



Data stewardship on the map:

A study of tasks and roles
in Dutch research institutes

LCRDM

The National Coordination Point Research Data Management (LCRDM) is a national network of experts on research data management (RDM) in the Netherlands. The LCRDM connects policy and daily practice. Within the LCRDM experts work together to put RDM topics on the agenda that ask for mutual national cooperation.



more information: www.lcrdm.nl



Contents

| | |
|---|----|
| Introduction | 5 |
| Tasks, responsibilities and roles as described in existing literature | 7 |
| Classification in task areas | 8 |
| What Dutch research institutes ask for in vacancies | 9 |
| National Data Stewardship survey | 13 |
| Interviews | 18 |
| Examples of (future/desired) embedding of data stewardship | 23 |
| Conclusion | 27 |
| Appendix 1: Links | 30 |
| Colophon | 31 |





Introduction

'Effective long-term scientific data stewardship touches on processes, standards, and best practices in multiple knowledge domains, including science, data management/preservation, and technology.'¹ (Peng, G. et al. 2018).

Good research requires good data stewardship. Data stewardship encompasses all the different tasks and responsibilities that relate to caring for data during the various phases of the whole research *life cycle*. The basic assumption is that the researcher himself/herself is primarily responsible for all data.² However, the researcher does need professional support to achieve this. To that end, diverse supportive data stewardship roles and functions have evolved in recent years. Often they have developed over the course of time. Their functional implementation depends largely on their place in the organization. This comes as no surprise when one considers that data stewardship consists of many facets that are traditionally assigned to different departments. Researchers regularly take on data stewardship tasks as well, not only for themselves but also in a wider context for a research group. This data stewardship work often remains unnoticed.

For research institutes, including universities, university medical centres, universities of applied sciences as well as other research organizations, it is becoming ever clearer that professionalization of data stewardship, or *research data management* as it is also known, is important; that data stewards are indispensable and that the various data stewardship roles must be placed within an organizational framework.³ This is due not only to a growing demand for support from the professional field. New requirements set by research funding organizations, rules and regulations and new policy that hold research institutes accountable for good data stewardship, also play a decisive role. For example, the *Gedragcode Wetenschappelijke Integriteit* (Netherlands Code of Conduct for Research Integrity) states that institutes are obliged to care for data management.⁴ Various international studies and recommendations underline the importance of professionalizing data stewardship. A report by the *High Level Expert Group on the European Open Science Cloud* of the European Commission asserts that 500.000 'data experts' are needed at short notice for research data management in academic Europe.⁵ This topic is also on the agenda at a national administrative level. In the context of the National Plan Open Science - a nation-wide platform that represents all administrative umbrella organizations - a project has been formulated to expand the proficiency of the data professional.⁶

As a consequence, different research institutes are setting up data stewardship programmes. This process of professionalization creates new roles, but positioning of data stewards within an organization still varies significantly: a data steward may occupy a central position (at a library or an ICT department) or be deployed at faculty level; sometimes he/she is assigned to a certain project or fulfils a discipline-specific role.⁷ Often it is a mix of the above.

¹ [Scientific Stewardship in the Open Data and Big Data Era — Roles and Responsibilities of Stewards and Other Major Product Stakeholders](#), Peng, G. et al.

² Among other publications described in [Handbook for Adequate Natural Datastewardship](#) (HANDS)

³ Peng, G. et al. 2018. [A Conceptual Enterprise Framework for Managing Scientific Datastewardship and Rol van de Datasteward in de organisatie van Data Management](#), Jacqueline Ringersma (WDC) et al.

⁴ <http://www.vsnul.nl/files/documents/Netherlands%20Code%20of%20Conduct%20for%20Research%20Integrity%202018.pdf>, p. 20–21.

⁵ [Realising the European Open Science Cloud](#) First report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud, p.12.

⁶ State of affairs April 2019. www.open-science.nl

⁷ As described in [A life sciences data steward function matrix](#), Scholtens, et al.



Support for researchers is increasingly delegated to specific staff members with a data profile. However, interpreting the role of data steward is not clear-cut. He/she is expected to perform a wide variety of tasks. The job title of this role can also vary greatly – from ‘data manager’, ‘core data expert’ to ‘data science researcher’.⁸ In addition, many of these names have a different connotation outside the academic world; some imply a strong focus on *compliance* control, for example, which can be confusing. It can cause potential job applicants to have inaccurate expectations and frustrate researchers who need data support. This can be detrimental to the image of the data steward, appreciation for their work and also their career prospects.

⁸ Among others EDISON [Data Science Framework: Part 1](#).

Another concern is that organizations find it hard to obtain suitable candidates for the role of data steward because of discrepancies between what a job function demands and what a potential employee is capable of or what he/she has been trained for. This can be resolved in part by offering training courses for data stewards but these are often marginal and incomplete. In the mean time, data stewards with specific experience are in great demand both in- and outside the research field. Institutes compete with each other to lure competent data stewards while financial enticements offered by the business sector also attract candidates.

*‘The researcher is increasingly supported
in his work by specific people with a data profile:
the data stewards’*

Approach adopted by the LCRDM Data Stewardship Task Group

During October 2018 through March 2019, a LCRDM task group, taking into account the prevailing uncertainty about the interpretation of data stewardship, worked on providing insight in what Dutch institutes currently demand of data stewards and also what has been implemented and developed by and for them. This offers a basis for a clearer job description for data stewardship roles.

This report was realized by studying (inter)national literature in this field, collecting existing task surveys and charting current function requirements and competencies for diverse data stewardship roles at Dutch institutes. A total of 22 vacancies were analysed. In addition, a nation-wide census was held. Staff from more than 30 research organizations completed a questionnaire. To evaluate the poll findings and ask more in-depth questions, a round of interviews was held with eight data experts from the professional field. This gave a clear picture of how data stewardship was being implemented at Dutch institutes. Appendix 1 contains a link to the LCRDM Data Stewardship page on LCRDM’s online platform.



TASKS, RESPONSIBILITIES AND ROLES AS DESCRIBED IN EXITING LITERATURE

The international field acknowledges that data stewardship is characterized by ambiguous tasks, responsibilities and roles. In past years various initiatives have published extensive surveys of tasks that fall under good data stewardship.⁹ These task surveys can help to draw up a function description that fits the requirements of a certain institute.¹⁰

Papers have also been published that each, according to their own (scientific) domain (library, archive, ICT, life sciences), outline different roles and functions for data stewards.¹¹ Often a distinction is made between roles that focus on technical workflows, specific use of data in research and roles concerned with policy, data management planning and general commitment and training. In other words: we see a difference between the role of an 'embedded' data steward and a 'generic' data steward.

The role of an embedded data steward as described in specialist literature

The embedded data steward is directly involved with research being carried out and offers support where necessary. He/she is familiar with the specific needs of fellow researchers within the research unit and the relevant domain and translates generic data policy so it can be practically implemented. An embedded data steward has expertise in certain research-related and domain-specific ways of working. For more generic information he/she will call in the help of a generic data steward. This role should preferably be interpreted as closely related to operations already well-known to a staff member, for example data administrators in a laboratory. The embedded data steward for instance helps with software code, scripts and algorithms to analyze data.¹² Positioning: often within a research unit.

The role of a generic data steward as described in specialist literature

The generic data steward helps researchers with all kinds of data related questions or refers them on; he/she supplies information and training with regard to policy requirements and guidelines and helps to draw up data management plans. In other words, the data steward as a *'centralized knowledge and communication hub for researchers'*.¹³ The generic data steward sometimes has specific knowledge of a certain domain but generally does not have adequate time to give advice on particular situations.¹⁴ Positioning: often with a support service or at faculty level.

⁹ [Data Science Framework: Part 1](#) EDISON project; [Skills and Capability Framework](#), EOSCpilot project;

¹⁰ See appendix 1

¹¹ Peng, G. et al. 2018. [A Conceptual Enterprise Framework for Managing Scientific Datastewardship](#); Sapp Nelson, Megan R., 'Pilot Data Information Literacy Competencies Matrix Scaffolded Across Undergraduate, Graduate and Datasteward Levels' (2016). Libraries Faculty and Staff Scholarship and Research; [Shifting to Data Savvy: the Future of Data Science in Libraries](#), Matt Burton et al.; [Data Governance and Stewardship: Designing Datastewardship Entities and Advancing Data Access](#) Sara Rosenbaum, [A life sciences data steward function matrix](#), Scholtens, et al.

