

Data as a New Research Publication Type: What could be the Role of Research Libraries as Service Providers?

Mari Elisa Kuusniemi

Helsinki University Library, University of Helsinki, Helsinki, Finland, mari.elisa.kuusniemi@helsinki.fi, <https://orcid.org/0000-0002-7675-287X>

Susanna Nykyri

Tampere University Library, Tampere University, Tampere, Finland, susanna.nykyri@tuni.fi, <https://orcid.org/0000-0002-5018-5176>

Abstract

This article examines the evolving role of research libraries in supporting the recognition of datasets as legitimate academic outputs through data publishing. Although the academic community increasingly acknowledges the value of treating research data as standalone contributions, there remains a lack of comprehensive frameworks and services to support this shift. Research libraries are well-positioned to lead in data curation and publication by collaborating with researchers, institutions, and other stakeholders.

Using a qualitative, multi-method approach—including a literature review, an exploratory survey of university libraries in the Nordic and Baltic countries, and professional experience—we investigate current practices, challenges, and institutional perspectives on data publishing. Our findings highlight inconsistent terminology in data policies and evolving services for data appraisal and visibility. We differentiate data publishing from general data sharing, emphasizing critical aspects such as data citability, quality control, and ethical reuse.

The article discusses various publishing pathways—such as data journals, repositories, and article supplements—and their respective implications. We identify key service gaps in libraries, particularly in data evaluation and discoverability, and propose strategies for libraries to promote data journals and

domain-specific repositories. Ultimately, we advocate for libraries to expand their role by developing integrated services for data appraisal, curation, and preservation, and by strengthening staff competencies in data management. Such efforts are essential for increasing the visibility, credibility, and scholarly impact of research data.

This paper is a continuation to a presentation provided in Liber Conference 2022. The presentation paper was acknowledged with the Innovation Award.

Keywords: Data publishing; Data publications; Research libraries; Service models; Research support

1. Introduction

Research libraries could play a key role in the process of curating and refining datasets into formal data publications (Kuusniemi & Nykyri, 2022). For over a decade, there has been ongoing discussion about the benefits of recognizing research data as a research output—one that can also serve as a merit to its contributors (see e.g. DORA;¹ Lawrence et al., 2011). However, a solid foundation to support such practices and collect related information is still lacking. Establishing the necessary conditions for producing data publications—and for gathering and analysing information based on them—requires the development of new types of practices and services. In this paper, we examine the current role of research libraries by analysing which services are already commonly offered, which are not, and how their role in data publishing could be expanded.

We argue that libraries are well-positioned to take on and coordinate this responsibility, provided they do so in close and well-defined collaboration with other stakeholders. As a foundation, we present research data as a form of academic publication—one that involves both curation and peer review.

LIBER has an open science roadmap, which recommends the following open science activities to libraries (Ayris et al., 2018):

- Provide a certified data repository.
- Create a data catalogue.
- Publish content with a machine-readable licence.

- Use open APIs to provide access to library services.
- Develop intelligent tools to automate metadata production and support FAIR data management during the entire data life-cycle.

These recommendations were made already in 2018. The goal was that libraries would provide these services by 2022. Not all of these are research data publishing services, but most are related to data publishing in one way or another. The topics presented in the roadmap have been further studied and discussed e.g. from the perspective of libraries' roles and how to improve the quality, integrity, reliability and reproducibility of research (Rantasaari, 2022; Schmidt et al., 2023).

The purpose of the article is to define the role of research libraries in supporting research data publishing. As a result, we propose concrete objectives and potential service models for research libraries to support data publishing. We illustrate various solutions like institutional repositories, long-term preservation and archiving practices, and the establishment of data journals. This study adopts a primarily descriptive approach, although it also offers some recommendations derived from the findings.

2. Research Material & Methods

In its nature the study is qualitative. It employs theoretical sampling, where cases are selected based on their potential to inform the research questions rather than through randomization (Seale, 1998, p. 329). This approach aligns with Strauss and Corbin's (1990, pp. 176–179) definition, where sampling is guided by concepts of proven theoretical relevance—those that recur or are notably absent across data and that evolve into categories through coding. Theoretical sampling is cumulative, deepens focus, and supports the identification of variation and process. It is typically used in qualitative research and continues until data saturation is reached.

