

PROJECT DELIVERABLE: LIFE-SCIENCES DATA STEWARD FUNCTION MATRIX, VERSION 1.1

DOI: [10.5281/zenodo.2561723](https://doi.org/10.5281/zenodo.2561723)

Date stamp: February 8, 2019

RIGHTS



Rights and permissions: copyright © 2019 The Authors.

This work is licensed under a [Creative Commons Attribution 2.0 Generic License](https://creativecommons.org/licenses/by/2.0/).

AUTHORS

[Salome Scholtens](#) (UMCG), Petronella Anbeek (UMCU), [Jasmin Böhmer](#) (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. All documents related to this project can be found on the Zenodo Collection “[Towards a community-endorsed data steward profession description for life-science research](#)”.

This publication contains the first project deliverable: a matrix, that may function as the basis for a common job description of a data steward that is broadly supported within the Dutch life-sciences community. In the next phase of the project, this matrix will be complemented by knowledge, skills and competencies of a data steward, which will be translated into concrete learning objectives. These 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.

KEYWORDS

Data management, data stewardship, job description, function matrix, training, skill building, competencies, life-sciences

BACKGROUND

Professionalising the data stewardship function and including sufficient data expertise and capacities in research projects is essential in today's research, in particular in data-intensive research areas such as genomics and personalised medicine, which also involves highly sensitive data. 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 on a national level is also lacking. 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 institutes and research projects.

PRINCIPLES

- The focus of this project is the life-sciences domain. Existing knowledge on data stewardship competencies from different domains is further developed and tailored to the life-sciences. However, the outcomes of this project could be beneficial to other domains as well, as in our experience many data stewardship responsibilities, tasks, skills and competencies can be shared and used across domains. Furthermore, as these data stewardship responsibilities, tasks, skills and competencies are often stronger related to the various types of data than to specific disciplines, the function description can be used for data stewards from other scientific domains as well.
- Although we focus on the life-science domain, the existing reports we used as a base are mainly domain-agnostic as well, focusing on open science and data science in general. Our project builds on and aligns with ongoing activities on defining these skills and competencies, both from the domain-agnostic side ([EOSC](#), [EDISON](#), [Purdue](#), [DAMA](#)), as well as the domain-specific side (ELIXIR).
- Where the existing approaches to data stewardship ([EOSC](#), [EDISON](#)) are primarily top-down, our approach is bottom up, starting in the practical field of data stewardship practices in the life-sciences in the Netherlands by analysing vacancy texts and job profiles for data stewards (and alike) and complementing this with our own experience in data stewardship (4 members of the core team fulfil a data steward role in their organisation). In this way we add to existing activities, reports and knowledge the 'knowledge, needs, and experiences from the daily job of being a data steward in the domain of the life-sciences'. This bottom-up approach is already mentioned in the reports ([EOSC](#), [EDISON](#)) we build upon as a useful contribution and next step.

- In addition, we also map these practices (responsibilities, tasks etc.) - as a top-down approach - to the existing principles and approaches that are relevant for the field of data stewardship, such as the [data life cycle](#) approach¹, the [FAIR principles](#), the [Purdue](#) competence areas² and the [DAMA](#) knowledge areas³. We thus align both the bottom-up approach (practices) and the top-down approach (existing approaches and principles).
- Consequently, our research method and sample is rather pragmatic, aimed at creating a first version definition of data stewardship (exploratory research; sensitizing concepts), instead of developing and testing a model ‘that fits all needs’ (saturation). We do realize that our deliverables require iterative processes, so we invite others to contribute to/bring our first version (see the matrix below) to the next level.
- This is even more relevant since the first two deliverables, namely the data steward function matrix and the outline of knowledge, skills & competencies, are not the end goal: the common job description(s) and an agreement on knowledge, skills and competencies of a data steward in the life-sciences domain help us to map the needs (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.
- With our project, we thus also ensure sustainable implementation, alignment with existing education, and continuous (re)evaluation of data steward function and training.