¹² Among others [Open Data, Grey Data, and Stewardship: Universities at the Privacy Frontier](#) Berkeley Technology Law Journal 33:2 (2018); Peng, G, et al. 2018. [A Conceptual Enterprise Framework for Managing Scientific Datastewardship](#)

¹³ [Data stewardship: a new UM Library service](#), Maastricht University; [a life sciences data steward function matrix](#), Scholtens et al.

¹⁴ Among others [Open Data, Grey Data, and Stewardship: Universities at the Privacy Frontier](#) Berkeley Technology Law Journal 33:2 (2018); Peng, G, et al. 2018. [A Conceptual Enterprise Framework for Managing Scientific Datastewardship; Role of data stewards and data stewardship community](#), Alastair Dunning, Delft University of Technology.

'The difference between the two roles cannot be clearly articulated in specialist literature or in practice'

CLASSIFICATION IN TASK AREAS

In practice, the division between the two above roles depends strongly on where a data steward is stationed, i.e. what kind of institute. Even if there are major differences between institutes regarding positioning and definition of roles, we often see that the same sort of tasks are performed. That's why this report has chosen to focus on classification according to task areas: defined here as 'embedded and operational' and 'generic and advisory'.

Different descriptions of good data stewardship often include tasks that do not fit these task areas self-evidently. These are tasks about developing policy and strategy, coordination and fine-tuning with managers and setting up appropriate support on an institute-wide scale. With regard to the associated responsibilities, these tasks differ in gravity and are often assigned to other controllers, a 'data stewardship coordinator' for example. For that reason a third task area has been added.

This allows us to make the following classification in task areas:

- 1) Embedded and operational;
- 2) Generic and advisory;
- 3) Policy, strategy and coordination.

See also figure 1. This is not a question of 'one size fits all'.¹⁵ Someone who fulfils a generic data steward role often takes on tasks from the *embedded and operational* task area and vice-versa. Therefore there is a fluid transition between task areas when one looks at the interpretation of specific roles or functions.

15
[5-MODELS FOR DATA-STEWARDSHIP](#) SAS best practices whitepaper, en [Open Data, Grey Data, and Stewardship: Universities at the Privacy Frontier](#) Berkeley Technology Law Journal 33:2 (2018)

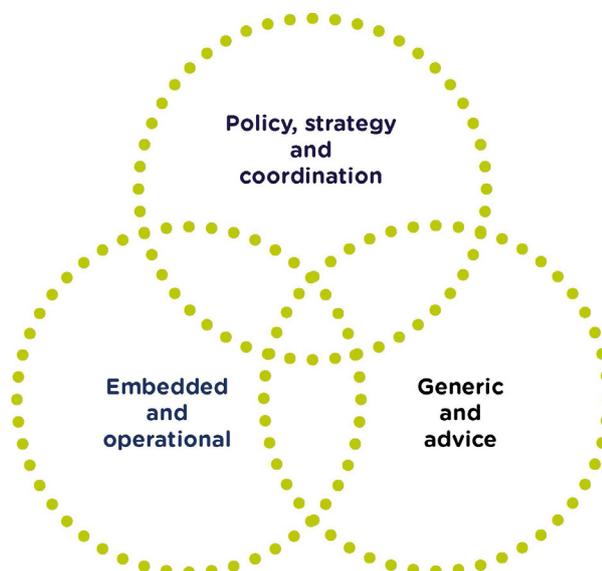


Figure 1: data stewardship task areas

WHAT DUTCH RESEARCH INSTITUTES ASK FOR IN VACANCIES

To gain an idea of the most frequently requested tasks, texts of 22 recently published data stewardship vacancies in the Netherlands were analysed.¹⁶ Not one vacancy required applicants to perform tasks from purely one task area. Sometimes it was difficult to determine what was meant with certain tasks. On enquiring into the nature of specific tasks, it sometimes transpired that the intended focus differed from the text in the vacancy. This is evidence of the different ways in which the role of data steward is established and also indicates the need for better formulated task descriptions for vacancies. Appendix 1 contains a link to the list of vacancies mentioned above.

16 June 2018 through
February 2019

A survey of function titles cited in vacancies is created that focus on a certain task area. It also lists the corresponding wage scale and function profiles insofar these are known.

'Not one vacancy required applicants to perform tasks from purely one task area'

Embedded and operational task areas

Tasks most frequently mentioned under the Embedded and operational task areas are:¹⁷

- Performing statistical analyses;
- Processing data;
- Building data bases;
- Caring for structured and secure data storage;
- Building scripts to assist data control and data cleaning;
- Optimizing processes;
- Correct data extraction;
- Data capturing/data harvesting;
- Data enrichment/linking datasets;
- Re-using and developing (machine learning) algorithms;
- Visualising data;
- Securing provenance and audit trail;
- Advising on the improvement of the data architecture;
- Working on complex data (processing) issues, providing input to improve processes or systems;
- Developing new workflows for ingestion, aggregation and export of data;
- Finding optimal solutions for use of existing data infrastructure;
- Extending existing data infrastructure;
- Monitoring up-time and security breaches of servers and services;
- Ensuring that research data from various domains is adapted in accordance with FAIR principles;

17 A large portion of
vacancy texts was written
in English.



In seven vacancies, the focus of tasks to be performed was located within the *Embedded and operational* task area. In two of these - developing or giving training courses – was also part of the job description.

Function titles used mainly by the Embedded and operational task area are:

- 1) Data steward
- 2) Data manager
- 3) Research Data Manager
- 4) Consultant Data Management & Designer
- 5) Data Scientist
- 6) Data Scientists: statistical analysts, machine learning specialists & scientific programmers
- 7) Bioinformatician

Insofar a wage scale is known, scale 9, 10 or 10/11 applies.

The recognized UFO-profiles⁸ that relate to these vacancies are: *Consultant ICT 3 or Education/Research Assistant 3 (2)* and *Specialist Scientific Information 1 or 2*.

18 UFO-profiles are part of the [function ordering system](#) used at Dutch universities. Other function profiles apply for universities of applied sciences (hogescholen) as listed in the [cao](#) of the association for universities of applied sciences (Vereniging Hogescholen).

‘On enquiry, it sometimes turned out that the focus of a specific function was different from how it was described in the vacancy text’

Generic and advisory task area

Tasks most frequently mentioned in the Generic and advisory task area are:

- Support researchers and students to store, (re)use and analyse research data and information;
- Advise researchers on careful management of research data and about procedures and technical aspects that are important for the quality of (meta)data;
- Advise researchers about how to deal with privacy sensitive data in accordance with the (GDPR) guidelines;
- Take the initiative to establish contacts with discussion partners from both research- or specialist domain groups and colleagues at faculty or discipline level in order to consult on subjects from the data management field;
- Develop or give training courses that relate to the RDM field;
- Is pro-active in knowledge dissemination, for example, by organizing events;
- As spokesman for the faculty has a mission to convince researchers of the added-value of good research data management;
- Monitors a process, system or the data architecture to optimize it;
- Takes the lead in developing further service provisions at research group-, discipline- or faculty level;
- Contributes to scientific articles and helps to draft subsidy applications;



- Deploys his/her knowledge of (inter)national developments in the field of data management and open science;
- Works to build up an (inter)national network in the RDM field;

Ten vacancy descriptions place extra emphasis on the *Generic and advisory* task area. Job descriptions of six of these also include developing and offering training courses. Tasks related to training seem to be linked more frequently to generic tasks than to embedded and operational tasks.

Function titles most frequently associated with the *Generic and advisory* task area:

- 1) Data steward (4x)
- 2) Research Data management Specialist (3x)
- 3) Research Data manager junior or medior
- 4) Research Data Management Consultants
- 5) Data Science Research Assistant

Insofar the wage scale is known, scale 10 or 10/11 applies. The recognized UFO profiles that relate to these vacancies are: *Information specialist /information specialist 2*, sometimes combined with *Functional Manager ICT*.

