous, but the participants who chose to participate in the giveaway were required to list identifying information such as name and telephone number.

Out of the 60 targeted participants, there were 19 respondents. Since the recommended response rate is around 65% in order for the results to be statistically relevant, the data from the questionnaire was used in an explorative manner, as proof of an existing interest in data science within Danish libraries.

3.2. Interviews

As a research method, interviews provide researchers with a personal way to collect large amounts of rich data in a format that often enables the participants to speak freely on a certain topic, which may often lead to the researchers getting unexpected insights. Another advantage of this approach is that responding verbally (as opposed to in writing) makes it easier for the participants to share their thoughts, particularly if the research topic is of a confidential or otherwise sensitive nature.

On the other hand, since anonymity is lost, this approach may cause the interviewee to omit relevant facts which he or she might have been comfortable sharing in an anonymised, written form. Gorman points out that factors such as the approach and personality of the interviewer have a significant impact on the direction of an interview – which he warns makes interview data open to bias. (Gorman & Clayton, 2009, p. 126)

One of the positive attributes of interview methodology is that it allows for a certain degree of adaptability. The types of interview range from structured, to semi-structured, to unstructured. Structured interviews are characterized by predetermined questions, and have answers which fall under predetermined categories; they are widely used as part of a quantitative methodology, and their resulting data has the advantage of being easily comparable. On the other end, unstructured interviews are best characterized as "an interactive conversation" in which the discussed issues "are explored in as much length as necessary or available." (Gorman & Clayton, 2009, p. 127) Due to their exploratory aspect, an inexperienced interviewer might have difficulty with leading the discussion and synthetizing the resulting data.

Semi-structured interviews (Cohen & Crabtree, 2006) offer an in-between alternative to the above, which is suitable for less experienced interviewers due to containing interview guides; these allow the interviewer to prepare and define questions and topics of discussion

ahead of time, yet still leave the conversation open enough to allow interviewees to speak freely, which can provide useful and unexpected insights. The question guide offers some of the benefits of both structured and unstructured interviews – it can be built in such a way as to facilitate easy comparison of responses, yet the questions can be open enough as to encourage informal conversation around the topics.

Two semi-structured in-depth interviews were conducted as part of the research for this thesis. The purpose of the interviews was to provide two different perspectives towards answering the secondary research question "What challenges do librarians face when dealing with data science in their work?" This approach allowed for a certain degree of adaptability, taking into account the domain differences between the two interview subjects.

3.2.1. Participants

In order to get two different angles of data science librarianship in Denmark, a certain degree of stratification (Gorman & Clayton, 2009, p. 128) was involved. One of the participants was chosen due to his professional involvement with data science strategy and products within Danish libraries. His experience as Lead Data Scientist at the Danish Bibliographic Center (DBC) is very relevant when discussing Danish librarians' competences around data and IT. The second participant is an information specialist, and was chosen due to his experience as a coordinator for a data lab within the research library of a major Danish university. His direct experience as a coordinator offered an inside perspective on the processes, competences and challenges involved in coordinating a data lab.

3.2.2. Materials

The two interview guides have similar areas of focus. The experience of the participants offers complementary perspectives on those areas, upon being discussed from both perspectives – for example, defining data science and expectations of librarians' skills.

In order to ensure the validity of the questions, the interview guides were validated with the help of my supervisor, and a librarian.

The interviews started with simple introductory topics which allowed the participants to discuss areas of high familiarity (their professional background and experience). The interview guide was constructed so that the interview started with general, concrete topics, and moved on towards more abstract notions such as roles and perspectives of data librarianship. Prompting questions were also included in the interview guide, with the purpose of getting the participants comfortable at the beginning of the discussion.

The interviews were recorded using a phone; the transcription was done by using the Artificial Intelligence-powered service Trint[14]. The automated transcription resulted in

[14] https://trint.com/

verbatim transcription, which I further edited to produce the final, less cluttered versions. The researcher considered it important for the final version of the transcriptions to contain as much of the original discussions, as due to the semi-structured character of the interviews, the discussions got quite rich around certain topics.

The duration of the interviews was between 40 minutes and two hours. Due to this, the size of the interview recordings was quite large, and the interview recordings and transcriptions had to be cut into half an hour sections. This allowed for easier referencing back to the transcribed materials, during the discussion.

As a way to ensure an ethical data collection process and facilitate a clear understanding of the interview process, a consent form was developed. It included a description of the project, information on data collection and data storage, as well as a section which clarified the level of citation the participants were comfortable with. They could select whether they were comfortable with their name and professional title to be mentioned in the discussion, and whether direct quotes from the interview could be included in the final version of the thesis. The consent forms can be viewed in the appendix section of the thesis.

With concern for an ethical data storage process, the interview materials –recordings, transcriptions, and consent forms – are stored on a secure online cloud platform. The consent sheet contained information of this storage process; the interview recordings will be stored on the platform for one year from the completion of the thesis, after which they will be destroyed. The consent forms and interview transcriptions will be kept after this period, being directly related to the contents of this thesis.

3.2.3. Procedure

The participants were contacted through email and informed of the purpose and duration of the interviews. Upon their agreement, a time and place was set for the interviews.

Interview A was conducted in a meeting room at the participant's workplace, and interview B was conducted in the participant's office; both settings provided a quiet and familiar space in which the participants could relax. For interview B, the location offered an additional unanticipated advantage – during the interview, the participant was able to show and refer to certain materials that came up in the discussion – such as research papers, an organigram, and multiple websites; this seemed to bring further insights into the conversation, and offer a more natural flow.

The sessions started by explaining the procedure to the participants, and asking them to look over and sign the consent sheet. Following that, the recording device was pointed out to the participants, and upon their agreement the session began to be audio recorded. Afterwards, the interviewing process began, following the interview guide which was printed out. To encourage a smooth discussion, no notes were taken.

The absence of time constraints encouraged a flexible discussion that followed the interview guide. By using clearly assembled questions together with prompting questions in areas where further understanding was needed, the discussion had a natural progression around the interview topics.

The natural narrative flow of some of the answers meant that the discussion inadvertently answered the next question as well – when this occurred, the questions that had been answered were skipped.

Overall, the triangulation of methods used for this thesis proved a good choice for the topic at hand. By using the materials from the official Data Science group questionnaire, the researcher was given insight into the experiences and wishes of a highly relevant demographic of information professionals in Denmark, a demographic that she could not have reached through her personal network. Further, the two semi-structured in-depth interviews uncovered highly relevant narratives and opinions from two complementary perspectives: on one hand, the more theoretical data scientists' perception of data librarianship and libraries' involvement with data science in Denmark, and on the other hand, the first-hand perspective of an information specialist who is directly involved in coordinating a data lab at a major research library.

IV. Data Science practices in Danish research libraries

As a way of answering the main research question, "What are the current data science initiatives in Danish research libraries?" from different perspectives, two secondary research questions were created.

4.1. Data competencies and skills of Danish librarians

Questions: What data competencies do Danish librarians possess? What skills do they wish to develop?

