

Data Stewardship @ WUR: advice on a role for Data Stewards

Jacqueline Ringersma Coordinator Data Management WDCC
Paulien Adamse Researcher and Coordinator Data Management RIKILT

April 2019

1 Introduction

Under the influence of Open Science developments, the new Code of Conduct for Research Integrity, global academic discussions on FAIR data, funder requirements *and* the introduction of a new data policy at Wageningen University & Research (WUR) the pressure on researchers for safe and sustainable management of their research data is increasing. At the start of projects Data Management Plans need to be made, during research prescribed safe data storage needs to be used and after research data need to be archived and registered. New developments at WUR are foreseen in the near future in data sharing procedures.

Although within WUR there is a functional Data Desk (at the Wageningen Data Competence Center) with a well-trained data support unit from the Library, IT and Legal, this support unit cannot always deliver the domain specific support required. With the ambition WUR, as presented in the strategic plan 2019-2022, to focus on FAIR data handling and support for the National Plan Open Science, structural and domain specific data support will be urgently necessary to remain efficient as a leading scientific institution.

Already, we observe that in some (if not most) research units (either a chair group, a clustering of chair groups, or a business unit) research staff starts to take up data management tasks, in order to support their fellow researchers, to overcome the gap between the demand for support and the support which the Data Desk can deliver (both time and domain wise).

WUR is not the only University in which data management has gained serious attention with impact on the data support organisation. In other Dutch Universities the same processes are observed. In 2018 Barend Mons (GO FAIR) introduced Data Stewardship as a pre-requisite for Open Science and the implementation of the FAIR Data principles. Karel Luyben, at TU Delft, created an official function of Data Steward (1.0 fte for each faculty, see <https://openworking.wordpress.com/data-stewardship/>), with currently around 9 Data Steward positions filled (on appr. 3000 researchers). Likewise in the Radboud University the first full time Data Stewards have been appointed at faculty level. In 2018 ZonMW initiated a project to develop a VSNU Hay profile for Data Stewards in the University Medical Centres for which the first result were published in February 2019 on Zenodo (<https://zenodo.org/record/2561723>). The National Coordination Point Data Management (LCRDM) started a Data Stewardship task group. This group is currently undertaking an inventory on Data Stewardship developments within the Netherlands (results of this Task group will be published at the end of April).

In 2018 the Wageningen Data Competence Center (WDCC) initiated a program in order to provide an advice on how to implement "Data Stewardship @ WUR". In this program the WDCC works together with RIKILT (WFSR) (Paulien Adamse) and the Library (Anna Besse), since both RIKILT (WFSR) and the Library have already gained experience in Data Stewardship. One of the first actions in the program was to bring together the research staff members at WUR in the first Data Steward Network meeting. This meeting was held in November 2018 and over 50 people with interest in Data Stewardship attended the meeting. One of the recommendations from the meeting was to define roles, tasks and responsibilities for the Data Stewards in WUR.

In January 2019 a small working group consisting of: Joshua Dijkman (AFSG), Anna Besse (Library), Jantien van Berkel (SSG), Maarten Storm (ESG), Paulien Adamse (RIKILT-WFSR), Peter van Baarlen (ASG), Theo Viets (ASG), Daoud Urdu (SSG), Sylvia Brugman (ASG), Gorkem Simsek (AFSG), Joeri Kalter (AFSG), Klaas-Andries de Graaf (AFSG) and chaired by Jacquelijnn Ringersma (WDCC) started working on this advice. We used the preliminary outcomes of the ZonMW project and the LCRDM Task group as a starting points for our discussions. In total the group met three times, each time discussing a new version of this document. Some discussions were held over e-mail.

In this document we present a lists of tasks which are part of Data Stewardship. The tasks are described in categories (which we took from the ZonMW report on Data Stewardship, and which are also used in the (concept) report of the National Task Group on Data Stewardship in the Netherlands). For each category we have defined a minimal set of tasks (minimal task scenario) and a possible extension (optimal scenario). Also for each category of tasks we defined responsibilities. It is important to realize that the Data Steward is not the final responsible person for the different tasks, sine his (her) role is to motivate and support researchers and not to control them.

2 Data stewardship organization @ WUR

In the data stewardship organization, the Wageningen Data Competence Center is the connecting organization between the strategic level (policy and strategy), the generic support organization and the research community, which includes the Data Stewards providing operational support in the domain.

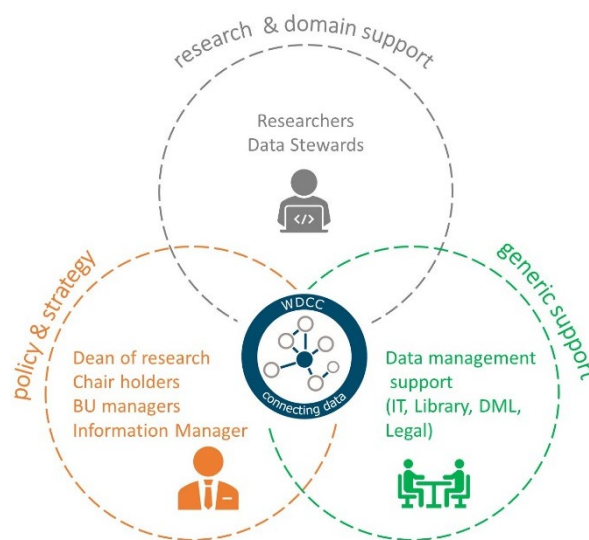


Figure 1: Stakeholder diagram Data Stewardship organization @ WUR

There are many stakeholders within the Data Stewardship Organisation, with the Dean of Research as the final responsible for RDM policy and services within WUR and the WDCC as coordinator for all RDM related activities. Table 1 gives a list of stakeholders and their roles and tasks.