Policy, strategy and coordination task area

The most frequently mentioned tasks in the *Policy, strategy and coordination* task area are:

- Operates on a strategic and tactical level;
- Development and operationalization of products and services in the RDM domain;
- Translates international developments into policies and practices at the university;
- Leads the development and implementation of the faculty's data management policy;
- Liaison function at policy level/connects with stakeholders in the research domain and acts as sparring partner for leading scientists and board of directors;
- Responsible for the coordination of the innovatory information agenda as part of the complete research life cycle;
- Bears final responsibility for the quality and protection of (meta)data;
- Safeguards the data architecture.

Five vacancy texts placed emphasis on *Policy, strategy and coordination*. In a number of these it was difficult to determine where the exact emphasis lay because many tasks from other task areas were also included. That's why interviews held as follow-up to the questionnaire specifically addressed the positioning of these tasks. Some vacancy texts described tasks not readily associated with data steward duties. This is consistent with the observation that in many cases tasks are performed by someone who also undertakes other management and policy assignments.



Function titles of which the tasks are located mainly in the *Policy, strategy and coordination* task area:

- 1) Research information manager
- 2) Data stewardship coordinator
- 3) Research data management coordinator
- 4) Specialist in scientific information for natural sciences
- 5) Senior research data manager

Insofar the wage scale is known, scale 11 or 12 apply. The recognized UFO profiles connected with these vacancies are: *Specialist Scientific Information 1*.

NATIONAL DATA STEWARDSHIP SURVEY

To gain a topical picture of how Dutch research institutes deal with data stewardship, an online survey was launched in December 2018. Goal of the survey was to gain insight in how data stewardship is currently implemented and positioned in Dutch research institutes and so contribute to the professionalization of research data management and data stewardship functions. What has already been developed and what is still needed to complete the role of data steward with regard to tasks and skills?

Completed by 94 people, the survey was worded in English because many data support staff are foreign nationals. The institutes at which the respondents are active, represent a reasonable crosssection of Dutch research institutes. From the survey it is clear that most respondents play a role in data support.

| Sort of institute | Number | Role | Number |
|-------------------------------|--------|-------------------------------------|--------|
| University of Applied Science | 22 | Research (data) support coordinator | 29 |
| UMC | 17 | Research (data) support assistant | 35 |
| University | 37 | Researcher | 06 |
| Other | 18 | Other | 24 |

‘The institutes at which the respondents are active, represent a reasonable cross-section of Dutch research institutes’

The survey offers an in-depth insight of desired data stewardship qualities and characteristics in various existing data support roles and the kind of support required for research data management.

This has been made apparent through questions about:

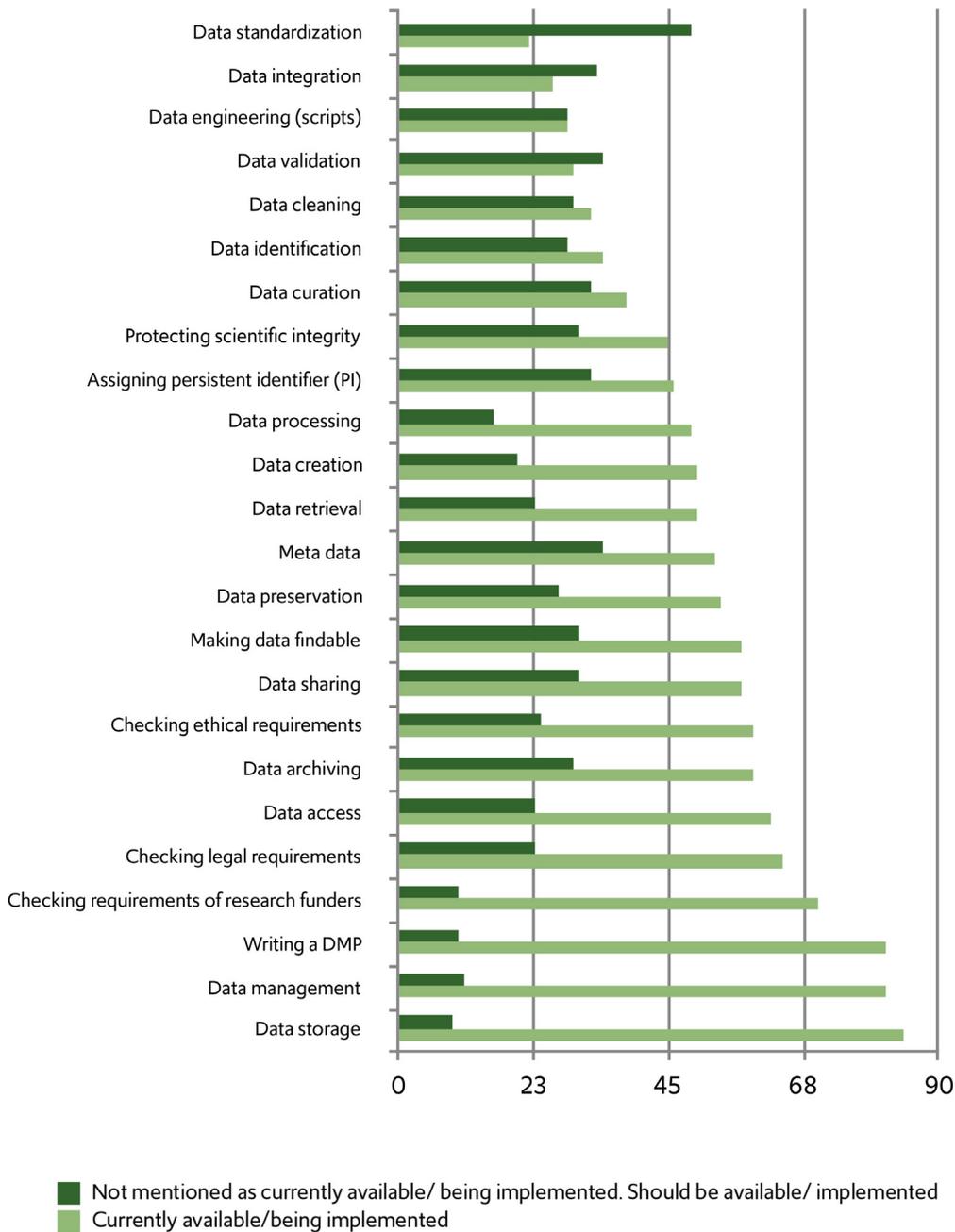
- Research support tasks offered and requested by an institute;
- Desired qualities, responsibilities and skills of a data steward;
- Embedding or positioning of data stewardship in an organization (specific role and distribution of tasks);
- Recent recruitment drives for data stewards or for positions with data stewardship tasks;
- The manner in which data stewardship is organized, broken down into how it is financed, the availability of policy, procedures and job descriptions and the responsibilities and quality requirements that apply to data stewards and research data management;
- Topical developments in the field of data stewardship and research data management support.

The following is a brief description of the findings. Appendix 1 contains a link to all questions and answers of the survey.

Research data management support tasks

Respondents could indicate in the survey if the organization where they work, in their opinion, carries out one or more of 24 pre-selected data management support tasks. In addition they could state if they thought these tasks *should* be performed.

