

Using Personas to Visualize the Need for Data Stewardship.

Live Kvale Oslo Metropolitan University, Norway

Abstract

There is a current discussion in universities regarding the need for dedicated research data stewards. This article presents a set of fictional personas for research data support based on experience and requests by experts in different areas of data management. Using a modified Delphi study, twenty-four participants from different stakeholder groups have contributed to the skills and backgrounds necessary to fulfill the needs for data stewardship. Inspired by user experience (UX) methodology different data personas are developed to illustrate the range of skills required to support data management within universities. Further as a competency hub for data stewards the development of a research data support center is proposed.

Introduction

Data are the entities researchers draw conclusions from, and essential for fellow researcher to examine and criticize results. Transparency and access to data, the analysis applied, and the conclusions drawn are part of what defines research.¹ Data sharing and data archiving is expected to resolve the reproducibility crisis in research and provide new insight². Consequently, academic journals and research funders are increasingly requiring research data to be made available.³ Along with requirements for sharing data in academic research, there has been a growing need for new skills for data managers, data stewards, data librarians, and data scientists.⁴ These new roles are professionals who assist researchers in managing research data, avoiding data loss during the research process, and preparing the data for archiving and public access. Digital research data are easily lost, and steps to preserve data must be taken in all stages of the research process.⁵ Consequently, skills to maintain and curate data are required, but which skills are needed? And where in the universities should curation services be offered? These questions are currently being explored⁶ and debated in libraries and among infrastructure providers.⁷ This paper draws on a study of stakeholders involved in research data management in Norway involving policy makersⁱ, national infrastructure providersⁱⁱ⁸ and researchers and research support staffⁱⁱⁱ from the four oldest universities in Norway. By using persona templates adapted from user experience (UX) methodology⁹ this paper explores how the data stewards are described by different stakeholders. The aim with the making of the personas has been to visualize how a data steward team could respond to the various necessary competencies and skills needed for data management support.

Internationally, “data steward” is one of several terms used in the literature and among practitioners to describe a person working with research data management (RDM). “Data librarian”, “data manager”, and “data curator” are examples of other titles with somewhat overlapping responsibilities.¹⁰ The term data steward is used in this article, as it is less domain specific than “librarian,” “curator,” or “scientist.” The usage of “data steward” is intended to include all the different requirements for data management

The research question investigated is:

ⁱ Representatives from the Norwegian Ministry of Education and Research, the research council of Norway and the rectorate of one of the included universities

ⁱⁱ In Europe and Norway there is a strong tradition for national custodians of research data.

ⁱⁱⁱ University IT, library and research office

Who are the data stewards in the universities?

- a. What roles should data stewards play
- b. What services should data stewards provide as part of these roles
- c. What skills do data stewards need to carry out these services

By developing a set of data personas, it becomes possible to illustrate and exemplify one possible response to each research questions; it is not to be interpreted as a universal solution, but rather as an example of how roles, skills, communication, and services for data management may be organized. The findings also focus on potential obstacles and what to be aware of when developing data steward services.

Literature Review

A broad range of literature on RDM skills were identified through searches for "data steward", "data librarian", "data manager", and "data curator" in Web of Science and Scopus. These articles were supplemented by searching relevant journals that are not indexed in these databases, such as *JesLib* and the *International Journal of Digital Curation*, and adding other relevant documents. The different articles highlight the skills required in data management and the different roles of data professionals. The articles were grouped into three categories according to how the data steward was described: 1. new responsibilities of the librarian, 2. the embedded data steward in the research environment, and 3. other approaches to data management services. In addition, the literature review contains a section on the usage of personas related to data management services.

In the library and information science literature a majority of articles on data stewardship aim at clarifying which skills are needed for the data professional librarian offering data management support to researchers at the university.¹¹ Both Brown and Federer emphasize that "support for researchers' data needs is a moving target"¹² that needs to be supported by a skills development program in libraries.¹³ The most important skills identified by Federer relate to communication, presentation, relationship with researchers, teamwork, and one-to-one training.¹⁴ This argument is supported by Kennan, who finds that communication skills in many forms were the most in demand for RDM positions; she further emphasizes the need for "boundless curiosity", including both the willingness and ability to learn new things.¹⁵ Kennan identifies four different roles that ensure data management in the different stages of the data life cycle: "the data librarian/data manager", "the data IT and systems experts," "the data scientist", and "the data creator".¹⁶ Cox and Corral illustrate the role of the "research data manager"¹⁷ in the breach between the faculty and the academic library, connecting the institutional repository manager role in the library with the research produced by the faculty. The data librarians described can either be skilled generalists in data management or be specialized in a particular discipline. Disciplinary specialization can be achieved through engagement with subject specialists and researchers.¹⁸

A data steward working in a research group or similar research environment with data management is here referred to as embedded data steward. These domain-specific data stewards are primarily used in data-intensive research within health sciences¹⁹ and natural sciences²⁰ and specialize in data management in a single discipline. An editorial from *Nature Genetics*²¹ starts with a clear statement regarding data stewardship, asserting that, "professional data stewards be trained and employed in all data-rich research projects, [which] raises the exciting prospect they will conduct research on data-intensive research itself."²² Some articles describe solutions for data management within national research institutes²³ or data centers.²⁴ The articles on embedded data stewards are discipline specific and involve a high degree of specialization with a focus on the development of best

practices and domain specific standards.²⁵ This illustrates how embedded data steward needs to understand the methods and data they are working with in addition to preservation and metadata. None of the articles describing data stewards in research environments are from the humanities or the social sciences. These disciplines have traditionally been less data intensive and research are often conducted without data sharing among collaborating researchers. Also, the humanities and social sciences cater to needs differently; such as the trend of digital scholarship centers run by university libraries that explicitly serve the field of the digital humanities.²⁶ These are some possible explanations to why the experiences with embedded data stewards in the humanities and social sciences are fewer and newer which again could explain why examples of embedded data scientists in humanities and social sciences and have not yet reached the literature.

While embedded and library-centric were the two large categories to be found in the literature, there are other approaches to data management services. One example is the one-stop research support described by Clements²⁷ where someone can find answers to all questions regarding research data in one place, possibly a web portal. Another approach by Delft University in the Netherlands places domain-specialized data stewards within the faculty departments.²⁸ The service is coordinated by the library but aims to integrate the services of the data steward in each faculty. Still, their goal is to provide “more granular disciplinary experts.”²⁹ The report from Research Libraries UK and Matt Greenhall exploring digital scholarship in UK libraries argues for a “mixed economy of digital scholarship support”³⁰ whereby the library partner supplies other research support facilities at the universities with complementary expertise in data management. In literature on data scientists the term “Data unicorns”³¹ is used in the meaning an unrealistic skillset for one person. Kennan transfers this to the idea of the data steward.³²

What the literature on data stewards has in common is the exploration of professional domains and services new to librarianship. The primary challenges described include the targeting of the right level of specialization versus the general knowledge of data management and communication and collaboration between the different levels within the organization.³³

In the context of RDM, there are three examples of the usage of personas³⁴ within the literature. Lage builds on the usage of personas to improve institutional repositories for publication.³⁵ Crowston includes “Abby the Science data librarian”³⁶ in the group of users of research data repository. Both focus on the researcher, presenting five³⁷ and eight³⁸ researcher personas with different needs in regard to data management and different interests in using institutional archives for research data. A recent report on education for data stewards from Denmark³⁹ presents the use of personas to illustrate the different needs and skill sets requested for data stewards in both the corporate and research sectors.

