

# RDA Professionalising Data Stewardship – Data Stewardship Landscape Initial Report



RDA Supporting Output

**DOI:** 10.15497/RDA00076

**Authors:** Lorna Wildgaard, Jukka Rantasaari

**Editors:** Claudia Engelhardt, Jaana Pinnick

**Published:** 4 October 2022

**Version:** 1.0

**Abstract:** This report is based on a literature review published by the Research Data Alliance Professionalising Data Stewardship Interest Group (RDA PDS-IG) Training sub-group in October 2022. A framework for the literature collection was devised before the literature search began, based on common concepts of the division of the research disciplines. The literature search was conducted between March 2021 and July 2021. The selection was verified in a cross-walk exercise, where common terminologies were tracked across different ontologies to learn more about concepts and disciplinary differences and commonalities in research data management (RDM). In the reference management software Zotero, folders were created to structure the literature into the following categories: data stewardship; education and training of data stewards; RDM education; training materials; data management plan templates; documents about the skills needed to support EOSC and FAIR principles; data management in the humanities, industry, life sciences, natural sciences, public institutions and the social sciences. Further, the aim was to identify papers concerning the RDM needs of researchers and data stewards and their role in the RDM life cycle. The screening resulted in 125 papers that are used as the knowledge base for this report.

**Impact:** The main take-home messages from this review are: 1) Collaborate internationally on certified training; 2) Enable researchers to share their data from the planning stage of their project and 3) Address the increased need for technical skills in data stewardship. The report offers 197 solutions compiled from literature to bridge gaps in data stewardship. The output can be used to identify and recommend learning pathways for data stewards, and to identify use cases to concretise the solutions in order to provide a valuable tool for data stewards. The results of the report could be used to contribute to the Competency Hub pages on Data Stewardship, the continuing work of the EOSC Taskforce or the creation of new educational materials for RDA task group developing curricula. The report will also give valuable information to all who are interested in developing their data stewardship skills needed to better serve the needs of researchers.

**Language:** English

**License:** [Attribution 4.0 International \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

**RDA webpage:** <https://www.rd-alliance.org/groups/professionalising-data-stewardship-ig>

**Citation and Download:**



# Gaps in data stewardship: What kind of needs for training do data stewards face in supporting research?



Picture credit: Patrick Hochstenbach in <https://zenodo.org/record/1212496>

## Authors

Lorna Wildgaard, 0000-0002-3900-5058  
Jukka Rantasaari, 0000-0001-5927-3781

## Editors

Claudia Engelhardt, 0000-0002-3391-7638  
Jaana Pinnick, [0000-0002-7718-5133](https://orcid.org/0000-0002-7718-5133)

Date	Version	Changes	Responsible
30/05/2022	001	Corrections i preparation for RDA Plenary	LOWI
04/10/2022	002	Peer review completed	LOWI



## Contents

<b>Contents</b>	3
Executive summary	4
1. Introduction	6
1.1 Research questions	8
1.2 Definitions	9
2. Methods	14
<b>3. Results and Discussion</b>	14
3.1 Gaps	14
3.1.1 General level gaps in data sharing culture, processes and infrastructure	14
3.1.2 Towards the FAIRness of data	14
3.1.3 Gaps in training materials (discipline, coverage, capabilities)	15
3.1.4 Gaps in DSs' skills	17
3.1.5 Gaps in researchers' skills	23
3.1.6 What do we not yet know about researchers' RDM needs and practices	26
3.2 How to meet these gaps (solutions)	27
3.2.1 Solutions in Governance of Data Stewardship	27
<b>4. Summing up: The Gaps, Solutions, and Sources</b>	32
5. Limitations	59
6. Conclusions and Next Steps	61
8. References	61
<b>Appendices</b>	69
Appendix 1: Books, Guides, Interest groups, Programs, Tools, Trainings, and Webinars on Data Stewardship	69
Books and reports	69
Guides	71
Interest groups	72
Programs	72
Tools	73
Trainings	74
Webinars	76

## Executive summary

This paper is a literature review which will inform the continuing work of the RDA Professionalising Data Stewardship Interest Group PDS IG<sup>1</sup>. The IG group aims, amongst other tasks, to identify and recommend learning pathways for data stewards. The results of the report could be used to contribute to the Competency Hub pages on Data Stewardship<sup>2</sup>, the continuing work of the EOSC Taskforce<sup>3</sup> or the creation of new educational materials for RDA task group developing curricula<sup>4</sup>. The report will also give valuable information to all who are interested in developing their data stewardship skills needed to better serve the needs of researchers.

The review is made up of 6 sections. Section 1 frames the challenges in professionalising data stewardship and poses the research questions. Section 2 describes the methodology we used to collect and analyse the text corpus, followed by section 3 which presents the gaps identified in current training materials, gaps in the skills sets of data stewards, and gaps in researchers' skills in research data management (RDM) as identified in the literature. Section 4 sums up the gaps, solutions and resources identified in section 3. In Section 5 we briefly state limitations to the scope of the report and finally section 6 provides the conclusions and the next steps.

The main take-home messages from this review are:

1. *Collaborate internationally on certified training.*  
Use international fora such as EOSC Skills & Training WG Outputs<sup>5</sup> or Train4E<sup>6</sup> as vehicles to provide certified training. Coordinated and accredited training will ensure implementation and the adoption of the best practices and standards. Coordinated training will also ensure that organisations, whatever their size and resources, can have access to shared resources. Furthermore, disciplinary differences in data stewardship can be made visible and hence provided support.
2. *Enable researchers to share their data from the planning stage of their project.*  
The key to get researchers to share their data and actively embrace FAIR data stewardship is motivating and teaching them how to share data appropriately, safely, and securely within an organisation and in projects as well as externally. Training materials should address the ethics, the societal and practical benefits of data sharing as well as the technical aspects. Such materials could be based on a gap analysis of researchers' practices, which firstly investigates barriers to sharing, secondly reveals how researchers currently share and would like to share, and thirdly investigates the support they need from funders and publishers regarding sharing,

---

<sup>1</sup> <https://www.rd-alliance.org/groups/professionalising-data-stewardship-ig>

<sup>2</sup> <https://competency.ebi.ac.uk/framework/datasteward/1.0>

<sup>3</sup> <https://www.eosc.eu/task-force-faq>

<sup>4</sup> <https://www.rd-alliance.org/groups/professionalising-data-stewardship-ig>

<sup>5</sup> <https://www.eoscsecretariat.eu/working-groups/skills-training-working-group/eosc-skills-training-outputs>

<sup>6</sup> <https://4euplus.eu/4EU-31.html>

the infrastructure they need to share and protect the confidentiality of data, and how to safeguard intellectual property from being stolen or data being misinterpreted or misused . Thereafter, develop materials about how to build models for sharing into day-to-day workflow, including the creation of metadata and the uploading of data to repositories, and “automating” the process.

3. *Address the increased need for technical skills in data stewardship.*

Future data steward competencies border on and partly overlap with computer science technical competencies. The future of data stewardship is becoming more automated, hence the data steward community needs increased collaboration with computer scientists to define and build solutions. Though a data steward needs subject knowledge on an expert level to be able to advise on ontologies, the technical skills are vital to be able to exploit linked data, metadata and machine-readable syntax. Such a technical shift in stewardship to be able to support open and FAIR science also means a shift in the placement of data steward service in the organisation, where the service encompasses teaching, technical and research skills.



## 1. Introduction

Even though the role of the data steward (DS) varies hugely both in job title and job content, data stewardship as an activity in itself is regarded as essential for good data governance. This recognition provides a solid foundation and motivation for defining data stewardship and ultimately the data steward profession. Work is underway to provide internationally agreed definitions of a data steward and professionalise data stewardship education (Jetten et al., 2021; Barker et al., 2021; Wildgaard et al., 2020; Whyte et al., 2018 ). The Research Data Alliance's (RDA) interest group "Professionalising Data Stewardship" recognises that the lack of consensus on the responsibilities, knowledge and skills of data stewardship results in confusion about the role of a data steward and hampers adequate data steward capacity (Jetten, 2019). As an RDA group, we are therefore working together to find viable solutions to data steward curricula and career paths. However, to professionalise data stewardship, there first needs to be an understanding of what a data steward is, the roles data stewards have in an organisation, and accordingly their responsibilities and expectations with regard to skills and service provision. Mapping definitions and roles of data stewardship is a huge undertaking that falls outside the scope of this report. Therefore, this current report builds heavily on the foundational work accomplished by the NPOS F project<sup>7</sup>, which in turn built on the ZonMw data stewardship project<sup>8</sup> and the LCRDM data stewardship project<sup>9</sup> to identify profiles and training in data stewardship (primarily in the life sciences) and consequently the capabilities and learning objectives for data stewards.

The description of a data steward used in the aforementioned work is as follows (Jetten et al., 2021):

*"A person responsible for keeping the quality, integrity, and access arrangements of data and metadata in a manner that is consistent with applicable law, institutional policy, and individual permissions. Data stewardship implies professional and careful treatment of data throughout all stages of a research process. A data steward aims at guaranteeing that data is appropriately treated at all stages of the research cycle (i.e., design, collection, processing, analysis, preservation, data sharing and reuse [and reproducibility])."*

Hence, a data steward implements policies and standards for data management according to data management best practices. The data steward ensures that the standards are followed at unit and operational levels and acts as "humanware", i.e. a liaison between the IT department, the data producer and their data, and the organisation. We talk about "humanware" as the transformation to a data-driven culture requires far more than

---

<sup>7</sup> NPOS F project:

<https://www.openscience.nl/en/projects/project-f-professionalising-data-stewardship-competences-training-and-education>

<sup>8</sup> ZonMw data stewardship project:

<https://www.openscience.nl/en/projects/project-f-professionalising-data-stewardship-competences-training-and-education>

<sup>9</sup> LCRDM data stewardship project: <https://zenodo.org/record/2669150>

technology (Aiken, 2016). Processes and people are needed too, and Aiken further points to data stewards as such persons who can develop, provide, and connect IT infrastructure, policy, best practices, analytical processes, and data. A data steward is described as having different profiles with regard to their area of stewardship expertise, be that extensive knowledge of policy, infrastructure, data analysis, or disciplinary data (Jetten et al., 2021; Wildgaard et al., 2020). The symbiose of the data stewards' expertise with the needs of the data producer means that they can take on roles that are embedded in an organisation or individual project, or they can provide generic services across an organisation. The key, according to Mons (2020), is to build capacity, enable groups within and across organisations (and countries) to work together and share good practices – so that good data stewardship becomes the rule, not the exception. Simply put, a data steward is a person who is responsible for managing data on behalf of the person with ultimate responsibility for the data, often referred to as the data owner. However, defining the responsibilities, skills, and needs of a data steward in reality is anything but simple.

When we use the term “organisation” in this report, we are talking about any type of organisation that produces data – public or private, university or industrial sector. When we refer to data, we refer to any digital research data. Consequently, the stewardship of physical research materials is not addressed in this report. However, there are different approaches to data stewardship in the university and industrial sector in practice that we acknowledge in this report. Data stewardship in an industrial context has for many years been seen as a strategic organisational asset. In-house training programmes and apprenticeships in, amongst others, master data stewardship, business architecture or governance, and customer data stewards specialising in purchasing experience and product data are established profiles (Wildgaard et al., 2020). Data stewardship in the industrial sector leverages the value of an organisation's collected data's unique characteristics to qualify product development, foster innovation and competition, and inform investment decisions (Smith, 2021). On the other hand, historically, at universities, data has not necessarily been treated as an strategic “business asset” but rather as a research by-product, whereas the publications, findings, and application of knowledge were regarded the principal “products” (Rantasaari, 2021). There are few upskilling opportunities or formal education programmes in data stewardship in academia. The acute need for data stewards at universities has been explored in Wildgaard et al. (2020) and Pergl et al. (2019), concluding that data stewardship strategies and programmes need to educate data stewards with expertise in business to bring organisational data management practices in research institutions up to par with advances in policies and requirements to data management and IT.

To our knowledge, a coherent approach and dedicated training for data stewardship on the (inter)national level is lacking. In the “Turning FAIR into Reality” report by the European Commission FAIR Data Expert Group (2018), there is a call for “new job profiles [...] to be defined and education programs put in place to train the large cohort of data scientists and data stewards required to support the transition to FAIR [data management]” (Cruz et al., 2019). Demchenko et al. (2021) propose a competence framework for FAIR data stewards that can be mapped to learning outcomes for defining academic curricula. The framework provides a considerable step in the right direction towards defining data stewardship training needs, as it combines competencies, skills and knowledge topics from previous frameworks with a market analysis of data steward job profiles and key competencies. Combined with the

lack of sufficient funding, the lack of educational opportunities also limits vital developments such as recruiting procedures and the definition of career paths for current and future data stewards. In the meantime, data stewards exist and requests for their help continue to increase. In 2020, Barend Mons estimated that Europe will have at least 10 million serious data producers among its 70 million science and technology professionals and 1.7 million researchers (Mons, 2020). He argued that, as a consequence, there is a need to "...educate about 500,000 data stewards of various kinds to support researchers through experimental design and data capture, curation, storage, analytics, publication and reuse." Hence, data management and facilitation is also an expense that needs to be budgeted for in research projects. His recommendation is that 5% of overall research costs should go towards data stewardship, which means that "with €300 billion (US\$325 billion) of public money spent on research in the European Union, we should expect to spend €15 billion on data stewardship" (Mons, 2020).

As the need for data stewards and their professionalisation increases, the acute need to professionalise and to some extent standardise data steward education has arisen. The community around data stewardship (i.a., Jetten et al., 2021; Jetten, 2019 and NPOS/ELIXIR, 2019) suggest that such education should be standardised across countries and practices and aligned to support training infrastructures and sustainable career paths for data stewards. As Dyche and Polsky (2016) state: "the promise of data stewardship is the inherent problem with data stewardship: it's not specific enough".

At present, it is not only unclear what a data steward is but also how to become an accredited data steward. Across the literature, various capabilities are flagged. This paper will not cover these capabilities in depth, rather it will provide an outline of capability "gaps" by using the definitions of data stewardship as provided in published literature and comparing these definitions with the needs of researchers, data stewards, and current training materials available in the broad arena of research data management. The purpose is to highlight competency gaps amongst professionals. As such we aim to provide a footing from which progress can be made towards coherent training and education and possibly national if not disciplinary alignment.

## 1.1 Research questions

What kind of needs for training do data stewards face in supporting research?"

*In answering this overarching research question, we attempt to describe discipline agnostic and discipline-specific gaps<sup>10</sup> and the capabilities needed to fill these gaps, specifically:*

- What kind of skills gaps do data stewards have?
- What kind of skills gaps do researchers have?
- What kind of gaps are in existing training materials?
- What kind of gaps are in our knowledge of the data stewards' and researchers' needs?

---

<sup>10</sup> In terms of discipline-specific gaps, please note that - due to the fact that relevant literature is only available for some disciplines - we cannot give a comprehensive account for all.

Consequently we aim to identify solutions already published by disciplinary experts, and recommend building on their original research. Hence, we address the aforementioned gaps and map these gaps to solutions, we pose the following question:

What kind of **solutions** to the aforementioned gaps are identified in the literature?

## 1.2 Definitions

This report uses terminology and concepts from research data management and data science to describe gaps in the coverage of data stewardship capabilities in data management curricula for data stewards, researchers, and other professionals. The scaffolding of data curricula and learning outcomes depends on a common terminology and understanding of concepts and activities in data management. Goben and Griffin (2019) and Sapp Nelson (2017) provide comprehensive glossaries, which we strongly recommend to be used in the development of the curricula needed to fill the gaps identified in this report. In this section, we list the common terms used throughout the present report and clarify how we understand them with the aim to improve the accessibility of the report for readers not well-versed in data management.

Term	Definition
Capability	The extent of the data stewards competencies (in a certain area).
Competence	“Knowledge, skills and abilities relating to a topic e.g., ‘workflow setup and management’” (Jetten et al., 2021, p. 62).
Data steward, general definition	“A person responsible for keeping the quality, integrity, and access arrangements of data and metadata in a manner that is consistent with applicable law, institutional policy, and individual permissions. Data stewardship implies professional and careful treatment of data throughout all stages of a research process. A data steward aims at guaranteeing that data is appropriately treated at all stages of the research cycle (i.e., design, collection, processing, analysis, preservation, data sharing and reuse” [and reproducibility] (Jetten et al., 2021, p. 62).
Data steward, embedded	“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. The embedded data steward for instance helps with software code, scripts and algorithms to analyse data. <sup>12</sup> Positioning: often within a research unit” (Verheul et al., 2019, p. 7).
Data steward, generic	“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 ‘centralised knowledge and communication hub for researchers.’ <sup>13</sup> The generic data steward sometimes has specific knowledge of a certain domain but generally does not have adequate time to

	<p>give advice on particular situations.<sup>14</sup>  Positioning: often with a support service or at faculty level” (Verheul et al, 2019, p. 7).</p>
Data Protection Officer (DPO)	<p>The data protection officer (DPO) ensures that the organisation</p> <ul style="list-style-type: none"> <li>● processes the personal data of its staff, customers, providers or any other individuals (also referred to as data subjects) in compliance with the applicable data protection rules.</li> <li>● advises the institution about the interpretation or application of the data protection rules</li> <li>● handles queries and complaints</li> <li>● draws the institution's attention to any failure to comply with the applicable data protection rules.</li> </ul> <p>(European Data Protection Supervisor, n.d.)</p>
FAIR	<p>A set of guiding principles to make research data findable, accessible, interoperable and reusable (Wilkinson et al., 2016).</p> <p>We do not differentiate between data stewardship and FAIR data stewardship. FAIR skills are but one of a set of skills data stewards should have. We do not single FAIR out or focus just on FAIR.</p>
Infrastructure	<p>The authors’ definition of “infrastructure” as referred to in this report can be: human, organisational or technical.</p> <ul style="list-style-type: none"> <li>● Technical infrastructure: IT networks, systems, hardware, software, development, and maintenance.</li> <li>● Organisational: the interconnection among organisation elements, divisions, departments, research laboratories, universities, and industries through fast and reliable communication networks.</li> <li>● Human: the network of people providing data steward support, activities and facilities.</li> </ul>
Researcher	<p>A scientist or researcher in any field of academic study or science, employed at a university, or in the public or private sector.</p>

## 2. Methods

This paper is based on a literature review. A framework for the literature collection was devised before the literature search began, based on common concepts of the division of the research disciplines. This division was verified in a [cross-walk exercise](#), where common terminologies were tracked across different ontologies to learn more about concepts and disciplinary differences and commonalities in research data management (RDM)<sup>11</sup>. The final framework guiding the search and dividing responsibilities between the authors of this report is presented in Figure 1. In the reference management software Zotero<sup>12</sup>, folders were created to structure the literature into the following categories: data stewardship; education and training of data stewards; RDM education; training materials; data management plan templates; documents about the skills needed to support EOSC and FAIR principles; data management in the humanities, industry, life sciences, natural sciences, public institutions and the social sciences. Further, the aim was to identify papers concerning the RDM needs of researchers and data stewards and their role in the RDM life cycle.