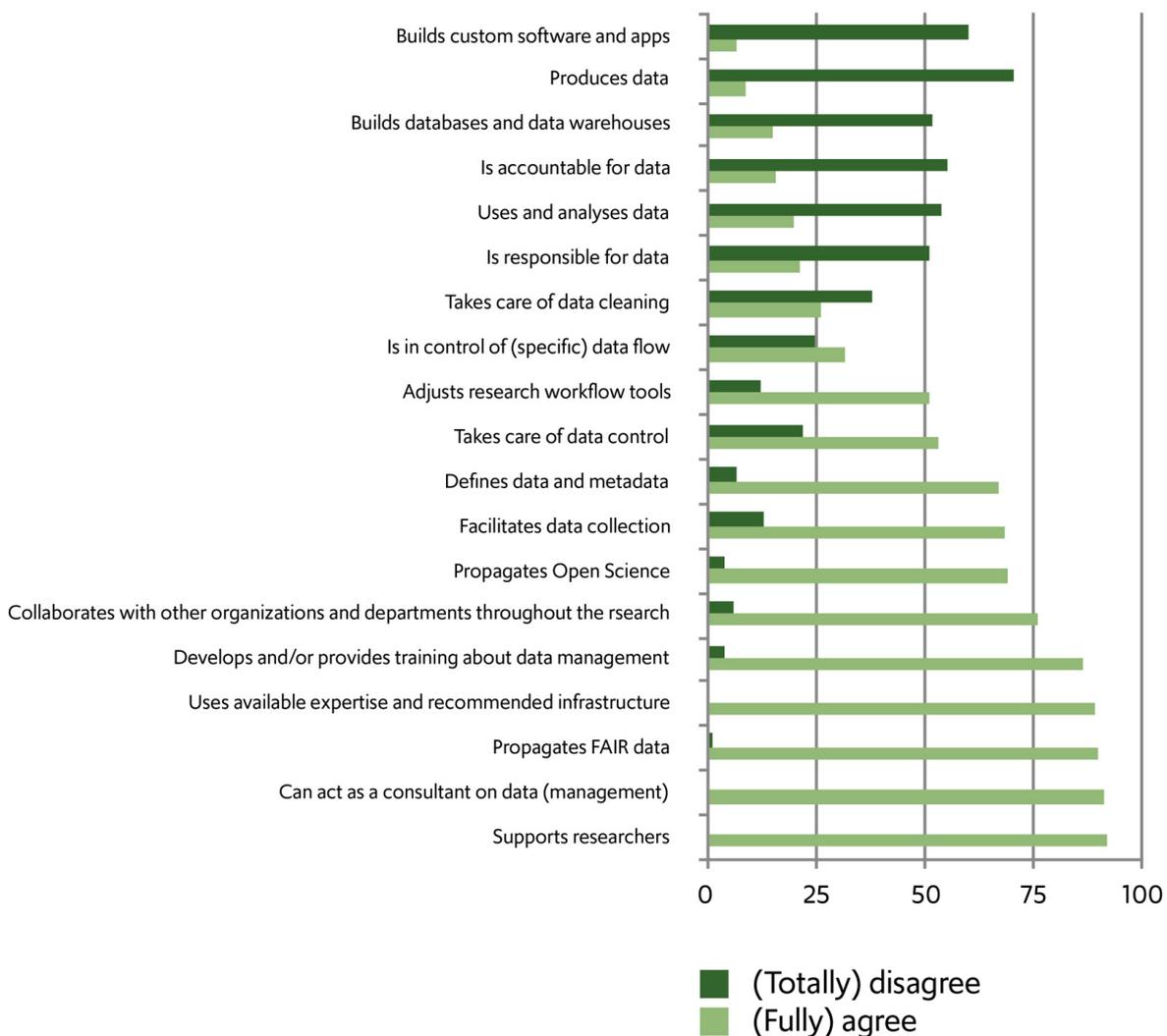
This is a review of the support tasks, listed in order of most cited as 'presented/performed':



The most frequently mentioned 'already presented/performed tasks' belong mostly to the generic and policy task area. 'Data standardization' is most often mentioned as a main concern for which as yet support is lacking (49 of the 94). Data standardization constitutes a very broad task that has strong links with the I of FAIR data (interoperability). Implementing data standardization demands a considerable amount of (international) coordination between numerous parties, both at a policy and an operational level.

Least named as answer to the question what is needed, were tasks from the embedded and operational task area especially 'data validation', 'data identification', 'data cleaning', 'data integration' and 'data engineering (scripts)'. The majority of UMC respondents indicated that data support was well-organized but also that there is still need for support. The least number of respondents to endorse this viewpoint were from universities of applied sciences.

'Tasks from the embedded and operational task area were least cited as answer to the question, what is needed'



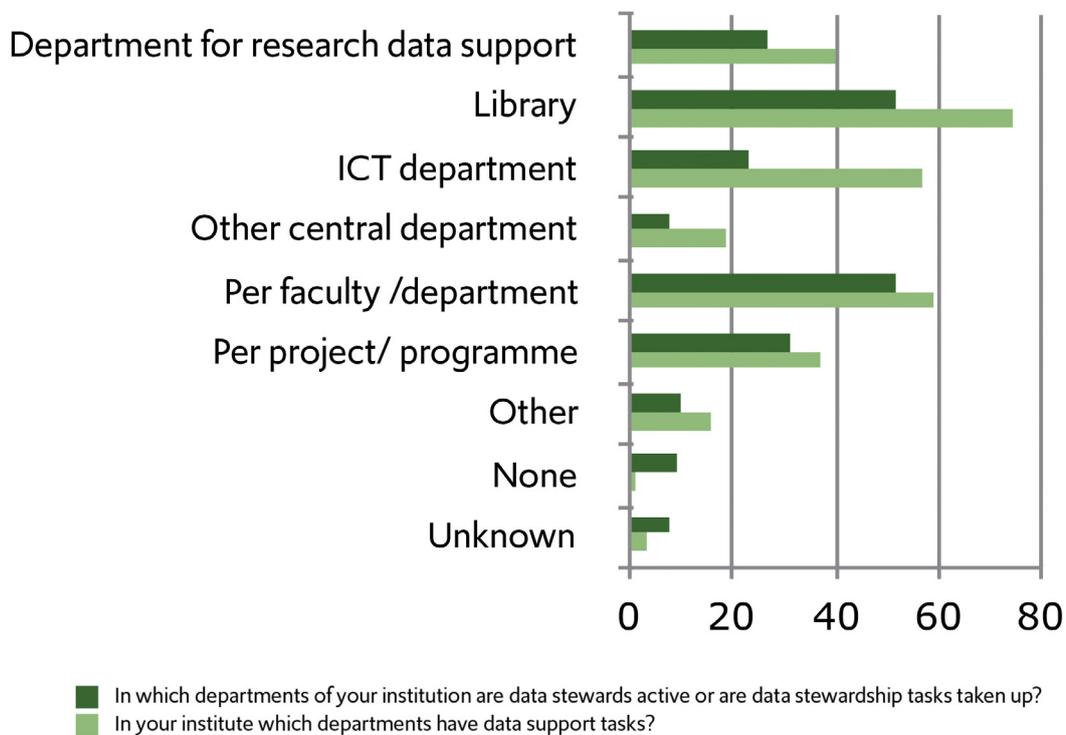
Qualities, responsibilities and skills of data stewards

Respondents could indicate if they thought data stewards should be competent in 19 pre-selected qualities, responsibilities and skills, on a scale of 1 (totally disagree) to 5 (fully agree). The following is a survey of the requisite qualities, responsibilities and skills arranged according to which scored most frequently as 'fully agree'. For clarity the answers 'agree' and 'fully agree' were combined, just as the answers 'disagree' and 'completely disagree'. (Diagram page 15.)

The qualities, responsibilities and skills most often flagged as 'agree' and 'fully agree' belong predominantly to the generic data stewardship and policy, strategy and coordination task areas.

Organization

Where data stewards work and where data support tasks are allocated can be depicted as follows:



The way in which research data management or data stewardship is organized is very diverse. Within an organization data support, -financing, -policy, -procedures, -roles and responsibilities are sometimes arranged centrally; sometimes functions are decentralized but a combination of both options may also occur. There is no visible relation between central or decentralized organized support and the support tasks named by respondents as being necessary or available.

Topical developments

Nearly all cited topical developments tie in with generic data stewardship tasks and/or policy-oriented data stewardship tasks.

| | |
|--------------------------------------|----|
| ORGANIZING SUPPORT | 36 |
| POLICY IMPLEMENTATION | 22 |
| INFRASTRUCTURE, TOOLS AND FACILITIES | 22 |
| POLICY DEVELOPMENT | 15 |
| AWARENESS AND KNOWLEDGE DEVELOPMENT | 5 |
| WORK FLOWS | 4 |

Embedded data stewardship tasks

When indicating necessary support tasks and qualities of data stewards, those tasks and qualities that fall within the scope of the embedded data stewardship task area are less frequently mentioned. When organizing data support, it turns out that such duties have nonetheless already been coordinated in one way or another per department or project. The collected vacancies also make frequent mention of tasks from the *Embedded and operational* task area. The reason for this might be that the survey gives less prominence to embedded data stewardship tasks because the data stewards who perform them are often active within a department or project and therefore less visible to the rest of the organization. Currently there is also more focus on generic tasks and a more policy-oriented data stewardship role. This is inherent to the fact that many institutes are still considering how to organize and professionalize their data stewardship (support).

To investigate these assumptions in greater detail and further assess the rough classification in task areas and also as follow-up to the online survey, a number of interviews were held with people who work as data stewards..