Methodology

RDM is a rapidly developing domain. In order to grasp some of the changes and developments, a Delphi study with an expert group and multiple rounds of data collection was found suitable.⁴⁰

The expert group of participants in a Delphi study provided the possibility of bringing preliminary findings back for discussion contributed to the understanding of the perception of roles through negotiation, testing, and learning. A group of twenty-four stakeholders participated in the study (Table 1). The group contained representatives from policymakers and national infrastructure providers in addition to researchers and research support staff from four universities in Norway. Recruitment of participants from different stakeholder groups was to include different aspects of the development of the sociotechnical infrastructure for research data and potentially uncover gaps or disagreements. The data steward was one element highlighted from multiple stakeholders as a gap

or missing link. The four universities are the oldest in Norway, are all multidisciplinary and have well-established collaborations on administrative and technical infrastructure. From the policy makers, rectors of research at the four universities were invited to participate^{iv} in addition to representatives from the Norwegian ministry of knowledge and research and the research council of Norway. The infrastructure providers represent different organizations that offer data archiving services to universities in Norway^v. The researchers were invited based on the receipt of European Union (EU) funding. The EU requires data management plans from the projects they fund. Researchers were identified through the cordis web page^{vi}, of 25 invited researchers eight participated in the study. This way of identifying and recruiting researchers was done to avoid potential biases related to engagement with data management as a topic. It also gave a pool of researchers with different disciplinary backgrounds (biology, musicology, science studies, economics, neuroscience, psychology, philosophy, gender studies). The grouping of researchers as working either individually (RI) or collaboratively (RG) was done during the analysis of data from the first round of collection, as the needs described corresponded with how the researchers collaborated with other researchers on data, rather than with disciplinary backgrounds. The research support staff were recruited with the focus of including representatives from three types of research support services (library, IT, and the research office). Five of the research support participants also had previous experience as researchers, and two provided IT services that were offered both locally and nationally. In some cases, the participants did not want their statements to be identified with them, these have been marked as “off-record”.

Table 1. The participants organized according to role

Role/stakeholder category		Individual Participant codes			
Researchers working individually	RI	RIZ	RIJ	RIL	RIB
Researchers working in groups	RG	RGV	RGD	RGA	RGW
Policymakers	PO	POU	POS	POK	
Infrastructure service providers	IN	INH	INO	INR	
Research support IT	IT	ITE	ITY	ITI	
Research support, research office	RO	ROC	ROX	ROT	
Research support, library	L	LM	LP	LG	LN

Within UX methodology, personas are commonly used to describe users of computer systems.⁴¹ The development of personas builds on data collected through interviews or surveys, with the aim of creating fictional characters, either based on the participants or to fill the roles they describe. By creating personas, system developers flip the focus from the system to the user, aiming to create a product that fits well for some users, rather than merely adequately for everyone.⁴² With the ongoing changes in the data management landscape and current infrastructure development, the data stewards will be central users. Still, who will fill these roles is not clear.

By using the expert group of the Delhi study to develop data stewardship personas, this study is not claiming to offer a universal solution but provides an illustration of how the different roles could be

^{iv} Unfortunately, only one of the four invited rectors agreed to participate.

^v These three are all publicly funded and offer different archive services.

^{vi} <https://cordis.europa.eu/en>

distributed. Data steward personas can be useful both to system developers and to the universities that employ data stewards and develop data management services.

Figure 1 illustrates the different phases of the study. In the exploration phase (January 2018), open interviews approximately one hour long were conducted with the participants. Data stewardship and skills for data management were but two of the several themes brought up in the interviews.

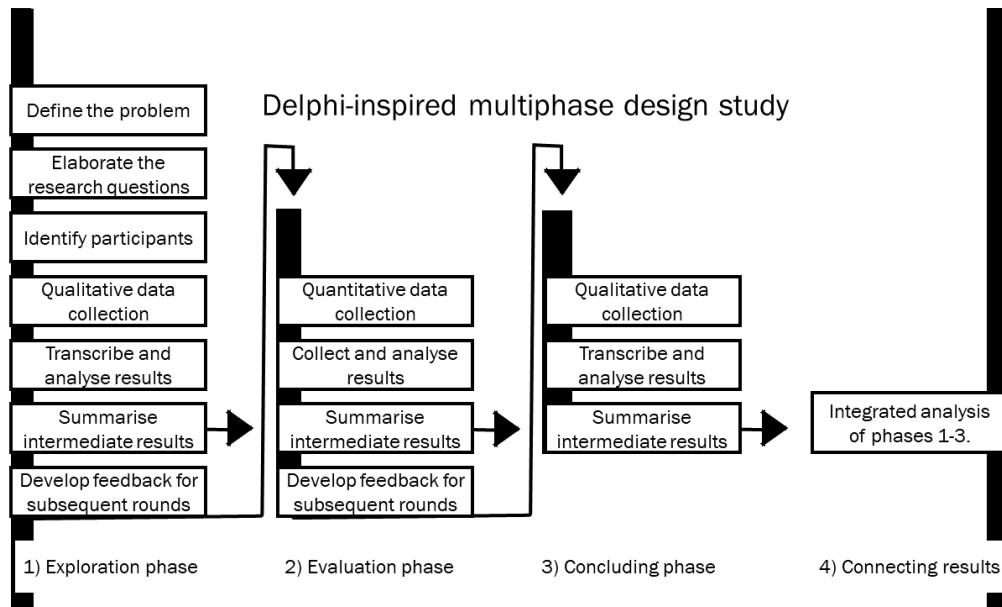


Figure 1 Delphi-inspired multiphase method⁴³

To further explore the expectations the different stakeholders interviewed held regarding data stewards, the second round of data collection (September 2018) had a section dedicated to the data steward (Appendix 1) inspired by UX-persona design. All answers were given in free text and were optional.

In the concluding phase (March 2019), a first draft for three personas was developed based on the preliminary findings. This draft was presented and discussed with the participants in open interviews lasting about 30 minutes. The persona drafts were shared with participants prior to interviews as part of the interview guide. The findings presented in this article are from all three rounds of data collection and include an integrated analysis⁴⁴ of the results. Quotes presented in this paper are marked with a participant code and 1 or 2, referring to first or second interview e.g. “RGA2”. Data from the survey are not linked to participant.

The collected data were, for the most part, qualitatively coded and analyzed thematically,⁴⁵ initially using the software NVivo and later using XML for thematic coding and Python script for extraction. Some of the results from the surveys, such as background, education, and skills, were counted and treated quantitatively.