Additionally, the study employs triangulation to enhance the validity of its findings. In line with Denzin's typology (1978, as cited in Janesick, 2000, p. 391), two of the four recognised forms of triangulation are utilised: data triangulation and methodological triangulation. Neither theoretical nor investigator triangulation is applied in this context. These approaches are employed in a complementary manner, as no single form of triangulation would be

sufficient in isolation. However, when combined, they serve to reinforce the robustness and credibility of the study's conclusions.

This study applies a qualitative, multi-method approach to examine the phenomenon of research data publishing and the associated support services from the perspective of university libraries. The investigation draws on three complementary sources.

- 1) First, a literature review was conducted to map existing research and policy developments related to research data publishing.
- 2) Second, an exploratory survey was carried out among university libraries in the Nordic and Baltic countries to gather insights into current practices and experiences (see Kuusniemi & Nykyri, 2025a, Kuusniemi & Nykyri, 2025b). The survey served to identify prevailing trends, challenges, and service models within this specific regional context.
- 3) Third, the study includes a qualitative analysis of institutional data policies using ATLAS.ti, a software tool designed for the management and qualitative analysis of textual data. This analysis focused on the terminology employed by respondents' organisations, with particular attention to language concerning data openness, sharing, and publishing.

In addition to these empirical components, the study incorporates the authors' extensive professional experience in the field. This experiential knowledge provides valuable contextual depth and informs the interpretation of findings.

Taken together, these methods enable a broad yet grounded analysis of the evolving service landscape. Particular emphasis is placed on the conceptual frameworks and institutional responsibilities that underpin support for research data publication. The study thus contributes to a deeper understanding of how university libraries navigate and shape the complex terrain of open research data practices.

The survey was disseminated primarily through communication channels targeted at research data management professionals. These included the DataCite user community mailing list in Estonia, the Dataverse network in Norway, the DataSupport chat and DMP consortium mailing list in Finland,

and the Swedish National Data Service mailing list in Sweden. In addition, the survey was circulated via mailing lists intended for research libraries, such as the library mailing list in Denmark and the Lithuanian Research Library Association mailing list in Lithuania. The LIBER newsletter was also used to promote the survey more broadly.

Responses were specifically asked from university libraries across the Nordic and Baltic countries. The survey remained open from November 2024 to February 2025 and yielded a total of 24 responses. All Nordic countries were represented among the respondents. However, participation from the Baltic countries was limited: while some responses were received from Estonia, no responses were obtained from Latvian libraries, despite the invitation being distributed via the LIBER mailing list. Latvia was the only country for which a suitable, targeted mailing list could not be identified.

The overall number of responses from university libraries was relatively low in comparison to the total number of universities in the region. Calculating an exact response rate proved challenging due to variations in how universities are defined across countries. For instance, Finland has 13 universities, Norway 27, and Estonia hosts 15 public and 8 private institutions of higher education. Given these definitional differences, a precise response rate was not calculated. Nonetheless, it is estimated that the response rate remained below 20%.

Table 1 presents the number of responses received from each country. All responses originated from distinct libraries or, at the very least, from separate organisational branches. In Denmark, for instance, several responses were submitted by units operating under the Royal Danish Library; however, each

Table 1: *Number of responses per country.*

Country	Responses
Denmark	6
Estonia	2
Finland	6
Lithuania	1
Norway	5
Sweden	4
Total	24

referred to different data policies and institutional websites. Consequently, these have been treated as separate entities for the purposes of this study.

Respondents were asked to describe their institution's data policy with regard to research data publication. Their responses were subsequently imported into ATLAS.ti for detailed qualitative analysis. In addition, respondents were invited to provide links to their organisation's official data policy documents. These documents were reviewed, and, where appropriate, supplemented with direct quotations from the survey responses to capture relevant statements concerning data openness, sharing, and publication. In cases where the data policy was not available in English—two such instances—Google Translate was used to produce working translations for the purposes of analysis.