'It could well be that the survey gives less prominence to embedded data stewardship tasks because embedded data stewards active within a department or project are less visible to the rest of the organization'



INTERVIEWS

The following people were interviewed:

- Joke Bakker, software developer and research data management coordinator at the Groningen Evolutionary Life Sciences institute (GELIFES) of the University of Groningen,
- Alastair Dunning, head of Research Data Services, TU Delft Library and head of the 4TU Centre for Research Data,
- Christina Elsen, consultant at the university library and research data office at the University of Groningen,
- Joeri Kalter, data steward at Wageningen University & Research, Department for Humane Nutrition and Health,
- Wouter Kool, data manager for several specific (archaeological) projects, including Nexus1492, Leiden University,
- Lena Karvovskaya, central data manager at the Utrecht University Library,
- Renate Mattiszk, data librarian, Saxion Library, Saxion University of Applied Sciences,
- Dick Vestdijk, information specialist at the library of the University of Applied Sciences Utrecht.

The interviews show that data stewardship at institutes is always divided between different levels: often centrally and at faculty level and sometimes also at project or institute level. Organization of data stewardship is still being developed at most institutes. The ambition is to increase the number of data stewards; especially those with specific know-how who can perform embedded tasks. The division into task areas is generally recognized. However, some of those interviewed suggested a further subdivision of embedded tasks or different names for the task areas.

Renate Mattiszk, data librarian at the Saxion University of Applied Sciences, is pleased with the clear division into three task areas. She concerns herself mainly with tasks from the *Generic and advisory* task area. At Saxion University of Applied Sciences, the library serves as headquarters for data stewardship. Saxion Research Services is a network organisation which draws on knowledge and support from all kinds of sources. Policy at Saxion dictates that researchers must register a project they want to initiate. Researchers ask Renate Mattiszk for general advice: about a data management plan, the GDPR, archiving and how to use the Saxion Research Cloud Drive etc. Renate does not perform *embedded* tasks that demand other competencies and training, like writing scripts and data analysis. In her opinion one should not try to allocate all data stewardship tasks to one person. That is an underestimation of the role of data steward. Ideally data stewards should work as a team, each with their own competency, in order to provide made-to-measure RDM support.

At Utrecht University the library also serves as a hub for first-line support provided by research data consultants: they help researchers with questions that can be solved relatively easily. In addition there is a pool of data managers who is assigned to projects for an extended period; a sort of secondary support. These data managers help with different issues throughout the entire research cycle of the project. The support they offer is very broad:

from writing down procedures to data engineering. The data manager pool allows the division of tasks among different people who each have their own particular competency. The interviewee, **Lena Karvovskaja**, central data manager at the Utrecht University Library, says that she for instance works on 'meta data management' by setting up work flows and searching for precedent cases that help to create the right preconditions for research projects. Other data managers are allocated more specific tasks within a certain project. There are also data managers at faculty level at Utrecht University. Their role depends on what is required within the faculty. They often have specific knowledge of the faculty domain but are not attached to specific projects. The difference between the various data stewardship roles at Utrecht University is somewhat blurred. Work is underway to improve all the different issues mentioned in the survey; no point has been overlooked. The distinctive task areas as described in the report are recognizable but instead of *embedded* and *operational* Lena regards *domain-specific* tasks as a counterpart to generic tasks.

Wouter Kool has been specifically recruited as data manager for several (large-scale) archaeological projects led by Leiden University. Much of his work lies within the *embedded* and *operational* task area, but he also offers more generic advice, for example, about data management plans for new projects. According to him the added value of an embedded data manager is that he or she has knowledge of data used in a certain domain, in his case, 3D models. Data stewards must nonetheless maintain a certain distance to preserve a sense of perspective of the working method that has evolved in a domain over time. Proper training can help in realizing this. It is also very important that researchers know where they can get help when it comes to research data management. Even if a data manager has been assigned to a project, during the research process, researchers often ask questions (too) late. Currently a lot of pioneering work is still being done, but with the availability of ever more professional support, the situation will surely improve: the more people who learn from practice that it pays to ask for help at an early stage, the less this will prove an obstacle in the future. In Wouter's opinion, data steward pools like the one in Utrecht, where data stewards' specific skills are taken into account when allocating projects, are a good way of providing operational help. It is also important that communication about the availability and importance of this support is well-organized.

Joke Bakker, software developer and research data management coordinator at the University of Groningen (RUG), says that embedded data stewardship works best at three different levels; centrally in a data office and also at institute and faculty level. At the RUG, the Data Federation Hub (DFH) is very effective. This is a network that connects and informs all research data support staff and services. Many researchers find it difficult to find information about RDM. Often the data steward acts as *people's manager* and does a lot of missionary work.

Christina Elsenga, who works as consultant for the university library and the Research Data Office at the RUG, also says that community building is essential. It also helps to highlight the availability of support and to make clear how to get support. Christina remarks that data stewardship tasks and responsibilities need to be divided evenly among all research participants. This implies that as data steward you must consider which tasks you can delegate to the researcher and which you should adopt yourself.

Dick Vestdijk, information specialist at the University of Applied Sciences Utrecht (HU), recognizes the task areas outlined in this report. He also has generic tasks but sees himself mainly as an embedded librarian. The more general role is taken up when the HU acts as coordinator of a project or research assignment, or when health data is used for research. At the HU there is a senior IT staff member who coordinates policy and is authorized to take financial decisions. This senior staff member usually decides on the purchase of systems that require financing. The *embedded* and *generic* task area focuses mainly on support. In Dick's view it therefore makes sense to take the obvious step of adding 'financing' to the former task area. Following policy directives is generally not a problem but with regard to financing the question is always: who pays for it? The researcher? The knowledge institute – the organization level above every lectorate – or the university of applied sciences? Dick also remarks that because the term 'generic' is often used in other contexts it would be helpful to make the term more specific for data management. For example, domain-oriented or domain-transcending, or: discipline-specific and discipline-transcending data management..

One of **Joeri Kalter's** tasks as data steward at Wageningen University and Research (WUR), department for Humane Nutrition and Health, is to check if policy is well-implemented. At the WUR data stewardship is often not the only task of data stewards. Many data stewards also undertake research or other supplementary tasks. WUR policy is developed at a different level to that of data steward but Joeri is also member of a workgroup that helps to structure policy initiatives. The WUR is working to draw up a good design for the embedding of data stewardship, among other ways by establishing a network for data stewards. The organization of data management can certainly be improved according to Joeri. Clear policies are needed – awareness is not enough – documents should make clear who you can ask for advice depending on which stage your project finds itself in.

Alastair Dunning, head of Research Data Services at the TU Delft Library and head of the 4TU.Centre for Research Data, says that at the TU Delft, *embedded* data stewardship tasks should be mainly allocated to data managers as *local subject specialist*. The data stewards, who have clear methodological expertise and knowledge of a certain discipline, are stationed at faculty level. *Coordination and strategy* is taken up by people in already existing management roles at the TU Delft. These roles are specifically focussing on data management..

Alastair adds that good coordination is essential if data stewardship has been organized in a decentralized way like in Delft. TU Delft is one of the trendsetters when it comes to competent organization of data stewardship. In addition to central support, there is also support at faculty level and data stewards are embedded at faculty level and in external networks, for example in the field of infrastructure. Even so, there is still much work to be done in the transition to Open Science. For example, as yet there are no data managers as *local subject specialists* yet: these will be recruited shortly.

Summarized

There is general agreement that good basic training for data stewards is a prerequisite for professional data stewardship. Many data stewards have completed the (basic) training course offered by Research Data Netherlands (RDNL). But knowledge of the *Netherlands Code of Conduct for Research Integrity*, knowledge of institute and faculty policy plus more specific know-how of the more embedded tasks is seen as essential.

In general everyone agrees that a mix of tasks works well for data stewards. Full-time policy officers run the risk of becoming paper-pushers. On the other hand, when working solely on one project, it's important to maintain an overall view of goings-on at an institute. Internal networks of data stewards or good coordination are seen as effective ways of achieving this, in addition to working with a pool of data stewards who can each deploy their own expertise.

Community forming within such a pool is essential - between generic and embedded data stewards mutually, but also in a domain-transcending way. It is also important to connect with the strategy and policy being pursued in one's own institute so that policy officers and management staff can be made aware of experiences from actual practice. But above all, sufficient data stewards must be recruited in order to offer competent service, working in close cooperation with researchers and their specific data.