Most participants granted permission to share the whole or parts of the data with directly identifiable information such as names removed. They all had the opportunity to review data they contributed ahead of publication and to indicate if there were parts they did not want published, or only to be identified “off the record”. The data, including the XML codebook, Python script, interview guides, transcripts, survey and consent forms can be accessed through Zenodo.⁴⁶

Findings

The findings first present the need for data stewardship before exploring in greater detail the skills and background requested for data stewards, which are used in the development of the personas.

The need for data stewards

Several participants pointed towards a need for data stewardship. The vocabulary used to describe this need varied along with expectations as to how this role should be filled. The practical challenges of data planning, data management, and data curation were explored, along with collaborative skills between existing research support services, data stewards, and researchers.

Data management does require knowledge of research. Respondent ITE (research support, university IT cf Table 1) emphasized that, “the researchers know their data so only they will be able to describe their data, but they need help from the data curators” (ITE1). The data curator role described by ITE is defined as working with the departments to create continuity and preserve valuable digital data. ITE also believed in employing data curators to avoid data loss when temporary staff leave and further describes the need for data stewardship and data management as a consequence of data-intensive research. Among the researchers, RIZ feared that the general data steward would not be able to understand the context: “To do this type of job you must know the context, and to do this on an industrial scale might work in some cases but probably not in all” (RIZ1). As RIZ pointed out, some data types might be easy to structure and organize with a lower degree of specialized knowledge, whereas other types require a higher degree of specialization and domain-specific knowledge.

Understanding when different types of knowledge are required is also an issue, as well as understanding what one can expect researchers to do themselves versus what they need additional expertise to do, such as making data interoperable and creating a data management plan (DMP). Participant ROC addressed the long-term perspective: “I don’t believe any researcher can have the responsibility to follow the data from collection [...] until they are ready to be stored for maybe 1000 years.” Making decisions on what should be selected for long-term storage itself requires expertise in addition to performing the actual data preservation.

Two of the researchers working in larger collaborations have hired, or are in the process of hiring, data stewards. RGA works on a multidisciplinary project that generates a large volumes of data from a variety of sources, while RGD collects social science data from previous projects for reuse in a new context. Both agreed on the need for data management: “The largest need is for human recourse to manage data” (RGA1) and, “For us, research data means how to integrate data from all these sites, how to harmonize, standardize, and integrate them, and then how to analyze them in a way that something new comes out of that” (RGD1). RGD also described how several people are working with different aspects of data management, from data cleaning and access control to the re-collection of consent from participants.

Collaboration is also suggested as a challenge: “I believe it is important with such a holy trinity that IT, library, and administration could become if they would work together” (RGA1). She pointed to the need to combine different people with different skills and different backgrounds to solve complex issues and to create robust data management services. Also, one of the research office staff noted that collaboration is key: “There is not one such person, one that knows everything, it is more than a Kinder egg, more than three things, at least four or five things you need to have thorough knowledge of” (ROT1). Metaphors such as “Kinder egg” or “Holy Trinity” have similarities with the “data unicorn,”⁴⁷ indicating that expectations for research support services for research data ought to be collaborative to deliver the complexity of skills required. In the survey, one participant from the

library explained that, rather than a person, she saw as the best solution a team of people with different competencies complementing each other.

Researcher RGW explained that data management was the responsibility of the professor in charge of the lab: "In our group we don't actually have a data manager, but it is mostly the job of the professor. The data type has been fixed a couple of years ago that the data should be analyzed in such and such a way, so it has the same data structure. The professor acts like a data manager also. But because we are temporary researchers, and we have our own style, [the] professor should decide the data structure" (RGW2). As the majority of the researchers in the lab are there temporarily, it is the responsibility of the lab, and the professor who decides the structure and formats of the data, to "act like a data manager." Still, since RGW's description pointed to a high level of awareness, there is likely a formal or informal protocol for data management in the lab, and the responsibility belongs to the principal investigator. Also, among research support staff, it is agreed that the researchers themselves should be responsible for knowing basic data management: "I think that as a researcher, I would not say you are obliged to, but you should know basic data management" (LM2). This does not, however, exclude the need for dedicated data managers and further highlights the need for available training.

The participants described the following needs for data stewardship:

- As research is becoming increasingly data intensive, larger research groups may need to hire data managers.
- Data loss from PhDs, postdocs, and other temporary staff when leaving the university is a challenge.
- To find the right balance between the generalist and the specialist is important in terms of playing the right data management support role.
- A closer collaboration between IT, library, and research offices is needed.
- All researchers cannot be expected to do data management on their own, yet it is the responsibility of the researcher to ensure good data management in his or her research.

Collaboration and communication between the different support levels

In the survey, there were 20 responses to the question regarding the workplace of the data steward, which showed a general agreement that closeness to the research environment is essential. In particular, the researchers emphasized that the employee should work in the research groups. The research group or/and departments (10) were mentioned most frequently, but the research administration (4) and the library (5) were also suggested as appropriate work environments for the data steward. One suggested the national research data infrastructures as the appropriate place to employ the data steward. In the interviews, several participants elaborated on this by emphasizing how collaboration and communication between different levels of support within the universities are crucial:

You need to create a system where these people actually work together and are able to interact in a good way. [.....] There is a pulverization of responsibilities absolutely everywhere, and with such a research data center it might be possible to avoid this, given the entry points and the information flow and such. (RGA2)

Both responses show how the workplace of the data steward is one issue to consider, but many challenges are related to organizational culture, organization, and information flow. One of the researchers suggested coordination between the universities to ensure standardized and high-quality

services: “You need a way to assure that you even out the pressure from place to place, so you don’t end up in one bubble, each with the development of strange subcultures; this is important to avoid, difficult to avoid but very important” (RIJ2). The workplaces of the data stewards need to be interconnected in networks of information and skills exchange locally and, perhaps, nationally and internationally. As RIJ notes, hiring data stewards without facilitating knowledge exchange can easily create dysfunctional subcultures rather than interoperable data.

Speaking the same language

The respondents (21) mentioned several educational backgrounds in different combinations; the results have been split and grouped in Figure 1. Further, four mentioned the master’s level, and five mentioned the PhD level as the appropriate educational level. Others suggested higher education without specifying the degree level. The respondents all suggested that the ideal candidate would be a highly educated person preferably with research experience, often in combination with a background in data stewardship, IT, or library and information science (LIS).

