

The Data Steward – an introduction

By Live Håndlykken Kvale, ORCID 0000-0001-5264-0382



1. TERM

Welcome to an introduction to the data steward

My name is Live Håndlykken Kvale. I work with research data management at the University of Oslo Library. I am also a PhD candidate in Library and Information science at Oslo Metropolitan University where I work on a project on research data infrastructures including the knowledge around these.

“Data steward” is one of multiple terms used to describe a person working with data management.

“Data librarian”, “data manager”, and “data curator” are other examples of titles with - somewhat overlapping responsibilities 20/04/2021 08:12:0020/04/2021 08:12:00(York, Gutmann, and Berman 2018).

The primary task of a data steward is data governance. Data governance include the overviews, overall plan and infrastructure to support that data can be used and accessed across domains. It is also the direct management of data.

The goal of the data steward is to ensure data quality.

2. RESPONSABILITES

Data stewards work in both private and public sector to ensure that the organizations make the most of their data.

By identifying the different steps of the data life cycle in an organization, data stewards enhance the value of the data trough facilitating archiving, retrieval and reuse of data.

The data stewards ensure high quality on the data in the organization - by planning data selection, storage, preservation, annotation, metadata and provenance which is documenting and describing the origin of the data, what the data have been through and not.

There is a growing need for skilled data stewards along with requirements for sharing data from research funders and governments (German Council for Scientific Information Infrastructures 2019). Both open governmental data and sharing of research data require data stewardship. Data stewards assist in planning data collections and preparing the data for sharing “as open as possible as closed as necessary”.

This requires knowledge of ethics and legal frameworks in addition to a dialog with IT developers to ensure that appropriate solutions for data are available.

Consequently, skills to maintain and curate data are required, but which skills are needed?

3. SKILLS

There is not ONE data steward that fits all the data stewardship needs. Metaphors such as the “kinder egg”, the “wholly trinity” and the “data unicorn” are used to illustrate the complexity of skills and knowledge expected from data stewards (Kvale 2021): A three in one solution, some kind of superhuman, being an in-depth expert of everything. I’ll go through some of the data steward skills, before splitting them up in different “types” or clusters to illustrate different types of data stewards.

The data stewards must balance technical skills such as programming and work-flow optimization with personal skills such as teaching, communication, networking and project management, along with disciplinary knowledge and experience with different data management practices, including ethics, policy and archiving. Data stewards make sure knowledge of what, when and where is documented in a way that translates beyond the discipline where the data are collected. Working as a data steward is not to know everything, but to learn from different disciplines, and transfer data into different workflows into sharable data (EOSC Executive Board. 2021).

This word cloud visualizes skills for data stewards in research described by researchers, research support, infrastructure providers and policy makers questioned about data steward skills as part of my own research project (Kvale 2021).

But the Data stewards fill different roles in an organization including those of an Administrator, Analyst, Developer and Agent of Change. Consequently, different skills are needed for the different roles.

4. HOW YOU AS PHD BECOME A DATA STEWARD

A Danish report on data stewardship training splits the data stewards in three clusters:

The Ethics Expert with a can-do attitude towards ethics and the responsible conduct of research.

The Administrator a fast learner, being challenged with understanding processes and procedures.

And then finally the Developer which is passionate about optimizing processes and systems. (Wildgaard et al. 2020)

This split highlights that different people with different backgrounds and interests are needed for data stewardship. Different organizations will need different profiles of data stewards to complement the skills they have in house. The Danish report collected data from LinkedIn and job description analysis. The request in job descriptions suggest that all three skills clusters are needed for good data stewardship. In my own data material using qualitative methods I divide skills in six categories: Archiving skills, technical skills, law and policy, research skills, general research data management skills, and personal skills

Data stewards are recruited from a broad range of disciplines including, humanities, law, social sciences, STEM and life sciences. The requests suggest that experience with ethics, is just as relevant as an IT background.

Complementing your PhD degree with a data stewardship training makes you a candidate that stands out. But I would also encourage you to use your PhD project to test, learn and practice data management.

Good luck!

References

- EOSC Executive Board. 2021. "Digital Skills for FAIR and Open Science: Report from the EOSC Executive Board Skills and Training Working Group." EU: European Commission. <https://data.europa.eu/doi/10.2777/59065>.
- German Council for Scientific Information Infrastructures. 2019. "Digital Competencies – Urgently Needed! Recommendations on Career and Training Prospects for the Scientific Labour Market." Göttingen: German Council for Scientific Information Infrastructures (RfII). <http://www.rfii.de/?p=4015>.
- Kvale, Live. 2021. "Using Personas to Visualize the Need for Data Stewardship." *College & Research Libraries*. <https://doi.org/10.5281/zenodo.3992572>.
- Wildgaard, Lorna, Evgenios Vlachos, Lars Nondal, Asger Væring Larsen, and Michael Svendsen. 2020. "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). <https://doi.org/10.5281/zenodo.3609516>.
- York, J., M. Gutmann, and F. Berman. 2018. "What Do We Know about the Stewardship Gap." *Data Science Journal* 17. <https://doi.org/10.5334/dsj-2018-019>.