For the first secondary research question, "What data competencies do Danish librarians possess? What skills do they wish to develop?" empirical evidence was used in the form of questionnaire responses from the Danish Union of Librarians' Data Science group; of interest to this thesis are the first three sections. The responses were used as explorative data alongside material gathered from domain literature and library websites. The two data sources have been used to uncover patterns in the competences librarians have around data science.

In the discussion of the questionnaire responses, certain terms have been emphasized; these terms are categories defined in the "areas of competence" table in chapter 2.5.3., and have been emphasized as proof that the librarians' self-reported knowledge falls under the area of data science competences.

The response rate was 31%, and as the responses provide plenty of qualitative insights, it allows an insightful discussion around the questionnaire responses.

Demographically, almost half (47.1%) of the respondents are aged between 31-40; 29.4% are aged between 41-50; 17.6% are under 30 years old, and 5.9% over 60; their gender is not considered relevant to this study; a good majority (77.8%) of them reside in the Capital Region, while several others reside in Zealand (16.7%) and Central Jutland (5.6%). All of the participants have a professional or educational background in librarianship, with over half (57.8%) professing as librarians, and the other half holding librarian degrees. Over 70% of librarians are employed in the public sector.

The professional and educational backgrounds are not surprising due to the sampling of the questionnaire; the other demographic factors help paint a profile of librarians involved with data science in Denmark: the age range of the respondents suggests that both novice and experienced librarians have a professional interest in developing data science competences.

Their residence around the Capital Region, Zealand and Central Jutland is explained by the presence of major research libraries in that area; the library websites exploration presented in chapter 2.7. confirms the existence of multiple initiatives around data science within that geographical space. The respondents work at the Royal Danish Library, Copenhagen University Library, Roskilde University Library, DTU Library, Aarhus Library, the Danish Institute for International Studies, and the Danish Bibliographic Centre. The first five libraries listed match with the libraries that mentioned having data labs and other research data support opportunities.

The fact that only a few libraries advertise their data science-oriented services says more about the library as an organization than about the librarians' competences. It is possible that librarians work with data science independently or are still be in a stage of competence acquisition, which might motivate them to not advertise their expertise until they consider it has reached an appropriate level.

Therefore, the questionnaire's demographic data together with the advertised data services offer a starting point in the investigation of data science competences. To get closer to the research question, section 2 and section 3 of the questionnaire offer insights into the areas of competence that librarians possess and the ones they wish to develop.

Data science-specific programming skills are listed as one of the elements that make up data science competencies. Sections two and three asked respondents about their knowledge of data science-specific tools and topics.

When questioned upon the programming languages that they had experience with, most of the respondents listed Python. JavaScript, HTML and SQL were mentioned by several respondents. Turbo Pascal, Flask, Django and XSL were also mentioned, each just once. Regarding data science software, the answers were comprised of several categories: data cleaning tools (OpenRefine), network analysis and visualization software (Gephi, Vos Viewer), code editors (Atom, Oxygen XML Editor) data visualization software (Tableau, Plot.ly), collaborative platforms (GitHub) and web scrapers (Netvizz, NCapture). These are quite diverse areas of competency, which *add value* to the librarian profile.

Furthermore, the respondents admitted having experience in the following areas: open data, data visualization, data cleaning, web scraping, research data management, data management platforms, metadata handling. One respondent felt that while they were not a machine learning specialist, they did feel like they knew enough about how code works. This shows an *awareness of role* and supports a previous argument presented in the competencies chapter: there are different roles around data science within a library; around these different roles, different levels of competence mean that while a librarian might now know *everything* about a data science method, they can reach a point where their skills are just enough to be useful within the library.

The above responses show that the respondents are already knowledgeable around different data science-relevant tools and topics; they also confirm their involvement in the areas of Stanton's "four A's of data" (data architecture, data acquisition, data analysis, and data archiving). (Stanton, 2012)

The way in which data science methods are used within the library provides insights into the levels of competence librarians have around the different areas of data science, and can be discussed from the perspective of the "three levels of data librarianship".

Below are some answers to the question "How do you work with data science in everyday life?" The following quotes exemplify the existence of different levels of data librarianship within Danish libraries:

Data-savvy librarians: "I use the group to gather knowledge and hopefully learn something so I can be at the forefront of my researchers' needs in the field"; "Data validation in PURE and documentation of research"; "Does not work actively with data science or data manipulation"

Data librarians: "We are building a teaching offer that will cover [data] harvesting, analysis and visualization"; "Support for colleagues about understanding Shell and DOS environments"; "Teaching in NVivo";

Data science librarians: "Through development projects"; "I try to take tasks - such as data cleaning, export data [...] and enrich and clean data"; "Data management of unstructured data."

As can be seen, the different levels of data librarianship affect the involvement with data science; some respondents answered that they do not actively work with data science yet were members of the DS group, which suggested an interest in the area and a wish to be "savvy" around this topic. Contrastingly, other respondents exemplified the ways they already use data science, either by supporting the researchers or being an active part of data science work and developments within the libraries. Responses such as the latter point towards awareness of their own *performance management* competences.

One of the main roles of the DS group is to promote competence development in the area; an appropriate course of action regarding planning future activities can be decided by adapting to the members' level of experience, and finding out in which areas they wish to develop competences. By studying their responses in relation to their wishes, one can get an understanding of which areas of competence librarians consider important enough to pursue.

The questionnaire respondents' interests towards increasing their programming competences can be summed up by the following comment: "more Python and Javascript, but much can be solved through Python if you understand." There was a recurring desire to increase skills in statistical tools such as SQL, Excel, SPSS and R. Respondents also wished to increase their skills in data cleaning (OpenRefine), visualization, (Tableau), analysis (Excel) business intelligence (Metabase), and database management (SQL, VBA). General areas in which the respondents wished to develop were areas in which they already had skills such as data visualization and analysis and open data, but also areas which they were not yet knowledgeable in, such as machine learning, data modeling or gamification of data. The responses suggest an *awareness of the librarians' role* within the organization and a wish to *add value* by increasing their I.T. skills.

Overall, respondents stressed their wish for hands-on opportunities to learn data science; some of the responses suggested that they faced a skills gap when participating in I.T. targeted data science workshops: "my challenge is usually that many participants have much more insight than I do in programming." Some wished for "more inspiration" and expressed a need to "move the limits of what library people can do", which might be explained as a wish to acquire data science-oriented thinking and falls under the "*innovation*" area of competency. Certification was also mentioned, suggesting that a formal acknowledgement of their skills could help their ideas and initiatives be heard more clearly within their organizations. The librarians' wish for the DS group to facilitate exchange of ideas and networking opportunities is important due to its community-building implications, and it falls under the "*collaboration*" area of competence.

By relating the above to information analysed from the libraries' websites, one can get a further understanding of the data competences of Danish librarians. It is clear that librarians' data competences are developed around areas relevant for data science. They have knowledge of many of the tools data scientists use, and an ambition to further extend their competences in a technical sense. These competences are reflected in the way data science-oriented services and spaces have been built within libraries: whether assisting users, instructing researchers, or equipping data labs, librarians (of all three 'levels' on the data librarianship scale) bring their current understanding of data science into creating the service or product. As such, tools and methods which the respondents admitted having experience with are present in library spaces around Denmark: reference management, data collection and analysis software, programming languages and web crawlers, to name just a few.