This process resulted in a dataset that was analysed to identify various dimensions related to the publication of research data. The findings contribute to a deeper understanding of how universities conceptualise and articulate their approaches to data publication.

A selection of the data policy documents provided or referenced by respondents was imported into ATLAS.ti for systematic coding. The analysis focused on identifying recurring terminology, conceptual nuances, and variations in the ways key concepts—such as open data, data sharing, and data publication—were defined and operationalised. This approach enabled a more nuanced understanding of how institutions frame their responsibilities and expectations concerning research data.

The overarching aim of the qualitative text analysis was to explore whether universities regard research data as a form of scholarly publication, or whether the emphasis lies more strongly on principles of openness and data sharing.

3. Definition of Main Concepts

In order to engage in a more detailed discussion of data publishing, it is essential to first define the core concepts underpinning the topic. These include research data, data publication, data publishing, data journals, data peer review, and data curation.

3.1. Research Data

In 2023, the second edition of the Finnish National Policy and Executive Plan for Open Access to Research Data was published. This policy was co-authored by a broad coalition of stakeholders from the higher education and research community. Within the policy, research data is defined as follows:

“Research data is a resource used by a researcher or a research group during a research process, that is, the basic data of scientific or artistic research, in digital, analogue or physical form. Research data has been collected, observed, measured or created to confirm hypotheses and verify results.” (Open Science Coordination in Finland, 2023)

This definition positions research data as a broad and inclusive concept, encompassing not only datasets in the traditional sense but also research methods, software, research infrastructures, and related materials. Crucially, it is the context in which the data is used that determines its classification as research data. In other words, data becomes research data when it is employed within a research process to generate or validate scholarly or artistic knowledge.

3.2. Data Publishing

Data publishing refers to the formal process of making research data publicly available in a structured, accessible, and citable form, typically through a trusted data repository or a data journal. It involves the application of persistent identifiers (such as DOIs), standardised metadata, and documentation to ensure that the dataset can be discovered, accessed, reused, and cited by others.

Unlike informal data sharing, data publishing treats datasets as legitimate research outputs, often subject to quality control or peer review. It supports transparency, reproducibility, and the broader goals of open science by enabling long-term preservation and scholarly recognition of research data.

“Research data publishing is the release of research data, associated metadata, accompanying documentation, and software code (in cases where the raw data

have been processed or manipulated) for re-use and analysis in such a manner that they can be discovered on the Web and referred to in a unique and persistent way. Data publishing occurs via dedicated data repositories and/or (data) journals which ensure that the published research objects are well documented, curated, archived for the long term, interoperable, citable, quality assured and discoverable – all aspects of data publishing that are important for future reuse of data by third party end-users.” (Austin et al., 2017)

3.3. Data Publication

The National Library of Medicine (NLM) defines data publication:

“A data publication is an article in a journal that is a description of the data itself, rather than an analysis of that data or findings based on that analysis. This term may also refer to a journal that only publishes such articles. Note that while putting data in a repository can be described as publishing that data, since that data can now be cited, the term data publication refers to a full article describing the data. Further, data in data publications undergo a peer review process.”²

According to NLM, a data publication can be either an article discussing research data or a journal that publishes articles about research data. Their definition also includes the requirement of peer review process.

It is important to note that open data is not synonymous with data publication. While both concepts relate to the accessibility of research data, data publication entails a more formalised process that aligns with established scholarly practices. For instance, data publication involves elements such as persistent identifiers and mechanisms for citation—features that are not necessarily present in all instances of open data. As outlined in *Data citation: A guide to best practice* (Publications Office of the European Union, 2022), the citability of data is a key criterion distinguishing data publication from more general forms of data sharing.

The concept of data publication has been discussed in scholarly literature since at least 2006. Klump et al. (2006) proposed two fundamental prerequisites for data to be considered formally published:

- A) There should be a persistent identifier, so that the data set can be referred to in a persistent way.
- B) The data must be usable and of high quality.