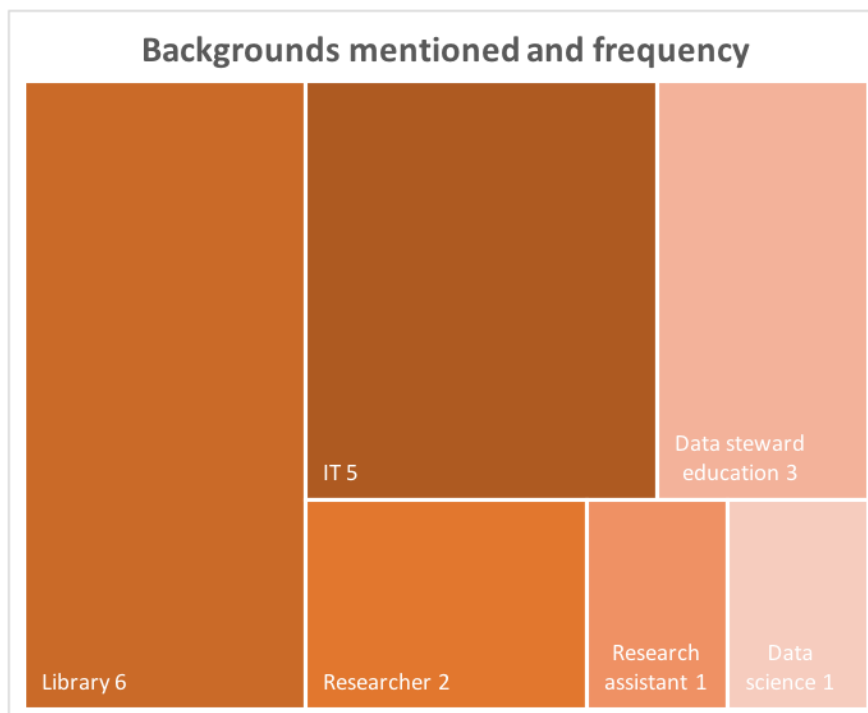


Figure 2 Preferred background for data stewards

As one of the staff members at the university explained, a PhD degree can be a gateway to communication with the researchers to “speak their language” and create trust:

It helps if they all speak the same language. That is part of the success in my department, where half of the staff have a PhD, so we can communicate with the researchers. [...] You first need some positive and some negative experiences in order to make the transition. Someone has done this internally [...] others who have not committed huge mistakes yet, they just continue to build, data on data on data and more data, without any control. (IT12)

The notion that experience of research can be one way of creating trust and knowledge of the data types and methods used in the field is another point of entry. However, ITI believed that the need for data management must often be experienced by the researchers before it can be taken seriously.

Similar backgrounds help in creating relational bonds and trust between researchers and data stewards:

I think this is also a kind of confidentiality. There is something like a role you trust, like that person is to be really trusted, so I think it would be, I don't know, but if I was a researcher I would be a bit, I don't know maybe awkward to contact somebody who is just a data manager and is not related closely to my field. (RGD2)

One of the researchers described a fear that the data stewards operate using their own agendas:

If you enter on the side in that mean of adding an additional agenda beyond solidity and such, that sometimes that might be an advantage for some types of projects, for others it might be alarming, both economic and in terms of work environment. (RIJ2)

Research experience among data stewards or similar disciplinary backgrounds are possible strategies to create common ground between data stewards and researcher. These strategies might also help to avoid additional agendas on the part of data stewards. Interest in research or the research topic might help to assure the researchers that solidity and reproducibility are the stewards' primary motivations for data management. There are, however, already several agendas present in the field of data management, such as economic interests,⁴⁸ and an interest to explore existing data in new ways through data science.⁴⁹ For the researchers, on the other hand, the purpose of data management is primarily to document and archive research data for their own reuse and for proof of reproducibility. One of the policymakers pointed to this conflict of interests and motivations: "A risk in this area, and what we have seen until now that the area suffers from, is that library and archive people, bureaucrats, and non-researchers have taken a strong role of leadership" (off record). When policymakers, archivists, IT developers, data scientists, and librarians all see different potential in research data, these interests might come to overshadow the core: the quality of research and the challenge in overcoming the reproducibility crisis.

The question of what motivates the data steward in doing their job becomes important for building relations between the researchers. Twelve participants answered this question. Ethical motivations and genuine engagement in research were seen as the most important motivations: "Engagement both with good research and ethical data management," "the enjoyment of assisting researchers in taking care for their data and sharing data in a safe way," and "[Contributing] to making research transparent and verifiable." Other responses described a methodical person with a genuine interest in research who can provide a valuable contribution by organizing, providing services, building something together as a team, and contributing to science

Dividing tasks but maintaining responsibility

When asked to write a short biography, nine participants responded. One of the descriptions given was that of a "technical and tidy person," and other characteristics included a good overall understanding of research and of the research data life cycle: "The person must be mature or experienced enough to understand the range of the field of data management and curation, and the limitations for what should be shared and [to] understand the whole lifecycle of research data in projects". Another participant described, "a service-minded person able to work closely with several research teams." Thus, both emphasize that technical and social skills are necessary, along with experience, knowledge, and the ability to provide professional guidance. One participant gave a longer description of a researcher who wanted to work in-depth with data and who enjoys both the service and the problem-solving aspects of data stewardship.

Balancing the interest in the research with motivations to keep the data structured and documented

to enhance the quality of the research results without adding additional agendas is important. Still, the involvement of the data stewards must be balanced in such a way that the responsibility of the research data is not completely transferred away from the researchers:

You hope that data management should become embedded in normal research practice, for much of this can, with fairly simple means, become part of existing routines. [...] Because the problem, if you get a data manager in the group, is that the others might not take as much responsibility for the data management (LM2)

The interviewed researchers shared the concern of LM. A data steward must provide support without creating an excuse to transfer the responsibility from the researchers; when data are deposited in an archive, a transfer of responsibility can take place:

The researchers themselves must be responsible [...] I realize this myself in part of these discussions, that one thinks 'yes, we create this role, and then everything is solved,' but it is not at all in that way. Because the researcher sits with the data set and needs to make sure this is in order, and then you need a curation function, but that again depends on the data set and where you are in your research process. [...] But from the moment we have a publication, with a corresponding data set, made available, then the data set will still need curation, but then you are more on the library side. First the researcher needs to sign off the responsibility, and then others take it on. (RGA2)

Another option proposed by RIZ is not to create data stewardship positions, but to distribute responsibility among existing researchers in a group:

I would say that the competency should be in the group, and not in an extra position; I believe there are other positions more important to prioritize, so I guess I am against all these, but the nearer the better. (RIZ2)

RGA and RIZ work in extremely different research environments: while RGA works in a collaborative and data-intensive environment, which employs its own data managers, RIZ is a theorist and collaborator with other researchers on publications. RIZ's point of assigning responsibility for ensuring data quality to the researcher is representative of the view of many researchers, in particular those working independently. She argued that the quality of your data is the quality of your research, and your responsibility as a researcher.

Fifteen participants listed different skills as being necessary for data stewardship; some skills were mentioned several times. The skills mentioned are analyzed and grouped in Table 3. Different labels, such as personal skills, general skills, research skills, knowledge of law and policy, technical skills, and archiving skills, differentiate the variety of skills listed. The label "general skills" is used for skills that are found to apply to more than one of the other categories. Knowledge of metadata are most commonly mentioned. However, none of the researchers mentioned metadata explicitly. There is one mention of "data management and storage for further use," while another writes about "coding, systematization and law"; this is the response that best reflects the feedback from the researchers, along with responses that emphasize personal skills, such as creativity, punctuality, and good communication skills.