Other initiatives: Wageningen University & Research and ZonMW

Currently there are two other initiatives being pursued with regard to data stewardship professionalization with which we have coordinated our findings.

The first initiative is the data stewardship project of Wageningen University & Research (WUR). At WUR they noticed that in many research groups, researchers or technical staff are performing data stewardship tasks to support their fellow researchers. Sometimes they just take care of the specific data management plans of the group. In other cases they advise on suitable and sustainable storage solutions, help with data analysis, provide for interoperable linked data or recommend on subject specific data repositories. These staff members invest time and expertise in good data stewardship but as yet receive little recognition, appreciation or reward.

WUR hopes to comply with the FAIR requirements within 5 years. Sustainable data stewardship is one of the prerequisites for attaining that goal. The WUR data stewardship project is working to create a data steward @ WUR networks. The project strives to gain over all recognition for the role of the data steward and data stewardship tasks in FAIR data handling. In addition, a training curriculum for FAIR data stewardship is being developed.

In April 2019 the project will deliver an administrative advice on Data stewardship @ WUR. The network for data stewards is a first step towards FAIR data governance at the WUR. By establishing this network the WUR shows it is 'serious about Data. The results of this LCRDM report will be included in the WUR-advice. Appendix 1 contains a link to the WUR project plan.



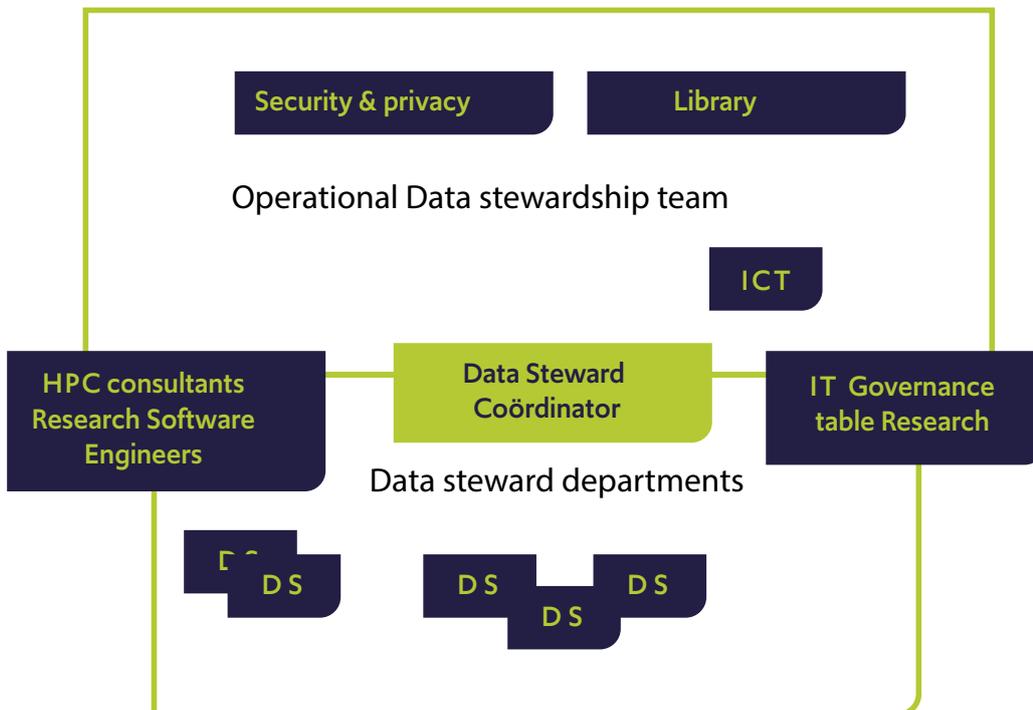
Also currently being developed is a project financed by ZonMW called *Towards FAIR Data Steward as profession for the Life Sciences*. This project is working to establish a common job description and a universally accepted overview of the knowledge, skills and competencies needed by a data steward who wishes to play a role in the Dutch life sciences community. These knowledge, skills and competencies are translated into concrete learning objectives, that in turn can be used to develop an education line and training material for data stewards (including a design for an e-learning module).

To assure sustainable implementation and fine-tuning with existing initiatives, the interim results of both projects have been shared. One of the first ZonMW project results, a matrix that can serve as a basis for a general profile of a life sciences data steward, serves as an important source of information for this LCRDM report. Appendix 1 contains a link to this matrix.

The ZonMW project with its explicit focus on the knowledge, skills and capacities of data stewards and the translation thereof into practical learning goals that can be implemented in data stewardship training courses, forms a significant supplement to this report.

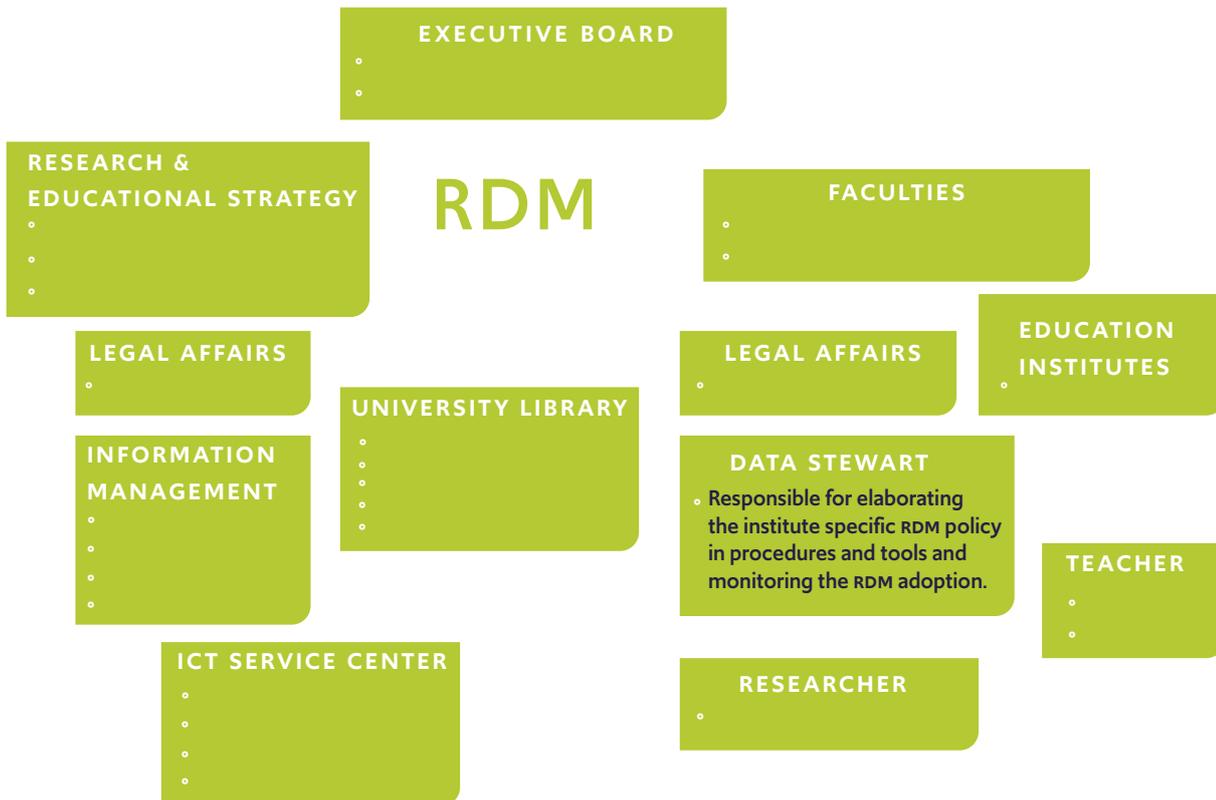
EXAMPLES OF (FUTURE/DESIRED) EMBEDDING OF DATA STEWARDSHIP

The survey results show that participants would like examples of the embedding of data stewardship at other institutes or examples of the desired but as yet not implemented situation. At many Dutch institutes this is indeed still being developed. Even so it makes sense to include, as inspiration, several examples of (future) embedding of data stewardship in this report. Numerous institutes have taken similar initiatives to realize the embedding of data stewardship. It has been decided to include some of the most diverse examples, in some cases still only rough outlines.

