

## Data Steward interviews - summary

Within this project, we have conducted four qualitative interviews with interested parties from the private and public sector, who – in one way or another – have significant experience in working with data stewardship. The interviews did not follow a rigid guideline, but comprised more of loose discussions based around the following questions:

- What tasks do employees in the organization perform that correspond to data stewardship?
- How are these tasks embedded within the organization?
- What competencies, skills and levels of experience are required for this work?
- What value does data stewardship have for the organization?

We have interviewed employees at four different institutions, representing the private sector, the public sector, research and research support, respectively. We present short summaries of each of these conversations in the appendix in order to illustrate commonalities and differences across the sectors. Note that they reflect the personal experiences and views of the interviewees and are not official statements of their respective institutions.

In conclusion, we have made the following observations:

- The term “data steward” is not well defined within the institutions, or not even used at all. Instead, tasks related to data stewardship complement other functions, such as data science, data analytics, data governance or data librarianship.
- In all cases, a distinct role of a data steward would be to “provide a link between the data and the user”, that means making data accessible to others at the right time and for a given purpose. A data steward should be able to identify and enhance the potential of data for application and communicate this to the users of the data.
- In practice, tasks within data stewardship include securing compliance and data quality, and applying the FAIR principles to the data.
- The work has a highly interdisciplinary character and usually reaches across many different departments within the institution. A designated data steward would often be assigned to participate in specific projects and also be included early on in the planning phase.
- A thorough understanding of the organization and the institution’s (research) activities is crucial for the work of a potential data steward.
- The actual need for a specific data steward role is somewhat unclear at this point. However, there is a strong demand for developing competencies in data stewardship among existing employees.

It became clear that “data stewardship” is an area under rapid development that is currently gaining big attention in all sectors. Though there is a wide spread in the terminology used, the underlying challenges and requirements for making better use of larger amounts of data are very similar across industry, public administration and academia.

We would like to thank all participants for their contributions and their interest in our work.



## Data Steward interview 1 – University administration

Nikolaj Helm-Petersen is working in the Department for Research and Innovation within the central administration of the University of Copenhagen. He is currently leading the installation of a number of data labs that will be located at different faculties and at the central IT department. The aims of the data labs are to enable more researchers to apply data science for their research, to increase their knowledge of data management and to actively support data science activities within research. The data labs will be staffed with data scientists and data stewards.

For Nikolaj, the ideal data steward is a person with a strong background in research, preferably an active researcher who dedicates around 50% of their time for activities in the data lab. These include instructing other researchers in how to apply methods from data science for their research and advising about available services, tools and infrastructure for data management, high performance computing, information technology, etc. The data stewards are also expected to join specific research projects and take an active role in all stages of the project, from preparing the grant application, over the planning of the experimental design, to the publication of results in scientific articles.

In many cases, the data steward will also help other researchers to get access to data from external sources, for example from the public Danish registers, and to combine those with other types of research data. This requires the data steward to be familiar with applicable legal and ethical restrictions and procedures, e.g. for working with personal data. Where the data stewards cannot advise on compliance with these regulations themselves, they should be able to refer to the respective specialists at the university.

On a general level, the data stewards will have an active role of promoting data science and data management services in the departments. An overarching network will ensure knowledge exchange and collaboration between the members of the different data labs, with the overall goal to enable more researchers to use data science for novel, interdisciplinary research.



## Data Steward interview 2 – Public Agency

The Agency for Modernisation under the Ministry of Finance develops and manages common databases and systems for administrative data from a large number of public authorities. According to Frederik Lindgren Greisen, Head of Division and coordinator of the Network for Advanced Analytics, and Thea Harresø Houden, Head of Section and GDPR office coordinator, the Agency has two focus areas related to data stewardship: data governance and data analytics. The former is about designing a unified framework and formalizing procedures for:

- a) ensuring legal compliance,
- b) securing and assessing data quality, and
- c) deciding how the data can be used and for which purposes.

The latter deals with improving and automatizing workflows for processing data and exploring new applications for the data by using e.g. machine-learning techniques and advanced analytics.

The work with data analytics is embedded in a unit with specialized data scientists. The area of data governance is currently being developed further and implemented within the agency, but it has not yet been fully established. At this point, it is unclear if there will be a need for a specific data steward role – and if so, whether this will be in a full-time or part-time position.