These criteria underscore the notion that data publication is not merely about making data available, but about ensuring that it meets standards of discoverability, reliability, and scholarly utility. In this sense, data publication positions research data as a legitimate and citable research output, akin to traditional publications.

The concept of data publication has been examined more closely in an article published by Parsons & Fox (2013): “Is data publication the right metaphor?” The article comprehensively explores the complexity of the concept and considers alternative terms that may better describe the phenomenon. Terms mentioned as closely related or at least partially overlapping include data citation, data quality, data preservation, data archives and research infrastructures.

3.4. Data Journals

Data journals publish articles about existing datasets, databases or other kinds of data collections. Some journals also serve as a platform for publishing data, while some have articles that refer to data located elsewhere (e.g., in a domain-specific data repository).

3.5. Data Article

Data articles are a distinct form of scholarly publication dedicated to the detailed description of research datasets. They are also referred to as data papers, data descriptors (as used in *Scientific Data*)³, dataset articles, or metadata articles. While terminology may vary across journals, the underlying concept remains consistent.

As Callaghan et al. (2012) describe, a data paper “would describe the dataset, providing information on the what, where, why, how and who of the data. The data paper would contain a link back (a DOI) to the dataset in its

repository, and the journal publishers would not actually host the data. This means that even in situations where the data paper might be restricted access, the dataset could still be open.”

Similarly, Data in Brief⁴ (Elsevier) defines a data article as “a short description of research data that have been made publicly available through a repository that makes it easier to comprehend and reuse. It does not offer conclusions or interpretive insights. Data articles give scientists the opportunity to describe and share their raw data and hence participate in open science and satisfy funder requirements.”

Regardless of the term used, the purpose is the same. Unlike traditional research articles, data articles do not present hypotheses, results, or theoretical interpretations. Instead, they focus on providing comprehensive metadata, contextual information, and technical validation to support the reuse, reproducibility, and proper citation of the dataset.⁵

Typically, data articles include:

- A description of how the data were collected or generated;
- Information on data structure and format;
- Details on quality control and validation;
- Persistent identifiers (e.g. DOIs) linking to the dataset in a repository;
- Licensing and access conditions.

Data articles are peer-reviewed and citable, offering researchers formal recognition for their data contributions.

3.6. Data Peer Review

It is generally understood that research data requires a peer review process if it is to be considered comparable to traditional scholarly publications (Callaghan et al., 2012). However, peer review of research data differs in important ways from the peer review of conventional journal articles. In the context of data, the emphasis shifts towards aspects such as curation, archiving, and data quality (Parsons & Fox, 2013).

Notably, the terminology and practices surrounding data peer review remain unsettled. Despite growing recognition of the distinct nature of data review,

even recent formal definitions—such as those in Standard Terminology for Peer Review (National Information Standards Organization, 2023)—continue to focus primarily on traditional article-based peer review. In practice, data peer review varies considerably depending on the context, whether it occurs in relation to traditional journal articles, data papers, or open-access data repositories.

Mayernik et al. (2015) identify three key challenges in the development of effective data peer review: variability in data accessibility, the need for specialised expertise, and the balance between pre- and post-publication review. Addressing these challenges is essential for enhancing scientific integrity and ensuring that data publication processes scale effectively with increasing data volumes.

Peer review of research data may occur as part of the review process for traditional research articles, where publishers may ask reviewers to assess the underlying data (e.g. Springer Nature's research data policy). However, this practice remains relatively uncommon and, as noted by Pop and Salzberg (2015), supplementary data are rarely reviewed—even when explicitly included in reviewer guidelines.

Dedicated data journals, such as *Scientific Data* (Springer Nature), offer a more structured approach. In these venues, peer review not only assesses the clarity and completeness of the data description but also considers other aspects such as usability, accessibility, and compliance with metadata standards. As Carpenter⁶ notes, further guidance is emerging to support consistent and transparent peer review of data, though practices remain diverse.