4.2. Perceived challenges in working with data science within the library

Questions: What challenges do librarians face when dealing with data science in their work?

For answering the above question, two interviews were conducted with professionals involved in the field, with the intention of getting two different perspectives on the topic - that of an information specialist who is coordinating a data lab, and that of the data scientist that is developing data science products for Danish libraries.

The following discussion provides arguments from both interviews, which are tied together with insights provided by theoretical materials presented in previous chapters. In order to facilitate comparison and analysis between the two interview transcriptions, the researcher used the topics defined in the interview guides to follow a "red thread" through the participants' answers. The interview responses can best be "tied together" and explored by discussing the following areas:

Roles of data science in the library

Facilitating data science services and spaces

Librarians' acquisition of data science skills and competences

4.2.1. Roles of data science in the library

Both participants agreed that data science is applied within the library primarily for the benefit of the user: [it's] "a lot about how can we use data science to make things easier and cheaper to build and make it better for our users" (Boesgaard, 2018), "it could be a good idea if the library could develop competencies". Data labs are an attempt to "try to support a meaningful and interesting environment for students [...] and could be [...] a stepping stone for some people into new methods" (Kjær, 2018)

When used to improve access to the library collection, Christian considered that data science provides a way to modernize library processes. For example, by making interactions with the library catalogue "as easy as Netflix," data science creates opportunities to attract new categories of users. Regarding the placement of data science services and spaces within the library space, it came as a natural decision to the KU leaders: the positioning made sense because of organizational (budgeting, service acquisition) reasons "but it's never been the idea that we should put the library lab [...] at the university campus." (Kjær, 2018) There are conflicting opinions around the positioning of competency-building services within the libraries: some "think that the libraries should provide information, but the universities should do all the teaching and using the tools and so on."

As to the way data science changes the librarians' role within the organization, Christian considered the librarians' practical experience within the library as advantageous towards developing data science products within the library – however sees their approach to service improvement as too speculative: "You can ask them some really obvious questions and they really can't answer, like 'What is a good search' and 'Is our search engine good?' [...] I think it's problematic that we don't have good answers for those obvious questions. Because libraries are expensive - and we should be able to say why."

Lars considered that "it's much better that the people in libraries do [data science]. Because they already have this instructional capacity - within humanities we know how to

teach, we are open and have an understanding of how these methods and tools can be applied in different subjects as well." However, just because librarians can provide an instructional role and teach researchers and students data literacy, doesn't mean that all librarians need to be equally involved with data science.

The "three levels of data librarianship" were reflected during the interview by Lars' observation that "...in a library not everybody is interested in working at this level [teaching data literacy]. Some of my colleagues say 'that's nothing for me'." He suggested that the first two levels, of "data-savvy librarian" and "data librarian" are enough for a data lab coordinator: "but the way the software are today you are able to teach how to use a database, to build a database. You don't need a data scientist to do that, it's just simply to too trivial for them." Even if, for example, reference librarians do not have a working knowledge of programming and data wrangling software, having an awareness of these tools can help refer students to the local library specialist, should they need his assistance. Within the KU library, regular meetings are held between staff to create a shared awareness towards which data services the library provides: "we try to tell them what we do in the labs. So when they meet students then they know they can send people in our direction."

Lars stressed that even though his role within the library was different to his colleagues', he still had 'normal librarian' tasks outside his "HUMlab hours". His involvement in the area didn't make him any more special than his colleagues, certainly not "a golden librarian". However, he did believe that his role within the library was important enough to deserve a more accurate job title than "information specialist."

He also believed that perception and recognition of their data science capabilities play an important role in the relationship between librarians, their skills and their institutions: "you have aggressively go out in the world and talk about what you do, not be afraid to show people what you're doing, and then you change the job titles. That's what you see in most countries as well. You see in USA they call them 'data librarians'." As it is an area that is still under development, the labels and titles are still changing; Christian pointed out that the term "data science" itself may be "just a term" and that maybe it should have its own name inside library science.

Without prompting from my side, Lars acknowledged a distinction between "data-savvy librarians", "data librarians", and "data scientist librarians" in the following ways: he believed that librarians who support staff by assuming an instructional role within the library can be defined as "data librarians." He considered it useful for all library staff to have an awareness around data science topics; he suggested that data scientists would not find it challenging enough to work inside a research support service.

4.2.2. Facilitating data science services and spaces

Within DBC, the use of data science started out as a way to improve the efficiency of library systems and so, to save money and time, and "now it's hard to see how you can compete in five or 10 years if you don't do this." (Boesgaard, 2018) Inside the university

libraries, data science was used to create services and spaces that could support researchers and students in their data projects, and in acquiring data competences.

At KU, the initiative to create data labs was inspired by the data labs in the US; among university leaders, there was a growing interest in "library-developed labs" and developing competencies within the library – budgets were soon assigned to this area and starting with 2014, KU built three data labs.

The opportunities that data science spaces such as HUMlab provide for the users range from providing convenient access to equipment, to enabling acquirement of digital competences, to creating networking opportunities and improving the students' job prospects upon graduation. A typical user of a data lab is "a person who wants to strengthen their digital profile. To add something to their profile that they can't get or they don't have an opportunity to add to the profile at the university; something extra which can make them more interesting later on in the job career."

A collaborative and action-oriented mindset is essential when establishing services and spaces that support researchers and students: Lars admits that the rapid growth of HUMlab over its first two years can be attributed to his and his team's spirit of "getting out there, talking to people, and making things happen."

Within the HUMlab team, collaboration is achieved by pooling together all the librarians' skills. Their areas of knowledge balance each other and their backgrounds are varied, which helps them establish diverse opportunities for students.

Reaching out to the teachers is a way in which data librarians promote collaboration inside the university. The lab can provide instruction to teachers, as a way to *teach them how to teach* data science skills in the curriculum of their courses. For example, Lars assisted some teachers in finding a solution to teach their students text analysis. In situations like this, having the support of a data lab means "the students are happy, and the teachers are happy because they got around to it. So that's the way that library actually can develop a new service around data and around information."

Running a data lab doesn't always go smoothly. Lars admitted that there have been occasions when the effort put into creating events didn't pay off – sometimes students were not interested, otherwise their schedules meant they had no time for extracurricular activities. His solution to that issue was developing the initiatives of the data lab in close collaboration with the students, and deciding lab purchases and future areas of focus is done by consulting the students first.

Also, Lars believes that it's the students' "own engagement and their own interests that actually are running" the lab. He sees unsuccessful events as factors that waste the lab's resources: by having 20 hours of HUMlab work each week and a set financial budget per year, the lab's ultimate task is to be used as much as possible. This is done through coordinating events, giving research groups a space to work individually, and encouraging peer-to-peer learning initiatives.