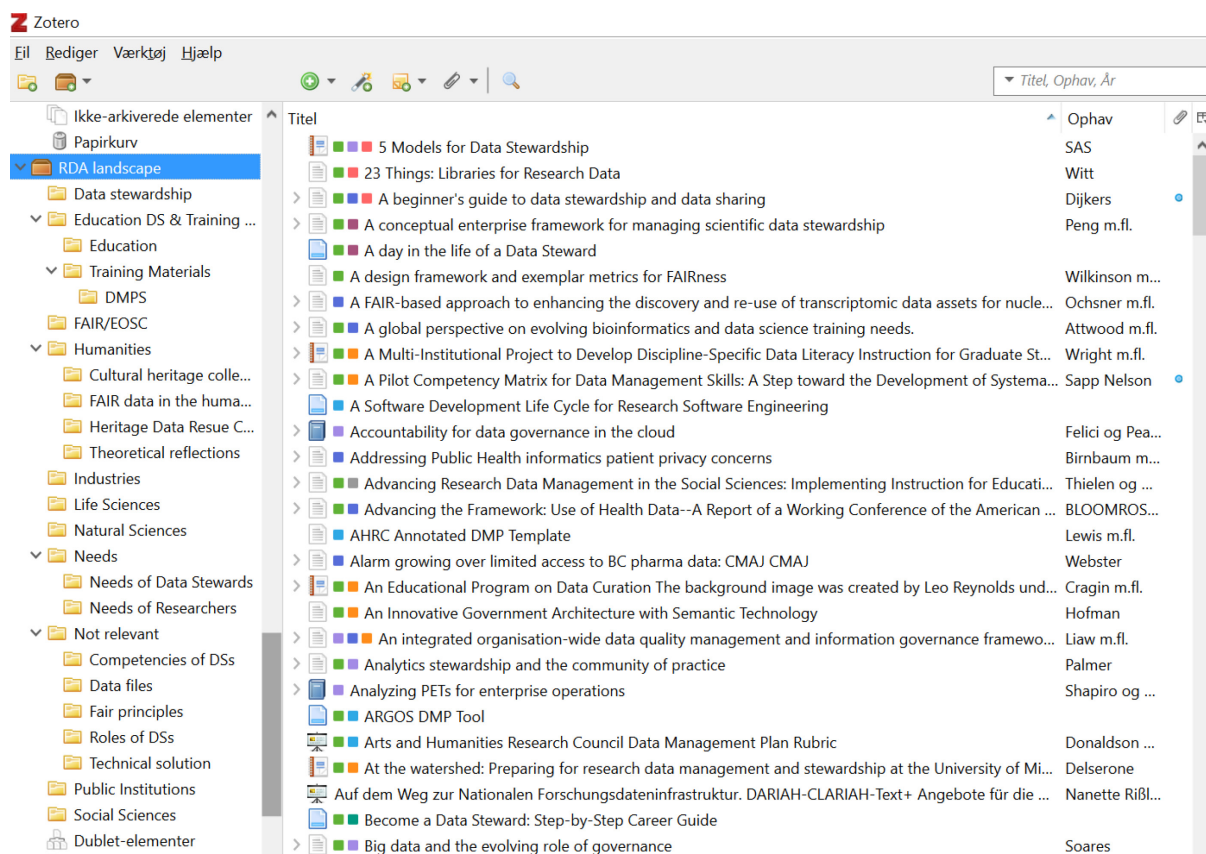


Figure 1: Categorisation of the literature

<sup>11</sup> Link to Cross-walk exercise: [https://docs.google.com/spreadsheets/d/1xL9yIPi-U90q3GY6mRIGDn1\\_UuJR9bt7gqJ3KI\\_wyw/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1xL9yIPi-U90q3GY6mRIGDn1_UuJR9bt7gqJ3KI_wyw/edit?usp=sharing)

<sup>12</sup> <https://www.zotero.org/>

Responsibilities for the search were divided between the authors, each searching for literature within their domain expertise and interest. The search was conducted between March 2021 and July 2021.

The search concerning data stewards was conducted in Google Scholar and aimed to identify original scientific works, white papers, presentations, initiative papers and reports. Multiple search strategies were used such as inTitle:"data steward" or inTitle: "data stewardship" Humanities. Citation chasing was also employed to identify related papers. As such, the search also identified published bibliographies that were incorporated in our Zotero library as well, including [LIBER skills for open science](#) and [Data Management Practices in the Humanities](#). We kept the search limited to results from the mid 2000's to keep the term "data stewardship" as close to our definition (see section 1.3) as possible. Including results from before 2006 resulted in a lot of noise in the search. Data stewardship responsibilities and profiles have evolved over time from the 1980s and 1990's, where the data steward is described as a person who has data warehousing skills, to the present day, where their profile includes skills in governance and policy, data analysis and processing, teaching, technical support, system development and more.

The search concerning researchers' RDM needs and competence gaps was conducted in Web of Science, Scopus, and relevant EBSCOhost databases (e.g. Academic Search Premier, Business Source Complete, and Information Science & Technology) using a combination of search words and phrases such as "research data management", "RDM", "education", "training", "teaching", "needs", "competence", and "skills". Truncation, wild cards and delimiters such as publication years (2012-2021) and language (English) were used.

Identified references were organised into categories and sub-categories (see Figure 1). Each reference was read and colour-coded to indicate which discipline, needs and topics were discussed in the paper. The majority of the references were placed in more than one category, as indicated by the colour-coding, as the paper addressed more than one topic relevant for our study.

In total, 276 resources were collected, grouped into categories. With irrelevant ones removed or omitted, this resulted in 257 potentially relevant abstracts, of which 151 went further for full text screening for inclusion in the report. Abstracts were excluded if the full text could not be sourced, did not address the stewardship of data, or focused on publishing of data underlying scientific works. Further, abstracts concerning an evaluation of research support services in the broader context were considered too generic to be useful, as were reports or slides of events intended to lead to book sprints or further project work. In such cases, we attempted to find related work published at a later date. The 151 full text papers were screened for inclusion based on the following criteria: the user groups' needs discussed in the paper, gaps identified, recommendations to fill these gaps, and the identified challenges in developing the required competences or training materials<sup>13</sup>. The screening resulted in 125 papers that are used as the knowledge base for this report.

---

<sup>13</sup> A matrix of the screening process is available here: <https://docs.google.com/spreadsheets/d/1xL9yIPi-U90q3GYp6mRIGDn1UuJR9bt7gqJ3KIwyw/edit?usp=sharing>



## 3. Results and Discussion

### 3.1 Gaps

#### 3.1.1 General level gaps in data sharing culture, processes and infrastructure

It has been estimated that because of missing agreement on sound and coherent RDM practices, low-quality data has become very expensive in terms of research renewing, strategic opportunities, stock prices, profits, and so on in academic organisations and corporations around the world (Aiken, Allen, Parker, & Mattia, 2007; European Commission, 2018a; Lucas 2010). Although valuable methods and tools for quality control and assessment of data have been developed, they have not sufficiently been taken up for example by academic organisations (Lucas, 2010). Moreover, because of the absence of or unclearly defined data stewardship roles, the problems with data have typically been treated as “IT problems” (Lucas, 2010).

Besides the low quality of data, the problem is that the data on which research results are based are usually not linked to the publications, including the most frequently cited articles (Hardwicke & Ioannides, 2018). This precludes the re-analysis and reuse of the data and leaves the reader no choice than to trust the arguments of the authors without the possibility to verify them (Hardwicke & Ioannides, 2018).

Meanwhile, a plethora of software applications are saving huge amounts of personal and even sensitive data to cloud services without any guarantee of safe and secure storage, backup, and long-term preservation of the data (Estrada-Galinanes & Wac, 2018).

#### 3.1.2 Towards the FAIRness of data

Research outputs can contain data, software, code, and other digital objects equipped with persistent identifiers (PID), metadata, and contexting documentation that help to find, comprehend, cite, and reuse the objects (Turning FAIR into reality, 2018). The goal in FAIR science is that all research outputs are maximally findable, accessible, interoperable, and reusable within the scope of the motto: as open as possible, as closed as necessary (e.g. Mons, 2018).

To reach this goal, we need collaboration between researchers, infrastructure developers, and research supporting professionals to develop a common view of research practices and standards on how to document the data life cycle and produce at least FAIR metadata defining the reusability of the data (Demchenko et al, 2021; European Commission Expert Group on FAIR Data, 2018; HLEG on EOSC, 2018; Mons, 2018; Schuster, 2021). Moreover, to get to a point where data sharing and reuse practices are well managed and in good control, we need a mutual, cross-disciplinary understanding of what data and their elements are, how data should be described and cited, what data stewardship is and which functions it comprises, and what kind of infrastructure and workflows need to be in place to ensure data privacy, security, quality, and user-friendliness (Callahan et al., 2017; Dijkers, 2019).

Though not all data can be shared and reused, it should always be findable by machines and – when possible – reusable by humans (Mons, 2018). To ensure the FAIRness of data, we need an easy-to-use data storage and sharing infrastructure which is seamlessly embedded in research processes, research data services (RDS) and research data

management (RDM) training for researchers, as well as commonly defined protocols, standards, and formats (Dijkers, 2019; DSCC, 2022; European Commission, 2018; Latif, Limani, & Tochtermann, 2021; Sansone et al., 2019; Weller & Monroe-Gulick, 2014). The location of the infrastructure can be distributed but they will be coordinated comprehensively (Latif, Limani, & Tochtermann, 2021). Although many funders and initiatives such as European Open Science Cloud (EOSC) and GO FAIR already develop common standards (e.g., FAIR) and criteria for describing, sharing and reusing data, there are still challenges in harvesting and combining multifaceted and multi-format metadata from dispersed systems (Latif, Limani, & Tochtermann, 2021). Hence, the ability to use EOSC-Core and EOSC-Exchange services for data publication and preservation as well as to continue the development of an infrastructure to support data discovery, curation, preservation and sharing is a core skills gap for data stewards, researchers, infrastructure support professionals and organisations as a whole that needs to be addressed (Barker et al, 2021).

### 3.1.3 Gaps in training materials (discipline, coverage, capabilities)

#### *Lack of a universal data steward body of knowledge reference-framework*

Training materials will never be complete and the gaps identified in the following provide a snapshot of current concerns expressed in the literature. The key takeaway is that each research project is so specialised with unique requirements to data stewardship that it is impossible to prepare for all eventualities in training materials. Likewise, requirements from funders, organisations and other stakeholders change over time. Of course, general and generic training can be interpreted and tailored to specific data management situations, but expert knowledge from managing specific types of data in projects and expectations to data management skills for specific user groups is not routinely stored and shared among the broader data steward community. Therefore, the greatest gap in training materials results from a lack of infrastructure that enables knowledge sharing and best practices that ultimately are used as a reference-framework, i.e. informal and internal training materials that are not communicated to the wider community.

#### *Synchronisation*

Ideally, data management plans should be synchronised with the project, the funder, and other stakeholders, such as institutions and project partners, so that the DMP translates the funders' requirements into a DMP template and any updates are communicated to the funder and vice versa (Williams et al., 2017). Williams et al. argue that the technological infrastructure to support this communication and feedback between funder and project is not yet functional. Until such an infrastructure is in place, training materials on how to best manage synchronisation and communication are needed.

#### *Awareness of other projects*

Research data management and data stewardship are complex, and identifying the usefulness of data adds another layer of complexity. Williams et al. (2017) and York et al. (2018) investigate the production of non-reusable data and what data practitioners, researchers, and data stewards need to be aware of to ensure the reusability of data. Both papers point to a "reuse gap". The reuse gap is the "gap" between the total amount of data that are shared and made available for reuse and the proportion of these that are actually reusable. Reuse is effected of course by restrictions that ensure the integrity of the data and

data subjects, technical features of the data set, as well as ways to facilitate data sharing, where to publish data and make them findable, accessible and interoperable. However, another skill data stewards need, is the ability to identify and discuss the extent to which data produced by others could be relevant and “useful” to a new research project. Training related to how to assess and ensure the usefulness of data in a new research project, where and how to identify datasets that can potentially answer research questions and the limitations, validity, and quality of these datasets is needed. Kühn & Streit (2017) point to a lack of advanced materials about how to evaluate data quality, synthesise data for all skill levels, and target professional groups. Thus the parameters of “redundancy” and “replication” need to be addressed in training materials, including how data stewards and researchers can best work together to ensure the scientific integrity of reused data.

#### *Sharing and promoting inclusivity across research disciplines*

Science Europe, amongst others, report a lack of inclusivity in the language used in data management policies and guidelines that could alienate researchers of non-STEM fields (Goben & Griffin, 2019). Likewise, Williams et al. (2017), based on an analysis of funders’ requirements for data management plans, identified a need for a common vocabulary around aspects of data management planning as a way to embed data management practices in the local culture. Creating terminology would train data stewards in the dissemination of data management practices in a local context. In their summary of research data management trends, needs, and opportunities, Goben and Griffin cite Partlo (2014) who explains that “humanities or non-STEM scholars do not claim to collect or work with data, despite potentially performing text and image analysis, gathering artefacts or recordings, or need to archive other products of research”. Drawing on observations from questionnaires and case studies from 43 public and private institutions, Goben and Griffin recommend using more inclusive language such as “products of research” or “research objects” rather than “data” to improve training materials and the uptake of RDM from “the bottom-up”.

#### *Applying metadata to improve the findability of data*

Institutional repositories, libraries, and data support staff within an organisation are mentioned as having the opportunity and responsibility to create an environment that facilitates researchers’ data-sharing behaviour. Sharing builds on making datasets findable. Kühn & Streit (2017) present a skills gap analysis of training materials that were mapped to the EOSC competence framework and their own experiences with the science Demonstrator group<sup>14</sup>. They identified a lack of knowledge about and training on how to apply the metadata required to ensure findability. On the other hand, they found ample training material about the application of metadata to ensure accessibility and reusability, including licensing information. Providing an easy-to-use, self-generating metadata option when designing data management systems, such as data repositories, can also facilitate the adoption of metadata standards (Joo & Peters, 2019), which in turn requires training in how to do so.

#### *Data management and the research lifecycle: data literacy*

There is a need for training materials for researchers that foster robust data management

---

<sup>14</sup> Science Demonstrator Group: <https://eoscpilot.eu/news/eosc-and-science-demonstrator-strengthening-their-relationship>

from the start of a research project, rather than materials that emphasise post-publication data management (Lefebvre et al., 2020; Williams et al., 2017). Lefebvre et al. echo the findings of Kühn & Streit, who identified the need for materials for researchers that cover all stages of the research lifecycle, particularly information on how to appraise and preserve research data, how to find data, as well as how to ensure the findability of one's own data. Furthermore, materials on a higher technical level are needed for domain specialists, data scientists, and engineers, specifically on how to govern data, give access to data, advise on, and enable data management (Demchenko et al, 2021).

For students and educators, Sapp Nelson (2017) summarises recommendations from Calzada Prado & Marzal (2013), Carlson et al. (2011), Piorun et al. (2012), Qin & D'Ignazio (2010b), and Schneider (2013) in a matrix of data management skills which shows that there are gaps in explicitly measurable, observable and transferable learning objectives in data curricula. She suggests a scaffolding of data curricula to enable learning paths throughout an academic career. The gaps identified relate to grounding communication between instructors regarding which learners have been taught what, and by whom; and assessing the level at which those learners are successful in meeting set educational goals for data management education.

However, at which point in their education the student needs to be not necessarily proficient in but aware of data management, is unclear and needs to be addressed in the provision of training materials, e.g. should data management be first introduced at the Bachelor or the PhD level? Should students only be expected to conduct data management when they are responsible for managing data as part of research projects? Accordingly, the ability to determine when, how, and to which extent to train students in data management is a special competence that data stewards and instructors need to acquire. Materials need to be developed that help them tailor information and scaffold data management and learning activities to different student groups within an educational strategy. Such a strategy is suggested to include training in transversal – also often called soft – skills such as how trainers can best communicate and assess learning outcomes and evaluate if the education is having an effect (Sapp Nelson, 2017).

#### 3.1.4 Gaps in DSSs' skills

The main challenge of data stewardship – and one that is repeatedly discussed in the literature – is the ability of the data steward to adapt to the specific needs of the researcher in the context of the project they are working on. Recommendations to prepare the data steward to provide a generic service are well documented and include descriptions of the data steward's capabilities that can be widely applied in broad research application contexts, competence with communicating and implementing policy, technical expertise with developing solutions, and also the importance of developing domain specific knowledge and soft skills (Jetten et al., 2021; Wildgaard et al., 2020). However, the skills needed to provide the above service are many and diverse, so diverse that as Jetten et al. and Wildgaard et al. suggest specific data steward career paths are the solution, where the data steward specialises in one or a combination of profiles and hence is able to apply data steward knowledge in both generic and embedded contexts: as a research steward, policy steward, agent of change, infrastructure profile, or in combination. The profiles are characterised by

their own expertise and terminology, each with responsibilities and tasks within competence areas that can be seen as different aspects of the function that the data steward has. A limitation of the report by Jetten et al. is the focus on data steward competencies in the life sciences. It is yet to be confirmed if the identified competencies are transferable and whether they can be implemented by a data steward in the context of the, e.g., humanities and social sciences or in other research contexts, commercial, public sector, and industry.

There is no consensus on the responsibilities and tasks of data stewards, or on formal profiles (including knowledge, skills and abilities (KSAs)). We present in the following common themes and expectations of capabilities identified in the surrounding literature and build further on the competence profiles already verified in Jetten et al. (2021) and Wildgaard et al. (2020).