Mayernik et al. (2015) further distinguish between four models of data publication—traditional journals, data journals, data repositories, and hybrid approaches—and examine how each contributes to data quality assurance. Despite differences in review mechanisms, all models share core requirements: data accessibility, adequate documentation, and clear reviewer guidance. Data journals, often in collaboration with repositories, provide the most formalised peer review structures. In contrast, traditional journals and repositories frequently lack consistent standards. Tools such as metadata validation, visualisation platforms, and statistical checks can support reviewers, though challenges persist, particularly with large or complex datasets.

To strengthen data peer review, Mayernik et al. recommend that scientific societies update author and reviewer guidelines, promote collaboration with repositories, and endorse established data citation principles. Broader adoption of practices such as integrated submission pipelines and rating systems could further enhance transparency and data quality. Looking ahead, key priorities include ensuring accessibility, engaging qualified reviewers, and balancing pre- and post-publication review in ways that scale with the growing volume and diversity of research data.

3.7. Data Curation

Data curation is a multifaceted concept that can be approached from at least two complementary, yet distinct, perspectives:

- 1) Lifecycle Management Approach: This perspective situates data curation within the broader research data lifecycle, encompassing activities from data creation and active use to long-term preservation and reuse. It emphasises continuity and stewardship throughout the data's existence.
- 2) Appraisal and Quality Assurance Approach: This view focuses on the selection, evaluation, and enhancement of research data to ensure its long-term value, usability, and readiness for publication. It highlights the importance of assessing data quality, relevance, and documentation.

Together, these perspectives illustrate that data curation is both a data management process and a strategic activity. It involves not only managing data over time but also making informed decisions about what data should be preserved, how it should be enhanced, and under what conditions it should be shared or published (Johnston et al., 2024; Lee & Stvilia, 2017).

Curated repositories for research data refer to carefully managed and quality-controlled collections, often organised around a specific discipline, theme, or data type. Unlike raw or minimally processed data repositories, curated archives involve active human oversight to ensure that datasets are accurate, well-documented, consistently formatted, and suitable for reuse. Curation typically includes the verification of metadata quality, the usability of file formats, the clarity of data structures, and the presence of persistent identifiers, licensing information, and citation guidance.

However, there is no single curated repository that accommodates all types of research data. Each service defines its own standards for curation, appraisal, and selection. Since the introduction of the FAIR principles (Wilkinson et al., 2016), many repositories have adopted these as benchmarks, particularly regarding machine readability and interoperability.

An illustrative example of a curated data archive is the Finnish Social Science Data Archive (FSD), which applies rigorous appraisal criteria focused on technical quality, documentation, and legal rights (see FSD, Operational guidelines)⁷. Another widely recognised example is RefSeq (Reference Sequence Database), which hosts curated DNA, RNA, and protein sequences. RefSeq employs a two-stage curation process: an initial phase of automated validation followed by expert-led quality assurance. This hybrid model is supported by a broad international network, enabling the rapid release of data while ensuring continuous improvement as new information becomes available. The curation status of each dataset is clearly communicated to users through standardised indicators, with explicit labels such as “Reviewed”, “Provisional”, “Model”, or “Predicted”. These labels help users assess the reliability and level of curation applied to each record.

4. Why Publish Research Data?

The publication of research data serves multiple purposes. According to the academic publisher Springer Nature:

“Your research data are valuable – without it, other researchers cannot learn from or build upon your work. Recognition cannot be attained if others are unable to locate and cite your data. Files stored on a desktop or USB device may contain knowledge of significant value to the wider research community.

Sharing research data facilitates:

- The advancement of others’ research based on your findings.
- Progress in your discipline and contributions to the public good.
- Compliance with funder or institutional mandates regarding data sharing.

Furthermore, it may enhance your research profile by enabling credit for data production and increasing the visibility of associated publications.”⁸

The Environmental Data Initiative (EDI), a U.S.-based organisation supporting ecological and environmental scientists in data stewardship, underscores that:

“Prior publication of data in a repository can help protect a data author from having their data misattributed. Researchers who publish data are more likely to be invited as co-authors on publications using the data and can increase research impact and citation rate.”⁹

It should be noted that making data openly accessible does not necessarily equate to its formal publication. The concept of publication implies adherence to scholarly standards, such as ensuring that data are citable (CODATA-ICSTI Task Group on Data Citation Standards and Practices, 2013).

