Attention, Prof: You need a data steward for your team.

Lars Schöbitz

2024-02-14

**Note:** This blog post was published is a re-published version from the Global Health Engineering website. For attribution, please see the [Citation information](https://global-health-engineering.github.io/website/3-blog/posts/2023-11-14-change-publishin/#citation) at the end of this post.

## Lead

You are a professor. It’s long working hours, often after 5 pm. It’s working weekends. It’s sitting in committees, and being stuck in meetings that could have been emails, or even better, not happened at all. It’s teaching, and grading, and supervising, and mentoring. And now, on top of all that, you are supposed to make all your research data public, applying FAIR principles.

## What’s a data steward?

In simple terms and in the context of research, a data steward is defined as “someone who does not generate their own research data, but has the responsibility for the management of research data of others.” (ref)

## What profile do they need?

Few established programmes specifically qualify people to be a data steward (e.g. REF). At the moment, it’s a skill that’s mostly learned on the job. Do not look for academic degrees (e.g. a PhD) or for someone on an academic career path (long list of publications). Look for someone who has an affinity for IT and has worked in various organizations (government, private sector, research) and with different types of data (lab experiments, observations, quantitative & qualitative surveys). A Research Software Engineer, which is a more established term, could be a good fit.

## What can you offer them?

Whether it’s data stewardship or research software engineering, both career paths still need to be fully recognized and established within the scientific community. However, at ETH Zurich, there is an attractive career path for someone without or needing a PhD. It falls under the “administrative, technical, IT and laboratory staff” category. The assigned function level XXX offers an attractive salary for someone with >5 years of experience, the yearly appraisal interview for mechanisms of a performance-bound annual increase, and also the potential for a permanent position.

## What are the tasks?

A data steward will support you in defining your strategy for Research Data Management (RDM). Ideally, this is a strategy that promotes the concepts of open research data and computational reproducibility, both practices that are increasingly important to make research more rigorous, transparent, and impactful. When we established our group in 2021, we defined this role more generally as “Open Science Specialist,” and you find the published job description.

## How to support the group?

On a group level, a data steward hosts strategic meetings. They organize workshops to identify current file and data management practices within a group. They don’t prescribe one way of doing things but rather help determine the best practices for the group, allowing individual researchers to use the workflows and tools that suit them best to maintain scientific autonomy.

## How to support the individual?

On an individual level, a data steward offers weekly RDM support. They become the bridge from abstract concepts like FAIR principles, version control, and literate programming to the actual intellectual work needed.

## Data stewardship in practice

At Global Health Engineering, we aim to make all our research outputs reproducible and foster an Open Science culture. In our workflow from designing a research study that aims at producing a scientific article, we categorize our data into three stages:

* Stage 1: unprocessed raw data
* Stage 2: processed analysis-ready data
* Stage 3: data supporting publication results

The unique work required to manage the collected data is recognized in each of these stages. We assign a digital object identifier (DOI) to the code that transforms the unprocessed raw data into analysis-ready data. Together, they form a data package, fully documented with relevant metadata and citeable and reusable by the research community. It’s a stage of data cleaning and curation, which a data steward is uniquely positioned to support so that the researcher can focus on applying their expertise to the data rather than on the data itself. The actual work that goes into authoring a manuscript, fine-tuning tables, and preparing figures and models rests with the researcher and is what we consider getting data from Stage 2 to Stage 3. The data steward remains available to support the researcher in applying best practices for data management. Still, the researcher is the one who is responsible for the analysis, discussion, and conclusions.

## ETH Library Data Stewardship Network

The ETH Library initiated a programme funded by swissuniversities to build a data stewardship network at ETH. It has identified six people at ETH who already consider themselves in this role and dedicate 20% of their time to data stewardship activities beyond the tasks they already perform for the research group. Through these activities, data stewards can further explore and develop their skills and competencies in data stewardship and, at the same time, support other researchers in applying best practices for data management and sharing.

## Hire now

Hiring a data steward for your group who can connect and engage with other data stewards at ETH and beyond will significantly advance the impact of your research. If you are interested in speaking about hiring a data steward, please get in touch.