Table 3. Data stewardship skills (times mentioned in prentices)

Personal Skills:	General skills:	Research skills:	Law and policy:	Technical skills:	Archiving skills:
Structured and organized (4)	Knowledge of research (4)	Knowledge of discipline specific terminology (2)	Understanding and interpretation of policies (3)	Programming, Coding, Scripting (4)	Metadata related (6) (here under: metadata demands, standards, documentation, descriptive metadata)
Accurate (5)	Research ethics (3)	Ability to understand discipline specific needs (1)	Knowledge of law and juridical aspects (2)	Technical aspect of data management (1)	Familiarity with organizing and planning for different types of research data (2)
Dialog with end user/ communication (2)	Knowledge of the FAIR ^{vii} principles. (1)	Statistics and methodology (1)	Define policies (2)	Ability to work with large databases and LIMS (2)	Systematization (2)
Creative (1)	Data management and storage for further use (1)		Personal privacy (1)	Digitization (1)	
Flexible (1)	Ability to work with guidelines and documentation (1)		IP-law (1)	User interface (1)	Search (1)
A problem solver able to think outside of the box (1)			Familiar with DMP procedures (2)	Data transformation (1)	Data archives (1)
Good listener (1)					Archival standard for curation and secure long-term archival storage (1)

The Personas

Based on the analyses, the placement of a support service in the right context, and with appropriate channels of communication and collaboration, appears to be one of the major challenges of delivering appropriate services. As a workplace for two of the data steward personas, the Research Data Service Center (RDSC) has been developed. The RDSC draws on inspiration from the development of digital scholarship centers, however with a multidisciplinary approach and with the emphasis on strengthening collaboration between the different research support services within a university. Several participants requested better collaboration in order to provide better data management support, the suggested RDSC is one response to this. In the RDSC, the library, IT, and the research administration are aligned in a partnership for coordinated research data support. Further, three different data steward personas filling different roles and levels of support are

^{vii} Findable, Accessible, Interoperable and Reusable: FAIR Guiding Principles for scientific data management and stewardship

presented: the RDM service coordinator, the data curator, and the data manager. Again, it is necessary to emphasize that personas are fictive entities, and real people could be filling these roles. The number of data stewards will vary depending on institution size. The survey responses gave a mix of male, female and gender-neutral names and the personas have been carefully constructed to reflect this. The author selected illustration photos to give the personas more of an identity by providing them with a face, care have been taken to avoid stereotyping. The names and photos were presented to the participants in the final interview, none of the participants presented any opinions on either, but focused on the roles and skills embedded in each persona while referring to each with the names.

The Research Data Service Center

The RDSC is run collaboratively by IT, the library, and the research office at the university. The RDSC has been established to solve issues of RDM support and training but also espouses other related research skills, such as data visualization, data analysis software, support on statistics, etc. The services they offer are divided into core services provided by RDSC staff and coordinated services where the RDSC is the host for related networks and courses. RDSC is designed to be user-centered and responsive to current needs among researchers who are testing and offering the latest in technologies for research data. By having an approval function for data management plans and, by coordination, network meetings for of data managers, they map and respond to the knowledge level and needs of their local environment. The RDSC are up to date on challenges and needs in their community. Further, they collaborate closely with different departments at the university to ensure that data management training is offered to researchers and graduate students.

Core services

- DMP review and consultancy
- One-to-one data management support for PhDs and researchers
- Courses in data management
- Coordination of the “peer-support network” of data managers

Coordinated services

- Hosting courses focusing on skills for research (Python, poster design, R and other courses provided by the Carpentry community).
- Hosting other peer support networks (Carpentry study group, R-ladies etc.)
- FAIR training courses

There are three groups of staff at the center: permanent staff, student staff and associated staff. In addition, they collaborate closely with the data protection officer and with a network of data managers hired by a research group.

The permanent staff includes one RDM service coordinator Kim and data curators of which David is one.

Based on requests, student staff are hired from a pool of data science students and PhD candidates. This offers students interested in data management an opportunity to practice and brings new

expertise into the center. Some of these students end up being hired as data managers in data-intensive research groups upon graduation.

Associated staff work at the research office, library, and in IT but have some tasks at the RDSC. Typically, expertise on data analysis software and statistics are offered by IT staff along with support on writing. DMPs and grant fulfillment are offered by the research office, and metadata and data archiving are offered by library staff. In addition, each individual brings their own skills—some with graphic design, others with ontology building, artificial intelligence, interaction design, or semantic web technologies. This renders the center an interdisciplinary environment that focuses on collaboration and RDM, as well as the proliferation of skills for data-centered research.

The RDM Service Coordinator – Kim Smith

Kim Smith is the coordinator and communicator with the RDM service. She has a master's degree in LIS and several years of experience at the university library. Kim works as RDM Service Coordinator at the RDSC and is responsible for the data management services at the university. She has the overview and coordinates everyone involved at the RDSC. Kim enjoys teaching and presides over several of the RDM training courses offered at the university. Through a series of workshops held at the center, she has given several researchers and master's students their first RDM course. She also advises on privacy and copyright issues, and while she does not have a background in law, experience has made her able to advise on many of the issues that occur. When in doubt, she consults the data protection officer. Kim is also responsible for the review and approval of DMPs. The workload is, however, shared, and the plans are reviewed collaboratively at the center. Through DMP reviews, Kim, David and other staff at the RDSC are able to identify potential challenges at an early stage and offer support. In addition, Kim is active in the international coordination work done with the Research Data Alliance:

- Communication and interpretation
- Policy expertise
- Research ethics and personal privacy
- Intellectual property law
- Data management plans
- Metadata



Photo 1 Kim Smith, Ill. from Colourbox

Motivation: Contribute to making research transparent and verifiable and build new knowledge in the organization

Kim believes that proper data management can solve the reproduction crisis and help rebuild trust in research in society in general. With a background as a librarian, she is focused on data quality and longtime curation. Kim is also concerned about maintaining the legacy of prominent researchers at her university. Her colleagues describe her as structured and strategic.

The Data Curator – David Carpenter

David holds a PhD in computational linguistics and many years of experience with data-intensive research. Recently, he has taken a course in data stewardship. David has a scientifically oriented, analytical mindset. He had been engaged for several years in data-driven research, but he became more interested in the challenges related to ontologies and metadata definitions, and less interested in scientific topics and final publications over time. David is good at convincing researchers that a by-product of proper data management is an increased number of citations, leading to more accreditations.

- Systematization
- Making data FAIR
- Metadata, documentation, and provenance
- Data archives and archiving
- Coding
- Data mining
- Formatting and data transformation



Photo 2 David Carpenter, III. from Colourbox

Motivation: David enjoys translating between disciplines, understanding researchers' needs, and solving problems.

David loves research and the university as a work environment, but he prefers working with the data rather than publishing. He is described as accurate and systematic.

The Data Manager – Kari Anderson

Data manager Kari Anderson is the disciplinary specialist, while the staff at the research support center are the generalists. She is one of the data managers working in the data-intensive research groups at the university. The data managers meet monthly at the peer support network at the RDSC to exchange experiences and solve concrete problems. Kari makes sure there is an agreement on standards and protocol for data management within the research group. When new staff is hired or if students are participating, she makes sure they are briefed in data management before touching anything. Kari identifies with the other researchers in the group. She is good at picking up on potential issues at an early stage, and if someone has problems with conversions, transfer, or the merging of data, she loves the challenge. She is also focusing on deleting what is obsolete, rather than keeping every version of everything.