Research funders and the broader scientific community increasingly regard research data as equivalent to traditional publications. This perspective is reflected in various frameworks, including the San Francisco Declaration on Research Assessment (DORA, 2012), which acknowledges the diversity of research outputs—ranging from articles and datasets to software and trained personnel. Funders and institutions are encouraged to assess the value and impact of all research outputs, including datasets and software. This approach has gained traction, particularly within European universities, which are currently developing responsible research assessment practices under the Coalition for Advancing Research Assessment (CoARA).¹⁰

Research infrastructures play a pivotal role in the generation and dissemination of research data. Their importance in data sharing is increasingly recognised. These infrastructures may either publish data directly or support researchers in doing so. In both scenarios, the foundational work carried out by the infrastructure significantly influences the quality of data publication. Consequently, recent recommendations and guidelines have been directed specifically at infrastructures. For example, the OECD & Science Europe publication Optimising the Operation and Use of National Research Infrastructures (2020), in collaboration with Science Europe, strongly advocates for the development of data management plans and the adoption of FAIR principles.

The growing prevalence of data publication is also driven by evolving publishing requirements. Increasingly, it is difficult to publish research articles without accompanying data. Many publishers, including Springer Nature

and Wiley, now mandate a “Data Availability Statement,” in which authors must indicate where the data can be accessed or justify any restrictions on sharing (e.g., third-party ownership). Data are also published alongside methods and code, reinforcing principles of good scientific practice and result verification.

From the perspective of open science, no single type of research output—whether method, code, data, or article—is considered superior. Rather, the comprehensive publication of all relevant outputs is encouraged. This principle is articulated in the UNESCO Recommendation on Open Science (Unesco, 2021), which advocates for openness and inclusivity across the research lifecycle.

In support of this, the European Commission conducted a comprehensive Cost-Benefit Analysis for FAIR Research Data (PwC EU Services, 2018), which estimated the economic and scientific losses associated with poor data management and the absence of FAIR practices. The report concluded that significant inefficiencies, such as duplicated research efforts, lost data, and reduced innovation potential, could be mitigated through structured data publication and adherence to FAIR principles.

5. The Many Faces of Data Sharing and Publishing

Research data can be shared openly through several overlapping approaches:

- as supplementary material accompanying a traditional text publication,
- as an independent dataset deposited in a trusted data repository,
- through a descriptive article published in a dedicated data journal, or
- by releasing descriptive metadata records while restricting access to the underlying data, typically due to ethical or legal constraints.

While all four approaches contribute to data openness, the first—sharing data as supplementary material—is generally not considered as robust or sustainable a method of data publishing as the latter three.

Sharing research data as **supplementary material** to scientific articles is a common method of enabling data access, and is supported by many research

journals. In this model, the data is included directly alongside the article, often in the form of additional tables, figures, or code. However, such supplementary materials do not always meet the criteria for formal data publication as outlined by Klump et al. (2006). Supplementary files often lack persistent identifiers, standardised metadata, and long-term preservation guarantees, which can limit their discoverability, usability, and scholarly value. In some cases, the data may not even be usable—for example, when it is provided as an image of a table rather than in a machine-readable format. According to Pop and Salzberg (2015), supplementary data are rarely subject to peer review, and in some cases, journals explicitly advise reviewers not to evaluate them.

Depositing data in a repository is a widely recognised method of publishing research data as an independent scholarly output. In this model, datasets are submitted to institutional, disciplinary, or general-purpose repositories, where they are assigned persistent identifiers (such as DOIs), accompanied by metadata, and made openly accessible or available under specified conditions. Repositories often support version control and citation, thereby enhancing the discoverability, usability, and scholarly value of the data. This approach aligns closely with the FAIR principles and is supported by funders and research institutions as part of open science practices. For example, Horizon Europe mandates that research data generated in funded projects must be managed in accordance with the FAIR principles and, where possible, deposited in trusted repositories. This is outlined in the Horizon Europe Model Grant Agreement and associated guidance documents.