### **Profile 1: Data steward research – project and research focused**

*Focus on students, researchers and data scientists who produce data and work with the data on a daily basis. The responsibilities and tasks of this data steward concern translating the scientists' and projects' needs with regard to data to infrastructural, service, and policy requirements. The DS can be embedded in the research team or provide generic advisory services. Jetten et al. (2021) Appendix 3 provides the detailed description of the function (responsibilities and tasks), competencies (knowledge, skills and abilities) and learning objectives for DS role structured by the eight competence areas, p.36-41.*

#### *Gaps in data steward research profile*

In this report, we identified that specific competencies needed for the support of citizen science projects (Goudeseune, 2020) were missing from the matrix in Jetten et al. (2021). These competencies include:

- supporting the link between the citizens and the project,
- nurturing the link and culture between the citizens and the data infrastructure,
- mediating between the project steering group or project group, and collaboration partners
- advising on relevant current policies and protocols that apply to citizen science.
- the development and use of platforms and tools designed to support citizen-driven data collection, including documentation of data collection methods
- practical application of models for validation, quality assessment and management of data to maximise usability
- integration of data, and skills in the preparation of appropriate data workflows, and
- the visibility of the project in data repositories so that it can be discovered and reused by others, including best practices in how data and results are made available to the public.

Citizen science is a data stewardship perspective that approaches data quality from an IT perspective and at the same time is sensitive to planning work with sensitive data and ensuring data quality in a data set collected by multiple contributors. In addition, Rickards and Ritsert (2012) point out that the data stewards must concern themselves with the

relationship between the IT tools and data quality with regard to the impact this relationship can have on reporting.

The data steward research will be expected to work with new types of research data such as data that can be collected from social media and other web platforms across academic disciplines. While the researchers will be developing new research methods and standards relating to the representativeness and reproducibility of such data, the data steward can contribute to fostering a culture of transparency in data analysis in both quantitative and qualitative approaches (Aguinis & Solarino, 2017) as well as focus on data protection concerns, research ethics, and sharing data (Weller & Watteler, 2021). Challenges for the data steward include the growing number of sources, awareness of these, integration and connectivity between them, and expectations to the value and accuracy of data in these sources as well as legality and quality issues (Stickel & Vandervalk, 2014). The competencies of the data steward are expected to be impacted in the future by the changing nature of data and the variety of sources and their entanglement in broader complexities of data governance.

Just as data types and sources are becoming more diffuse, so is the notion of “publication”. Articles, datasets, codes, models, working papers, and other supplementary materials are published to support transparency and replicability, but they are also used in evaluations in hiring, tenure, and grant proposals. Hence, the same or similar content, often with the same or similar titles and authors, may appear as preprints, postprints, working papers, slide decks, and as the formal “official” version of a publication. The publication versioning problem intersects with the data stewardship problem in at least two ways (Borgman, 2018). One is determining the relationship between a dataset and a publication or other document that describes the dataset, and these competencies are detailed in Appendix 3 of Jetten et al. (2021). The second problem, and identified as a gap, is the application of the data steward’s linguistic and domain-specific competencies in documenting the differing degrees of validation and of permanence of publications and datasets, and making the data useful to anyone beyond the original data collectors.

The above sections point to gaps in the knowledge and capabilities of the data steward research profile that affect their work throughout the entire research data lifecycle. Closing these gaps requires resources and skills development at the leadership level in an organisation to ensure that the data steward can deliver effective data stewardship. York et al. (2018) summarise these gaps as gaps between an organisation’s commitment to sustainable stewardship. Suggested solutions include:

- increased investment in infrastructure and capabilities in the long-term preservation of valuable data
- investigating and reducing the gaps between curation, management, and preservation goals and the ways data are actually managed and prepared for preservation and reuse
- how to budget for data stewardship, discussing the funding needed for effective stewardship in research projects and the actual available funding.

## Profile 2: Data steward policy – institute and policy focused

*Focus on policy development, policy implementation, procedures, regulations and principles relating to data. DS interacts with policy makers, managers, board of directors, deans, funders, financial and legal experts in advisory and coordination roles and initiates and oversees implementation and monitoring. This group of stakeholders has a say in, and are ultimately responsible for, how data should be handled. Jetten et al (2021) Appendix 3 provides the detailed description of the function (responsibilities and tasks), competencies (knowledge, skills and abilities) and learning objectives for DS role structured by the eight competence areas, p.30-35.*

### *Gaps in policy steward profile*

Professional data management is complex. Research data policy and support need to be integrated in a wide diversity of research projects within an organisation. Long-term preservation, persistence, accessibility and legibility of data are already established priorities in RDM policy where the data steward has established skills and expertise. New skills include the data steward being able to contribute to the identification and implementation of an organisation-wide definition of data quality, data quality measurement of data content, and nurturing a culture where data is seen as an asset or resource (Stickel & Vandervalk, 2014). Thus the data steward is expected to not only have the technical prowess to assist in data management but also, a common skill missing in data stewardship educational profiles (Wildgaard et al., 2020), to have the negotiation and literacy skills to inform policymakers, and communicative skills to facilitate various parties involved in the creation and curation of data to ensure proper hosting and preservation of content, including ensuring compliance with privacy laws, such as GDPR or other data regulation for example the forthcoming data governance act<sup>15</sup>. Political tools in organisational political mediation and conflict resolution, eg. in budgetary negotiations about organisational priorities in data stewardship when there is competition about available resources, are needed for the data steward to reach solutions and gain a deeper understanding of the workplace.

Data stewards are employed in many different organisations, each with different data architectures. Specifically for institutional policy stewards, a data stewardship competency is documenting the role of that architecture and its capabilities, increasing the data flow within and between organisations, and accordingly increasing data management maturity levels and coordinating data between organisations, individuals, and systems. Aiken et al. (2007) point out that future data stewards need to be able to conduct self-assessment of data architecture and provide a roadmap for an organisation's or funder's data management policy. Self-assessment includes the ability to define performance measures to support policy, planning, and operations and the data to support those measures (Stickel & Vandervalk, 2014). The gaps in skills in self-assessment need to be addressed so that the data steward can address the gaps between current regulations and policies that govern data stewardship and reuse and those that would maximally facilitate stewardship and reuse (York et al., 2018). If not, progress towards reliable data management and ultimately efficient and reliable science may be impeded (Lefebvre et al., 2018).

---

<sup>15</sup> Data Governance Act: <https://datacharter.hypotheses.org/>



### **Profile 3: Data steward agent of change**

*Focus on being the data contact or go-to-person similar to a “data champion”. Passionate to implement solutions via project and change management, customer oriented with a deep understanding of users, and focused on execution regarding policy and strategy awareness, approaching the tasks with an agile mindset and enthusiasm. Strong communication capabilities and understanding of user needs. Tasks include giving advice on compliance, motivating a cultural change in data practices and workflows in a client oriented approach through a greater understanding of processes and operations (Wildgaard et al., 2020).*

#### *Gaps in agent of change profile*

Data stewardship needs strong executive backing to ensure organisational support. In an organisational context, “backing” is identified as financial (Soares, 2012) and communal (Sapp Nelson, 2017). Data stewardship requires continued investment to grow and be executed at a high level (Stickel & Vandervalk, 2014). Decision-makers focus typically on quantifying the financial value of company programmes and hence the data steward as an agent of change needs to be able to document and maximise the business value and impact of good data management (Soares, 2013; 2012). York et al. (2018) point to gaps in the agent of change profile, specifically to those arising between different research cultures and practices that the steward needs to be able to diplomatically navigate and mediate. Such cultural awareness for stewardship is essential for engagement in the human-technical infrastructure stewardship provides. Therefore networking and collaboration skills lift stewardship across an organisation. As such, the data steward needs the qualities of a “business data steward”.

With the qualities of a “business data steward”, an agent of change extends the role of data in their organisation, and through their subject-matter expertise lead data as a business facet of the organisation. The role includes the skills and vision to govern data whatever its size, big or small, simple or complex, market data management and data as a product strategy across the organisation, network with business partners on data strategy, and document return on investment in data stewardship to evaluate how well time and money used in data management is performing for the organisation (Soares, 2013; Aiken, 2016). However, a business data steward, sometimes referred to as “chief data officer”, is an emerging role in federal agencies and universities, with as of yet no known formal training pathway.

Data governance needs communal support in an organisation. As an agent of change, the data steward needs the skills to convince data creators and users, researchers, organisational colleagues, and decision-makers of the value of good data management. Thus, the agent of change has a highly motivational role, providing data creators and users with the freedom to unfold their interests without getting “weighed” down in data administration whilst at the same time applying data management policy and data science skills in real-world projects. Data management is becoming increasingly complex and within an organisation the data steward has the job of helping individuals organise data management work as a common endeavour. Data stewards have “the buck stops here” levels of responsibility for the data (Sapp Nelson, 2017), and so they not only need to know how to manage the data cooperatively, but also how to teach and lead others to manage the



data. Pournaras (2017) discusses the need for an increased focus on learning theories to ensure the data steward can tutor data users and creators, increase their confidence and self-efficacy. Learning theories require the data steward to possess the skills to collect and use knowledge of an organisations' learners, habits, and points of view to be able to provide relevant data management training, supervision, and services (Sapp Nelson 2017).

#### **Profile 4: Data steward infrastructure - data and e-infrastructure focused**

*Focus on liaising with data and IT infrastructure providers, e.g. IT staff, technicians, and application managers inside and outside of the organisation. This role translates the requirements of policies and science into suitable IT solutions and provides advice. It also facilitates the implementation of IT infrastructure and gives access to data and software. DS can perform data analyses, ensure the FAIRness data, or appropriate use of ontologies in the project. This stakeholder group devises and provides tools to enable the implementation of certain data policies and tools.*

*Jetten et al.'s (2021) Appendix 3 provides the detailed description of the function (responsibilities and tasks), competencies (knowledge, skills and abilities) and learning objectives for the DS role structured by the eight competence areas (p.42-46).*

#### *Gaps in infrastructure profile*

Leaning heavily on skills from data science, the infrastructure steward is the most complete profile, but perhaps also the most complex as multiple skill sets come into play when transforming data into valuable analytical content and providing analytics (Palmer, 2014). Palmer continues with a common description of the skills of a data steward as one that focuses on integrity, quality, and definition of data as well as the technical, analytical capabilities fitting the profile of the data steward for infrastructure. In this scenario, the data steward ensures that best practices have been followed and that the analytical content, infrastructure, and tools, not just the data, are reliable. Consequently, this data steward needs skills in leadership, so as to be able to collaborate across an organisation to achieve better data quality on the enterprise or concern level. Close cooperation between IT experts and managers of all areas and functions is necessary (Rickards & Ritsert, 2012). Thus, data stewardship is not only about actions and providing services but also the capability to manage a suite of responsibilities and an infrastructure that require a high level of planning, collaboration, and judgement, thereby binding people to practice (Steelworthy, 2014; Stickel & Vandervalk, 2014) and maximising DS capabilities (Demchenko et al, 2021; York et al., 2018).

However, as data management becomes increasingly complex and FAIR, gaps are appearing in the literature. For example, the data steward infrastructure is expected to have subject-specialist expertise and experience with the research process rather than a background purely in IT (Oliver, 2017). Increased subject specialist knowledge is identified as a critical asset in working with FAIR data (Mons et al., 2017; Barker, 2021) where the steward needs to use subject knowledge to be able to advise on ontologies and technical skills to exploit linked data, metadata and machine readable syntax, and ensure alignment with the CARE principles (Collective benefit, Authority to control, Responsibility and Ethics).

These gaps in a data stewardship knowledge and skill set demand the professionalisation of data steward education, require a higher level of subject-expertise included in the curriculum and new educational formats, such as hyflex PhD and graduate programmes, that are open and flexible to attract professionals back into education (Aiken, 2016). Alternatively, existing technical educations such as engineering, library science, or business degrees must be supplemented with a formal curriculum that introduces the data centric perspective, and teaches the challenges associated with implementing a formal organisational data management programme to encourage this new thinking about data stewardship (Aiken, 2016).

### 3.1.5 Gaps in researchers' skills

#### *Data management planning*

A widely documented finding is that early-career researchers and graduate students may have extensive responsibilities in everyday data management in research projects. Yet very few of them have had any education in RDM (Goben & Griffin, 2019; Krahe et al., 2020; Maienschein et al., 2019; Wiley & Kerby, 2018). When asked about educational needs of the researchers and graduate students in RDM, respondents have highlighted needs for training and support for making data management plans (DMPs); where and how to store, preserve, and share data; the creation of metadata and documentation; to become familiar with funders' mandates; managing sensitive data; intellectual property rights (IPR) issues, and citing data (Knight, 2013; Parsons et al., 2013; Rantasaari, 2021).

Still, there is a lack of understanding about the significance of RDM, and a DMP may be regarded as an administrative burden (Buys & Shaw, 2015; Schumacher & VandeCreek, 2015; Weller & Monroe-Gulick, 2014; Yu et al., 2017). In a study of health and medical researchers' RDM practices, it appeared that almost half of the academic staff did not have a DMP and 70% of the PhD or masters students were unsure whether they had a DMP or not (Krahe et al., 2020). Furthermore, researchers in arts and humanities are less aware of the funders' mandates for a DMP than researchers in other disciplines. This has much to do with the multifaceted nature of the research materials of arts and humanities and the fact that these researchers are typically less dependent on external funding compared to researchers in natural sciences (Akers & Doty, 2013; Edmond & Tóth-Czifra, 2018). Though there are still researchers who find a DMP to be more like a bureaucratic burden (Rolando et al., 2013), many researchers have expressed a wish for help in developing, improving, and standardising data management planning (Cox & Williamson, 2015; Goban & Griffin, 2019).

#### *Data storing*

According to several studies, researchers and graduate students need more information about the safe and secure storage and backup platforms for the data of their ongoing research (e.g. Koltay, 2017; Krahe et al., 2020; Lefebvre et al., 2020; Weller & Monroe-Gulick, 2014). There is a lack of knowledge or trust in institutionally recommended platforms following the usage of heterogeneous platforms instead (e.g. Cox & Williamson, 2015; Goban & Griffin, 2019). Especially in arts and humanities, researchers and graduate students typically use their laptops, external hard drives, and commercial cloud servers for storing their data (Akers & Doty, 2013).

### *Data documenting and organising*

When it comes to creating documentation for research data, researchers usually learn it ad hoc in research projects. Documentation is typically not standardised and will be left to researchers themselves, to carry out and publish in their ongoing study (Rantasaari, 2021; Rolando et al., 2013). Non-standardised documentation and organisation of the data causes problems in correlating and relating datasets and limits preservation, sharing, and reusing of the data (Carlson, 2011; Joo & Peters, 2019; Whyte & Ashley, 2017). Adopting sound documentation practices and the use of metadata standards during collection and description phase enable data discovery, access, analysis, and synthesis (Specht et al., 2015).

### *Legal and ethical considerations*

Training needs for managing intellectual property, copyright, and agreements issues as well as for ethical aspects affecting data processing, sharing and reuse have been reported in many studies (e.g., Cox & Williamson, 2015; Knight, 2013; Parsons et al., 2013; Rantasaari, 2021). What makes these legal and ethical aspects challenging to manage is that researchers experience conflicting pressures to open the data on one hand, on the other hand to limit the collecting, storing, processing and sharing of the data to minimise the risks of insecure data handling and protect study subjects' privacy and confidentiality (Akers & Doty, 2013; Borgman, 2018; Krahe et al., 2020; Thielen, 2017). Researchers also need more information on available IPR and data privacy support services to manage these issues.

### *Analysis and visualisation*

Studies on research data management have mainly focused on data management planning, documentation, and sharing of data. There is a gap in the literature that examines the earlier stages of the data lifecycle, such as data creation, processing, analysing, and visualising which all have a major impact on data quality, integrity, and usability (Krahe et al., 2020). This applies to services as well. The quantity and heterogeneity of data exceeds many researchers' abilities to properly analyse their data (Anderson et al., 2007). Especially researchers using quantitative, qualitative statistical, and experimental methods, and in particular health scientists and social scientists, have expressed greatest needs for assistance in analysis and visualisation as opposed to humanities researchers (Joo & Peters, 2019; Weller & Monroe-Gulick, 2014). Researchers may lack the information of the existing tools and resources, and time, funds, skills, or experience to efficiently utilise available tools and services and how to incorporate them into their research (Anderson et al., 2007; Lefebvre et al, 2018).

### *Technical needs*

Apart from consultative and informative RDM services (RDSs) such as guides, basics trainings, and support for writing DMPs, researchers have expressed that they also have more technically oriented data management needs like assisting with metadata; data cleaning; converting and integrating external data; software development, visualisation, and digitising sources (e.g. Knight, 2013; Parsons et al., 2013; Weller & Monroe-Gulick, 2014).

### *Quality control*

Unlike in many industries where data management (DM) processes are intrinsic, there are no agreed and standardised procedures, practices, and roles in universities to audit and

control research data quality (Lefebvre et al, 2018). Graduate students and early-career researchers may have big data management responsibilities without proper training nor agreed procedures to perform the tasks (Goben & Griffin, 2019; Krahe et al., 2020; Maienschein et al., 2019; Wiley & Kerby, 2018). As long as standardised procedures with proper quality control are missing, a culture of data sharing will not become the norm (Lefebvre et al, 2018). While retrieval of historical data is unrealistic and would demand enormous resources, Hardwicke and Ioannides (2018) recommend that data from the most highly-cited literature should be essentially made discoverable, retrievable, and accessible to the scientific community. Furthermore, they point to the “Data Ark Initiative” as a resource for preservation, improved data sharing, and a discussion forum on data stewardship<sup>16</sup>.

#### *Data preservation, sharing, and re-use*

In surveys and interviews, researchers have expressed a need for help in long-term preservation, sharing, publishing, reuse of data, and evaluating the costs of the aforementioned (Cox & Williamson, 2015; Dijkers, 2019; Joo & Peters, 2019). Researchers may have uncertainty or fears concerning IPR, confidentiality, and privacy issues of the data, or that the shared data will be misused or used without crediting the author (Chiarelli et al., 2021). There is also uncertainty about the length of the preservation requirements and what kind of procedures are needed (Knight, 2013).