Kari has a PhD in neuroscience and is fascinated by classification. Through statistical classification, she has developed an interest in AI. She was working closely with a research group during her master's and was later hired as a PhD. During her PhD period, her role gradually became more of a data manager, and when a new center for brain research was established, she was hired as a data steward. She is also taking some extra courses within data science to work with still more methods and disciplines as a data manager/data scientist.

Through the RDM network at the university, she learned of the Research Data Alliance and is now engaged in the health data interest group, where she keeps up to date. Still, her heart is most at home in the R-ladies network.

- Documentation
- Working with large databases
- Coding
- Systematization
- Data transformation
- Metadata standards
- Interoperability



Photo 3 Kari Anderson, III. from Colourbox

Motivation: She loves working in the creative environment of research while still clocking office hours.

At the lab, she is described as the right hand of the professor, the go-to person for the people working there, and a creative and hard-working part of the team.

Persona summary

By creating the personas Kim Smith, David Carpenter, and Kari Anderson, the aim has been to visualize and concretize one example of how both a team providing general support and a data steward working within a research group can function. What is crucial is that the data stewards have a genuine interest in contribution to research and a combination of the right soft skills and knowledge of research along with technical, law and policy, or archival skills. The personas can be applied both in the development of software solutions and as inspiration when creating better research data support at the institutions.

Conclusion

The findings from this study show that outreach, education, and problem-solving are only some of the keys to the creation of a functional service for data management. There are several concerns that must be taken into account as a service is developed.

Four primary challenges for providing data stewardship at universities are identified:

1. **Placement of responsibility:** Researchers must retain their responsibility for data throughout the research cycle. When depositing to a data archive responsibility can be transferred if the selected archive offers curation services.
2. **Communication:** Lines of communication between support levels must be established to avoid closed subcultures and to exchange best practices between domains.
3. **Knowledge of data and methods:** There is a need for local and specialized expertise within an increasing number of domains. It is necessary to find the appropriate degree of disciplinary knowledge to provide support. Knowledge of research is essential; however, the researchers are responsible for data management in their projects.
4. **Joint research support effort:** Research data management requires several different types of expertise that traditionally are spread among different research support departments at universities. The creation of a general research data support team or center with connection to the research office, IT, and the library is crucial to cover all aspects of data management.

One solution can never fit all and, while a general team will be able to solve and support a wide

range of issues, many larger research communities need dedicated staff with specific knowledge of the issues and concerns that are relevant for their research data. While data management is gradually becoming current practice within several data-intensive communities, it is also needed among researchers producing and collecting small heterogeneous datasets, referred to as the long tail of research data;⁵⁰ a research data support center is an attempt to resolve this. A general team will function as a professional network for discipline-specific research data staff and could potentially assist research groups in recruitment and transfer of skills and knowledge across disciplinary boundaries. Motivated by contributing to research, data stewards can be recruited among both graduate students and researchers; however, understanding of research and research methods is important.

References

- ¹ Robert King Merton, "The Sociology of Science: Theoretical and Empirical Investigations" (Chicago: University of Chicago Press, 1973).
- ² Christine L. Borgman, *Big Data, Little Data, No Data: Scholarship in the Networked World* (Cambridge, MA: MIT Press, 2015); Peter T. Darch, "Limits to the Pursuit of Reproducibility: Emergent Data-Scarce Domains of Science," in *Transforming Digital Worlds*, ed. Gobinda Chowdhury et al., vol. 10766 (Cham: Springer International Publishing, 2018), 164–74, doi:10.1007/978-3-319-78105-1_21.
- ³ Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences* (Los Angeles, Calif: SAGE Publishing, 2014).
- ⁴ Alma Swan and Sheridan Brown, "The Skills, Role and Career Structure of Data Scientists and Curators: An Assessment of Current Practice and Future Needs," Report to the JISC (Key Perspectives Ltd, 2008); Robin Rice, *The Data Librarian's Handbook* (London: Facet, 2016).
- ⁵ Kitchin, *The Data Revolution*.
- ⁶ Michael J Scroggins et al., "Thorny Problems in Data (-Intensive) Science," Publications (Los Angeles, California: UCLA: Center for Knowledge Infrastructures, 2019), <https://escholarship.org/uc/item/31b1z69c>; Marta Teperek et al., "Data Stewardship Addressing Disciplinary Data Management Needs," *International Journal of Digital Curation* 13, no. 1 (December 27, 2018): 141–49, doi:10.2218/ijdc.v13i1.604; German Council for Scientific Information Infrastructures (RfII), "Digital Competencies – Urgently Needed! Recommendations on Career and Training Prospects for the Scientific Labour Market" (Göttingen: German Council for Scientific Information Infrastructures (RfII), 2019), <http://www.rfii.de/?p=4015>.
- ⁷ German Council for Scientific Information Infrastructures (RfII), "Digital Competencies – Urgently Needed!"; Philipp Conzett and Lene Østvand, "Støttetenester for forskingsdatahandtering på UiT Noregs arktiske universitet – erfaringar og forslag til beste praksis," *Nordic Journal of Information Literacy in Higher Education* 10, no. 1 (May 31, 2018): 65–80, doi:10.15845/noril.v10i1.283.
- ⁸ Kristin R. Eschenfelder and Kalpana Shankar, "Of Seamlessness and Frictions: Transborder Data Flows of European and US Social Science Data," in *Sustainable Digital Communities: 15th International Conference, IConference 2020, Borås, Sweden, March 23–26, 2020, Proceedings*, ed. Anneli Sundqvist et al., vol. 12051, Lecture Notes in Computer Science (Cham: Springer International Publishing, 2020), 695–702, doi:10.1007/978-3-030-43687-2.

⁹ C. Lewis and J. Contrino, "Making the Invisible Visible: Personas and Mental Models of Distance Education Library Users," *Journal of Library and Information Services in Distance Learning* 10, no. 1–2 (2016): 15–29, doi:10.1080/1533290X.2016.1218813.

¹⁰ Mark D. Wilkinson et al., "The FAIR Guiding Principles for Scientific Data Management and Stewardship," *Scientific Data* 3 (March 15, 2016): 160018, doi:10.1038/sdata.2016.18; Sara Rosenbaum, "Data Governance and Stewardship: Designing Data Stewardship Entities and Advancing Data Access: Data Governance and Stewardship," *Health Services Research* 45, no. 5p2 (2010): 1442–55, doi:10.1111/j.1475-6773.2010.01140.x; Swan and Brown, "The Skills, Role and Career Structure of Data Scientists and Curators"; Jingfeng Xia and Minglu Wang, "Competencies and Responsibilities of Social Science Data Librarians: An Analysis of Job Descriptions," *College & Research Libraries* 75, no. 3 (May 1, 2014): 362–88, doi:10.5860/crl13-435; Anna Clements, "Research Information Meets Research Data Management ... in the Library?," *Insights: The UKSG Journal* 26, no. 3 (November 1, 2013): 298–304, doi:10.1629/2048-7754.99.