Finding a suitable data repository is not always easy. A data repository is a storage space for researchers to deposit data sets associated with their research. The publication of research data in data repositories may seem simple at first glance, but they are actually complex and varied. The diversity is particularly evident in appraisal and selection processes, with some repositories offering self-publishing options with minimal selection, while others perform extensive curation and quality assurance, leading to more complex workflows. Discipline-specific repositories often adhere to specialised metadata schemas, while more general repositories tend to use domain-agnostic schemas like the mandatory DataCite fields. (Austin et al., 2017)

A range of tools, guidelines, and certification frameworks have been developed to support the identification and use of trustworthy data repositories.

These include registries such as re3data, guiding principles like FAIR, and certification schemes such as the CoreTrustSeal (CoreTrustSeal Standards & Certification Board, 2022), which assess repositories against criteria for long-term sustainability and data stewardship. The European Commission's Data Quality Guidelines (Publications Office of the European Union, 2021) outline best practices for ensuring that open data and metadata meet high standards of quality and FAIR compliance. Similarly, the TRUST Principles (Transparency, Responsibility, User focus, Sustainability, and Technology), as proposed by Lin et al. (2020), provide a conceptual framework for evaluating the trustworthiness of digital repositories.

From the perspective of data publishing, it is essential to prioritise services that support persistent identifiers and standardised metadata schemas. Libraries can play a key role in this process by promoting tools and checklists that assist researchers in selecting appropriate repositories. One such resource is the DCC checklist for evaluating data repositories (Whyte, 2016), which is specifically designed to support data support professionals in guiding repository selection.

Publishing in data journals involves submitting a data article—a peer-reviewed manuscript that provides a detailed description of a dataset, including its context, collection methods, structure, and potential for reuse. The dataset itself is not hosted by the journal. Instead, authors are typically required to deposit their data in a trusted data repository—either generalist (e.g. Zenodo, Figshare) or domain-specific (e.g. GenBank, ICPSR)—prior to submission. The data article then includes a persistent identifier (such as a DOI) linking directly to the dataset in the repository.

This integrated model ensures that the dataset is preserved and accessible via a stable platform, while the accompanying article offers rich metadata and methodological detail that goes beyond what is typically included in a repository record. It also provides researchers with academic credit through a citable, peer-reviewed publication. In this way, data journals and repositories operate in tandem: the repository ensures long-term access and technical stewardship of the data, while the journal provides scholarly recognition and contextual framing.

Recent research by McGillivray et al. (2022) highlights the value of this model, particularly in the humanities and social sciences. Their study

shows that data articles—often published following the deposition of datasets in repositories—not only enhance the visibility and reuse of the data itself, but also positively impact associated research publications. Moreover, targeted dissemination strategies, such as the use of social media hashtags, were found to significantly increase views and downloads, thereby advancing transparency and supporting the broader open research agenda.

Restricted-access data with open metadata is a model of data publishing in which the descriptive metadata of a dataset is made publicly available, while access to the underlying data is limited due to ethical, legal, or confidentiality concerns. This approach is particularly relevant for sensitive data, such as those involving personal, medical, or culturally protected information.

In this model, metadata records—often deposited in a trusted repository or a data catalogue—include key information about the dataset: its scope, structure, collection methods, and conditions for access. Although the data itself is not openly downloadable, the metadata ensures that the dataset is discoverable and citable, and provides clear instructions on how qualified researchers may request access (e.g. through a data access committee or research permit process).

This approach supports the FAIR principles—particularly Findability and Accessibility—while respecting privacy and legal constraints. It is commonly used in fields such as health sciences and social sciences.

6. Support for the Publication of Research Data

A range of services is provided by libraries and research institutions to support the publication and management of research data. These services typically include data repositories or archives, long-term preservation infrastructures, data catalogues, and various consultative offerings related to data journals, data articles, and domain-specific repositories. In this chapter, the nature and scope of these services are examined in greater detail. Table 2 summarises the services described below.

Table 2: Roles and services of research libraries on data publishing.