The most common methods for researchers to share their data are still sharing them on request or as a supplement to the research paper (Joo & Peters, 2019). Researchers lack knowledge, trust, and skills to share and discover data more efficiently through relevant repositories. Barriers to data preservation, sharing, and reuse are the lack of easy-to-use infrastructures, missing academic incentives, and a non-prevalent data sharing culture (Goben & Griffin, 2019; Perrier et al., 2020; Rolando et al., 2013; Zenk-Möltgen et al., 2018). Moreover, the multiplicity of data and lack of metadata standards make data sharing a challenge (Whyte & Ashley, 2017). Researchers typically create the documentation for themselves and their present study, not for further use and users (Rantasaari, 2021). When educating researchers about the benefits of using metadata, Kim (2014) recommends highlighting the importance of metadata for preservation and access functions, such as data archiving, sharing, and reuse. As mentioned above, there is also much uncertainty about IPR, copyright and rights issues, and – especially in the social sciences and health sciences – how to treat data that contains personal and sensitive data (Akers & Doty, 2013; Joo & Peters, 2019). Many researchers do not know any open data repositories, and even if they find suitable external data, the possibility to integrate it with their own data depends on the quality of the metadata and usage rights (Specht et al., 2015). Irrespective of discipline, shared datasets are underused (Quarati & Raffighelli, 2020). One way to enhance the quality, reliability, and usage of data sharing and reuse would be by extending journals’ peer-review procedure to datasets (Edmond, 2020).

When it comes to disciplinary differences, researchers in the natural sciences, engineering, linguistics, and archeology appear to be more willing than researchers in other disciplines to share their data (Akers & Doty, 2013; Berez-Kroekel et al., 2022; Derudas et al., 2021; Forkel et al., 2018; Jacobs & Holland, 2007; Joo & Peters, 2019; Wilson, 2013). Social

---

<sup>16</sup> <https://osf.io/meetings/DataArk>

sciences' researchers often use external data (Borgman, 2008) such as data from different registries, although they may find identifying relevant data challenging (Weller & Monroe-Gulick, 2014). Quantitative and experimental researchers encountered less difficulties in acquiring external data than other researchers, whereas historians found acquiring access to materials the most challenging (Weller & Monroe-Gulick, 2014). Although in the social sciences, many journals have a data sharing policy, the supporting dataset is available only in a small number of the articles (Zenk-Möltgen et al., 2018). Still, a journal's data sharing policy together with sound support services from researcher's organisation, funders and scholarly societies increase data sharing, whereas infrastructure with insufficient data protection decreases it (Zenk-Möltgen et al., 2018).

Besides challenges with data multiplicity, researchers in arts, humanities and history also must reckon with the strong adherence of the data to context, viewpoint, and interpretation: It is often about comparing and interpreting different sources. To be able to share and reuse humanities' data, they should have metadata with versatile and rich description of the original contexts and provenance (Edmond, 2020). Besides, data in arts and humanities are usually not owned by researchers but housed in different archives and cultural heritage institutions with no standardised principles and practices for citing and reusing (Seillier et al., 2017).

### 3.1.6 What do we not yet know about researchers' RDM needs and practices

According to the literature, there is a shortage of research-based knowledge on the RDM needs and practices concerning

- 1) different research personnel groups;
- 2) different disciplines, research methods, and data types;
- 3) other research institutions than big research-intensive universities; and
- 4) the real everyday research work.

#### *Different research personnel groups*

According to the meta-analysis of data management surveys between 2007 and 2017, the focus was on faculties' needs (Goben & Griffin, 2019). At the same time, we do not know much of the needs of postdoc researchers, other staff, graduate, and under-graduate students though these are the groups of people who work most with the processing of data in research projects (Goben & Griffin, 2019).

#### *Different disciplines, research methods, and data types*

With regard to disciplines, RDM needs of the researchers in HSS disciplines are understudied compared to researchers' needs in STEM disciplines (Goben & Griffin, 2019; Tóth-Czifra, 2019). We also have a lack of knowledge in terms of the needs and practices influenced by different research methods and data types.

#### *Institutions*

Most of the research in RDM has focused on the needs of researchers at big research-intensive universities. Consequently, we need more studies of the needs and practices at

smaller research institutions like liberal arts colleges, small research institutes, and universities of applied sciences (Goben & Griffin, 2019).

#### *The everyday practices of researchers*

According to a scoping review of 301 research articles and 10 reports on RDM in academic institutions (Perrier et al., 2017), 80% of the studies are based on self-assessments of researchers or on the interpretations of outside observers. More empirical research is needed to uncover how researchers apply RDM in their everyday research practice (Perrier et al., 2017). Moreover, we need to know what the contributors and barriers of sound RDM practices are, and what the quality of the data deposited in repositories is (Perrier et al., 2017).

### 3.2 How to meet these gaps (solutions)

In the previous sections, we have sketched the skills gaps researchers and data stewards face in managing data responsibly. The core challenge is that both researchers and data stewards operate in a complex landscape where specialisation is needed to focus on stewardship activities. The second challenge is whether centralised services can be general enough to cater to all research projects and at the same time recognise subject-specific differences (Teperek et al., 2018). Below, we present solutions to improve data stewardship practices in research organisations as suggested in the literature.

#### 3.2.1 Solutions in Governance of Data Stewardship

##### *Measure success*

Measurement is suggested as the clearest and simplest way to demonstrate the success of data stewardship and hence to document the necessity of data stewardship in the organisation. Smith (2021) argues that all activities of data stewards should be aligned with specific metrics and measured regularly. However, meeting or verifying compliance with the requirements on data stewardship that set the parameters for these metrics requires assessing the current state, identifying gaps, and, if necessary, defining a roadmap for improvement of data stewardship programmes and services. Thus, models of maturity assessment are argued to be foundational for gaining strategic knowledge of standards, best practices, and ambition with data stewardship within an organisation. Such models rely on collaboration across multiple knowledge domains: IT, infrastructure, service providers, and organisational leadership. Therefore, Peng (2018) and Peng et al. (2016) explore existing maturity assessment models for data stewardship and evaluate each model's suitability for verification and improvement needs. They conclude that continuous evaluation of organisational stewardship capability and practices needs to be applied to individual data products, the quality of individual data products, and services throughout the entire data product lifecycle. Maturity assessment models help with working towards full compliance with federal, agency, organisational, and user requirements as well as other directives for data stewardship. Targets are identified and strategic steps on how to get there are planned. This formal approach includes verification, reporting, metrics, and evaluation. Further, the organisation gains a holistic view of processes, stakeholders, and other dependencies which are part of data stewardship processes. Lefebvre et al. (2018) conclude that organisations that negate the responsibility for coordinating the various data stewardship processes (data

governance, master data management, metadata management, data quality, enterprise data architecture, etc.) will not be able to provide guidance to a data stewardship effort on the implementation of these components. As a result, any efforts to improve the value of data and data stewardship will be unsuccessful.

#### *Formalise the data steward strategy*

The data stewardship strategy should be formalised in a comprehensive document. This document should address policies, guidelines, data programme alignment, data governance, and data stewardship (Stickel & Vandervalk, 2014). Such a strategy should commit resources to the alignment of strategic goals with data collection, reporting, analysis, and IT infrastructure as well as staff (subject matter experts, IT experts, data owners etc.) at each maturity level of the data governance process; maturity models and data stewardship evaluation requirements, priorities, risk assessment (risk assessment process from a life-cycle perspective that includes policies, standards, data repositories, and calculation processes), and bottle-necks and silos that are a detriment to the organisation (Brackett, 2003).

A formalised strategy is expected to result in better contact and coordination (Brackett, 2003). Lefebvre et al. (2018) advise central policies in data stewardship that need to be refined locally, as there is no evidence that central committees or task forces are the most optimal decision layer when it comes to managing research data, or elaborating on data steward responsibilities and tasks. Instead, the data creators within an organisation should be included in the formulation of the policy and the latter then refined locally through groups of stakeholders based on the type of data (e.g. observational, experimental, simulation and derived or compiled) to gain significantly more impact on services needed to plan and handle data.

#### *Promote joint governance*

Developing effective principles for governing data stewardship results in extensive deliberations between faculty, administrators, and students (Borgman, 2018). These deliberations have also proved to increase communication, understanding, and trust between different actors within an organisation. Borgman presents two positive outcomes of increased deliberation which these are 1) central administration can learn more about the challenges in data stewardship at different faculties and 2) data stewards can learn more about how to support data producers who have very different skills sets and access to very different technologies.

Likewise, organisations that simply assign data stewardship tasks to certain staff members without clearly defining the role and explaining the specific responsibilities of a data steward see poor results from those efforts. Hence, they do not realise the sustained benefits from a data stewardship programme that trained data stewards can provide (Smith, 2021).

#### *Collaborate in the provision of data stewardship*

Sound coordination is crucial, especially as solutions point towards a decentralised data steward strategy that allows close collaboration with data creators, collectors, and providers to cater to their specific support requests (Lefebvre et al., 2018; Smith, 2021; Verheul et al., 2019). Within an organisation, embedding data stewards in the existing organisational structure will promote the sharing of knowledge and findings between data stewards from

different domains and departments, policy officers, managers, data protection officers, lawyers, licence negotiators, and other staff involved in RDM in some way or another. The appeal of a “network of expertise model” for delivering data governance has been shown to be successful in addressing both specialised information needs and common problems (Erway, 2012; Gendron et al. 2015). Further, collaborative networks can encourage tiered staff models where different levels of participation and knowledge enable the full utilisation of an organisation’s resources as well as access to knowledge and services at partner institutions (e.g. academic libraries or general data repositories) that can contribute with staffing and funds to sustain and offer data steward services to potential users at affiliated organisations with limited or no data steward resource of their own. Conversely, an academic library with no data security expertise can – through a network approach – gain access to lawyers who can provide guidance around data rights requirements. A network approach to collaboration requires different levels of effort in providing data stewardship from the different stakeholders (Sansone et al., 2019). However, all stakeholders will gain access to data curation expertise in more disciplines and formats than locally available and contribute to a larger ecosystem of data curation practice (Johnston et al., 2018).

#### *Collaborate in the provision of training materials for data stewards*

Several European initiatives have begun to fill the gap of data professional profiles and to identify the competences that will be required in the development of data steward career paths, amongst them Barker (2021), Jetten et al. (2019), HLEG on EOSC (2018), Demchenko et al. (2017). There have also been investigations into the coordination of data steward education on a national level (Jetten et al, 2021; Wildgaard et al, 2020). Whilst courses exist in data stewardship, there are no related certification accreditation mechanisms, defined curricula or coordinated and coherent approaches to skills, competencies and training provision (Barker et al, 2021; Demchenko et al, 2021). Therefore, there is a lack of coordination regarding initiatives and training materials on a national, European and international level. This in turn hampers the identification of the best resources to use and implement within an organisations’ training programme or universities’ curricula (Stoy, 2019).

To create a baseline approach to data stewardship, the aforementioned initiatives all recommend that the first step to providing relevant training materials is to provide clear definitions of professional data steward profiles. These profiles will act as a baseline for career paths in these roles. Defining roles and formalising responsibilities of stewards will help organisations and stewards themselves understand their responsibilities (Peng et al., 2016). Since there are virtually no higher-education-based offerings focused on data stewardship, organisations cannot look to higher education to provide data stewardship profiles and expectations to what kind of knowledge, skills, and abilities these stewards should acquire (Aiken, 2016). Also, as the field of data stewardship is rapidly evolving, individual organisations need to commit to investing in continuous data stewardship education and upskilling their current staff to help them keep up with changing job requirements (European Commission Expert Group on FAIR Data, 2018). Therefore, recommendations in the literature point to two phases of collaboration in the provision of data steward training to reduce the burden of increased cost of continuous skill development. Firstly, collaborate in training people to train data stewards (“train-the-trainer” programmes) (Attwood et al., 2019). Train-the-trainer programmes will establish networks of expertise that can result in a collaborative approach to providing dynamic training for data stewards. Such programmes include interactions between those who have achieved best



practice in data stewardship and those who aspire to it, thus providing efficient and evidence-based approaches to training development. Secondly, collaborate in the design and execution of training programmes for data stewards. These programmes could be via a combination of lectures, workshops, hack events, conference sessions, webinars, tutorials, summer schools, podcasts, visiting scholars' programmes, or even collaborative research projects across organisations to not only reduce cost, but also increase inclusivity and currency of content. Hands-on courses where participants learn how to actually carry out specific tasks and are equipped to put these into practice over various disciplines and domains are particularly valuable (European Commission Expert Group on FAIR Data, 2018 and DSCC, 2022).

Collaboration around curriculum development is also encouraged in papers investigating network models for delivering data steward training. One example is establishing a collaborative network in curriculum development that specifically addresses data and metadata curation issues (Johnston et al., 2018). Recommendations for collaborative topic networks include allowing different levels of participation so that institutions can contribute according to their available resources and knowledge and allowing different partner institutions to enrich the collaboration (e.g. representatives from academic libraries, researchers, technologists, general data repositories, curation services, and funders provide expertise in the development of curricula which together add layers of knowledge, application contexts, and challenges (Parham & Murray-Rust, 2011). As a result, training materials become more efficient as they capture as much context and description of the topic as possible.

In conclusion, collaboration will allow effective cross-disciplinary communication and efficient resource allocation for data stewardship, supporting organisations in better meeting the challenges of stewardship (Rolando et al, 2013). Training materials from both approaches should of course be made available through open educational resources to enable reuse and peer review of the quality of the content.

#### *Implement organisational frameworks for data stewardship*

As discussed in the previous section, it requires targeted and continuous investment to ensure trained data stewards and scientific data experts. Financial resources are required for research data curation and preservation technology and services. Decisions about data, quality tools, data architectures, data standards, data management and data support must therefore be deliberate as stewardship must be budgeted for. As Verheul et al. (2019) point out: "Professional data stewardship needs professional data stewards, who receive proper recognition for the work they do". However, many stewardship initiatives fail to gain the attention they deserve or the momentum they need because they do not have organisational support (Rolando et al., 2013; Smith, 2021).

Establishing an organisational framework for data governance at the policy level in an organisation would explicitly connect data and data stewardship with business goals, resource allocation, capabilities and areas of responsibility. Even though embedding stewardship in the organisational policy would align stewardship to the organisations' aforementioned values, resources, and goals, it is not the only prerequisite for data stewardship being advocated as a sustained programme across the organisation. There also

needs to be an infrastructure in place which comprises both technological and human “on the ground” support (Smith, 2021). The human data steward infrastructure in an organisation should at a minimum, according to Verheul et al., at a minimum support the legitimacy of embedding stewards in projects, provide generic and advisory services across the organisation, and participate in (cross-organisational) advisory boards. These boards do not work directly with data but with the development of organisational policy, strategy, and coordination. Rolando et al. (2013) recommend that the framework supporting the responsible care for data at an institution should, at a minimum, include:

1. Data archiving provision
2. A research data stewardship group
3. A formal data stewardship marketing plan
4. A repository of data management plans
5. Provision of data management training
6. Partnership in the creation and update of the necessary and appropriate institutional policies

This leads us back to the case in point raised in the previous section, that decision makers within an organisation must be willing to enforce rigour and invest in data skills and, most importantly, need to adapt tactics arising from policy and infrastructure mandates in order to nurture a data-driven culture and value system (Dyché & Polsky, 2016).

#### 4. Summing up: The Gaps, Solutions, and Sources

General level gaps		Reference material (newest material to oldest)	Solutions
Lack of agreement on sound RDM practices	European Commission, 2018a; Lucas, 2010; Aiken et al., 2007	<ul style="list-style-type: none"> <li>• Collaborate and draw on best practices</li> <li>• Evaluate the current state of an organisation and provide risk and opportunity information in data management practices, data strategy, data quality engineering, data risk, and data training to inform application of organisation-wide DM practice</li> <li>• Provide a road-map for improvement of organisational DM practices</li> </ul>	
Lack of knowledge about quality control of data and metadata	Kühn & Streit, 2017; Lucas, 2010	<ul style="list-style-type: none"> <li>• Provide certified training targeting different domains of research</li> <li>• Devise and implement a framework for data governance including total quality management (TQM), concepts and practices to improve data and</li> </ul>	

		<p>information quality, set up data quality policies and guidelines, data quality measurement (including data quality auditing and certification), data quality analysis, data cleansing and correction, data quality process improvement, and data quality education</p> <ul style="list-style-type: none"> <li>● Identify data source metadata, characterisation, reuse and documentation</li> </ul>
How to link between data & publications (data upon which the research results are based is usually not linked to the publications).	Hardwicke & Ioannides, 2018	<ul style="list-style-type: none"> <li>● Preserve and liberate important scientific data</li> <li>● Address barriers to data sharing</li> <li>● Advance community discussions on data stewardship</li> </ul>
Lack of knowledge about ensuring data security in cloud services	Estrada-Galinanes & Wac, 2018	<ul style="list-style-type: none"> <li>● As the cloud is not the ideal place for personal (health) information, a collaborative platform for management and archival of personal health information that supports the individual, community and societal value of data is recommended</li> <li>● Focus on rich data rather than big data</li> <li>● Store data privately by default, but provide options to share at any time</li> <li>● Provide licences and succession policies</li> </ul>

		<ul style="list-style-type: none"> <li>● Involve citizens as data providers and provide functionalities to connect citizens, researchers and data</li> <li>● Support local storage and distributed storage to avoid single points of failure</li> <li>● State a long term retention period</li> </ul>
Lack of collaboration between different stakeholders, such as researchers, infrastructures' developers, and research supporting professionals.	Schuster, 2021; European Commission Expert Group on FAIR Data, 2018; HLEG on EOSC, 2018; Mons, 2018	<ul style="list-style-type: none"> <li>● Use international fora as vehicles to support implementation and the adoption of best practices/standards</li> <li>● Create career enhancement and funding incentives for users and providers and other engagement schemes</li> <li>● Develop strong synergy with cybersecurity</li> </ul>
<b>Gaps in implementing FAIR data</b>		
An understanding of FAIR	Demchenko et al., 2021; European Commission Expert Group on FAIR Data, 2018; Mons, 2018	<ul style="list-style-type: none"> <li>● Promote FAIR digital objects (metadata, persistent identifiers, documentation, and analysis procedures) and the FAIR ecosystem</li> <li>● Implement interoperability frameworks</li> </ul>