4.2.3. Librarians' acquisition of data science skills and competences

Section 4.1. presents the data science skills and competences of several Danish librarians. Within the interviews, I attempted to capture further insights into the process of skills and competences acquisition.

Lars considered that data librarians' competences can easily go unnoticed within the array of services of a research library; he felt like "sometimes [they] have to be a bit loud about it and say 'we do have these competences' because there can be people at the university that maybe think that it's not a good idea that the library develops [such competences]" Being vocal about their competences can go a long way in a library, offering visibility to the data librarians and their services.

Due to the recent creation of the data lab, HUMlab is still testing the waters and reaching out to as many departments as possible; the competence profile of the lab is still developing, and due to that, the librarians' decision towards future acquisition of skills is adapted to the requests of their users: "if there is an interest in a specific software, then we have to be able to honestly say 'I don't know much about it right now but, maybe in a year [...] we will both know much more.'"

As data services are created around the needs of the library users, librarians might feel like they always need "more exposure to different environments and tools", in order to cover the wide array of competences that users can expect of them. In a library where most of the users are researchers and students, librarians need to be up to date with their technological competences and "know the options and opportunities in different software" in order to guide and instruct users. At a large university library such as the one at KU, they need to know "how [to] apply software but in different fields." As Lars pointed out, it is the researchers and students that come with the domain-specific knowledge – which makes it the data librarians' duty to at least have "general knowledge about what software would be relevant to apply to this subject that they're using, to the research that they're doing."

Both Christian and Lars supported the importance of competences in the area of programming. While Lars supports ongoing self-education in that area, Christian believes that programming should be a part of librarianship curriculum. He considered it "not necessary to spend a year on programming", but thought several programming modules would be useful in a librarian's education, at the very least "so they can speak to an I.T. person." In his vision, a familiarity with programming may improve their solutions to library data problems. He believed that knowledge workers, no matter their area, "deserve to be able to know what the possibilities are", and acquiring an understanding of data science may bring totally new ways of looking at old problems.

A difficulty that librarians face in the area of skills acquisition is the lack of a clear professional path that one can follow, especially from a competency-acquiring viewpoint. "Identifying the specific knowledge and skills required to be a data professional" (Bern, 2005, p. 8) is just one piece of the puzzle. In the previous sections I have identified data-science specific skills; from the questionnaire responses it is apparent that indeed some Danish librarians have data science competences. It is unclear what levels of competency

they have in each of the areas they listed. Their indication that they both have and wish to get more experience with certain tools (e.g. Python) suggests that upon self-reflection, the respondents felt that they had not reached their desired competence level in certain areas.

But how can librarians know which of their competences are at a satisfactory level and which they should still develop?

The HUMlab team measure their competences by colour-coding the software the lab works with: "I coded them like the colours of a traffic light [...] what we are going to do next is start with Python and NVivo and we decided that's what we're going to focus on. And then when they are green, we're going to the next thing, then we go to photo scan and Gephi. It's an easy way to work around this, just an easier method to handle it. It's just not to be confused in where to start and where to end, and to have a way of talking about it; that way, we can decide in our team what should we focus on for the next half a year. We can look at this and realize where we would be in one year, and then know how we should start and organize it."

The team competence assessment sheet can be consulted in the appendix. This model is easy to apply within a library service, and can be used to validate the skills of data librarians. It can also be modified and used as a self-assessment tool for individual librarians along their data science journey.

V. Conclusion

Librarians have always been an important ally for researchers and organisations. The amount and complexity of data nowadays, together with progressive changes to the librarian profession mean their skills and competences make them good candidates for introducing data science as part of their library services.

One way in which they can do that is through the creation of services aimed to grow students' and researchers' data skills.

To do that, librarians need to be competent in the following areas: data architecture, data acquisition, data analysis, data archiving; when working with data science, these areas translate into the following skills: applied maths, statistics and scientific practice skills; data engineering and data manipulation skills; data science innovation skills; developing data science capability; domain expertise; data science oriented programming; understanding of data science analysis lifecycle.

I have distinguished three levels of librarianship, concerning involvement with data: data-savvy librarians, data librarians, and data science librarians.

The theoretical concept of three levels of librarianship around data science is reflected by the questionnaire and interview data. All three types of librarians can be identified among the respondents. While most of the respondents report having technical data-science skills, some admit that their interest in data science is more theoretical and comes from a need to learn to speak the researchers' and students' language when it comes to more technical aspects of data work. Within the interviews, the information specialist calls himself a "data

librarian", while describing his job role, which matches with the definition I have given of a data librarian; the profile of the data scientist also matches my assumptions, in that he works for libraries by developing data science products; his role is clearly distinguished from that of a data librarian by the lack of an instructional role.

Data librarians' competence acquisition process is influenced by the roles they hold within the library; due to the shifting role of data science in relation to librarianship, there is no clear universal path that all data librarians can use to improve their competences in this area.

Among both questionnaire respondents and interview participants, data science-oriented programming and communication competences were perceived to be the main competences a data librarian should possess. Danish librarians admitted to having knowledge of various programming languages, the most widely mentioned being Python.

Danish data librarians have knowledge of the following data science-specific areas: open data, data visualization, data cleaning, web scraping, research data management, data management platforms and metadata handling.

It is up to them to adapt to the specific requirements of their own library, and to communicate with their team in order to come up with a plan for competence development. Depending on where they are in their career, there are several available opportunities for further skills acquisition.

Librarians who are still in education are in a good position to acquire some of the required skills before graduation; for more experienced professionals, there is the option of self-education through online courses, or by participating in workshops such as DST4L.

Library and university leaders are starting to understand the importance that digital literacy instruction holds within the library, as can be seen by the existence of KU Datalabs, which means it might be a good moment for librarians to open a discussion around resource allocation towards further data science initiatives within libraries.

Danish research libraries are placed in a good position to support researchers and students with their research data; some libraries offer consultation-based services such as research data management, others have online guides available for the users, and others provide researchers access to tools for working with their research data – including support in the shape of instructional events. Some libraries offer all of the above, and some go further towards increasing their users' data competencies by creating physical spaces which give them the opportunity to do so.

Several large universities in Denmark have founded data labs in order to provide students and researchers with opportunities for improving their data literacy and to encourage better research data practices across their departments. The concept of data labs incorporated into research libraries is a relatively new concept in Denmark, inspired by similar initiatives in the US; these data labs have an important role within universities, and their placement in Danish research libraries provides opportunities for collaboration among students, researchers and librarians.

The creation of data labs and data-oriented services also requires Danish librarians to have the opportunity to increase their competences in this area. Within the Danish Union of

Librarians, the Data Science Professional Group provides librarians with an interest in this area to join a growing community of specialists in this field. The group provides librarians to increase their data science competencies by arranging events such as hands-on workshops and study groups. The existence of this group is proof that there are increasing opportunities around data librarianship in Denmark.

Despite many of the research libraries not using data science in a formal way in their institutions (by creating data support services or data labs), the Data Science group as well as the successful application of the Data Scientist Training for Librarians show a growing interest of Danish librarians in the area of data librarianship.