METHOD

Deliverable 1: Function description

The content of the function description was defined in multiple iterative cycles. First, it was based on an analysis of existing competency frameworks for data management and stewardship and recently published reports such as the [EOSC pilot](#), [EDISON](#), [Purdue](#) and [DAMA](#) competency frameworks, as well as the [FAIR principles](#) (round 1). This was complemented with a review of over 40 published data stewardship vacancy descriptions as well as with experiences of persons working as data experts in the Netherlands (round 2). These 40 function profiles came from the own institutes and an internet search on job openings. In a third round the function profile was presented to the Consultation

¹ **Data life cycle:** ① creating data; ② processing data; ③ analyzing data; ④ preserving data; ⑤ giving access to data; and ⑥ reusing data.

² **Purdue competence areas:** ① databases and data formats; ② discovery and acquisition of data; ③ data management and organization; ④ quality assurance; ⑤ data conversion and interoperability; ⑥ metadata; ⑦ curation and re-use; ⑧ cultures of practice; ⑨ data preservation; ⑩ data analysis; ⑪ data visualization; and ⑫ ethics, including citation of data.

³ **DAMA knowledge areas:** ① data governance; ② data architecture; ③ data modelling and design; ④ data storage and operations; ⑤ data security; ⑥ data integration and interoperability; ⑦ documents and content; ⑧ reference and master data; ⑨ data warehousing and business intelligence; ⑩ metadata; and ⑪ data quality.

Committee of the project with national experts and stakeholders on several aspects of data stewardship, to the National Coordination Point for Research Data Management and to a network of Dutch data stewards. In parallel, interviews were held with data stewards and national stakeholders on the current content and position of the data steward.

Up to now, in the project, two levels are distinguished.

- Data steward Level A: this is an institutional/department focused and a coordinating, policy oriented function.
- Data steward Level B: this is a project focused and an operational, supporting-oriented function.

On each level, nine different function areas are defined. For details see [the matrix](#) below.

NEXT STEPS

In the next phase of the project, for each level, we will further define the knowledge and skills needed and translate them 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 is ensured by collaboration with e.g. the national coordination point research data management ([LCRDM](#)), [NFU Data4lifesciences programme](#), the [HANDS handbook](#) for Adequate Natural Data Stewardship, the Dutch Techcentre for Life Sciences ([DTL](#)), and with data education at universities of applied sciences. With this project we contribute to a professionalization of the data steward function within the life-sciences domain so future research can be optimally supported.

MATRIX
Version 1.1

Data steward Level A - Institutional, coordinating, policy						
	Responsibilities	Activities / tasks	Mappings (see below for reference)			
			Research data life cycle	FAIR data	Purdue competence	DAMA knowledge
1) Policy / strategy	Responsible for advice on and development, implementation and monitoring of a research data management (RDM) policy and strategy for the research institute, which includes the complete research data life cycle, and supports FAIR data and Open Science, in alignment with the relevant stakeholders and within financial and legal constraints, within the institute and in the context of the institute. The policy is the basis for (project) data management plans (DMP).	<ul style="list-style-type: none"> Develops, implements and monitors the institute's RDM policy. Advises the institute's management on short- and long-term actions to advance RDM in the institute. Assesses and monitors institute's time and financial investments in relation to the institute's needs for RDM. Explores new needs, opportunities and trends in RDM. 	① ② ③ ④ ⑤ ⑥	F A I R	③ ④ ⑧ *	① ⑪ *
2) Compliance	Responsible for compliance of the RDM policy to the Netherlands Code of Conduct for Academic Practice and the General Data Protection Regulation (GDPR), as well as continuous alignment with legal and ethical standards.	<ul style="list-style-type: none"> Ensures compatibility of the RDM policy and monitors compliance. Contacts the institute's privacy officer, legal advisors or ethical board in case of questions regarding compliance. Translates policies from legal/privacy officer to the institutes practice. Develops and/or guides standard solutions for recurring data issues and for data classification, including input for the DPIA. 	① ② ③ ④ ⑤ ⑥	F A I R	③ ④ ⑧ ⑫ *	① ⑤ ⑪ *
3) Alignment with FAIR data principles	Responsible for compliance to FAIR data.	<ul style="list-style-type: none"> Ensures compatibility of the RDM policy and monitors compliance to the FAIR principles Pursues and advises on the findability (F) of data, including adequate infrastructures, persistent identifiers and rich (institute-specific) metadata standards. 	① ② ③ ④	F A I R	③ ④ ⑧ *	① ⑪ *

