

# Towards a community-endorsed data steward profession description for life science research

[Salome Scholtens](#) (UMCG), Petronella Anbeek (UMCU), [Jasmin Bohmer](#) (UMCU), Mirjam Brullemans (Radboudumc), Marije van der Geest (UMCG), [Mijke Jetten](#) (Radboud University), [Christine Staiger](#) (DTL), [Inge Slouwerhof](#) (Radboud University) and [Celia W.G. van Gelder](#) (DTL, ELIXIR-NL)

## Abstract

Sufficient, high quality data steward expertise and capacity in projects and institutes is one of the necessities for FAIR data management in Life Sciences and personalised medicine research. In a ZonMw funded project of UMCG, UMCU, Radboudumc, Radboud University and DTL, supported by the relevant national stakeholders, we are working to make the data steward function concrete, to create consensus on the function and required competencies and to develop tailored education.

The overall project aim is to professionalise the data steward function within the life-sciences domain, with a special focus on the implementation of the FAIR data principles. The project will work towards a common job description and an agreement on knowledge, skills and competencies of a data steward that is broadly supported within the Dutch life-sciences community. These knowledge, skills and competencies will be translated into concrete learning objectives, which in turn will be used to develop an education line and training material for data stewards (including a design for an eLearning module). Sustainable implementation and alignment with existing education will be ensured. This 1-year project will be executed by a core team in close collaboration with a consultation committee consisting of representatives of the main stakeholders in the domain.

## Background

Professionalising the data stewardship function and including sufficient data expertise and capacities in research projects is essential in today's research<sup>1,2,3</sup> in particular in data-intensive research areas as genomics and personalised medicine, which also involve highly sensitive data. Data stewards are being appointed at different [organisations](#), projects or teams, and training is developed at different institutes. Also, recently a number of initiatives started to develop and share knowledge and best practices, such as the data steward interest group of [DTL](#), [LCRDM](#) and [RDNL](#). It is recognized that this expertise should be easily accessible by the researchers, ideally within their own institute or faculty.

---

<sup>1</sup> [www.knaw.nl/nl/actueel/publicaties/big-data-in-wetenschappelijk-onderzoek-met-gegevens-over-persoon](http://www.knaw.nl/nl/actueel/publicaties/big-data-in-wetenschappelijk-onderzoek-met-gegevens-over-persoon)

<sup>2</sup> <https://www.leru.org/files/LERU-AP24-Open-Science-full-paper.pdf>

<sup>3</sup> EOSC High Level Expert Group, Oct 2016

[https://ec.europa.eu/research/openscience/pdf/realisingthe\\_european\\_open\\_science\\_cloud\\_2016.pdf#view=fit&pagemode=one](https://ec.europa.eu/research/openscience/pdf/realisingthe_european_open_science_cloud_2016.pdf#view=fit&pagemode=one)

However, currently no consensus exists on the function of a data steward in the life sciences domain, nor on the required skills and knowledge. Moreover, to our knowledge a coherent approach and dedicated training is also lacking on a national level. This limits the broad implementation of the data steward function in life-sciences research, and it also creates a bottleneck in funding, recruiting, appointing and training data stewards in institutions and research projects.

The aim of this project is to professionalise the data steward function in the life-sciences domain in the Netherlands. By doing so, we aim to increase the data steward capacity in the institutes in quality (by training the existing data stewards) and in quantity (by training new data stewards). The project is part of the [ZonMw programme Personalised Medicine](#) (part of the programme on [Goed Gebruik Geneesmiddelen, GGG](#)) and will engage with the stakeholders that are working on generic as well as life-science specific activities towards Data Stewardship. This includes educational institutes (universities, universities for applied sciences and UMCs), the [VSNU](#) and [NFU Data4lifesciences](#) programme and research infrastructures such as [Health-RI](#), LCRDM, [DANS](#), [SURF](#) and [GO-FAIR](#).

## Proposal description

The overall project aim is to professionalise the data steward function within the **life-sciences domain**. A special focus is on the implementation of the FAIR principles. The project will work towards (a) common job description(s) and an agreement on knowledge, skills and competencies of a data steward in the life-sciences domain that is broadly supported within the Dutch life-sciences community and can be translated to a NFU/VSNU function description. This applies to both data stewards at the institute level and at the project level. It will build on and align with ongoing activities on defining these skills and competences, both from the domain-agnostic side ([EOSCpilot](#), [EDISON](#)), as well as the domain-specific side ([ELIXIR](#)). These knowledge, skills and competencies will be translated into concrete learning objectives, which in turn will be used to develop a curriculum for data stewards (including a design for an eLearning module). Finally, we will ensure sustainable implementation, alignment with existing education, and continuous (re)evaluation of data steward function and training.

## Deliverables

This project delivers:

1. A function description for a data steward, specific for the life-sciences domain. The description includes the responsibilities and activities of a data steward and entails minimally two different function levels. This function is supported by NFU partners, DTL, LCRDM, and Health-RI and will be built based on existing efforts and descriptions (Deliverable 1).
2. A definition of the knowledge base, skills and competencies for life-science data stewards based on the function description (Deliverable 2).
3. A report on a blueprint for an education line (course trajectory) that describes the required education and training for data stewards in life-sciences, with reference to existing courses. Accompanying training material for a maximum of two courses-components will be developed.

All training materials will be made modular, and will be set up in such a way that they can be optimally reused and implemented by education institutes after the project ends (Deliverable 3).

4. Using the developed training materials, one two-day hands-on courses for a minimum of 20 data stewards will be organised, aiming for a national and discipline wide coverage. This course will function as a pilot to test the training and the training materials (Deliverable 4).
5. A design and script for an eLearning module will be developed as a starting point for full implementation in the following phase, after more funding has been acquired (Deliverable 5).
6. A sustainability plan to ensure national embedding of the expertise, to ensure further development of data stewards function and continuation of the training (Deliverable 6).

The function description, definition of knowledge base, skills and competencies for life-science data stewards, documentation on the education line and training materials will all be made available publicly, via the DTL website, LCRDM website and Data4lifescience [HANDS handbook](#) as well as the ELIXIR Training Portal TeSS, and through this Zenodo community.

### Funding Information

This project is funded by the ZonMw Personalised Medicine Programme under Dossier number: 80-84600-98-3007. The project period is Aug 2018 - July 2019.