Table 1: RDM stakeholders within WUR

Stakeholder	Role/Task
Researchers	Operational RDM during the research data cycle
Data Stewards*	Embedded (domain specific) operational RDM support & RDM advice in research groups; bridge between researchers and generic support
Dean of Research	Ownership, strategic and final responsible for RDM Policy
Chair holders, BU managers, Directors Science Groups, Information Manager WUR	Tactic level responsibility RDM
Data Management Support (IT, Library, DML, Legal)	Generic operational RDM support & RDM advice, Liaison with National Library RDM network (UKB – RDM)
WDCC (Coordinator Data Management)	Coordinating RDM activities, policy & services developments, organisation of Data Stewardship network, Liaison with external networks (LCRDM, GO FAIR, RDA).

*The role of Data Steward is currently fulfilled on a “voluntary basis” by researchers within the Chair groups or business units.

The WDCC is the one stop shop for all questions on data management (the data desk, data@wur.nl). It has direct lines with the researchers and data stewards on the operational and tactical side and with the Dean of Research, and management layers on the tactic and strategic side. WDCC coordinates the support unit with staff members from the IT, Library, DML and Legal.

Data Stewardship consists of a number of tasks which have as goal to provide disciplinary support to the members of a defined research community (researchers, PhD students, MSc students, management of group) for research data management throughout the data cycle. The tasks are presented in Ch. 3. The minimal scenario of these tasks should be performed by an appointed Data Steward within a group or clustering of groups. The optimal tasks can either be performed by the Principle Investigator or the Data Steward.

External stakeholders on the WUR data policy and services developments are: research funders (EU, NWO, ZonMW, EZK), (inter) national data developments (National Plan Open Science), and (inter) national data alliances (LCRDM, RDA, GOFAIR). The WDCC is the responsible organisational unit to keep WUR aligned with external developments.

3 Data stewardship tasks

We describe the Data Stewardship tasks *only* for the level of chair group or cluster of chair groups, research team or project (the Data Steward). Apart from Data Stewards there is a data management support unit (Library, IT, DML, Legal) and the Wageningen Data Competence Center. Both are also part of the Data Stewardship organisation, however, they are not within the scope of this advice. Data management support (DMS) provides services at a more generic level to the whole WUR. The Data Steward is a domain specialist and linking pin between DMS and the researchers.

We have divided the Data Stewardship tasks in 4 categories: (1) Policy implementation & compliance, (2) Services, (3) Archiving & Registration and (4) Infrastructure and tools. For each category we describe in bullet points what the task entails. It might be that not every bullet point is applicable to every Data Steward, since needs vary within domains, so each list of tasks consists of a minimal scenario and an optimal scenario. For the minimal scenario we recommend a time investment of 2 days per months for appr. 30 researchers (~0.1 fte per 30 researchers).

3.1 Policy implementation & compliance

Minimal scenario:

- communicates the WUR data policy to the members of the group. Ensures that they are well informed, and aware of the 'why' and 'how' of the policy.
- makes sure that the group or project has a Data Management plan (DMP) which is in line with the policy (including advice on the DMP template to use). The Data Steward is the initiator and creator of the group DMP and provides advice on this matter to the management of the group.

Optimal scenario:

- Follows the implementation of the Data Management Plans. Gives advice when adjustments and improvements need to be made in the implementation; and signals to the management when actual data practises are not in line with the agreed DMP.

Relations & responsibilities:

Task	Who is final responsible:	Who supports the Data Stewards:
Policy implementation	Chair of the group (WU) or management of the group (WR).	Data Management Support (Library)

3.2 Services during research (active data)

Minimal scenario:

- Connects the researchers in the group to the Data Management Support for generic data services, or advice/services the DS is unable to provide.
- directs to the local privacy officer in the case of Data Management of privacy sensitive data.
- creates a group Data Management Plan template, based on the template provided by DMS.
- assists in the development of a group DMP. Provides advice how to deal with the different aspects mentioned in the DMP.
- advises on data storage environments, data storage standards, directory structures, file formats, versioning, data documentation etc. during research.
- makes sure that the data of the group are findable (and knows how to use the generic DM support to actually do this).