		<ul style="list-style-type: none"> <li>● Implement incentives and recognition for working FAIR &amp; FAIR stewardship</li> <li>● FAIR for humans and machines</li> </ul>
Lack of knowledge and skills in how to secure data reusability	Schuster, 2021; European Commission Expert Group on FAIR Data, 2018; HLEG on EOSC, 2018; Mons, 2018	<ul style="list-style-type: none"> <li>● Establish interoperability frameworks</li> <li>● Select standards and community-endorsed best practices</li> <li>● Monitor data access and reuse</li> </ul>
How to manage sharing practices	Dijkers, 2019; Joo and Peters, 2019; Callahan et al., 2017	<ul style="list-style-type: none"> <li>● Collaborate with funders in describing how to share in DMP at project start</li> <li>● Jointly develop strategies to deal with data privacy, licensing, security and copyright</li> <li>● Build models for sharing into day-to-day workflow, including creation of metadata and the uploading of data to repositories, and “automating” it</li> <li>● Create user-centred research support services locally</li> </ul>
How to support the findability of data	DSCC, 2022; FAIR Data Austria, 2021; Mons, 2018	<ul style="list-style-type: none"> <li>● Jointly evaluate optimal technologies, software, open source code, repositories, etc. to build and maintain national IT infrastructure also usable within the new generation research workspaces (cloud, FAIR, VRE).</li> <li>● Create materials describing terminology systems, their strengths,</li> </ul>

		<p>weaknesses, limitations, and adoption in the community.</p> <ul style="list-style-type: none"> <li>● Implement joint data citation principles in order to make data reusable, citable and findable, so the organisation can do metrics on and reward researchers for sharing their data with others.</li> </ul>
<p>Lack of easy-to-use infrastructure, human and technological, to embed FAIR in research practice</p>	<p>DSCC, 2022; Barker et al., 2021; Latif et al., 2021; Dijkers, 2019; Sansone et al., 2019; European Commission, 2018</p>	<ul style="list-style-type: none"> <li>● Include a data management plan as part of a research proposal from the start, costs are limited, and grant makers allow these costs to be part of a budget.</li> <li>● Coordinate and align relevant skills, curricula, and training frameworks by generating a consensus on a core European higher education curriculum to deliver FAIR and open science skills at university level.</li> <li>● Create overviews of standard tools, metadata models, and initiatives in registries. Provide information about the interrelationships between the tools and specify what they do rather than how they do it.</li> <li>● Use FAIRsharing to explore what resources exist in which areas (and whether those resources can be used or extended), as well as enhance the discoverability and exposure of the resource.</li> </ul>

<p>Lack of mutual understanding about data, metadata, commonly shared protocols, standards, and formats between disciplines and stakeholders</p>	<p>Dijkers, 2019; Callahan et al., 2017</p>	<ul style="list-style-type: none"> <li>● Stress the ethics as well as societal and practical benefits of data sharing</li> <li>● Review any manuscript written, prior to its submission to a journal, to make sure there is no misunderstanding of the data or misrepresentation of those who collected them.</li> <li>● Bring together junior and senior level experts including researchers from academia and industry, data science, and other experts such as program staff representing federal and private funding agencies and address how to simplify the logistics and cultural barriers of data documentation and sharing.</li> </ul>
<p>Providing an easy-to-use, self-generating metadata option when designing data management systems</p>	<p>Joo &amp; Peters, 2019; Callahan et al., 2017</p>	<ul style="list-style-type: none"> <li>● Offer training opportunities for researchers to better organise their metadata for research data.</li> <li>● Highlight the importance of metadata for preservation and access functions, such as data archiving, sharing, and reuse.</li> <li>● Provide common templates for structuring metadata.</li> </ul>



Gaps in training materials		
How to synchronise between project, funder, and stakeholder	Williams et al., 2017	<ul style="list-style-type: none"> <li>● Funders should require DMP and demand DMP maintenance.</li> </ul>
How to avoid production of non-reusable data	Williams et al., 2017	<ul style="list-style-type: none"> <li>● Document and describe the origin of the data.</li> <li>● Provide an unambiguous definition of the data and associate the definition with data values and location of data values in data models used for the study.</li> <li>● Provide operations through which data should be traceable. Provides algorithms needed for traceability.</li> </ul>
Lack of inclusivity (of different disciplines) in training materials	Goben & Griffin, 2019; Williams et al., 2017	<ul style="list-style-type: none"> <li>● Conduct needs assessment determining institution type, target population, cross-institutional trends in topics such as storage, long term preservation, and access to data.</li> <li>● Assess staff, student, and non-researcher faculty needs.</li> <li>● Investigate non-STEM disciplines.</li> </ul>

<p>Lack of common terminology around data management</p>	<p>Goben &amp; Griffin, 2019; Williams et al., 2017; Partlo, 2014</p>	<ul style="list-style-type: none"> <li>● Find language specific enough for people to see their experience in it, but general enough to draw people out of their own disciplinary perspectives to allow conversation across disciplines.</li> <li>● Implement cross walks across terminologies and definitions of terms.</li> <li>● Create cross-cutting documentation and refer to industry standards.</li> </ul>
<p>Advise on and implement metadata to ensure findability</p>	<p>Joo &amp; Peters, 2019; Kühn &amp; Streit, 2017</p>	<ul style="list-style-type: none"> <li>● Highlight importance of metadata for preservation and access functions, such as data archiving, sharing, and reuse.</li> <li>● Provide cases focusing on challenges of handling data itself, i.e. handling of data types and metadata.</li> <li>● Establish training in handling of large data volumes and management of community requirements in terms of standardised workflow languages, data types, and metadata.</li> <li>● Provide guidance on 1) expectations to a minimal set of metadata to ensure suitability for the most research projects; 2) a desired set of metadata which would optimise discovery; and 3) choosing metadata which has the potential to improve findability,</li> </ul>

		accessibility, interoperability, and reuse for specific communities.
How to support data management from the start of the project for different stakeholders	Lefebvre et al., 2020; Williams et al., 2017; Sapp Nelson, 2017; Calzada Prado & Marzal Miguel, 2013; Schneider, 2013; Piorun et al. 2012; Carlson et al., 2011; Qin & D'Ignazio, 2010	<ul style="list-style-type: none"> <li>• Work with pre- and post award teams and provide data management support in grant application process.</li> <li>• Provide peer review of DMPs.</li> <li>• Advise on FAIR data management at project start.</li> <li>• Provide infrastructure and legal advice.</li> </ul>
How to scaffold data curricula throughout an academic career	Demchenko et al., 2021; Whyte, et al., 2018; Sapp Nelson, 2017	<ul style="list-style-type: none"> <li>• Implement a skills and capability framework to recognise skills demanded of researchers, support professionals, and the organisation.</li> <li>• Facilitate exchange between universities in implementing data steward curricula/programmes.</li> <li>• Align competencies according to application (personal, team, research enterprise) in such a way that the skills attained at the start of a career give persons moving up the career ladder greater familiarity with data management and therefore greater likelihood of success at the more senior levels.</li> </ul>

<p>How to use EOSGCore and EOSGExchange services for data publication and preservation</p>	<p>Whyte et al., 2022; Barker et al., 2021; European Commission Expert Group on FAIR Data, 2018; HLEG on EOSC, 2018</p>	<ul style="list-style-type: none"> <li>● Skills needed to provide front office support and enable researchers in the development and provision of a core data infrastructure for their research.</li> <li>● Navigate the EOSC catalogue and identification of relevant services.</li> <li>● User-level knowledge of relevant services of EOSC and operational support tools.</li> <li>● Generic data science skills (DMP, stewardship, statistics, etc.).</li> <li>● Data science skills depending on the discipline (data formats, validated algorithms, ethics, etc.).</li> <li>● Apply the services and data to generate new scientific findings.</li> <li>● Share the tangible outputs in a way that makes them visible in EOSC (data, workflows, new algorithms, papers).</li> </ul>
<p>How to continue the development of an infrastructure to support data discovery, curation, preservation, and sharing</p>	<p>European Commission Expert Group on FAIR Data, 2018</p>	<ul style="list-style-type: none"> <li>● Implement robust business processes for managing the data lifecycle, long-term preservation, and file format transformation; data protection and security where needed.</li> <li>● Create a value proposition and business model for sustainability and a handover plan in the case of discontinued service.</li> </ul>

<b>Gaps in data stewards' skills</b>		
How to adapt to the users' specific needs	Jetten et al., 2021; Wildgaard et al., 2020; TóthCzifra, 2019; Perrier et al., 2017	<ul style="list-style-type: none"> <li>● Provide local training and identify long-term education needs.</li> <li>● Use the case studies to set up a community survey to further define the needs of data stewardship training and identify gaps, taking into account institutional settings like type of institution, size, research disciplines, available resources.</li> <li>● Close collaboration in the organisation embeds the data stewards and aligns data management needs of the various stakeholder groups.</li> </ul>
Where to start, regarding the specialisation of the data steward and the competences they need	Jetten et al., 2021; Wildgaard et al., 2020; Mons et al., 2017; Aiken, 2016	<ul style="list-style-type: none"> <li>● Develop hybrid, flexible, educational opportunities that allow candidates with specific needs to augment their competence within certain disciplines.</li> <li>● Collaborate on cases and internships with local industry, departments and organisations.</li> </ul>

		<ul style="list-style-type: none"> <li>● Focus on foundational data management practices for all organisations, regardless of their organisational or data strategies.</li> </ul>
<p>What kind of skills data stewards need when working with new types of data: skills in citizen science, specifically IT tools and data quality</p>	<p>Demchenko et al, 2021.; Goudeseune et al., 2020; Rickards &amp; Ritsert, 2012</p>	<ul style="list-style-type: none"> <li>● Assessing sensitivity and data quality.</li> <li>● Legal frameworks (e.g. GDPR.regulation) and data ownership</li> <li>● Centrally-defined IT models that link to individuals.</li> <li>● Facilitate access and use of technical tools by adapting them to citizens.</li> <li>● Collaborate with skilled scientists within the research consortium and ensure proper training of scientists.</li> <li>● Provide multiple alternative strategies for working towards the same goal.</li> </ul>
<p>How to address the need for transparency in data analysis and ethics when working with new types of data</p>	<p>Weller &amp; Watteler, 2021; Aguinis &amp; Solarino, 2017; Stickel &amp; Vandervalk, 2014</p>	<ul style="list-style-type: none"> <li>● Data protection and legal frameworks on national, sub-national and specialised levels.</li> <li>● Link between data protection and research ethics in eg. social media research.</li> <li>● Set up transparency criteria covering the sequential aspects of the qualitative and quantitative research process.</li> </ul>

<p>How to work with publication versioning between datasets and documents, including validation of data</p>	<p>Jetten et al., 2021; Borgman, 2018</p>	<ul style="list-style-type: none"> <li>● Ensure persistent identifiers are able to “resolve to a location” online to facilitate access in addition to identification.</li> <li>● Ensure data and publications are released to the scholarly communication system with comparable status.</li> <li>● Peer-review datasets as rigorously as publications.</li> </ul>
<p>How self-assessment of stewardship can inform policy</p>	<p>Lefevbre et al., 2018; York et al., 2018; Stickel &amp; Vandervalk, 2014; Aiken et al., 2016; Soares, 2013; Soares, 2012; Aiken et al., 2007</p>	<ul style="list-style-type: none"> <li>● Documenting data architecture, data flow, maturity and quality levels.</li> <li>● Maximise the business values &amp; impact of good data management.</li> <li>● Address the gap between current regulations and policies that govern data stewardship and reuse, and those that would maximally facilitate stewardship and reuse.</li> <li>● Identify deficiencies that inhibit stewardship, access, and use.</li> <li>● Identify performance measures to support policy, planning, and operations and the data to support those measures.</li> <li>● Identify policy, planning, and operations needs for data – current and future.</li> <li>● Conduct policy review of data needs at local, district, state, and national levels.</li> </ul>

<p>The data steward's need for mediation skills to negotiate, inform, and facilitate</p>	<p>Jetten et al., 2021; Tasovac et al., 2021; Wildgaard et al., 2020; York et al., 2018; Pournaras, 2017; Aiken, 2016; Soares, 2013</p>	<ul style="list-style-type: none"> <li>● Facilitate local and national communities and networking activities.</li> <li>● People skills knowledge and skills in order to comply with relevant standards and policies and to facilitate good RDM practices.</li> <li>● Ability to communicate across different platforms, using different media to connect with different user-groups.</li> <li>● Proficiency to act as a “bridge” between data infrastructure, users, staff, customers, and the organisation.</li> </ul>
<p>Advice on commitment to sustainable data stewardship</p>	<p>Demchenko et al., 2021; European Commission Expert Group on FAIR Data, 2018; Teperek et al., 2018; York et al., 2018; Sapp Nelson, 2017; Stickel and Vandervalk, 2014; Soares, 2012</p>	<ul style="list-style-type: none"> <li>● Effective communication and collaboration between data stewards and the central research data support team.</li> <li>● Address the need for more granular disciplinary experts.</li> <li>● Coordination and collaboration across the organisation.</li> <li>● Pre-award stage data management statements considering resources and costs of data stewardship.</li> </ul>
<p>What kind of leadership skills data stewards need</p>	<p>York et al., 2018; Palmer, 2014; Steelworthy, 2014; Stickel &amp; Vandervalk, 2014</p>	<ul style="list-style-type: none"> <li>● Project management, sustainability planning, and liaison skills across diverse stakeholders.</li> <li>● Team management.</li> </ul>



		<ul style="list-style-type: none"> <li>● Collaboration and stakeholder engagement (eg. researchers, ethics boards, funder support, university leaders).</li> <li>● Business intelligence and organisation culture: customer insights, policy and strategy development ,and implementation, ability to leverage data for maximum value.</li> </ul>
Absence or unclearly defined data stewardship roles and profiles	Jetten et al., 2021; Wildgaard et al., 2020	<ul style="list-style-type: none"> <li>● Implement common competence profiles as a place to start: data steward as administrator, analyst, developer, policy support, research support, and agent of change.</li> <li>● Identify roles, tasks and responsibilities documenting, curating, and structuring data across the organisation.</li> <li>● Build on European frameworks for data stewardship.</li> </ul>
What kind of skills, tools, and environments do data stewards need to manage to ensure that data stewards are a sustainable profession and relevant in collaborations with researchers	Demchenko et al., 2021; Attwood et al., 2019; European Commission Expert Group on FAIR Data, 2018; Aiken, 2016; Rolando et al., 2013; Parham & Murray-Rust, 2011	<ul style="list-style-type: none"> <li>● Data management plans.</li> <li>● Write and maintain guidelines, resources, standards, and policies to properly care for their research data.</li> <li>● Metadata and documentation.</li> <li>● Conditions that apply to sharing data and where to publish and preserve.</li> </ul>

		<ul style="list-style-type: none"> <li>● Pre-award stage data management statements considering resources and costs of data stewardship.</li> </ul>
How to motivate students and researchers to use their time to learn RDM skills and adapt RDM as a natural part of their research work	Attwood et al., 2017; Sapp Nelson, 2017	<ul style="list-style-type: none"> <li>● Embed training early in academic career path.</li> <li>● Create stand-alone workshops and online resources at the point of need.</li> <li>● A concerted, worldwide response from all stakeholders to deliver a tangible action plan, with sustained investment, to transform university programmes and cultivate a new cadre of scientists.</li> </ul>
How to provide expert disciplinary support	Barker et al., 2021; Mons et al., 2017; Oliver, 2017; Aiken, 2016	<ul style="list-style-type: none"> <li>● Learn subject specific knowledge as well as IT skills on an expert level to be able to advise on ontologies and to exploit linked data, metadata, and machine readable syntax.</li> <li>● Understand the premise of and requirements to data management in the research lifecycle in the discipline.</li> <li>● Share knowledge and materials with other support professionals and stewards.</li> <li>● Coordinate disciplinary and cross-disciplinary initiatives.</li> </ul>

Gaps in researchers' skills		
<p>Knowledge or provision of guidelines, policies, standards, and agreed procedures in RDM for early career researchers including data management plans and funders' mandates</p>	<p>Krahe et al., 2020; Edmond &amp; TóthCzifra, 2018; Goben &amp; Griffin, 2019; Maienschein et al., 2019; Wiley &amp; Kerby, 2018; Cox &amp; Williamson, 2015; Akers &amp; Doty, 2013; Rolando et al., 2013</p>	<ul style="list-style-type: none"> <li>● Understand data management and sharing behaviours in the organisation</li> <li>● Develop localised intervention strategies for different contexts</li> <li>● Use Theoretical Domains Framework of behaviour change to investigate implementation problems</li> <li>● Guidelines that support reflective process that expose and tweaks existing behaviours</li> </ul>
<p>Knowledge about safe storage and preservation, including evaluating costs and procedures</p>	<p>Krahe et al., 2020; Lefebvre et al., 2020; Dijkers, 2019; Goben &amp; Griffin, 2019; Joo &amp; Peters, 2019; Koltay, 2017; Cox &amp; Williamson, 2015; Weller &amp; Monroe-Gulick, 2014; Akers &amp; Doty, 2013; Knight, 2013</p>	<ul style="list-style-type: none"> <li>● Provide information and support with regard to protecting the confidentiality and safety of data as well as sharing data ethically</li> <li>● Provide storage and preservation policies, procedures, and practical guides</li> <li>● Ensure that requirements are met for the information infrastructure to fit the needs of the organisation in order to ensure proper access, storage, and data recovery.</li> <li>● Connects long term storage and preservation to usefulness of data in</li> </ul>

		preparation for publication (immediate needs)
How to document research data including how to assess the quality of datasets	Rantasaari, 2021; Joo & Peters, 2019; Hardwicke & Ioannides 2018; Lefebvre et al., 2018; Whyte & Ashley, 2017; Johnson et al., 2016; Specht et al., 2015; Rolando et al., 2013; Williams, 2013; Wilson, 2013; Carlson, 2011	<ul style="list-style-type: none"> <li>● Standardise data management planning, including awareness of the intellectual property and agreements issues affecting data processing and sharing</li> <li>● Improve and standardise data documenting and describing, not only for the researcher themselves but especially for data preservation, sharing, and re-using</li> <li>● Explain quality controls and provide data quality assurance support throughout research project</li> </ul>
More information and guidance regarding the legal aspects of managing copyright and IPR, data processing, sharing, and reuse	Rantasaari, 2021; Krahe et al., 2020; Borgman, 2018; Thielen, 2017; Cox & Williamson, 2015; Akers & Doty, 2013; Knight, 2013; Parsons et al., 2013	<ul style="list-style-type: none"> <li>● Provide campus/organisation-wide collaboration in planning and implementing teaching on rights and responsibilities. Applying RDM competencies in different research settings requires multi-professional expertise by many specialists, eg. researchers, teachers, lawyers, data librarians, research IT professionals, biostatisticians, and repository specialists</li> </ul>