Due to data librarianship being an emerging area in Denmark, librarians experience challenges in implementing data services aimed at supporting students and researchers. Most of these challenges are explained by the fact that data science labs and services are not focused on one single niche in which they can excel at; by extending their reach towards many different departments at the same time, it is inevitable that some initiatives will have less success than others. As time passes, the focus of data labs and the roles of data librarians will clarify and concentrate.

As the Danish librarians increase their data science competencies, this will probably start to be reflected on a formal level, through the development of data-oriented services and spaces within Danish research libraries.

5.1. Limitations of the study

It is important to address the possible limitations that may have occurred through the design of this study.

Due to the large number of research libraries in Denmark, there was not enough time to make an in-depth analysis of all the research libraries in Denmark. I tried to overcome this by exploring their websites, and by focusing on the university research libraries. Most of the websites had an English version available; however, for future consideration, an understanding of Danish would have been helpful towards a better analysis.

Materials-wise, the questionnaire that was used to get insights about librarians' data science skills and competences received a relatively small amount of responses. I overcame this limitation by focusing on the quality rather than quantity of the responses, and by using the questionnaire data as explorative material alongside the library website findings.

Biases in data collection may have arised due to the methods used. However, the insights bought by the use of in-depth semi-constructed interviews confirmed some central points in my arguments, which I felt would overcome the amount of bias caused by using a qualitative interview method.

5.2. Perspectives

This thesis stands among the first theoretical initiatives on the topic of data science in relation to librarianship in Denmark.

It uncovers several areas of data science skills and competencies of Danish librarians, which are studied through the lens of their practical application within library data services and data labs. Gathering two specialists directly involved in data science from different angles within the library world enriches the analysis of the data, and creates a discussion around the roles of data librarians within the context of technological changes. This study is not exhaustive, but by gathering and listing some of the data science initiatives in Denmark, provides a good starting point and leaves ample room for further research in this area.

VI. References

[HUMlab]. (n.d.). [HUMlab] - Copenhagen University Library, South Campus: Software & Hardware. Retrieved July 23, 2018, from https://kub.kb.dk/humlab/hard_and_software

365DataScience. (2017). Can I Become a Data Scientist: Research into 1,001 Data Scientist Profiles. Retrieved June 28, 2018, from https://365datascience.com/research-into-1001-data-scientist-profiles/#2

Aalborg University Library. (2018). Software and webservices - Aalborg University Library (AUB). Retrieved July 27, 2018, from https://www.en.aub.aau.dk/software-webservices/

Affelt, A. L. (2015). *The accidental data scientist, big data applications and opportunities for librarians and information professionals*. Medford, New Jersey: Information Today, Inc.

AU Library. (2018). Events. Retrieved July 26, 2018, from http://library.au.dk/en/events/

AU Library. (2018). Find your personal liason librarian at AU library. Retrieved July 26, 2018, from http://library.au.dk/en/researchers/liaison-librarian/

Australian National Data Service. (2013). Information specialists and data librarian skills. Retrieved July 16, 2018, from http://www.ands.org.au/working-with-data/data-management/overview/data-management-skills/information-specialists-and-data-librarian-skills

Ball, A. (2013). Les métiers liés aux données de la recherche : Data Librarian. *43e Congrès Annuel de l'ADBU, 2013-09-18 - 2013-09-20, Le Havre.*, (Dcc), 1–10.

Beckett, S. (2015). What's the difference between skills and competencies? Retrieved June 28, 2018, from https://resources.hrsg.ca/blog/what-s-the-difference-between-skills-and-competencies

Bern, P. H. (2005). You ’ re A What ??? Taking Stock of the Data Profession. *Conference Presentations of IASSIST 2005*.

Blei, D. M., & Smyth, P. (2017). Science and data science. *Proceedings of the National Academy of Sciences*, *114*(33), 8689–8692. https://doi.org/10.1073/pnas.1702076114

Boesgaard, C. (2018). Personal communication, May 24, 2018. Copenhagen.

Borgman, C. L. (2010). Research Data: Who Will Share What, with Whom, When, and Why? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1714427

Burton, M., Lyon, L., Erdmann, C., & Tijerina, B. (2018). *Shifting to Data Savvy : The Future of Data Science In Libraries*. Pittsburgh, PA.

Callahan, D. R. (1991). The librarian as change agent in the diffusion of technological innovation. *The Electronic Library New Library World*, *98*(7), 13–15.

Carlson, J., & Johnston, L. R. (2015). *Data Information Literacy: Librarians, Data and the Education of a New Generation of Researchers*. *Journal of Academic Librarianship* (Vol. 41).

Choi, Y., & Rasmussen, E. (2009). What Qualifications and Skills are Important for Digital Librarian Positions in Academic Libraries? A Job Advertisement Analysis. *The Journal of Academic Librarianship*, *35*, 457–467. https://doi.org/10.1016/j.acalib.2009.06.003

Christensen-Dalsgaard, B., Castelli, D., Ammitzbøll Jurik, B., & Lippincott, J. (Eds.). (2008). Research and Advanced Technology for Digital Libraries, *5173*. https://doi.org/10.1007/978-3-540-87599-4

Cohen, D., & Crabtree, B. (2006). Qualitative Research Guidelines Project: Semi-structured Interviews. Retrieved July 29, 2018, from http://www.qualres.org/HomeSemi-3629.html

Copenhagen Centre for Social Data Science. (2016). About the Copenhagen Centre for Social Data Science.

Crowder, J. A., & Friess, S. (2015). *Agile Project Management: Managing for Success*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-09018-4

Crumpton, M. A. (2015). *Strategic human resource planning for academic libraries: information, technology and organization*. Waltham, MA: Chandos. https://doi.org/10.1016/B978-1-84334-764-4.09984-6

Danish Agency for Libraries and Media. (2011). The Public Libraries in the Knowledge Society. *Focus*.

Data Science Laboratory. (2017, November 18). About the Data Science Laboratory. Retrieved July 23, 2018, from https://datalab.science.ku.dk/english/about/

DATALAB. (2018). About. Retrieved July 26, 2018, from http://datalab.au.dk/about/

Dataquest. (2017). Dataquest - What is Data Engineering? Retrieved July 10, 2018, from https://www.dataquest.io/blog/what-is-a-data-engineer/

Davies, R. (1976). Danish Academic Libraries. *Journal of Librarianship*, *8*(4), 229–243. https://doi.org/10.1177/096100067600800402

DEFF. (2009). The Future of Research and the Research Library: A Report to DEFF Denmark's Electronic Research Library, 1–70.