Optimal scenario:

- all bullet points mentioned under minimal
- assists in the evaluation of the level of FAIR-ness of the data in the group and advises on FAIR storage tools and methods, and archiving methods in order to achieve FAIR data (see also previous bullet point).
- is the main guardian of data quality in relation to FAIR and the WUR data policy of the group
- advises on external data sources in the research domain of the group and knows which expertise exists on the use of these sources in the group

Relations & responsibilities:

Task	Who is final responsible:	Who support the Data Stewards:
Services	Data Management Support/WUR privacy officer (in case of personal data)	Data Management Support

3.3 Archiving & Registration

Minimal scenario:

- Aligns with data archiving facilities of DMS and directs researchers towards the archiving service of DMS
- Provides advice and/or assistance to the researchers in the group in the creation of metadata

Optimal scenario:

- all bullet points mentioned under minimal
- Monitors archiving of research data by researchers of the group department or project.
- Assesses whether data archiving facilities used for departments or projects meet the DMP and WUR data policy (and assesses the FAIR levels of data archiving of the group).

Relations & responsibilities:

Task	Who is final responsible:	Who support the Data Stewards:
Archiving	Library (registration) Chair holder or Management	Data Management Support

3.4 Technical/Infra

Minimal scenario:

- advises on which infrastructure and tools to use throughout the data/research life cycle
- connects group members to IT and WDCC services with respect to infrastructure and tools
- monitors the needs regarding tools and infrastructure for Research Data Management within the department
- evaluates the data storage quality (cost, sustainability, privacy) and/or can liaise with DMS to do this

Relations & responsibilities:

Task	Who is final responsible:	Who support the Data Stewards:
Technical/Infra	IT	IT, WDCC infra

4 Required skills

Although this working group acknowledges the need for the development of a training curriculum for Data Stewards, we did not include this in our advice. We have asked the Coordinator Data Science Education (WDCC) and the Education Support unit (Library) to start developing a training curriculum (see also <https://zenodo.org/record/2561723>) once there is consensus of the tasks and role of Data Stewardship@WUR.

5 Networking and knowledge exchange

The Data Steward functions at the level of the chair group(s) (WU) or business unit (WR). It is likely that there is more than one Data Steward within each Science Group. Data Stewards should actively participate in the WUR Data Steward network organized by the Wageningen Data Competence Center. The purpose of this network is to exchange knowledge, discuss new developments and provide input for new data policy developments. Network events are to be held twice a year. Given the high dynamics and fast changing developments in Data Management requirements from society, funders and WUR itself we recommend to include the participation in this network in the task of the Data Stewards.

6 Time investments, and costs

For the implementation of the minimal data stewardship scenario, the working group estimates that a time investment of 2 days per months for approximately 30 researchers will be required (0.1 fte). This proposed time investment is similar to investments currently being made in Delft and Nijmegen. Bear in mind that this is valid for the minimal scenario in data stewardship tasks.

When more tasks are attributed to the Data Steward, this requires more capacity. According to the [newly developed tool](#) by TU Delft, Data stewardship investment really depends on the amount of data to be handled and the level of protection required for the data (personal data or commercial interest).

Additional capacity might also be required for projects with collaborating partners. Within EU H2020 projects, the EU allows for the attribution of 5% of the budget for Data Management.

We expect that the organization will benefit from an efficient organization of data stewardship, as proposed in this advice, due to better reproduceability and traceability of research, more efficient data handling and use of organizational resources, new opportunities for data reuse, and reduced risks of losing data.

7 Future developments

As mentioned in the introduction the focus of the current WUR data policy is on the F and the A of the FAIR principles. Thus, the proposed role of Data Steward includes mainly the facilitation of these principles, with tasks from Data Management Planning to Archiving. New developments will certainly influence the tasks of the Data Stewards. We name the two most important ones:

1. Data sharing guidelines
Currently a working group, installed by the Rector, is working on guidelines for Data Sharing. The drivers for this working group are f.e. the opportunity that the WDCC has created for more collaboration on data intensive research, or the implementation of an Open Science Road Map within WUR. Both developments require that WUR thinks about how data can be shared.
2. Linked data and data exchange
The WDCC coordinates, through Kennisbasis and investment programs, projects which develop concepts and standards for linked (open) data and data exchange (and machine learning etc.). The results of these research programs will deliver new insights on data opportunities.

8 Our advice

1. Acknowledge the organization of Data Stewardship within WUR, and the emerging role of the Data Steward within this organization;
2. Appoint a Data Steward in each chair group (if not already), cluster of chair groups, business unit or project;
3. Provide the Data Steward with a time allowance of 2 days per month on approximately 30 researchers (PhD students etc.) and assign the minimal tasks scenario to the Data Steward, to be performed in this time (0.1 fte);
4. Attribute the final responsibility for the implementation of the Data Management Policy with the management of the group, and ask the WDCC to coordinate and evaluate the implementation of the program.
5. If a group decides to attribute tasks in the optimal scenario, provide extra time. If not, these tasks will have to be performed by the Principle Investigator (PI) or researchers themselves. The tasks remain equally important to support the notion that WUR is serious about data.
6. Realize that the role of Data Steward is dynamic and new developments will influence the role and required time investment.

7. It is important that there is a strong link between the generic Data Management Support unit, the WDC and the DataSteward@WUR network. It will be a task of the WDC to organize this network and organize communications.