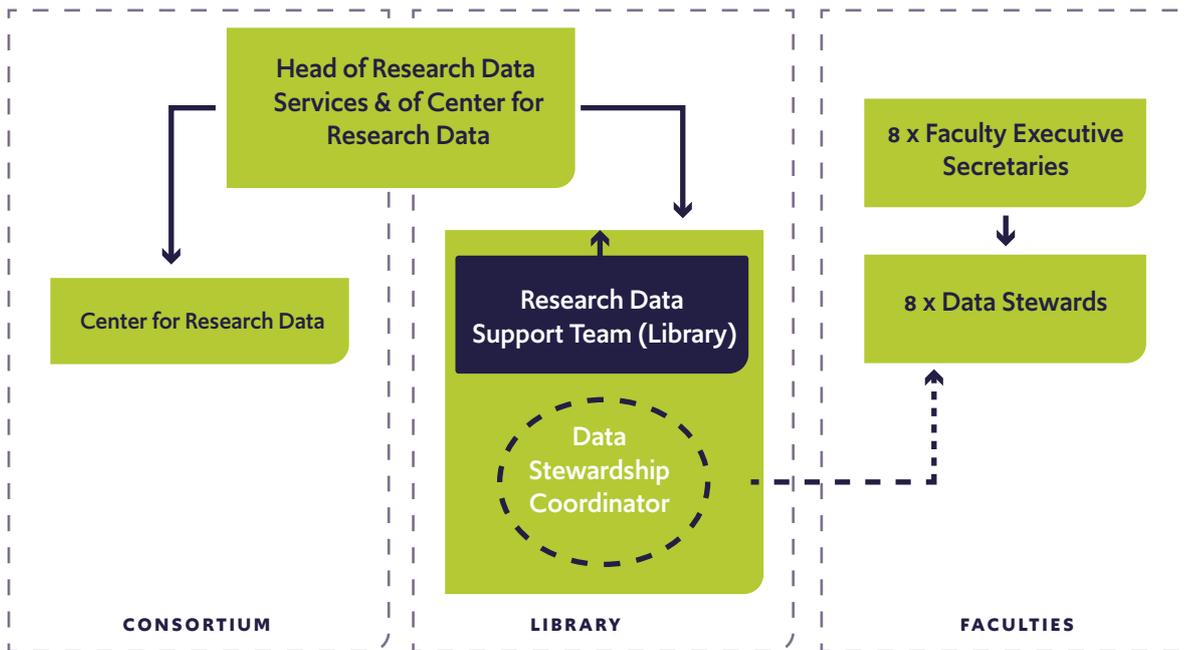


The data stewardship team in this example is a collaboration between security and privacy officers, library staff and ICT staff. A full-time data steward coordinator is responsible for managing the team. The data stewardship team provides backing for researchers, for example, by reviewing data management plans, publishing data sets and offering GDPR support.

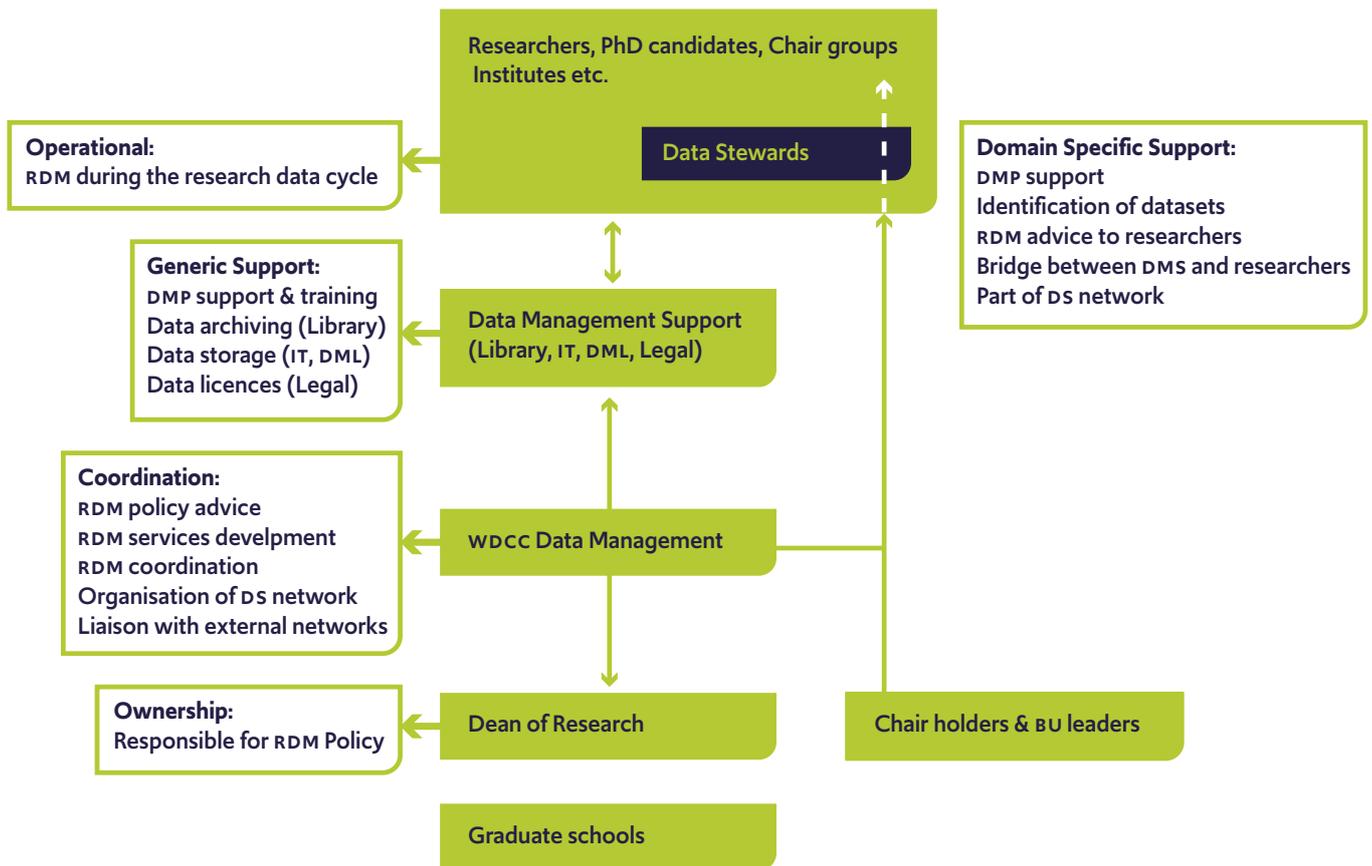
The faculties have been asked to appoint a data steward. In most cases this is senior researcher. In a number of faculties the function of data steward has been formalized. Data stewardship tasks are increasingly being performed by people in new functions such as Research Software Engineers or HPC consultants, sometimes as part of the ICT service (IMS) and sometimes within the faculties and research groups. 'IT-governance' is responsible for ICT-projects in the RDM field and also concerns itself with developing policy on data stewardship..



Since 2017, every research institute in this particular institution has appointed its own data steward. This may be an auxiliary function (in addition to, for example, laboratory manager or researcher) or in a separate role with a half-week task allocation. The data steward is embedded at the research institute and is involved with the full scope of research policy, IT and support services with a primary focus on RDM. Since 2013, the institution also has central data stewards at the university library who are known as *data librarians*. These are generic experts who support the data stewards and deal with everyday support tasks for researchers and students.



In this example, two teams provide support for research data: the central Research Data Support team in the library and the discipline-based data stewards deployed at the different faculties. The data stewards act as primary contacts for the researchers because of their specific research know-how. The central team serves as contact point for more general affairs, like use of the central data archive. The data stewardship coordinator, who for organisational reasons holds office in the library, is the *linking pin* between the central support team and the data stewards. A coordinator manages the team of data stewards.



In this example, all parts of the organisation are engaged with data management. Together they form the data management structure of this institute. The *Dean of Research* is ultimately responsible. Coordination of data stewardship is organized via a central data *competence centre*. Data management support based in the library performs more generic tasks. The data stewards are part of the research community.

CONCLUSION

This report investigates the provision of data stewardship tasks and roles in Dutch research institutes. The task group has given insight in the needs of data stewards in Dutch institutes in addition to creating awareness for what they develop and implement.