Department of Information Studies. (2017, March 12). About the department. Retrieved July 4, 2018, from https://informationstudies.ku.dk/about/

Det Kongelige Bibliotek. (2017). The new Royal Danish Library. Retrieved July 27, 2018, from http://www.kb.dk/en/kb/organisation/fusion/start.html

Det Kongelige Bibliotek. (2017). The new Royal Danish Library - The Royal Library. Retrieved July 20, 2018, from http://www.kb.dk/en/kb/organisation/fusion/start.html

Digital Data and Technology Profession Department. (2018). Data scientist: skills they need - GOV.UK. Retrieved June 27, 2018, from https://www.gov.uk/government/publications/data-scientist-skills-they-need/data-scientist-skills-they-need

Dougherty, W. C. (2010). Managing technology in libraries: A look to the past with hope for the future. *Journal of Academic Librarianship*, *36*(6), 543–545. https://doi.org/10.1016/j.acalib.2010.08.011

Driscoll, M. E. (2012). The data science debate: domain expertise or... Retrieved July 15, 2018, from http://medriscoll.com/post/18784448854/the-data-science-debate-domain-expertise-or

DST4L. (2017). Data Science Training for Librarians. Retrieved July 27, 2018, from http://www.dst4l.info/index.html

DTU Bibliotek. (2015). "Data Scientist Training for Librarians" DST4L, afholdes for første gang i Europa den 9 - 11 september 2015 - DTU Bibliotek. Retrieved July 27, 2018, from http://www.bibliotek.dtu.dk/nyheder/2015/06/dst4l?id=218dfe17-1f24-44b6-af75-e932a86a3b4d

DTU Library, & Lavanchy, P. M. M. (2018). Love Data Week. Retrieved July 27, 2018, from http://www.bibliotek.dtu.dk/english/news/2018/02/love-data-week-1?id=a1f66999-6303-4171-9f6d-d8dee1f4ebcc

DTU Library, & Refn, A. (2018). Guide to data management. Retrieved July 27, 2018, from http://www.bibliotek.dtu.dk/english/servicemenu/publish/research-data/guide

EDINA National Data Centre, & Edinburgh University Data Library. (2005). Conference Presentations 2005 | IASSIST Home. In *IASSIST 2005-Evidence and Enlightenment*. Edinburgh.

Edwards, H. (2017). Opportunities in Research Data Support: The Data Librarian's Handbook. *Refer*, *33*(1), 18–20.

Ekstrom, J., Elbaek, M., Erdmann, C., & Grigorov, I. (2016). The research librarian of the future: data scientist and co-investigator. *The Impact Blog*, 2–4.

Erdmann, C. (2014). Teaching Librarians to be Data Scientists. *Information Outlook*, *18*(3), 21.

Farkas, M. (2016). A learning organization: supporting professional development in libraries.(TECHNOLOGY: In Practice). *American Libraries*, *47*(3–4), 27.

Fisher, H. (2010). Competencies for Science Librarians. *Australian Library Journal*, *59*(3), 161.

Galluzzi, A. (2014). Contemporary challenges and public perception. *Libraries and Public Perception*, 97–118. https://doi.org/10.1533/9781780634258.97

Galluzzi, A. (2014). Wondering about the future of libraries. *Libraries and Public Perception*, 1–6. https://doi.org/10.1533/9781780634258.1

Galluzzi, A. (2013). Libraries and public perceptions: A comparative analysis of the European press. Methodological insights. *JLIS.It*. https://doi.org/10.4403/jlis.it-8987

Goodman, A. L. (2014). *Digital media labs in libraries / Amanda L. Goodman*.

Gordon-Murnane, L. (2012). *BIG DATA*.

Gorman, G. E., & Clayton. (2009). Qualitative research for the information professional: A practical handbook. *Library and Information Research*, *33*(104), 68–69.

Goulding, A. (2017). Weathering the perfect storm: Librarians as partners in teaching, learning and scholarship | SAGE Connection – Insight. Retrieved July 1, 2018, from https://connection.sagepub.com/blog/opinions/2017/04/12/weathering-the-perfect-storm-librarians-as-partners-in-teaching-learning-and-scholarship/

Griffiths, J.-M. (1995). The changing role of librarians: Managing new technologies in libraries. *Vistas in Astronomy*, *39*, 127–135. https://doi.org/10.1016/0083-6656(95)00065-U

Haendel, M. A., Vasilevsky, N. A., & Wirz, J. A. (2012). Dealing with Data: A Case Study on Information and Data Management Literacy. *PLoS Biology*, *10*(5), e1001339. https://doi.org/10.1371/journal.pbio.1001339

Heale, R., & Forbes, D. (2013). Understanding triangulation in research. *Evidence-Based Nursing*, *16*(4), 98. https://doi.org/10.1136/eb-2013-101494

Higgins, S. (2016). *Managing academic libraries : principles and practice*.

Hunt, D., & Grossman, D. (2013). *The librarian's skillbook : 51 essential career skills for information professionals*.

Iglesias, E. G. (n.d.). *An overview of the changing role of the systems librarian : systemic shifts*.

Ismail, S. (2008). Competency-Based Training for Digital Librarians: A Viable Strategy for an Evolving Workforce? *Journal of Education for Library and Information Science Fall*, *49*(4).

IVA. (2010). IVA Strategy. Retrieved from https://web.archive.org/web/20101124025952/http://iva.dk/english/aboutus/strategy/

IVA. (2010). IVA Vision. Retrieved November 24, 2010, from https://web.archive.org/web/20101124025200/http://iva.dk/english/aboutus/vision/

Johannsen, C. G., & Pors, N. O. (2010). The 2007 Structural Reform and the Public Libraries in Denmark. *BIBLIOTHEK Forschung Und Praxis*, *34*(3), 342–350. https://doi.org/10.1515/bfup.2010.051

Kajbergand, L., & Ole Pors, N. (1995). The Royal School of Librarianship, Denmark: present educational programmes and envisaged curricula. *Librarian Career Development*, *3*(4), 19–25. https://doi.org/10.1108/09680819510101371

Kang, K., & Kang, J. (2016). Data Science Curriculum in Library Information Science for School Librarians. *International Association of School Librarianship. Selected Papers from the ... Annual Conference*, 1–2.

Katrenko, S. (2015). Q&A: How to use bibliometrics to evaluate interdisciplinary research. Retrieved July 26, 2018, from https://www.elsevier.com/connect/q-and-a-how-to-use-bibliometrics-to-evaluate-interdisciplinary-research

Kelley, K., Clark, B., Brown, V., & Sitzia, J. (2003). Good practice in the conduct and reporting of survey research. *International Journal for Quality in Health Care*, *15*(3), 261–266. https://doi.org/10.1093/intqhc/mzg031

Kim, J. (2016). Who is Teaching Data: Meeting the Demand for Data Professionals. *Journal of Education for Library and Information Science Online*, *57*(2), 161–173. https://doi.org/10.12783/issn.2328-2967/57/2/8

Kim, J., & Angnakoon, P. (2016). Research using job advertisements: A methodological assessment. *Library and Information Science Research*, *38*(4), 327–335. https://doi.org/10.1016/j.lisr.2016.11.006

Kjær, L. (2018). Personal communication, July 24, 2018. Copenhagen.