		<ul style="list-style-type: none"> <li>● Base teaching materials and guides on the data sharing process (knowledge), and knowing how, where, by whom and with whom an organisation/research project shares (physical skills)</li> <li>● Discuss how to share and protect the confidentiality of data and to safeguard intellectual property from being stolen or data being misinterpreted or misused (both beliefs about consequences)</li> </ul>
<p>How to collect, create, process, analyse, and visualise data to secure data quality, integrity, and usability. Which tool to use when.</p>	<p>Krahe et al., 2020; Joo &amp; Peters, 2019; Lefebvre et al., 2018; Weller &amp; Monroe-Gulick, 2014; Knight, 2013; Parsons et al., 2013; Anderson et al., 2007</p>	<ul style="list-style-type: none"> <li>● Provide a toolbox, including forms of consent, ethical, and legal legislation regarding data collection, including anonymisation and pseudonymisation.</li> <li>● Provide cases for quantitative, qualitative, and mixed-methods studies w.r.t. the organisation, methods, and discipline</li> <li>● Guidelines need to express a certain level of domain knowledge in the discipline</li> <li>● Build on relationships between support personnel and researchers that are well established. These might be the ideal place for support of data collection and analysis</li> </ul>

<p>Enhancing skills and encouraging trust in sharing data and using others' data.</p> <p>Note: The epistemic cultures can differ not only between disciplines, but also between researchers using different research methods and collecting, processing, and analysing different data types.</p>	<p>General application: Rantasaari, 2021; Edmond, 2020; Perrier et al 2020; Quarati &amp; Raffighelli, 2020; Joo &amp; Peters, 2019; Goben &amp; Griffin, 2019; Zenk Möltgen et al., 2018; Whyte &amp; Ashley, 2017; Specht et al., 2015; Kim 2014; Rolando et al., 2013</p> <p>Social sciences Joo &amp; Peters, 2019; ZenkMöltgen et al., 2018; Weller &amp; Monroe-Gulick, 2014; Akers &amp; Doty, 2013; Borgman, 2008</p> <p>Health Sciences Joo &amp; Peters, 2019; Akers &amp; Doty, 2013</p> <p>Natural Sciences and engineering: Joo &amp; Peters, 2019; Akers &amp; Doty, 2013; Wilson, 2013</p> <p>Humanities: Edmond, 2020; Seillier et al., 2017; Welle &amp; Monroe-Gulick, 2014</p>	<ul style="list-style-type: none"> <li>● Identify the effects disciplinary, methodological, and data type specific differences have on trust.</li> <li>● Discuss expectations to data quality, inc. the nuances around the conditions, context, or materials that are not normally recorded in documentation</li> <li>● Clear communication of journals data sharing policies</li> <li>● Improved DM skills and infrastructure</li> <li>● Skills and support in application of metadata and how to share data safely and securely</li> <li>● Improve university-based storage</li> <li>● Methods to receive more credit for their data via publications</li> <li>● Integrate physical and digital records/documentation to improve meaningfulness of data over time</li> </ul>
<p>Guidance in more technically oriented data management needs</p>	<p>Weller &amp; Monroe-Gulick, 2014; Knight, 2013; Parsons et al., 2013</p>	<ul style="list-style-type: none"> <li>● Assist with metadata; data cleaning; converting and integrating external</li> </ul>

		<p>data; software development, visualisation, and digitising sources</p> <ul style="list-style-type: none"> <li>● Guidance on data management plans and complying with journal sharing requirements, inc. sharing agreements and other licence, storage and security, hard disk archiving processes and data encryption</li> <li>● Ensure funded researchers consider data management from project start and take appropriate steps to preserve data following its completion, inc. database development, persistent identifiers, metadata application, secure data transfer, secure data destruction, institutional SQL services</li> </ul>
<p>Standardised procedures and practices with agreed roles to audit and control research data quality</p>	<p>Lefebvre et al., 2018</p>	<ul style="list-style-type: none"> <li>● Establish data quality management services</li> <li>● Use FAIR principles to guide RDM infrastructure development and practices to generate reliable, quality data</li> <li>● Perform quality checks on the data as it is filled in by multiple people</li> <li>● Support the technical quality and sustainability of open data</li> <li>● Specify tasks and responsibilities in data quality control</li> </ul>

<b>Gaps in our knowledge</b>		
We do not know much of the needs of post-doc researchers,, graduate, and undergraduate students	Goben & Griffin, 2019	<ul style="list-style-type: none"> <li>● Help with storage, sharing, and long-term preservation of data</li> <li>● Data visualisation and curation services</li> <li>● Judge the cost of DM in funding applications</li> <li>● Align data management activities with promotion and tenure requirements</li> </ul>
RDM needs of the researchers in HSS disciplines are under-studied compared to researchers' needs in STEM disciplines.	Goben & Griffin, 2019; Tóth-Czifra, 2019	<ul style="list-style-type: none"> <li>● Understand and resolve challenges of applying the FAIR principles in HSS</li> <li>● Further develop FAIR to ensure detailed description of HSS data</li> <li>● How to build well-s-structured, searchable database from heterogeneous sources and records</li> <li>● Discuss the cultural, social, legal, ethical, technical, and economic reasons conditioning the findability and reusability of HSS data</li> <li>● Support tracing provenance and IPR</li> </ul>



<p>We need more studies of the needs and practices in smaller research institutions like liberal arts colleges, small research institutes, and universities of applied sciences.</p>	<p>Goben &amp; Griffin, 2019</p>	<ul style="list-style-type: none"> <li>● Investigate if funds are prioritised in favour of curricular and teaching needs rather than research infrastructure</li> <li>● Provision of dedicated funding to provide sustained midscale research infrastructure development and maintenance support</li> <li>● Collaborate between institutions to pool resources such as disciplinary repositories and aggregators</li> </ul>
<p>How researchers apply RDM in their everyday research practices.</p>	<p>Perrier et al., 2017</p>	<ul style="list-style-type: none"> <li>● Identify contributors and barriers of sound RDM practices, and quality of the data deposited in repositories</li> <li>● Effectivise and reward time spent on RDM</li> <li>● Provide resources and skills in data sharing</li> <li>● Provide guidelines for what can be shared, how and what is not worth sharing even if it can be shared</li> </ul>
<p><b>The Governance of data stewardship in research organisations</b></p>		

<p>How to measure the success of data steward and DM implementation in an organisation.</p>	<p>Smith, 2021; ,Lefebvre et al., 2018; Peng, 2018; Peng et al., 2016</p>	<ul style="list-style-type: none"> <li>● Establish an enterprise data management function that is responsible for coordinating the various data management disciplines (data governance, master data management, metadata management, data quality, enterprise data architecture, etc.)</li> <li>● Establish a culture that embraces the enterprise nature of data stewardship</li> <li>● Conduct and implement maturity assessment of repository process and procedures, asset management and stewardship practices</li> </ul>
<p>How to formalise the data steward strategy</p>	<p>Lefebvre et al., 2018; Stickel &amp; Vandervalk, 2014; Brackett et al., 2003</p>	<ul style="list-style-type: none"> <li>● Align strategic goals with data collection, infrastructure, assessment of stewardship, and risk assessments</li> <li>● Implement data governance that formalises a set of data policies and procedures across the organisation to encompass the full life cycle of data, from acquisition to use and to disposal</li> <li>● Professionalise stewardship and evaluate data programme governance</li> </ul>
<p>Promote joint governance of data stewardship and build trust in DS efforts</p>	<p>Smith, 2021; Borgman, 2018</p>	<ul style="list-style-type: none"> <li>● Invest in training designed for data stewards (foundations of data management, concepts of data governance, task-focused training in</li> </ul>

		<p>data definitions, and activities related to data quality)</p> <ul style="list-style-type: none"> <li>• Develop effective principles for governing privacy and information security through dialogue with organisational stakeholders (eg faculty, administration and students) - learning is passed down through generations of stakeholders through joint governance processes</li> </ul>
<p>Collaborate on the provision of data stewardship.</p>	<p>Smith, 2021; Verheul et al., 2019; Sasone et al., 2019; Lefebvre et al., 2018 ;Johnsten et al., 2018; Gendron et al., 2015; Erway, 2012</p>	<ul style="list-style-type: none"> <li>• Work together with data collectors, creators, providers, security and privacy officers, library and ICT staff to embed data stewardship in the organisation and still cater to local requests</li> <li>• Promote harmonisation of models, artefacts, repositories, infrastructure, and services through collaboration</li> <li>• Consider cross-institutional staffing models</li> </ul>

Collaborate and coordinate on the provision of training materials for data stewardship

DSCC, 2022; Demchenko et al., 2021; Barker et al., 2021; Jetten, 2021; Wildgaard et al., 2020; Attwood et al., 2019; Stoy, 2019; Verheul et al., 2019; European Commission Expert Group on FAIR Data, 2018; Johnsten, 2018; Demchenko et al., 2017; Aiken, 2016; Peng et al., 2016; Parham & MurrayRust, 2011

- Identify competencies and career paths for data stewards. Training materials become more efficient as they capture as much context and description of the topic as possible and appreciate multicultural differences
- Collaborate on training “market places”
- Cooperate with EUA to facilitate the implementation of data steward curricula and FAIR principles in university curricula
- Combining competences from different competence areas and using them as building blocks can allow flexible job-profiles definition. This enables the derived job-profiles to be easily updated by a changing set of competences related to profiles without the need to restructure the entire profile
- Certify training programmes and qualification

<p>Implement organisational frameworks for data stewardship</p>	<p>Smith, 2021; Verheul et al., 2019; Peng et al., 2018; Demchenko et al., 2017; Dyché &amp; Polsky, 2016; Rolando et al., 2013</p>	<ul style="list-style-type: none"> <li>● Targeted and continuous investment in data stewardship will ensure trained data stewards</li> <li>● Create frameworks that connect data policy with business goals, capabilities and areas of responsibility</li> <li>● Nurture cross organisational knowledge sharing culture</li> <li>● Outline roles and functions for data stewards, e.g roles that focus on technical workflows, specific use of data in research, and roles that concern policy, data management planning, and general commitment and training.</li> <li>● Identify where data stewards can work generically and embedded</li> </ul>
---	---	---

## 5. Limitations

The report is based on a representative sample of previous literature and is not a systematic, comprehensive search. Unique procedures and challenges regarding the stewardship of physical materials, such as biological samples and archeological artefacts are not addressed in this report. We recommend further investigation to the needs of researchers working with physical data and how best data stewards can to support their needs.

## 6. Conclusions and Next Steps

The report has highlighted the many facets of data stewardship and the needs for stewardship skills and training materials arising from both researchers' and data stewards' skills needs. The data steward is responsible for supporting data producers in collecting, collating, analysing, storing, and publishing data. They also evaluate issues and problems with data, and are increasingly responsible for providing technical solutions for the data producer. These responsibilities require many different skills, both soft and hard skills. This report has identified needs for and gaps in data steward skill development to inform the development of capabilities in providing support for data producers. Training needs that data stewards face in supporting research at the general level include knowledge about how to ensure the quality of data and metadata, how best to share data supporting a publication, and how to overcome cultural or technical barriers to sharing data. Particularly soft, perhaps mediation skills are needed to be able to engage in the political and sustainable infrastructure development work necessary to support best practices, improve incentives for data producers to manage data, and create synergies across an organisation to ensure collaboration between different stakeholders in the research data infrastructure.

Gaps in skills that data stewards need training for are how to implement FAIR in practice, in particular engaging researchers with the FAIR/EOSC ecosystem, securing reusability of data, and managing automated sharing practices. Further, the report points to gaps in training materials on how to conduct needs assessment in organisations. Such a needs assessment helps data stewards to understand the complex data culture of the organisation and to shape their role and profile accordingly. Researchers need training in the significance and planning of RDM, which can include (amongst others) safe and secure storing; documentation and description of the data life cycle; IPR, contract, and ethical aspects of data processing; analysing, visualising, converting, cleaning, and merging of data; controlling the quality of data; and the methods of sharing, preserving, discovering and reusing data.

In this report, we have identified many potential solutions and areas for future work in supporting data steward skills, training, and curricula development. However, there are still gaps in our knowledge with regard to the required skills and data stewardship in the non-STEM disciplines. There is a shortage of research-based knowledge on the RDM needs and practices concerning different research personnel groups such as post-doc researchers, graduate, and under-graduate students; the impact of different disciplines, research

methods, and data types on RDM needs; data stewards' and researchers' RDM needs in research institutions other than big research-intensive universities; and the needs in concrete, real everyday research work, i.e. in terms of how researchers apply RDM in their everyday research practice and what kind of barriers they face. Even more so, it appears from the literature that the role of the data steward seems to have an increasingly political profile in an organisation, which could perhaps be regarded as a further extension of the policy steward profile described in section 3.1.4. The solutions we presented in Section 4 point to the steward needing skills in measuring and validating the success or failure of their work in the organisation, aligning stewardship with strategic and business goals, promoting joint governance of data with diverse organisational stakeholders, and harmonising services through collaboration on a technical and human level.

Ultimately, in the development of new training materials we recommend the following:

1. Use international fora as vehicles to provide certified training. Coordinated and accredited training will ensure implementation and the adoption of best practices and standards. In addition, this approach will also provide organisations, whatever their size and capacity, with access to shared resources. Further disciplinary differences in data steward support can be made visible and hence provided for.
2. The key to getting researchers to invest in data sharing and actively embrace FAIR data is motivating them to and teaching them how to share data appropriately, safely and securely within an organisation or project as well as externally with reasonable effort. Training materials should address the ethics, the societal and practical benefits of data sharing as well as the technical aspects. Such materials could be based on a gap analysis of researchers' practices, which firstly investigates barriers to sharing, secondly reveals how researchers currently share and would like to share, and thirdly investigates the support they need from funders and publishers regarding sharing, the infrastructure they need to share and protect the confidentiality of data, and how to safeguard intellectual property from being stolen or data from being misinterpreted or misused (both common beliefs about risks of data sharing). Thereafter, materials could be developed of the models to include data sharing into day-to-day workflows, including the creation of metadata, depositing of data in repositories, and "automating" the process.
3. Future data steward competencies border on and partly overlap with computer science technical competencies. The future of data stewardship is becoming more automated, hence the data steward community needs increased collaboration with computer scientists to define and build solutions. Yes, the data steward needs subject knowledge on an expert level to be able to advise on ontologies but technical skills are vital to be able to exploit linked data, metadata, and machine-readable syntax. Such a technical shift in stewardship to be able to support open and FAIR science also entails a redefinition of the data stewards' role and service within the organisation, where they encompass teaching, research and technical skills.

The report will be sent to the RDA Professionalising Data Stewardship (PDS) Interest Group and EOSC Task Force Data Stewardship Curricula and Career Paths to support their drive to devise a strategy for the development of new training materials for data stewards and prioritisation of the work. The gaps and solutions identified throughout the report and particularly in section 4 will inform the work of the RDA PDS Models IG Models Task who are writing a report entitled 'Models of Data Stewardship' based on a survey they ran in October to November 2021, and the RDA PDS IG Training Task Learning Pathways Subgroup which is hoping to use these findings in their work to develop learning pathways to data stewards depending on their individual requirements.

## 8. References

- Aiken, P. (2016). Experience: Succeeding at data management: BigCo attempts to leverage data. *ACM Journal of Data and Information Quality*, 7(1–2). <https://doi.org/10.1145/2893482>
- Aiken, P., Allen, M.D., Parker, B., and Mattia, A. (2007) Measuring data management practice maturity: A community's self-assessment. *Computer*, vol. 40, no. 4, pp. 42-50. <https://doi.org/10.1109/MC.2007.139>
- Akers, K. G., & Doty, J. (2013). Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation*, 8(2), 5–26. <https://doi.org/10.2218/ijdc.v8i2.263>
- Anderson, N. R., Lee, E. S., Brockenbrough, J. S., Minie, M. E., Fuller, S., Brinkley, J., & Tarczy-Hornoch, P. (2007). Issues in biomedical research data management and analysis: Needs and barriers. *Journal of the American Medical Informatics Association*, 14(4), 478–488. <https://doi.org/10.1197/jamia.M2114>
- Aguinis H and Solarino AM (2017). Transparency and replicability in qualitative research: The case of interviews with elite informants. *Strat Mgmt J* 40:1291-1315
- Attwood, T., Blackford, S., Brazas, M.D., Davies, A., Schneider, V.A. (2019). A global perspective on evolving bioinformatics and data science training needs, *Briefings in Bioinformatics*, Volume 20, Issue 2, March 2019, Pages 398–404, <https://doi.org/10.1093/bib/bbx100>
- Barker, M., Manola, N., Gaillard, V., Kuchma, I., Lazzeri, E., Stoy, L., & Piera, J. (2021). *Digital skills for FAIR and open science*: Report from the EOSC Executive Board Skills and Training Working Group. European Commission.
- Berez-Kroeker, A. L., McDonnell, B., & Collister, L. B. (Eds.). (2022). The Open Handbook of Linguistic Data Management. In *The Open Handbook of Linguistic Data Management*. The MIT Press. <https://doi.org/10.7551/MITPRESS/12200.001.0001>
- Borgman, C. L. (2018). Open data, grey data, and stewardship: Universities at the privacy frontier. *Berkeley Technology Law Journal*, 33(2), 365. <https://doi.org/10.15779/Z38B56D489>
- Borgman, C. L. (2008). Data, disciplines, and scholarly publishing. *Learned Publishing*, 21(1), 29–38. <https://doi.org/10.1087/095315108X254476>