Many institutes are currently professionalizing their data stewardship. This is also a topic at a national, administrative level. The development of data stewardship is a general phenomenon; the number of data stewards will increase in the near future, especially those with domain-specific expertise who can perform embedded and operational tasks. Considering national and international developments, the professionalization of research data management support or in other words – good data stewardship – is essential.

Professional data stewardship needs professional data stewards with the right training possibilities and recognition for the tasks they perform. Furthermore, good coordination is crucial, especially in the case of a – much desired – decentralized format when data stewards work closely with researchers and cater to their specific support requests.

Policy tasks like formulating organisation strategies are often carried out by other staff (for example, a data stewardship coordinator or RDM coordinator), but nonetheless they make up an integral part of the organisational framework for good data stewardship. It is imperative that data stewards, policy officers and managers share knowledge and findings. Close cooperation is recommended – also among data stewards and the different domains. That's partly why embedded data stewardship in an existing organisation structure is so important. But above all it's crucial that enough data stewards are available to offer researchers close-range support.

'Professional data stewardship needs professional data stewards, who receive proper recognition for the work they do'

The results of this report show that the names and positioning of the different data stewardship roles within institutes vary greatly. The same tasks regularly crop up in different positions, functions and roles. Focussing on classification according to task areas therefore seems a sensible approach that can help in formulating new task descriptions.

The differences between the task areas *Embedded and operational*, *Generic and advice* and *Policy strategy and coordination* are a good indication of where the emphasis of a data stewardship function lies, although the survey results and the vacancy descriptions show that functions always consist of a mix of diverse tasks. The *exclusively generic* or



operational data steward does not exist. And that is desirable because flexibility to assume a role is essential.

Those interviewed endorse the differences in interpretation and positioning for data stewardship as revealed by the survey. The division into three task areas is widely recognized. Possibly the design of the task areas can be expanded with an overlapping ICT/infrastructure task area with better integration of the financial component. This could well be an interesting follow-up project.

A number of interviewees remarked that setting up a pool of data stewards, each with their own data and/or domain expertise, is a good way to offer (domain) specific support. No matter where data stewards are positioned in the organisation (centrally or in a research group) it does not stop them contributing to such a pool.

It has been said that a pre-requisite for professional data stewardship is sound basic training and recurrent vocational schooling (training programmes). This is something the project financed by ZonMW is looking into. This report has investigated embedded data stewardship tasks and roles in various organisations. The ZonMW project looks at the necessary knowledge, skills and competencies and how these can be translated into training courses for data stewards..

The survey also shows that the task 'data standardization' urgently needs to be regulated, a daunting mission that will need considerable (international) fine-tuning. Another noteworthy outcome is that tasks and qualities that fall within the embedded task area are less often mentioned than expected. This is inherent to the fact that many institutes are still searching for the right way to set up their data stewardship (support) and fully professionalize it. It also appears that research institutes start by defining generic tasks. We expect that in the near future there will be more call for embedded and operational support closer to the researcher.

Good community building is important. Of central importance when forming a data stewardship network is that available support should be simple to find. Staff members must be visible and easy to contact. It is therefore advisable to keep this in mind with regard to the positioning of data stewards. Optimal exchange of knowledge is essential, both at institute level, over the full breadth of the domain and nationally.

Towards new function profiles

When drafting this LCRDM report, it was stated that ideally the results could help to draw up new or improved function profiles for data stewards. At the moment data steward functions – as shown by the published vacancies – are classified according to different function groups within a function system. Greater uniformity is needed, also to ensure the acknowledgement of data stewards and data stewardship tasks. It is clear that data stewardship functions are no longer a niche function and that new data stewardship tasks are being generated.

When revising function profiles it is advisable to formulate them loosely so they incorporate all relevant responsibilities. Clearer recognition and accompanying career perspectives will make the role of data steward more appealing. It will promote their visibility in the organisation and prevent competitive recruitment by institutes. Possibly it will also help retain data stewards for the research community.

Also when researchers take on certain data stewardship tasks themselves, it is essential that they receive proper recognition. The current WUR project clearly focuses on this phenomenon. The fact that people in different positions deal with all kinds of data stewardship tasks doesn't reverse the fact that more resources and more recognition are needed to professionalize sustainable data stewardship at the different institutes.

Conclusion

Professional data stewardship needs professional data stewards. Everyone who performs data stewardship tasks should receive proper recognition including matching career perspectives and appropriate study opportunities or training.

In the field there is plenty of enthusiasm for establishing a pool of data stewards, each with their own expertise, as a way of offering the proper (domain data specific) support.

Each member of the pool deploys his/her own competencies to offer tailor-made support and solve specific questions posed by researchers. This in turn warrants a broadly formulated function profile for data stewards. A pool or network of data stewards with a clear job description and adequate recognition for the work they do, contributes to an optimal professionalization of services and helps to structure the research process to unburden the researcher and allow an institute to fulfil its duty of care.



APPENDIX 1: LINKS

- [Data Science Framework: Part 1](#) EDISON project;
- [Skills and Capability Framework](#), EOScpilot project;
- [Pilot Data Information Literacy Competencies Matrix Scaffolded Across Undergraduate, Graduate and Datasteward Levels](#) Sapp Nelson et al.;
- [A life sciences data steward function matrix](#), Scholtens, et al.;
- [Datastewardship binnen de Radboud Universiteit](#) Mijke Jetten, Robert van der Vooren;
- [Data Roles of the Future at TU Delft](#) Alastair Dunning, Marta Teperek.

LINKS TO BACKGROUND DATA

- [Vacancies](#) divided in task areas;
- [Questions survey](#)
- Results survey in [xlsx](#).
- Results survey in [sav](#)

LINKS TO DATA STEWARDSHIP PROJECTS

- [LCRDM task group Data Stewardship](#) LCRDM project;
- [Rol of the Data Steward in the organization of Data Management](#) WUR project (Dutch only);
- [Towards FAIR Data Steward as profession for the Life Sciences](#) ZonMW project;



COLOPHON

DATA STEWARDSHIP ON THE MAP

A study on tasks and roles in Dutch research institutes

PUBLICATION DATE | April 2019

DOI | 10.5281/zenodo.2669150

(DOI Dutch version: 10.5281/zenodo.2642066)

AUTHORS

Ingeborg Verheul (LCRDM), Melanie Imming (Imming Impact), Jacquelijnn Ringersma (Wageningen University & Research), Annemie Mordant (Maastricht University/MEMIC), Jan Lucas van der Ploeg (University Medical Center Groningen), Martine Pronk (Utrecht University)

CONTRIBUTORS

LCRDM Task Group Data Stewardship

Maarten van Bentum (University of Twente), **Boudewijn van den Berg** (LCRDM support), **Marije van der Geest** (University Medical Center Groningen, liaison ZonMW project), **Henk van den Hoogen** (Maastricht University, also LCRDM Advisory Group), **Melanie Imming** (Imming Impact, consultancy), **Mijke Jetten** (Radboud University), **Toine Kuiper** (Eindhoven University of Technology), **Annemie Mordant** (Maastricht University/MEMIC, also LCRDM Advisory Group), **Nicole Koster** (University of Twente), **Jan Lucas van der Ploeg** (University Medical Center Groningen), **Gera Pronk** (ISSeeTea), **Martine Pronk** (Utrecht University), **Saskia Rademaker** (The Hague University of Applied Sciences), **Jacquelijnn Ringersma** (Wageningen University & Research, also LCRDM Advisory Group), **Marta Teperek** (Delft University of Technology), **Marijke Verheij** (Groningen University), **Ingeborg Verheul** (LCRDM)

WITH CONTRIBUTIONS OF

Joke Bakker (Groningen University), **Vincent Boxelaar** (Groningen University), **Joldy Coenen** (Maastricht University/MEMIC), **Alastair Dunning** (Delft University of Technology), **Christina Elsenga** (Groningen University), **Joeri Kalter** (Wageningen University & Research), **Wouter Kool** (Leiden University), **Lena Karvovskaya** (Utrecht University), **Renate Mattiszik** (Saxion, University of Applied Sciences), **Paula Rinkens** (Maastricht University/MEMIC), **Dick Vestdijk** (Utrecht University of Applied Sciences)

LAY-OUT | Nina Noordzij, Collage, Grou

TRANSLATION | Gosse van der Leij

COPYRIGHT

all content published can be shared, giving appropriate credit

creativecommons.org/licenses/by/4.0



LCRDM



LCRDM supported by