		<ul style="list-style-type: none"> Pursues and advises on the accessibility (A) of (meta)data to potential (re)users. Pursues and advises on the interoperability (I) of data, including broadly applicable languages, vocabularies and other standards. Pursues and advises on the reusability (R) of data, including documentation and licenses with the conditions for reuse and IP rights. 	<p>⑤</p> <p>⑥</p>			
4) Services	Responsible for the availability of sufficient support on RDM, in staff or services, for the researchers and research support staff of the institute.	<ul style="list-style-type: none"> Advises management, researchers (including PhD students and students) and support staff of the institute on RDM. Initiates the set-up and update of suitable support facilities or services in the institute. Monitors the development and implementation of (standard) solutions for recurring RDM issues and questions. Monitors the availability of sufficient support staff. Initiates hiring or training of staff. 	<p>①</p> <p>②</p> <p>③</p> <p>④</p> <p>⑤</p> <p>⑥</p>	<p>F</p> <p>A</p> <p>I</p> <p>R</p>	<p>③</p> <p>*</p>	<p>①</p> <p>*</p>
5) Infrastructure	Responsible for the availability of adequate data and e-infrastructure and tools for RDM to comply with the institute's RDM policy and alignment to (inter)national data and e-infrastructures.	<ul style="list-style-type: none"> Requests and acquires tools and infrastructure for RDM within the institute. Monitors the need, use and availability of infrastructure and tools. Advises the management of the institute on infrastructure and tools. Explores new trends in infrastructure and tools for RDM. Is aware of and advises the institute on relevant (inter)national data and e-infrastructures. 	<p>①</p> <p>②</p> <p>③</p> <p>④</p> <p>⑤</p> <p>⑥</p>	<p>F</p> <p>A</p> <p>I</p> <p>R</p>	<p>①</p> <p>③</p> <p>*</p>	<p>①</p> <p>②</p> <p>*</p>
6) Knowledge	Responsible for an adequate level of knowledge and skills on RDM within the institute in order to comply with the institute's RDM policy.	<ul style="list-style-type: none"> Monitors RDM skills of the researchers (including PhD students and students) and research support staff within the institute. Identifies knowledge and skill gaps and ensures appropriate training. Ensures a sufficient level of awareness among researchers and research support staff of the institute. 	<p>①</p> <p>②</p> <p>③</p> <p>④</p> <p>⑤</p> <p>⑥</p>	<p>F</p> <p>A</p> <p>I</p> <p>R</p>	<p>③</p> <p>④</p> <p>⑧</p> <p>⑫</p> <p>*</p>	<p>①</p> <p>⑪</p> <p>*</p>