- Brackett, M.H. (2003). Business and IT roles for improving data resource quality. *Cutter IT Journal*. Volume 16, Issue 1, Pages 23 - 27
- Buys, C. M., & Shaw, P. L. (2015). Data management practices across an institution: Survey and report. *Journal of Librarianship and Scholarly Communication*, 3(2), 1–25.  
<https://doi.org/10.7710/2162-3309.1225>
- Callahan A, Anderson KD, Beattie MS, et al. Developing a data sharing community for spinal cord injury research. *Exp Neurol*. 2017;295:135–43
- Calzada Prado, J and Ángel Marzal, MI. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri* 63(2): 123-134.  
<http://dx.doi.org/10.1515/libri-2013-0010>
- Carlson, J., Fosmire, M., Miller, C. C., & Sapp Nelson, M. (2011). Determining data information literacy needs: A study of students and research faculty. *Portal: Libraries and the Academy*, 11(2), 629–657. <https://doi.org/10.1353/pla.2011.0022>
- Chiarelli, A., Loffreda, L., & Johnson, R. (2021). The art of publishing reproducible research outputs: Supporting emerging practices through cultural and technological innovation.  
<https://doi.org/10.5281/ZENODO.5521077>
- Cox, A., & Williamson, L. (2015). The 2014 DAF survey at the University of Sheffield. *International Journal of Digital Curation*, 10(1), 210–229. <https://doi.org/10.2218/ijdc.v10i1.362>
- Cruz, M., Imming, M., Shanahan, H., Teperek, M., van Gelder, C., & Whyte, A. (2019). Time to professionalise data stewardship. <https://doi.org/10.5281/zenodo.3420179>
- Demchenko, Y., Stoy, L., Engelhardt, C. & Gaillard, V. (2021). D7.3 FAIR Competence Framework for Higher Education (Data Stewardship Professional Competence Framework).  
<https://doi.org/10.5281/zenodo.4562088>
- Demchenko, Y., Belloum, A., and Wiktorski, T. (2017). Part 1. Data science competence framework (CF-DS). Data science competence framework (CF-DS) | Edison Project. Retrieved May 13, 2022, from <https://edison-project.eu/data-science-competence-framework-cf-ds/>
- Derudas, P., Dell'unto, N., Callieri, M., & Apel, J. (2021). Sharing Archaeological Knowledge: The Interactive Reporting System. *Journal of Field Archaeology*, 46(5), 303–315.  
<https://doi.org/10.1080/00934690.2021.1911132>
- Dijkers, M. P. (2019). A beginner's guide to data stewardship and data sharing. *Spinal Cord*, 57, 169–182. <https://doi.org/10.1038/s41393-018-0232-6>
- DSCC (Data Stewardship Competence Centers). GO FAIR. (2022, March 16). Retrieved May 2, 2022, from <https://www.go-fair.org/implementation-networks/overview/dscc/>
- Dyché, J., & Polsky, A. (2016). *5 models for data stewardship*. sas.com. Retrieved April 28, 2022, from [https://www.sas.com/th\\_th/whitepapers/5-models-for-data-stewardship-106846.html](https://www.sas.com/th_th/whitepapers/5-models-for-data-stewardship-106846.html)
- Edmond, J. (2020). Digital technology and the practices of humanities research. *Digital Technology and the Practices of Humanities Research*. <https://doi.org/10.11647/OBP.0192>
- Edmond, J., & Tóth-Czifra, E. (2018). *Open data for humanists, A pragmatic guide*.  
<https://doi.org/10.5281/ZENODO.2657248>

Erway, Ricky. 2012. You've got to walk before you can run: First steps for managing borndigital content received on physical media. Dublin, Ohio: OCLC Research.  
<http://www.oclc.org/research/publications/library/2012/2012-06.pdf>

Estrada-Galinanes, V., & Wac, K. (2018). *Visions and challenges in managing and preserving data to measure quality of life*. Retrieved February 14, 2022, from <https://www.healthit.gov/>

European commission expert group on FAIR data. (2018). Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data. *Directoriate-General for Research and Innovation*. Retrieved May 2 2022, from:  
[https://ec.europa.eu/info/sites/default/files/turning\\_fair\\_into\\_reality\\_1.pdf](https://ec.europa.eu/info/sites/default/files/turning_fair_into_reality_1.pdf)

European Data Protection Supervisor. (n.d.). Data Protection Officer (DPO).  
[https://edps.europa.eu/data-protection/data-protection/reference-library/data-protection-officer-dpo\\_en](https://edps.europa.eu/data-protection/data-protection/reference-library/data-protection-officer-dpo_en)

FAIR Data Austria (2021). Open Educational Resources Research Data Management.  
<https://fair-office.at/index.php/lernen-sie-mehr/?lang=en>

Gendron, H., Kirchner, J., Diaz, J., Fliss, S., Henry, G., Culshaw, J., and Cawthorne, J. E. (2015). The center of excellence model for information services (CLIR pub 163). *Library Staff Publications*. 3. [http://elischolar.library.yale.edu/yul\\_staff/3](http://elischolar.library.yale.edu/yul_staff/3)

Goben, A., & Griffin, T. (2019). In aggregate: Trends, needs, and opportunities from research data management surveys. *College and Research Libraries*, 80(7), 903–924.  
<https://doi.org/10.5860/crl.80.7.903>

Goudeseune, L., Eggermont, H., Groom, Q., Le Roux, X., Paleco, C., Roy, H.E., van Noordwijk, C.G.E. (2020). BiodivERsA citizen science toolkit for biodiversity scientists. BiodivERsA report, 44 pp. DOI: <https://doi.org/10.5281/zenodo.3979343>

Hardwicke, T. E., & Ioannidis, J. P. A. (2018). Populating the Data Ark: An attempt to retrieve, preserve, and liberate data from the most highly-cited psychology and psychiatry articles. *PLOS ONE*, 13(8), e0201856. <https://doi.org/10.1371/JOURNAL.PONE.0201856>

HLEG on EOSC. (2018). Final report and recommendations of the Commission 2nd High Level Expert Group on the European Open Science Cloud (EOSC). Prompting and EOSC in practice. Available online: <https://publications.europa.eu/en/publication-detail/-/publication/5253a1af-ee10-11e8-b690-01aa75ed71a1> (accessed: 18.07. 2019).

Jacobs, P., & Holland, C. (2007). Sharing Archaeological Data: The Distributed Archive Method on JSTOR. *Near Eastern Archaeology*, 70(4), 197–200.  
[https://www.jstor.org/stable/20361334?seq=1#metadata\\_info\\_tab\\_contents](https://www.jstor.org/stable/20361334?seq=1#metadata_info_tab_contents)

Jetten, M. Grootveld, M. Mordant, A. Jansen, M., Bloemers, M., Miedema, M. & van Gelder, C. W.G (2021). Professionalising data stewardship in the Netherlands. Competences, training and education. Dutch roadmap towards national implementation of FAIR data stewardship. Zenodo.  
<https://doi.org/10.5281/zenodo.4320504>

Jetten, M. (2019, October 24). *Professionalizing data stewardship in the RDA community*. RDA. Retrieved April 28, 2022, from <https://www.rd-alliance.org/professionalizing-data-stewardship-rda-community>

Johnston, L. R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R., Stewart, C., Blake, M., Herndon, J., McGeary, T. M., & Hull, E. (2018). Data curation network: A cross-

institutional staffing model for curating research data. *International Journal of Digital Curation*, 13(1), 125–140. <https://doi.org/10.2218/ijdc.v13i1.616>

Joo, S., & Peters, C. (2019). User needs assessment for research data services in a research university. *Journal of Librarianship and Information Science*, 1–14. <https://doi.org/10.1177/0961000619856073>

Kim, J. (2014). A study on research data management services of research university libraries in the U.S. *Journal of the Korean BIBLIA Society for Library and Information Science*, 25(3), 165–189. <https://doi.org/10.14699/KBIBLIA.2014.25.3.165>

Knight, G. (2013). *Research data management at London School of Hygiene and Tropical Medicine: Web survey report*. [https://researchonline.lshtm.ac.uk/id/eprint/2659425/1/Building%20a%20research%20data%20management%20service\\_GOLD%20VoR.pdf](https://researchonline.lshtm.ac.uk/id/eprint/2659425/1/Building%20a%20research%20data%20management%20service_GOLD%20VoR.pdf)

Koltay, T. (2017). Data literacy for researchers and data librarians. In *Journal of Librarianship and Information Science*. <https://doi.org/10.1177/0961000615616450>

Krahe, M. A., Wolski, M., Mickan, S., Toohey, J., Scuffham, P., & Reilly, S. (2020). Developing a strategy to improve data sharing in health research: A mixed-methods study to identify barriers and facilitators. *Health Information Management Journal*. <https://doi.org/10.1177/1833358320917207>

Kühn, E., and Streit, A. (2017). Interim report and catalogue of EOSC skills training and educational materials: Report from the EOSC Pilot. European Commission. [https://eoscipilot.eu/sites/default/files/eoscipilot-d7.2\\_0.pdf](https://eoscipilot.eu/sites/default/files/eoscipilot-d7.2_0.pdf)

Latif, A., Limani, F., Tochtermann, K. On the complexities of federating research data infrastructures. *Data Intelligence*, 3(1), 79-87 (2021). [https://doi.org/10.1162/dint\\_a\\_00080](https://doi.org/10.1162/dint_a_00080)

Lefebvre, A., Bakhtiari, B., & Spruit, M. (2020). Exploring research data management planning challenges in practice. *IT - Information Technology*, 62(1), 29–37. <https://doi.org/10.1515/itit-2019-0029>

Lefebvre, A., Schermerhorn, E., & Spruit, M. (2018, November 28). *How research data management can contribute to efficient and reliable science*. AIS Electronic Library (AISeL). Retrieved April 29, 2022, from [https://aisel.aisnet.org/ecis2018\\_rp/35/](https://aisel.aisnet.org/ecis2018_rp/35/)

Lucas, A. (2010). Corporate data quality management in context. In *Proceedings of the 15th International Conference on Information Quality*, Little Rock, AR (USA), 2010-11-13.

Maienschein, J., Parker, J. N., Laubichler, M., & Hackett, E. J. (2019). Data management and data sharing in science and technology studies. *Science Technology and Human Values*, 44(1), 143–160. <https://doi.org/10.1177/0162243918798906>

Mons, B. (2020). Invest 5% of research funds in ensuring data are reusable. *Nature*, 578(7796), 491–491. <https://doi.org/10.1038/d41586-020-00505-7>

Mons, B. (2018). Data stewardship for open science: Implementing FAIR principles. Chapman and Hall/CRC.

Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B., & Wilkinson, M. D. (2017). Cloudy, increasingly FAIR: Revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use*, 37(1), 49-56.

- NPOS/ELIXIR. (2019). Data steward competency hub. Retrieved April 28, 2022, from <https://competency.ebi.ac.uk/framework/datasteward/1.0>
- Oliver JC. (2017). Bioinformatic training needs at a health sciences campus. *PLoS ONE* 12(6): e0179581. <https://doi.org/10.1371/journal.pone.0179581>.
- Palmer, F. (2014). Analytics stewardship and the community of practice. *IBM Data Management Magazine*, 7.
- Partlo, K. "From data to the creation of meaning part II: Data librarian as translator," *IASSIST Quarterly* 38, no. 2 (2014)
- Parham, S. W., & Murray-Rust, C. (2011). Stewardship of research data. Retrieved 2 May 2022, from: <https://smartech.gatech.edu/bitstream/handle/1853/42091/STIC-dataCuration.pdf?sequence=1&isAllowed=y>
- Parsons, T., Grimshaw, S., & Williamson, L. (2013). *University of Nottingham research data management survey*. <http://eprints.nottingham.ac.uk/1893/>
- Peng, G. (2018). The state of assessing data stewardship maturity: An overview. *Data Science Journal*, 17: 7, pp. 1–12, <https://doi.org/10.5334/dsj-2018-007>
- Peng, G, Lawrimore, J, Toner, V, Lief, C, Baldwin, R, Ritchey, N A, Brinegar, D and Delgreco, S. A. (2016) Assessing stewardship maturity of the global Historical Climatology Network-Monthly (GHCN-M) Dataset: Use Case Study and Lessons Learned. *D.-Lib Magazine*, 22. DOI: <https://doi.org/10.1045/november2016-peng>
- Pergl, R., Hooft, R., Suchánek, M., Knaisl, V. and Slifka, J. (2019). "Data Stewardship Wizard": A tool bringing together researchers, data stewards, and data experts around data management planning. *Data Science Journal*, 18: 59, pp. 1–8. <https://doi.org/10.5334/dsj-2019-059>
- Perrier, L., Blondal, E., & MacDonald, H. (2020). The views, perspectives, and experiences of academic researchers with data sharing and reuse: A meta-synthesis. *PLoS ONE*, 15(2), 1–21. <https://doi.org/10.1371/journal.pone.0229182>
- Piorun, M., Kafel, D., Leger-Hornby, T., Najafi, S., Martin, E., Colombo, P., and LaPelle, N. (2012). Teaching research data management: An undergraduate/graduate curriculum." *Journal of eScience Librarianship* 1(1): e1003. <http://dx.doi.org/10.7191/jeslib.2012.1003>
- Pournaras, E. (2017). Cross-disciplinary higher education of data science—beyond the computer science student. *Data Science*, 1(1-2), 101-117.
- Quarati, A., & Raffaghelli, J. E. (2020). Do researchers use open research data? Exploring the relationships between usage trends and metadata quality across scientific disciplines from the Figshare case. *Journal of Information Science*. <https://doi.org/10.1177/0165551520961048>
- Qin, J., and D'Ignazio, J.. (2010). Lessons learned from a two-year experience in science data literacy education. *International Association of Scientific and Technological University Libraries*, 31st Annual Conference. Paper 5. <http://docs.lib.purdue.edu/iatul2010/conf/day2/5>
- Rantasaari, J. (2021). Doctoral students' educational needs in research data management: Perceived importance and current competencies. *International Journal of Digital Curation*, 16(1), 36. <https://doi.org/10.2218/IJDC.V16I1.684>

- Rickards, R. C., & Ritsert, R. (2012). Data governance challenges facing controllers. *International Journal of Business, Accounting, & Finance*, 6(1), 25–42.
- Rolando, L., Doty, C., Hagenmaier, W., Valk, A., & Parham, S. W. (2013). *Institutional readiness for data stewardship: Findings and recommendations from the research data assessment*. <https://smartech.gatech.edu/handle/1853/48188>
- Sansone, S. A., McQuilton, P., Rocca-Serra, P., Gonzalez-Beltran, A., Izzo, M., Lister, A. L., & Thurston, M. (2019). FAIRsharing as a community approach to standards, repositories and policies. *Nature biotechnology*, 37(4), 358-367.
- Sapp Nelson, M. (2017). A pilot competency matrix for data management skills: A step toward the development of systematic data information literacy programs. *Journal of ESience Librarianship*, 6(1), e1096. <https://doi.org/10.7191/jeslib.2017.1096>
- Schneider, R. (2013). *Research data literacy*. In Worldwide commonalities and challenges in information literacy research and practice, 134-140. Springer International Publishing. [http://dx.doi.org/10.1007/978-3-319-03919-0\\_16](http://dx.doi.org/10.1007/978-3-319-03919-0_16)
- Schumacher, J., & VandeCreek, D. (2015). Intellectual capital at risk: Data management practices and data loss by faculty members at five American universities. *International Journal of Digital Curation*, 10(2), 96–109. <https://doi.org/10.2218/ijdc.v10i2.321>
- Seillier, D., Baillot, A., Puren, M., & Riondet, C. (2017). *Survey on researchers requirements and practices towards cultural heritage institutions - Inria*. <https://hal.inria.fr/hal-0156286>
- Smith, A. M. (2021, March 29). *Foundations of data stewardship*. EWSolutions. Retrieved April 28, 2022, from <https://www.ewsolutions.com/foundations-data-stewardship/>
- Soares, S. (2013) Big data and the evolving role of governance. *IBM Data Management Magazine*, 1.
- Soares, S. (2012). Governance over critical data elements. *IBM Data Management Magazine*, 7.
- Specht, A., Guru, S., Houghton, L., Keniger, L., Driver, P., Ritchie, E. G., Lai, K., & Treloar, A. (2015). Data management challenges in analysis and synthesis in the ecosystem sciences. *Science of the Total Environment*, 534, 144–158. <https://doi.org/10.1016/j.scitotenv.2015.03.092>
- Steeleworthy, M. (2014). Research data management and the canadian academic library: An organizational consideration of data management and data stewardship. Partnership: The Canadian Journal of Library and Information Practice and Research, 9(1). Social Science Premium Collection. <https://search.proquest.com/docview/1700661330?accountid=13607>
- Stickel, J. R., & Vandervalk, A. (2014). Data business plans and governance programs: Aligning transportation data to agency strategic objectives. *Transportation Research Record*, 2460(1), 154-163.
- Stoy, L. (2019). *FAIRsFAIR focus group: University of Amsterdam*. FAIRsFAIR. Retrieved May 2, 2022, from <https://www.fairsfair.eu/articles-publications/fairsfair-focus-group-university-amsterdam>
- Tasovac, T., Romary, L., Toth-Czifra, E., Seillier, D., & Irena, M. (2021, February 9). *Heritage data reuse charter*. Heritage Data reuse charter [Blog post]. Retrieved May 2, 2022, from <https://datacharter.hypotheses.org/>



- Tasovac, T., Romary, L., Tóth-Czifra, E., & Marinski, I. (2021). Lexicographic data seal of compliance (Doctoral dissertation, ELEXIS; DARIAH).
- Teperek, M., Cruz, M. J., Verbakel, E., Böhmer, J., & Dunning, A. (2018). Data stewardship addressing disciplinary data management needs. *International Journal of Digital Curation*, 13 (1). <https://doi.org/10.2218/ijdc.v13i1.604>
- Thielen, J., & Hess, A. N. (2017). Advancing research data management in the social sciences: Implementing instruction for education graduate students into a doctoral curriculum. *Behavioral and Social Sciences Librarian*, 36(1), 16–30. <https://doi.org/10.1080/01639269.2017.1387739>
- Tóth-Czifra, E. (2019) The risk of losing thick description: Data management challenges Arts and Humanities face in the evolving FAIR data ecosystem: ffhalshs-02115505. <https://halshs.archives-ouvertes.fr/halshs-02115505/document>
- Verheul, Ingeborg, Imming, Melanie, Ringerma, Jacquelij, Mordant, Annemie, Ploeg, Jan-Lucas van der, & Pronk, Martine. (2019). Data stewardship on the map: A study of tasks and roles in Dutch research institutes. Zenodo. <https://doi.org/10.5281/zenodo.2669150>
- Weller, Katrin, & Watteler, Oliver. (2021, February 8). GDPR & ethical issues with social media data. Zenodo. <https://doi.org/10.5281/zenodo.4558694>
- Weller, T., & Monroe-Gulick, A. (2014). Understanding methodological and disciplinary differences in the data practices of academic researchers. *Library Hi Tech*, 32(3), 467–482. <https://doi.org/10.1108/LHT-02-2014-0021>
- Whyte, A., Ashley, K., Kalaitzi, V., Cavalli, V., Sipos, G., Kuehn, E., Thorat, R., & Vries, J. de. (2018, April 4). D7.3: Skills and capability framework. Retrieved April 28, 2022, from <https://eoscipilot.eu/content/d73-skills-and-capability-framework>
- Whyte, A., & Ashley, K. (2017). *D7.1: Skills landscape analysis and competence model*. [www.eoscipilot.eu](http://www.eoscipilot.eu)
- Wildgaard, L. E., Vlachos, E., Nondal, L., Larsen, A. V., & Svendsen, M. (2020). *National coordination of data steward education in Denmark: Final report to the National Forum for Research Data Management (DM Forum)*. (1 udg.) Danish e-Infrastructure Cooperation. <https://doi.org/10.5281/zenodo.3609516>
- Wiley, C. A., & Kerby, E. E. (2018). Managing research data: Graduate student and postdoctoral researcher perspectives. *Issues in Science and Technology Librarianship*, 89. <https://doi.org/https://doi.org/10.5062/F4FN14FJ>
- Williams, M., Bagwell, J., & Zozus, M. N. (2017). Data management plans: The missing perspective. *Journal of Biomedical Informatics*, 71, 130-142.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific data*, 3(1), 1-9.
- Wilson, J., Jeffrey, P., Patrick, M., Rumsey, S., & Jefferies, N. (2013). *Results of the 2012 University of Oxford research data management survey*. Mendeley. <https://doi.org/10.17632/7C9CP6KKMJ.1>
- York, J., Gutmann, M., & Berman, F. (2018). What do we know about the stewardship gap. *Data science journal*, 17. <http://doi.org/10.5334/dsj-2018-019>