Københavns Universitetsbibliotek. (n.d.). Digital social science lab: vision. Retrieved July 23, 2018, from https://kub.kb.dk/c.php?g=205510&p=3063885

Københavns Universitetsbibliotek. (2018). Index: KUB Datalabs: Home. Retrieved July 22, 2018, from https://kub.kb.dk/kubdatalabs/home

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

KUB Nord. (n.d.). Data Lab - KUB Nord: Hjem. Retrieved July 23, 2018, from https://kub.kb.dk/datalab

Larsen, G., & Statens bibliotekstjeneste (Denmark), G. (1996). *State-of-the-art of information technologies in Danish libraries*. Statens bibliotekstjeneste.

Lauersen, C. (2016). Librarians as Data Scientists? Damn straight! | The Library Lab. Retrieved July 11, 2018, from https://christianlauersen.net/2016/01/11/librarians-as-data-scientists/

Lavrakas, P. (2008). *Encyclopedia of Survey Research Methods*. 2455 Teller Road, Thousand Oaks California 91320 United States of America: Sage Publications, Inc. https://doi.org/10.4135/9781412963947

Loukides, M. (2012). The unreasonable necessity of subject experts - O'Reilly Radar. Retrieved July 15, 2018, from http://radar.oreilly.com/2012/03/subject-matter-experts-data-stories-analysis.html

Marcum, D. B. (n.d.). Do Librarians Need PhDs? *Information Outlook*, *16*(5).

Martin, E. (2013). Highlighting the Informationist As a Data Librarian Embedded in a Research Team. *Journal of EScience Librarianship*, *2*(1), 1–2. https://doi.org/10.7191/jeslib.2013.1044

Mason, H., & Wiggins, C. (2010). A Taxonomy of Data Science. Retrieved June 25, 2018, from http://www.dataists.com/2010/09/a-taxonomy-of-data-science/

McKnight, M. (2010). *The agile librarian's guide to thriving in any institution*. Libraries Unlimited.

Minton-Eversole, T. (1991). Additional Reading. *Training & Development*, *45*(8), 78. https://doi.org/10.1016/B978-0-323-08691-2.15007-0

Moed, H. F., & Luwel, M. (2014). *Scientometrics as Big Data Science: On integration of data sources and the problem of different types of classification systems*. OECD Workshop.

Molaro, A., & White, L. L. (2015). The library innovation toolkit : ideas, strategies, and programs, 194.

Morawski, J., Stepan, T., Dick, S., & Miller, J. (2017). A Fuzzy Recommender System for Public Library Catalogs. *International Journal of Intelligent Systems*, *32*(10), 1062–1084. https://doi.org/10.1002/int.21884

National Science, B. (2005). NSB-05-40, Long-lived digital data collections enabling research and education in the 21st century.

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

Ogunsola, L. A. (2011). The Next Step in Librarianship : Is The Traditional Library Dead ? *Library Philosophy and Practice*.

Ole Pors, N. (2005). Changing perceptions and attitudes among Danish library managers and directors. *New Library World*, *106*(3/4), 107–115. https://doi.org/10.1108/03074800510587327

Ole Pors, N. (2003). Managing change in Danish libraries. *The Journal of Academic Librarianship*, *29*(6), 411–415. https://doi.org/10.1016/j.jal.2003.08.008

Ørom, A. (2000). Information science, historical changes and social aspects: a nordic outlook. *Journal of Documentation*, *56*(1), 12–26. https://doi.org/10.1108/EUM0000000007133

Ozdemir, S. (2016). *Principles of Data Science* (1. ed.). Birmingham: Packt Publishing.

Popp, M. P. (2012). Changing World, Changing Libraries: New Literacies, New User Needs, and Leadership for Change. *Reference & User Services Quarterly*. American Library Association. https://doi.org/10.2307/refuseserq.52.2.84

Pors, N. O. (1994). The Changing Labour Market of the Information Professional:: Challenges\nfor Library School Education. *Librarian Career Development*, *2*(3), 14–21. https://doi.org/10.1108/09680819410066930

Press, G. (2013). A Very Short History Of Data Science. Retrieved June 25, 2018, from https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/#1951e3cd55cf

Ramzan, M., & Singh, D. (2010). Factors affecting librarians' attitudes toward IT application in libraries. *The Electronic Library*, *28*(2), 334–344. https://doi.org/10.1108/02640471011033675

Rice, R. (2016). *The data librarian's handbook*. London: Facet Publishing.

Roskilde University. (2018). Roskilde University Library Main Page. Retrieved July 27, 2018, from https://ruc.dk/en/roskilde-university-library

Roskilde University's Digital Media Lab. (2018). About – Digital Media Lab. Retrieved July 27, 2018, from https://digitalmedialab.ruc.dk/about/

Scupola, A., Klinger, K., & Tosheff, L. (2008). E-Services in Danish Research Libraries : Issues and Challenges at Roskilde University Library, 204–217.

Shahbazi, R., & Hedayati, A. (2016). Identifying Digital Librarian Competencies According to the Analysis of Newly Emerging IT-based LIS Jobs in 2013. *The Journal of Academic Librarianship*, *42*(5), 542–550. https://doi.org/10.1016/J.ACALIB.2016.06.014

Stanton, J. (2012). An Introduction to Data Science. *Syracuse University*, 1–157. https://doi.org/10.1007/978-3-319-50017-1_1

Statistics Denmark. (2018). Forsk1: Research libraries with special obligations by region and activity. Retrieved from www.statbank.dk/FORSK1

Statistics Denmark. (2018). Research Libraries: Statistical presentation - Statistics Denmark. Retrieved July 19, 2018, from https://www.dst.dk/en/Statistik/dokumentation/documentationofstatistics/research-libraries/statistical-presentation

Svendsen, M., & Lauersen, C. U. (2015). Digital Social Science Lab: incubating data information literacy. In *Danish Research Library Association's Annual Meeting*. Aarhus. https://doi.org/10.5281/zenodo.30985

Svendsen, M., & Lauersen, C. U. (2015). Digital Social Science Lab - incubating data information literacy (Poster, presented at Annual meeting of the Danish Research Library Association, sept. 17, 2015). https://doi.org/10.5281/zenodo.30985

The Royal Danish Library. (2018). Royal Danish Library - Organization Chart. Retrieved July 27, 2018, from http://www.kb.dk/export/sites/kb_dk/en/kb/organisation/dir/Organisation-1-5-2018-ENG.pdf

Torstensson, M. (1993). Is There a Nordic Public Library Model? *Libraries & Culture*, *28*(1), 59–76. https://doi.org/10.2307/25542505

Toy, B. J. (1978). The Role of the Academic Librarian: A Symposium. *Journal of Academic Librarianship*, *4*(3), 128–130.

Tsang, D. C., & Renaud, J. P. (2014). Challenges in Developing A New Library Infrastructure for Research Data Services. *IASSIST Conference 2014*.

Tukey, J. W. (1962). The Future of Data Analysis. *The Annals of Mathematical Statistics*, *33*(1), 1–67. https://doi.org/10.1214/aoms/1177704711

University of Southern Denmark. (2018). FabLab - University of Southern Denmark, SDU. Retrieved July 27, 2018, from https://www.sdu.dk/en/fablab

Winter, M. F. (1983). The Professionalization of Librarianship. *University of Illinois Graduate School of Library and Information Science*, (160), 52.