¹¹ Xia and Wang, "Competencies and Responsibilities of Social Science Data Librarians"; Rebecca A. Brown, Malcolm Wolski, and Joanna Richardson, "Developing New Skills for Research Support Librarians," *Australian Library Journal* 64, no. 3 (July 3, 2015): 224–34, doi:10.1080/00049670.2015.1041215; Lisa Federer, "Defining Data Librarianship: A Survey of Competencies, Skills, and Training," *Journal of the Medical Library Association* 106, no. 3 (July 2018): 294–303, doi:10.5195/jmla.2018.306; Lyn Robinson and David Bawden, "'The Story of Data': A Socio-Technical Approach to Education for the Data Librarian Role in the CityLIS Library School at City, University of London," *Library Management* 38, no. 6/7 (August 15, 2017): 312–22, doi:10.1108/LM-01-2017-0009; Mary Anne Kennan, "'In the Eye of the Beholder': Knowledge and Skills Requirements for Data Professionals," *Information Research-an International Electronic Journal* 22, no. 4 (December 2017): 1601; Fabian Cremer, Claudia Engelhardt, and Heike Neuroth, "Embedded Data Manager - Embedded Research Data Management: Experiences, Perspectives and Potentials," *Bibliothek Forschung Und Praxis* 39, no. 1 (April 2015): 13–31, doi:10.1515/bfp-2015-0006; Andrew M. Cox and Sheila Corral, "Evolving Academic Library Specialties," *Journal of the American Society for Information Science and Technology* 64, no. 8 (August 2013): 1526–42, doi:10.1002/asi.22847.

¹² Federer, "Defining Data Librarianship."

¹³ Brown, Wolski, and Richardson, "Developing New Skills for Research Support Librarians."

¹⁴ Federer, "Defining Data Librarianship."

¹⁵ Kennan, "'In the Eye of the Beholder.'"

¹⁶ Ibid.

¹⁷ Cox and Corral, "Evolving Academic Library Specialties."

¹⁸ Minglu Wang, "Supporting the Research Process through Expanded Library Data Services," *Program-Electronic Library and Information Systems* 47, no. 3 (2013): 282–303, doi:10.1108/PROG-04-2012-0010; Ricardo L. Punzalan and Adam Kriesberg, "Library-Mediated Collaborations: Data Curation at the National Agricultural Library," *Library Trends* 65, no. 3 (WIN 2017): 429–47, doi:10.1353/lib.2017.0010; T.P. Bardyn, T. Resnick, and S.K. Camina, "Translational Researchers' Perceptions of Data Management Practices and Data Curation Needs: Findings from a Focus Group in an Academic Health Sciences Library," *Journal of Web Librarianship* 6, no. 4 (2012): 274–87, doi:10.1080/19322909.2012.730375.

¹⁹ Meredith N. Zozus et al., "Analysis of Professional Competencies for the Clinical Research Data Management Profession: Implications for Training and Professional Certification," *Journal of the American Medical Informatics Association* 24, no. 4 (July 2017): 737–45, doi:10.1093/jamia/ocw179; Gabriele Schnapper et al., "Data Managers: A Survey of the European Society of Breast Cancer Specialists in Certified Multi-Disciplinary Breast Centers," *Breast Journal* 24, no. 5 (October 2018): 811–15, doi:10.1111/tbj.13043; Martin Dugas and Susanne Dugas-Breit, "Integrated Data Management for Clinical Studies: Automatic Transformation of Data Models with Semantic Annotations for Principal Investigators, Data Managers and Statisticians," *Plos One* 9, no. 2 (February 28, 2014): e90492, doi:10.1371/journal.pone.0090492; R. Esser, "Biostatistics and Data Management in Global Drug Development," *Drug Information Journal* 35, no. 3 (September 2001): 643–53, doi:10.1177/009286150103500302; George Hripcsak et al., "Health Data Use, Stewardship, and Governance: Ongoing Gaps and Challenges: A Report from AMIA's 2012 Health Policy Meeting," *Journal of the American Medical Informatics Association* 21, no. 2 (March 2014): 204–11, doi:10.1136/amiajnl-2013-002117; Hong Huang et al., "Prioritization of Data Quality Dimensions and Skills Requirements in Genome Annotation Work," *Journal of the American Society for Information Science and Technology* 63, no. 1 (January 2012): 195–207, doi:10.1002/asi.21652; Crystal Kallem, "Data Stewardship," *Journal of the American Health Information Management Association* 79, no. 9 (2008): 58-59;63.

²⁰ John Cartwright, Jesse Varner, and Susan McLean, "Data Stewardship: How NOAA Delivers Environmental Information for Today and Tomorrow," *Marine Technology Society Journal* 49, no. 2 (April 2015): 107–11, doi:10.4031/MTSJ.49.2.11; T. A. Boden, M. Krassovski, and B. Yang, "The AmeriFlux Data Activity and Data System: An Evolving Collection of Data Management Techniques, Tools, Products and Services," *Geoscientific Instrumentation Methods and Data Systems* 2, no. 1 (2013): 165–76, doi:10.5194/gi-2-165-2013; Aj Barrett, "Socioeconomic Aspects of Materials Data - Serving the User," *Journal of Chemical Information and Computer Sciences* 33, no. 1 (February 1993): 22–26, doi:10.1021/ci00011a004; Kenneth R. Knapp, "Scientific Data Stewardship of International Satellite Cloud Climatology Project B1 Global Geostationary Observations," *Journal of Applied Remote Sensing* 2 (2008): 023548, doi:10.1117/1.3043461; Kenneth R. Knapp, John J. Bates, and Bruce Barkstrom, "Scientific Data Stewardship - Lessons Learned from a Satellite-Data Rescue Effort," *Bulletin of the American Meteorological Society* 88, no. 9 (September 2007): 1359–61, doi:10.1175/BAMS-88-9-1359; Xin Li et al., "Toward an Improved Data Stewardship and Service for Environmental and Ecological Science Data in West China," *International Journal of Digital Earth* 4, no. 4 (2011): 347–59, doi:10.1080/17538947.2011.558123; R.R. Downs and R.S. Chen, "Designing Submission and Workflow Services for Preserving Interdisciplinary Scientific Data," *Earth Science Informatics* 3, no. 1 (2010): 101–10, doi:10.1007/s12145-010-0051-6; R.R. Downs et al., "Data Stewardship in the Earth Sciences," *D-Lib Magazine* 21, no. 7–8 (2015), doi:10.1045/july2015-downs; Helena Karasti et al., "Knowledge Infrastructures: Part I (Guest Editorial)," *Science & Technology Studies* 29, no. 1 (2016), http://ojs.tsv.fi/index.php/sts/article/download/55406/pdf_1; E.A. Kihn and C.G. Fox, "Geophysical Data Stewardship in the 21st Century at the National Geophysical Data Center (NGDC)," *Data Science Journal* 12 (2013): WDS193–96, doi:10.2481/dsj.WDS-033; T.P. Lauriault, P.L. Pulsifer, and D.R.F. Taylor, "The Preservation and Archiving of Geospatial

Digital Data: Challenges and Opportunities for Cartographers,” *Lecture Notes in Geoinformation and Cartography*, no. 9783642127328 (2010): 25–55, doi:10.1007/978-3-642-12733-5_2; D.J. Lowe, “The Geological Data Manager: An Expanding Role to Fill a Rapidly Growing Need,” *Geological Society Special Publication* 97 (1995): 81–90, doi:10.1144/GSL.SP.1995.097.01.10; R.D. McDowall, “Understanding Data Governance, Part I,” *Spectroscopy (Santa Monica)* 32, no. 2 (2017): 32–38; C.J. Moore and R.E. Habermann, “Core Data Stewardship: A Long-Term Perspective,” *Geological Society Special Publication* 267 (2006): 241–51, doi:10.1144/GSL.SP.2006.267.01.18; T. Nadim, “Data Labours: How the Sequence Databases GenBank and EMBL-Bank Make Data,” *Science as Culture* 25, no. 4 (2016): 496–519, doi:10.1080/09505431.2016.1189894.