Yu, F., Deuble, R., & Morgan, H. (2017). Designing research data management services based on the research lifecycle – a consultative leadership approach. *Journal of the Australian Library and Information Association*, 66(3), 287–298. <https://doi.org/10.1080/24750158.2017.1364835>

Zenk-Möltgen, W., Akdeniz, E., Katsanidou, A., Naßhoven, V., & Balaban, E. (2018). Factors influencing the data sharing behavior of researchers in sociology and political science. *Journal of Documentation*, 74(5), 1053–1073. <https://doi.org/10.1108/JD-09-2017-0126>

## Appendices

### Appendix 1: Books, Guides, Interest groups, Programs, Tools, Trainings, and Webinars on Data Stewardship

#### Books and reports

- Cross-disciplinary higher education of data science – beyond the computer science student. (Pournaras, Data Science, vol. 1, no. 1-2, pp. 101-117, 2017)  
“This paper illustrates experience and lessons learnt from the design and teaching of a novel cross-disciplinary data science course at a postgraduate level in a top-class university.”  
<https://doi.org/10.3233/DS-170005>
- The Data Librarian’s Handbook. (Rice, R. and Southall, J., 2016. Facet Publishing).  
“The Data Librarian's Handbook, written by two data librarians with over 30 years' combined experience, unpicks the everyday role of the data librarian and offers practical guidance on how to collect, curate and crunch data for economic, social and scientific purposes.
- Data Management for Researchers. (Briney, K. 2015. Pelagic Publishing).  
“A comprehensive guide to everything scientists need to know about data management, this book is essential for researchers who need to learn how to organize, document and take care of their own data.”
- “Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance (Plotkin, 2014, Elsevier Science).  
“Data Stewardship explains everything you need to know to successfully implement the stewardship portion of data governance, including how to organize, train, and work with data stewards, get high-quality business definitions and other metadata, and perform the day-to-day tasks using a minimum of the steward's time and effort.”
- Data Stewardship for Open Science: Implementing FAIR Principles. (Mons, B. 2018. CRC Press, Taylor & Francis Group).  
“The book has been written with the intention of making scientists, funders, and innovators in all disciplines and stages of their professional activities broadly aware of the need, complexity, and challenges associated with open science, modern science communication, and data stewardship. The FAIR principles are used as a guide throughout the text, and this book should leave experimentalists consciously incompetent about data stewardship and motivated to respect data stewards as representatives of a new profession, while possibly motivating others to consider a career in the field.”
- Delivering Research Data Management Services: Fundamentals of Good Practice.. (Pryor, G. et al., 2014, Facet Publishing).



“This groundbreaking guide will lead researchers, institutions and policy makers through the processes needed to set up and run effective institutional research data management services. This book will provide a step-by-step explanation of the components for an institutional service - effectively a 'how to guide'.”

- Engaging Researchers with Data Management: The Cookbook. (Clare et al., 2019, Open Book Publishers).  
“This book contains 24 RDM case studies, each describing an innovative activity used by a research institution to engage with its researchers about research data. These case studies, collected from research institutions worldwide, illustrate the diversity of feasible initiatives that could be implemented in other institutional settings.” Different Data Stewardship / Data Manager models on pages 90 to 110.  
<https://www.openbookpublishers.com/product/1080>
- Exploring Research Data Management. (Cox, A. and Verbaan, E. 2018. Facet Publishing).  
“This book will be useful reading for librarians and other support professionals who are interested in learning more about RDM and developing Research Data Services in their own institution. It will also be of value to students on librarianship, archives, and information management courses studying topics such as RDM, digital curation, data literacies and open science.
- Managing and Sharing Research Data. (Corti, L. et al., 2014, SAGE).  
“Written by experts from the UK Data Archive with over 20 years experience, this book gives post-graduate students, researchers and research support staff the data management skills required in today’s changing research environment.”
- The Objectives, Scope and Activities of a Possible GO TRAIN Implementation Network (Hodson et al., 2017).  
“This report is the outcome of a workshop hosted by CODATA at the International Council for Science in Paris on Friday 3 February 2017. The objective of the workshop was ‘to inform the activities and focus of a possible GO TRAIN Implementation Network (IN) as part of the wider GO FAIR initiative.’”  
<https://doi.org/10.5281/zenodo.1168503>
- “Open a GLAM Lab: Digital Cultural Heritage Innovation Labs. (Mahey, M., et al., 2019).  
“This book has been inspired by the International GLAM Labs Community, that was born in 2018 at the event on global 'Library Labs' held by the British Library. The event was attended by over 70 people from 43 institutions and 20 countries and followed up by a second global GLAM Labs meeting at the Royal Danish Library in Copenhagen in Spring 2019.” <https://doi.org/10.21428/16AC48EC.F54AF6AE>
- Towards FAIR data steward as profession for the lifesciences: Report of a ZonMw funded collaborative approach built on existing expertise.  
<https://doi.org/10.5281/zenodo.3474789>

## Guides

- 23 Things: Libraries for Research Data. (Witt, 2016).  
“An overview of practical, free, online resources and tools that you can begin using today to incorporate research data management into your practice of librarianship.” <https://dx.doi.org/dx.doi.org/10.15497/RDA00005>
- Arts and Humanities Research Council Data Management Plan Rubric. (Higman & Donaldson, 2018).  
“This rubric is designed as a checklist or marking aid for those reviewing data management plans for submission to the Arts and Humanities Research Council (AHRC). The Data Management Plan should outline the project’s approach to managing data.”  
<https://doi.org/10.5281/zenodo.1745532>
- Become a Data Steward: Step-by-Step Career Guide (BestAccreditColleges, 2021).  
“Data stewards manage the quality and security of an organization's information. They act as liaisons between users and information technology (IT) departments to ensure successful management of company data. This occupation is typically an office job with a regular schedule.”  
[https://study.com/articles/Become\\_a\\_Data\\_Steward\\_Step-by-Step\\_Career\\_Guide.html](https://study.com/articles/Become_a_Data_Steward_Step-by-Step_Career_Guide.html)
- DARIAH Pathfinder to Data Management Best Practices in the Humanities.  
“This resource list brings together tools, videos, short articles and other training materials that might be relevant when reflecting on your data management processes both in the immediate context of your research and in their broader disciplinary context.”  
<https://campus.dariah.eu/resource/dariah-pathfinder-to-data-management-best-practices-in-the-humanities>
- Data Governance & Data Steward Certification: Data related certifications and why do you need certification (Seiner, 2015)  
<https://www.slideshare.net/Dataversity/data-governance-data-steward-certification>
- Finnish DMP evaluation guidance (Tuuli working group, 2021).  
“This guide gives some general tips for evaluators. It can be used when evaluating DMP by students, peer reviewing or when evaluation is conducted by a data steward. The working group hopes you develop the guidance further in order to meet your specific needs and policies.”  
<https://doi.org/10.5281/zenodo.4729831>
- How to identify and assess Research Data Management (RDM) costs. (OpenAIRE, 2022).  
“In this guide you can find a tool listing, explaining and estimating the cost of possible expenses of data management. Estimates for quantifying amounts are only indicative

of the order of magnitude."

<https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs>

- Open Data for Humanists, A Pragmatic Guide.

<https://zenodo.org/record/2657248#.YkPrTjdBy-o>

In the arts and humanities, digital data production is still expensive, challenging and time-consuming. We all know this, and yet the results of these processes often in the end can't be reused by other researchers, meaning that we reinvent (or redigitise) the wheel far too often. This resource is aimed at giving practical advice for arts and humanities scholars who are willing to take their first steps in research data management but don't know where to begin. Our approach to data management views it as a reflective process that exposes and tweaks existing behaviours, rather than one that introduces specific tools. It is intended to encourage awareness of one's own processes and mindfulness about how they could be more open and how and how small changes across three points in your research workflow can make big differences.

- Planning to meet the costs of managing research data to be FAIR (Whyte et al., 2021).

"Training resources to adapt to your institutional context, helping researchers to do the following:

- Understand why they should budget for the costs of making data FAIR, and keeping it FAIR, and include these costs in grant applications
- Appreciate the benefits that services may provide to justify their costs
- Know about the different kinds of data management costs, including costs that funding bodies may allow to be charged to projects
- Apply a costing guide to help budget for the costs that may arise in preparing data to be FAIR
- Share experiences and expectations about costing the preparation of FAIR data"

<https://doi.org/10.5281/zenodo.4518900>

## Interest groups

- Data Stewards Interest Group.

"Providing a platform for data stewards and like-minded in the Netherlands (and abroad) to share experiences."

<https://www.dtls.nl/about/community/interest-groups/data-stewards-interest-group/>

## Programs

- Certified Data Steward (CDS) Program. (eLearningCurve).

"Your organisation cannot continue to waste IT spend on systems and projects that are hamstrung by poor quality data."

[https://ecm.elearningcurve.com/Certified\\_Data\\_Steward\\_Program\\_CDS\\_s/136.htm](https://ecm.elearningcurve.com/Certified_Data_Steward_Program_CDS_s/136.htm)

- Information Culture and Data Stewardship: Master of Library and Information Science Online. (University of Pittsburgh School of Computing and Information).  
<https://www.icds.pitt.edu/degree-programs/master-of-library-and-information-science-online-mlisonline/>

## Tools

- Data Stewardship Wizard.  
“Create Smart Data Management Plans for FAIR Open Science.”  
<https://ds-wizard.org/>
- LEGO® Metadata for Reproducibility game pack (LEGO, 2019).  
“The LEGO® Metadata for Reproducibility game is an interactive game for 424 players, using LEGO® to help researchers explore the metadata they might need to record to aid reproducibility. The game addresses issues including planning for metadata, formats of metadata recording, standards and automation.”  
<http://eprints.gla.ac.uk/196477/>
- Life sciences data steward function matrix (Scholtens et al., 2019)  
“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.  
<https://zenodo.org/record/2561723>
- Parthenos data management plan template - draft. (the PARTHENOS)  
“The PARTHENOS DMP builds on the Horizon2020 DMP template, enriched and tailored with specification from the humanities.”  
[https://www.rd-alliance.org/sites/default/files/attachment/PARTHENOS%20DMP\\_draft.pdf](https://www.rd-alliance.org/sites/default/files/attachment/PARTHENOS%20DMP_draft.pdf)
- Plan and follow your data. (Argos)  
“Create machine actionable DMPs.  
Configure to best fit your discipline.  
Link to EOSC components out of the box.  
Share easily in your repository.”  
<https://argos.openaire.eu/splash/index.html>
- Professionalising data stewardship in the Netherlands. (Jetten et al., 2021).  
Competences, training and education. Dutch roadmap towards national implementation of FAIR data stewardship. Fairly extensive focus on the role of a data

steward including function descriptions, competencies, training provisions.

<https://doi.org/10.5281/zenodo.4623713>

- Research Data Management Competencies Self-Assessment (Davis, M. et al., 2021).  
“A tool for assessing the competencies of individuals who support Research Data Management (RDM). The tool was developed to help academic libraries bolster skills and services surrounding RDM. This assessment allows the library to better understand and visualize the strengths and gaps in knowledge necessary to effectively run an RDM team.”  
<https://doi.org/10.18122/dataservices.8.boisestate>
- SSHOC: Training.  
“As we build the SSH area of the European Open Science Cloud, a key focus for SSHOC is providing training, advice, and educational resources for producers, users, and curators of Social Sciences and Humanities data. Our training activities are led by a team of expert trainers cherry picked from partner organisations recognised for their world-class know-how in the discovery, use, and management of research data. Browse the overview of our activities on this page and follow the links to detailed information.”  
<https://sshopencloud.eu/training>
- The SSH Training Discovery Toolkit.  
“The SSH Training Discovery Toolkit is a resource for researchers, service providers, data stewards, and trainers in the Social Sciences and Humanities, and provides a vast inventory of educational materials across a range of topics including research data management, FAIR data, Open Science, programming and didactics.”  
[https://training-toolkit.sshopencloud.eu/entities?search=&f%5B0%5D=collections%3ATraining%20Discovery%20Toolkit&f%5B1%5D=content\\_type%3Asource](https://training-toolkit.sshopencloud.eu/entities?search=&f%5B0%5D=collections%3ATraining%20Discovery%20Toolkit&f%5B1%5D=content_type%3Asource)

## Trainings

- Carpentries-based workshop (FAIR, GIT, PYTHON, R): FAIR data and software (2018).  
“This workshop focussed on the application of the FAIR Principles on scientific data and software. Because it covered a variety of examples, it did require a basic knowledge of the tools listed in the schedule. If you are interested in learning these basics, please consider applying for one of the (non-experimental) Software and Data Carpentry workshops, or work through their material in a self-paced manner.”  
<https://tibhannover.github.io/2018-07-09-FAIR-Data-and-Software/>
- CESSDA Data Management Expert Guide training toolkit.  
“The Data Management Expert Guide is designed by European experts from the CESSDA Training Working Group to help social science researchers make their

research data Findable, Accessible, Interoperable and Reusable (FAIR).”

<https://training-toolkit.sshopencloud.eu/source/43>

- Collibra data steward learning path for business steward and data custodian:  
[https://university.collibra.com/learn/public/learning\\_plan/view/46/collibra-data-steward-learning-path](https://university.collibra.com/learn/public/learning_plan/view/46/collibra-data-steward-learning-path)
- Credit bearing courses on data analytics:  
<https://catalog.canisius.edu/graduate/courses/dat/>
- Data analysis & stewardship.  
"These courses offer you the chance to become industry-ready by providing hands-on lessons in four major pillars of data analysis and data stewardship: FAIR data stewardship, statistics, omics data analysis, and machine learning. The types of training available must be tailored to the needs of the surrounding companies. The volume of information and data will only increase as more developments are made in life sciences, a clear indication of the importance of data analysis and stewardship for companies."  
<https://helisacademy.com/en/data-analysis-stewardship>
- Data stewardship Training.  
"The DTL/ELIXIR-NL community identifies data training needs and develops new data-related courses for the life science research community: FAIR Data Training; Bring Your Own Data Workshops (BYODs); Training in Research Data Management; Introduction to FAIR Data Stewardship".  
<https://www.dtls.nl/training-and-education/data-related-training/>
- Datatree - Data Training Engaging End-users by Natural Environment Research Council (NERC).  
"A free online course with all you need to know for research data management, along with ways to engage and share data with business, policymakers, media and the wider public. The self-paced training course will take 15 to 20 hours to complete in eight structured modules. The course is packed with video, quizzes and real-life examples of data management, along with valuable tips from experts in data management, data sharing and science communication. The training course materials will be available for structured learning, but also to dip into for immediate problem solving."  
<https://datatree.org.uk/course/>
- Introduction to the Course Data Science essentials. Part 1 and Part 2. From Open Data to Open Innovation (Sempreviva, 2019).  
"This is the introductory lecture to the course "Data science essentials for the energy sector: the case of wind energy: Harness the power of data in the digitalized energy sector. The course was held from 8th to 12th of April 2019 at DTU RISØ Campus, Denmark. The course is designed for future energy domain professionals, that wish to be equipped with data science essentials and tools to address the challenges of the renewable energy transition to renewables. The course aims at giving the

competences and knowledge to identify the opportunities hidden in the big data and trigger innovative thinking to make significant, strategic changes that minimize costs and maximize efficiency outcomes and values."

<https://doi.org/10.5281/zenodo.3580532>

- Training 4 Data Support.  
"Essentials 4 Data Support is an introductory course for those people who (want to) support researchers in storing, managing, archiving and sharing their research data. Essentials 4 Data Support is a product of Research Data Netherlands."  
<https://datasupport.researchdata.nl/en/>

## Webinars

- Data Management Plans for Humanities and Social Sciences.  
"This workshop will begin with an overview of RDM, and will then delve into a Data Management Plan template specifically geared toward humanities and social sciences research. We'll walk through the DMP and discuss questions that arise in these research fields, and will wrap-up by looking at some resources and tools to help support the management of your data."  
<https://hss-series.netlify.app/dmp/>
- Meeting funders' requirements - archiving and data sharing.  
"Aims to raise awareness about relevant key data management practices for sharing, specifically regarding data documentation, gaining consent, and data anonymisation. Addressing each of these three topics, it provides a short theoretical introduction, including what FAIR means and how it is implemented, as well as practical illustrations drawing on a large-scale cross-national survey (the European Social Survey). It also provides some practical tips with respect to data archiving, in particular how to choose an appropriate archive or repository."  
<https://www.youtube.com/watch?v=vgOsDudNTRQ>

## Appendix 2: Tag and Crosswalk matrix

The matrix read-only. There are three sheets in the appendix: an about sheet, the terminology crosswalk and the sheet documenting the literature-screening process.

[https://docs.google.com/spreadsheets/d/1xL9yIPi-U90q3GYP6mRIGDn1\\_UuJR9bt7gqJ3KI\\_wyw/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1xL9yIPi-U90q3GYP6mRIGDn1_UuJR9bt7gqJ3KI_wyw/edit?usp=sharing)