7) Network	Responsible for obtaining and maintaining a network of aligned expertise areas and relevant departments and organisations inside and outside the institute with regard to RDM.	<ul style="list-style-type: none"> Refers researchers (including PhD students and students) to other RDM related facilities and services (legal, financial and/or operational), inside and outside the institute. Liaise with experts inside and outside the institute. Maintains a network with colleagues and other relevant departments and organisations. Explores needs, opportunities and new trends in RDM. 	<p>① ② ③ ④ ⑤ ⑥</p>	<p>F A I R</p>	<p>③ ④ ⑧ ⑫ *</p>	<p>① ⑪ *</p>
8) Data archiving	Responsible for FAIR and long-term archiving of data of the institute, stored internally as well as externally, and for sustainable and legitimate access to data sources of the institute, for the required period.	<ul style="list-style-type: none"> Develops, implements and monitors the institute's internal and/or external data archiving and access policy. Monitors the internal and/or external archiving of data by researchers (including PhD students and students) of the institute. Assesses whether internal and/or external data storage and archiving facilities meet the applicable requirements. 	<p>④ ⑤ ⑥</p>	<p>F A I R</p>	<p>① ③ ⑦ ⑨ ⑫ *</p>	<p>① ④ ⑪ *</p>
9) Supervision	Responsible for the supervision of staff commissioned with the execution of and/or support for the institute's RDM policy (data managers, data steward level B, etc.), if applicable.	<ul style="list-style-type: none"> Supervises RDM staff and coordinates activities of the RDM team, if applicable. 	<p>① ② ③ ④ ⑤ ⑥</p>	<p>F A I R</p>	<p>③ ④ ⑧ ⑫ *</p>	<p>① *</p>

Data steward Level B – Project focused, operational, supporting						
	Responsibilities	Activities / tasks	Mappings (see below for reference)			
			Research data life cycle	FAIR data	Purdue competence	DAMA knowledge
1) Policy / strategy	Responsible for the development and implementation of a data management plan (DMP) for departments, projects or data collections within the institute, in line with the institute's RDM policy and within financial and legal constraints, that supports FAIR data and Open Science.	<ul style="list-style-type: none"> Develops DMP templates tailored for the departments, projects or data collections within the institute. Writes and/or supports researchers (including PhD students and students) in writing a DMP for departments, projects and data collections, in line with the institute's RDM policy. Implements RDM as a regular aspect of doing research. 	① ② ③ ④ ⑤ ⑥	(F) (A) (I) (R)	③ ④ ⑧ *	① ⑪ *
2) Compliance	Responsible for monitoring compliance of the project or data collection with the DMP, and the Netherlands Code of Conduct for Academic Practice, the General Data Protection Regulation (GDPR), as well as continuous alignment with legal and ethical standards.	<ul style="list-style-type: none"> Monitors and supervises the execution of a project or data collection in line with the DMP. Identifies gaps and takes action if needed. 	① ② ③ ④ ⑤ ⑥	(F) (A) (I) (R)	③ ④ ⑧ ⑫ *	① ⑤ ⑪ *
3) Alignment with FAIR data principles	Responsible for facilitating and supporting FAIR data.	<ul style="list-style-type: none"> Advises, supports and provides guidelines to researchers (including PhD students and students) on the findability (F) of data, including adequate infrastructures, persistent identifiers and rich (institute-specific) metadata standards. Advises, supports and provides guidelines to researchers (including PhD students and students) on the accessibility (A) of (meta)data to potential reusers. Advises, supports and provides guidelines to researchers (including PhD students and students) on the interoperability (I) of data, including broadly applicable languages, vocabularies and other 	① ② ③ ④ ⑤ ⑥	(F) (A) (I) (R)	③ ④ ⑧ *	① ⑪ *