²¹ “European Open Science Cloud,” *Nature Genetics* 48, no. 8 (2016): 821–821, doi:10.1038/ng.3642.

²² *Ibid.*

²³ Cartwright, Varner, and McLean, “Data Stewardship.”

²⁴ Boden, Krassovski, and Yang, “The AmeriFlux Data Activity and Data System”; Li et al., “Toward an Improved Data Stewardship and Service for Environmental and Ecological Science Data in West China”; Kihn and Fox, “Geophysical Data Stewardship in the 21st Century at the National Geophysical Data Center (NGDC).”

²⁵ Cartwright, Varner, and McLean, “Data Stewardship”; Knapp, Bates, and Barkstrom, “Scientific Data Stewardship - Lessons Learned from a Satellite-Data Rescue Effort”; Huang et al., “Prioritization of Data Quality Dimensions and Skills Requirements in Genome Annotation Work.”

²⁶ Rikk Mulligan, *SPEC Kit 350: Supporting Digital Scholarship (May 2016)*, SPEC Kit (Association of Research Libraries, 2016), doi:10.29242/spec.350.

²⁷ Clements, “Research Information Meets Research Data Management ... in the Library?”

²⁸ Teperek et al., “Data Stewardship Addressing Disciplinary Data Management Needs.”

²⁹ *Ibid.*

³⁰ Matt Greenhall, “Digital Scholarship and the Role of the Research Library,” The result of the RLUK digital scholarship survey (London: RLUK, 2019), <https://www.rluk.ac.uk/wp-content/uploads/2019/07/RLUK-Digital-Scholarship-report-July-2019.pdf>.

³¹ Saša Baškarada and Andy Koronios, “Unicorn Data Scientist: The Rarest of Breeds,” *Program* 51, no. 1 (January 1, 2017): 65–74, doi:10.1108/PROG-07-2016-0053; Kennan, “In the Eye of the Beholder.”

³² Kennan, “In the Eye of the Beholder.”

³³ Teperek et al., “Data Stewardship Addressing Disciplinary Data Management Needs.”

³⁴ Kathryn Lage, Barbara Losoff, and Jack Maness, “Receptivity to Library Involvement in Scientific Data Curation: A Case Study at the University of Colorado Boulder,” *Portal: Libraries and the Academy* 11, no. 4 (2011): 915–937; Kevin Crowston, “User Personas” (DataOne - Data Observation Network for Earth, 2015), <https://www.dataone.org/personas/abby-science-data-librarian>; Lorna Wildgaard et al., “National Coordination of Data Steward Education in Denmark: Final Report to the National Forum for Research Data Management (DM Forum)” (National Forum for Research Data Management (DM Forum), 2020), <https://doi.org/10.5281/zenodo.3609516>.

-
- ³⁵ Jack M. Maness, Tomasz Miaskiewicz, and Tamara Sumner, "Using Personas to Understand the Needs and Goals of Institutional Repository Users," *D-Lib Magazine* 14, no. 9/10 (2008): 15.
- ³⁶ Crowston, "User Personas."
- ³⁷ *Ibid.*
- ³⁸ Lage, Losoff, and Maness, "Receptivity to Library Involvement in Scientific Data Curation."
- ³⁹ Wildgaard et al., "National Coordination of Data Steward Education in Denmark."
- ⁴⁰ Erio Ziglio, "The Delphi Method and Its Contribution to Decision-Making," in *Gazing into the Oracle - The Delphi Method and Its Application to Social Policy and Public Health*, ed. Michael Adler and Erio Ziglio (London: Jessica Kingsley, 1996), 3–33.
- ⁴¹ H. Rex Hartson and Pardha S. Pyla, *The UX Book: Process and Guidelines for Ensuring a Quality User Experience* (Amsterdam ; Boston: Elsevier, 2012).
- ⁴² *Ibid.*, 266.
- ⁴³ Live Kvale and Nils Pharo, "Understanding the Data Management Plan as a Boundary Object through a Multi-Stakeholder Perspective," Submitted for publication <https://doi.org/10.2218/ijdc.v15i1.729>.
- ⁴⁴ John W. Creswell and Vicki L. Plano Clark, *Designing and Conducting Mixed Methods Research*, 3rd ed. (SAGE Publishing, 2018), 80.
- ⁴⁵ Johnny Saldaña, *The Coding Manual for Qualitative Researchers*, 3rd ed. (London: SAGE Publishing, 2016).
- ⁴⁶ Data from a three-phase Delphi study used to investigate Knowledge Infrastructure for Research Data in Norway, KIRDN_Data, 2020, <http://doi.org/10.5281/zenodo.3673053>.
- ⁴⁷ Kennan, "'In the Eye of the Beholder'"; Baškarada and Koronios, "Unicorn Data Scientist."
- ⁴⁸ OECD, ed., *Data-Driven Innovation: Big Data for Growth and Well-Being* (Paris: OECD Publishing, 2015), <http://dx.doi.org/10.1787/9789264229358-en>.
- ⁴⁹ Chris Anderson, "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete," *Wired*, June 2008, <https://www.wired.com/2008/06/pb-theory/>.
- ⁵⁰ Bryan p. Heidorn, "Shedding Light on the Dark Data in the Long Tail of Science," *Library Trends* 57, no. 2 (2008): 280–99, doi:10.1353/lib.0.0036.

Appendix 1.

Questions describing the data steward in the survey.

Your ideal data person

In several interviews the need for a data person of some kind (Data Steward, Data Curator, Data Scientist, Data Librarian, “Datarøkte”) was mentioned. In order to get a better understanding of who this is or could be, I would like you to spend some minutes creating an image of an ideal person.

I am here asking you to create an imaginary character so please use your imagination.

- a. Position/job title
If you do not see the need for such a position, please give a short explanation on why there is no need for this.
- b. Name
- c. Workplace - Where does this person work and who are they employed by?
- d. Background – brief description of work experience and educational background
- e. Bio - Please provide a short description of who this person is.
- f. Skills - please add minimum three words that describes what this person is particularly good at.
- g. Motivations – please describe what makes this person enjoy their work
- h. Other things - Feel free to add additional information about this person