Writers, S. (2012). Making computer data storage cheaper and easier. *Space Daily*.

Xia, J., & Wang, M. (2014). Competencies and Responsibilities of Social Science Data Librarians: An Analysis of Job Descriptions. *College & Research Libraries*, *75*(3), 362–388. https://doi.org/10.5860/crl13-435

VII. Appendix

Item 1 - Areas of competency comparison

Description of competency area	**Data scientist competency areas**	**Research librarian competency areas**	**Data science librarian competency areas**
Innovation	Openness to change and improvement	Continuous professional development and capacity to innovate within the library, according to the user needs (Stern, 2009, p. 14)	Building an innovation-oriented spirit within the library; embracing change and taking up on new learning opportunities
Collaboration	Collaborative, team player mind-set	Outreach to faculty, and students, building a support network (Stern, 2009, p. 21)	Establishing themselves as a link "between research challenges and their solutions." (Affelt, 2015, p. 190)
Performance management	Delivering timely performance; taking responsibility for outcomes	Ability to manage collections and carry out other responsibilities in the scholarly environment	Making sure "results are thoroughly analyzed, data meets the highest standards, the sources are reliable, and alternative interpretations of results are considered" (Affelt, 2015, p. 191)

Adding value	Seeking out the best quality solutions	Assessment of collections; expert knowledge and ability to evaluate content and format of information resources (Stern, 2009, p. 14)	Helping the data science team "consider all aspects of possible solutions, as well as identify pitfalls." (Affelt, 2015, p. 191) Knowledge of core data science competencies (enough programming skills to test hypotheses and algorithms) (Affelt, 2015, p. 192)
Communication	Effective communication and showing pride and passion in their work	Ability to communicate and reach out to faculty, administration, students; pedagogical competencies (Stern, 2009, pp. 20-21)	Ability to speak both "programmers' language" and "clients' language"; storytelling abilities;
Awareness of role	Seeing the bigger picture of their role in the organisation	"Deep understanding of the mission, vision and objectives of their institution" (Stern, 2009, p.15)	Understanding of their role in the data science process;

Item 2 - Questionnaire design

Section 1 – Profile

Gender	○ Male ○ Female
Age	○ Under 30 ○ 31-40 ○ 41-50 ○ 51-60 ○ Over 60
Region	○ Capital Region ○ Northern Jutland ○ Southern Denmark ○ Central Jutland ○ Zealand
Education	
Place of employment	
Work sector	○ Public ○ Private
Spoken language	○ Danish ○ English

Section 2 – Experience

What programming languages do you have experience in (Python, JavaScript ...)?	

What tools / software do you have experience in (OpenRefine, Tableau, Git ...)?	
What topics do you have experience in (data visualization, open data, machine learning ...)?	
How do you work with data science in everyday life?	

Section 3 – Interests

Which programming languages do you want to learn to use (Python, JavaScript ...)?	
What tools / software do you want to learn to use (OpenRefine, Tableau, Git ...)?	
What topics are you interested in learning about (data visualization, open data, machine learning ...)?	
What would you like to have from the DS professional group (meetings, workshops ...)?	

Section 4 – Contact

Where would you prefer to receive information related to the DS group?	○ Facebook ○ Twitter ○ LinkedIn ○ Email ○ On the group's webpage
Do you have any other wishes or comments regarding communication and social media in the professional group?	

Section 5 – Giveaway information registration

Based on the submitted responses and to show our appreciation of our members, we will issue a small gift. If you wish to participate in the draw, please provide your name / email and phone number.	

Item 3 - Interview Guide for Interview A – Christian Boesgaard, Lead Data Scientist at DBC

The purpose of this interview was to get Christian's perspectives on data science in Danish libraries as well as uncover opinions related to librarians' roles and competencies around data science.

Research topic	**Interview question**

Introductory questions; establishing a starting point, data science history at DBC	• Tell me about your work at DBC.
	  ○ What motivated you to choose this path?
	  ○ How was the data science department created?
	• What product are you developing right now?

Discussing the role of librarians;	• I see you have been working with statistics and machine learning to help offer library users better search options. Do you think this might also change the tasks of library staff? If so, in what ways? ○ Will this cause librarians to be more or less involved with assisting users' information search processes?
Defining data science in a library context	• What is data science, and what does it mean for librarians and DBC's work? • In a library context, do you think there is a difference between "data" work and "IT" work?
Librarians' data skills; expectation of librarians' skills, from IT specialist perspective	• Considering this changing environment, what are some ways librarians can keep up to date with their IT/data skills? • There have been some initiatives (DST4L, the Data Science Faggruppen) to train librarians with the purpose of becoming more data-savvy. Still, being a librarian can be quite a specialized area of study, and many feel like they already "have a lot

	on their plate" in their daily jobs to begin with. Is there really a need for a librarian to be data-savvy? Or is it better to leave all data problems to IT specialists?
Future perspectives of data science in the library world	• What role do you think data science will play in the library world in 5 years?
End of interview	• Thank you for your time, Christian. Is there anything else you'd like to add?

Item 4 - Interview Guide for Interview B - Lars Kjær, Information Specialist and HUMlab coordinator at KU's Humanities Library

The purpose of this interview was to get insights into the role that data labs can play within a library, as well as get an understanding of the placement of data librarians in relation to students and researchers, by having the information specialist reflect on his experience and challenges within the library and the data lab.

Research topic	**Interview question**
Introductory questions; establishing a starting point, understanding the background of the data lab and its placement in relation to the library.	● Tell me about your work at [HUMlab]. ○ How was the lab created? ○ What inspired you personally to get involved with [HUMlab]? Why is [HUMlab] positioned inside the library space? · What do you think is the most frequent usage of [HUMlab]'s space (e.g. individual or group study sessions, workshop attendance)
Discussing the role of information specialists in a data lab.	· I see that you have established some initiatives at [HUMlab], such as creating study groups, holding data sprints and generally providing support for digital humanists using your lab. How does the team collaborate and divide between these tasks? · How does [HUMlab] work interfere with your other tasks within the library?
Defining data science in a library context	● What is data science, and what does it mean for librarians? ● Do you think acquiring knowledge of data science methods is a direction any librarian would benefit from? Why or why not?

Librarians' data skills; expectation of librarians' skills, from information specialist perspective	• What are some ways that already established librarians can keep up to date with their IT/data skills? • There have been some initiatives (DST4L, the Data Science Faggruppen) to train librarians with the purpose of becoming more data-savvy. Still, being a librarian can be quite a specialized area of study, and many feel like they already "have a lot on their plate" in their daily jobs to begin with. Is there a need for librarians to take up this extra role of "data literacy instructor" – or would it be better to assign such a role to instructors with an IT background?
Future perspectives of data science in the library world	• What role do you think data science will play in the libraries in 5 years?
End of interview	• Thank you for your time, Lars. Is there anything else you'd like to add?