		standards. <ul style="list-style-type: none"> Advises, supports and provides guidelines to researchers (including PhD students and students) on the reusability (R) of data, including documentation and licenses with the conditions for reuse and IP rights. 				
4) Services	Responsible for delivering sufficient support and services on RDM for the involved researchers (including PhD students and students) on RDM.	<ul style="list-style-type: none"> Advises and supports researchers (including PhD students and students) on RDM. Advises researchers (including PhD students and students) on all aspects of the data life cycle in the department, project or data collection. Develops and implements (standard) solutions for recurring issues and questions. Provides guidance and instruction on discovery, acquisition and (re-)use of data (in the public domain). 	① ② ③ ④ ⑤ ⑥	F A I R	③ *	① *
5) Infrastructure	Responsible for monitoring the department's or project group's needs regarding e-infrastructure and tools for RDM, including supporting adequate access, in accordance with the DMP.	<ul style="list-style-type: none"> Monitors the needs regarding tools and e-infrastructure for RDM within the department, project or data collection. Supports the access to infrastructure and tools for RDM. Assesses adequate infrastructure and tools for RDM. Explores new trends in infrastructure and tools for RDM. Is aware of relevant (inter)national IT-, data- and e-infrastructures. 	① ② ③ ④ ⑤ ⑥	F A I R	① ③ *	① ② *
6) Knowledge	Responsible for an adequate level of knowledge and skills on RDM within the department or project group.	<ul style="list-style-type: none"> Monitors RDM skills in the department or project. Identifies knowledge and skill gaps, including understanding of the DMP, and takes action if needed. Initiates and provides training on RDM, tailored to needs of the researchers (including PhD students and students) involved in department or project. Creates awareness on RDM among researchers (including PhD students and students) and explains the added value of RDM. Introduces to new employees the institute's RDM policy as well as the department's or project's DMP. 	① ② ③ ④ ⑤ ⑥	F A I R	③ ④ ⑧ ⑫ *	① ⑪ *
7) Network	Responsible for liaison and alignment of RDM within the department or project group and with relevant stakeholders outside the department or project group.	<ul style="list-style-type: none"> Refers researchers (including PhD students and students) to other RDM related facilities and services (legal, financial or operational), inside and outside the 	① ②	F A	③ ④	① ⑪

		<p>department or project.</p> <ul style="list-style-type: none"> • Liaise with experts inside and outside the department or project. • Maintains a network with colleagues and other relevant departments and projects. • Explores needs, opportunities and new trends in RDM. 	<p>③ ④ ⑤ ⑥</p>	<p>① R</p>	<p>⑧ ⑫ *</p>	<p>*</p>
8) Data archiving	Responsible for FAIR and long term internal and/or external archiving of data of the department or project group by researchers (including PhD students and students), including selection of data, and sustainable and legitimate access to data sources of the department or project group, for the required period.	<ul style="list-style-type: none"> • Monitors the department's, project's or data collection's internal and/or external data archiving and access policy. • Monitors the internal and/or external archiving of data by researchers (including PhD students and students) of the department or project. • Assesses whether internal and/or external data storage and archiving facilities meet the applicable requirements. 	<p>④ ⑤ ⑥</p>	<p>F A I R</p>	<p>① ③ ⑦ ⑨ ⑫ *</p>	<p>① ④ ⑪ *</p>
9) Supervision	NA	NA	-	-	-	-

Mappings to existing data stewardship principles and approaches:

(*: indicates that this topic involves potentially all areas, with emphasis on the indicated areas)

Research data lifecycle: (<https://www.ukdataservice.ac.uk/manage-data/lifecycle>)

① creating data; ② processing data; ③ analysing data; ④ preserving data; ⑤ giving access to data; ⑥ reusing data.

FAIR principles: (<http://www.nature.com/articles/sdata201618>)

Ⓕ findable; Ⓐ accessible; Ⓛ interoperable; Ⓡ reusable data

Purdue competence areas (https://docs.lib.purdue.edu/lib_fsdocs/136)

① databases and data formats; ② discovery and acquisition of data; ③ data management and organization; ④ quality assurance; ⑤ data conversion and interoperability; ⑥ metadata; ⑦ curation and re-use; ⑧ cultures of practice; ⑨ data preservation; ⑩ data analysis; ⑪ data visualization; and ⑫ ethics, including citation of data.

DAMA knowledge areas (<https://dama.org/content/body-knowledge>)

① data governance; ② data architecture; ③ data modelling and design; ④ data storage and operations; ⑤ data security; ⑥ data integration and interoperability; ⑦ documents and content; ⑧ reference and master data; ⑨ data warehousing and business intelligence; ⑩ metadata; and ⑪ data quality.