Initial thoughts on desired skills and competencies for potential data stewards in the agency include a profound understanding of the institution, a structured and systematic approach to the work, the ability to recognize the potential value of the data and a strong personal interest in creating this added value. Practical experience with collecting and organizing data would be useful. Finally they need to be able to work in interdisciplinary teams across different departments and complement the specialized skills of technical experts and the broad knowledge of the generalists in public administration.

A Data Stewardship training would build competencies for understanding the circumstances and processes, under which the data were created, identifying potential sources of error and knowing the implications of working with different types of data. Frederik and Thea think that a solid theoretical foundation given by their academic education within economics and political science, respectively, is prerequisite for their work in the agency. In addition, future data stewards could also benefit from including practical elements in the courses at university – or give them the opportunity to gain the required additional skills on the job.



### Data Steward interview 3 – Data Domain Specialist

Anna Maria Sempreviva is a senior scientist at DTU Wind Energy who has more than 35 years of experience of working in a data-intensive research area. When she describes her work, she considers herself to hold three roles at the same time: the data scientist, the data steward and the domain scientist.

The data scientist develops software and tools for data analysis and implements mathematical models. The data steward organizes and structures the data and makes them accessible. The domain scientist interprets the data as basis for writing scientific articles and doing innovation. Conducting research – at least in engineering – involves all these tasks, and they are often executed by the same researcher. However, as data are becoming much more complex and their amounts are growing rapidly, researchers are forced to spend more of their time on the collection and organization of the data than on analysis and production of results.

In larger projects, many of the tasks of the data scientist and the data steward can be delegated to designated specialists within the team by the researcher. Some of the data steward's tasks are to make existing data findable, accessible and interoperable for the researchers to reuse, by cataloguing and interlinking data, software and workflows and applying taxonomies for metadata – similar to the role of a librarian.

Anna Maria sees a large demand for data stewards in her field of research. For her, it is important that the data steward has a good understanding of the research. Therefore, a one-year education on top of a bachelor in engineering would be preferable. She also believes this model to be very attractive for students, since they could gain hands-on experience in research by working in a team and either continue with a master or easily find a job in industry afterwards.

A data steward should be trained in knowledge management, text and data mining as well as organizing and structuring data. The data steward should have a scientific understanding of the data and be able to communicate how these can be used.

For Anna Maria, having a data steward embedded in a research team and including this person early on in the planning of a new project would help to “build up the project on solid ground”. However, all researchers need to have basic knowledge of both data stewardship and data science as well, since they are the ones to generate value from the data in the end.



#### Data Steward interview 4 – Pharmaceutical Company

Anders Didriksen has a background in Information and Library Science and has for many years worked with different aspects of data management and data stewardship in the private sector. Currently, he is employed as an information specialist in an international pharmaceutical company with a strong focus on research. There is no official definition of a data steward within the company (and neither has this been the case at his previous workplaces), but he estimates to be occupied with tasks related to data stewardship corresponding to around 20 % of his working hours, e.g. one day a week. A data steward forum with quarterly meetings provides a network for all employees with equivalent functions in the company.

Anders' present role as data steward is mostly administrative and advisory, while some of his colleagues have a stronger focus on analytics and technical aspects. In general, a data steward is the main contact person for all questions related to data, in particular how to store the data, which systems to use to manage the data, how to name files and add metadata and how to provide access to the data. Typically, the data steward is involved early on in projects in order to help designing the workflows for the data. Anders considers the data steward to be the administrative link between the data producers and the end users, ensuring that all data are managed in line with the internal rules set up in the company so that they can be used, where needed. This implies that the work reaches across the different departments and branches at various locations around the world. In addition to enforcing data governance, the data steward needs to give general advice on how to comply with the General Data Protection Regulation and be able to identify potential issues and determine, whether legal assistance will be required.

Anders describes the ideal set of competencies for an administrative data steward as a combination of librarian skills (organizing, structuring and sharing knowledge) and organizational insight (communication with stakeholders and understanding of user needs). For technical data stewards, whose work is more strongly embedded within research projects, experience within the field of research conducted in the company is beneficial. Therefore, Anders sees great potential in educating and training existing employees in data stewardship.

Especially with the rise of technologies for text and data mining and artificial intelligence, the need for data stewards will be growing in the private sector in order to ensure that more data of better quality will be available in a well-organized manner to support the company's research activities and to provide evidence for the generated results. According to Anders, improving the practices for structuring the vast amounts of available data can also give financial and commercial benefits for the company.