Service	Role of the library	Service models
Data repository	Make it possible to publish data in a data repository.	Create a institutional repository OR recommend to use a specific data repository.
Long-term preservation service	Make it possible to preserve valuable data sets in a long-term preservation archive. Provide services related to assessment of the value of research data or metadata.	Create an institutional service OR use national or international data archive.
Data catalogue	Collect metadata of published data sets.	Create an institutional data catalogue OR use national or international catalogue.
Support of data journals	Promote data journals and data papers.	Part of university press OR use outsourced OA journal platforms.
Support for domain specific repositories	Ensure support for domain specific repositories or data collections in your organisation.	Provide support by yourself OR find experts outside the organisation.

6.1. Data Repository and Data Archive Services

In the context of research data, the terms data repository and data archive are closely related and often used interchangeably. Formal definitions of these terms are absent from the scholarly literature, and discussions concerning their potential distinctions are found only in blog posts.^{11,12} The terms may carry slightly different connotations depending on the context and institutional practices.

A data repository typically refers to a platform or service where research data can be stored, managed, shared, and accessed. Repositories emphasise active use and discoverability, and they often support metadata standards, persistent identifiers (such as DOIs), and access control mechanisms. Examples of widely used repositories include Zenodo, Figshare, Dryad, and Harvard Dataverse.

In contrast, a data archive generally implies a stronger focus on long-term preservation, with an emphasis on ensuring the durability, integrity, and future accessibility of data. Archives may involve more rigorous curation processes and preservation strategies. Data archives provide a variety of benefits including the preservation, discovery, control, reuse and repurposing of

research data (Rieger, 2007). Notable examples include the UK Data Archive, ICPSR, and the CESSDA network of archives.

While the distinction between data repositories and archives is not always clear-cut, repositories are generally associated with enabling immediate access and reuse of research data, whereas archives are more closely aligned with its long-term stewardship and preservation. In our survey, however, the terms data archive and data repository were frequently used interchangeably to refer to repositories—that is, platforms designed for data sharing and citation—rather than to long-term preservation services, which will be discussed in the following section.

6.2. A Long-Term Preservation Service

A long-term data preservation service refers to an infrastructure or system designed to ensure the sustained accessibility, usability, and integrity of selected research data over extended periods—long beyond the duration of the original research project. These services typically implement recognised preservation standards, conduct regular integrity checks, apply format migration strategies, and maintain secure storage environments to protect data from technological obsolescence and degradation (Albani et al., 2020).

Importantly, not all data can or should be preserved indefinitely. Due to resource constraints and the need for curation, data selected for long-term preservation is typically assessed based on its scientific value, legal or ethical obligations, potential for reuse, and relevance to institutional or disciplinary priorities. This selection process is often guided by appraisal criteria and may involve collaboration between researchers, data stewards, and archivists. Such services are frequently provided by institutional or national archives and may be certified under frameworks such as the CoreTrustSeal, ensuring adherence to best practices in digital preservation (CoreTrustSeal Standards & Certification Board, 2022).

6.3. Data Repository vs. Long-Term Data Preservation Service

A data repository and a long-term preservation service serve distinct roles in research data management. A data repository focuses on the dissemination,

discoverability, and reuse of data over the short to medium term. It supports metadata standards and often allows versioning to facilitate open science and data citation. In contrast, a long-term preservation service ensures the integrity, authenticity, and accessibility of data over decades or even centuries. It employs preservation strategies such as format migration and checksums, and typically adheres to frameworks like the Open Archival Information System (OAIS)¹³. While repositories promote visibility and accessibility of research outputs, long-term preservation services safeguard the longevity and usability of data for future generations.

6.4. Data Catalogues

Data catalogues are structured, searchable indexes that provide standardised metadata descriptions of research datasets, thereby facilitating their discovery, citation, and potential reuse. Rather than storing the data itself, a data catalogue functions as a centralised metadata repository, aggregating information from multiple sources such as institutional repositories, disciplinary archives, and national data services. Typical metadata elements include dataset content, provenance, format, access conditions, and responsible parties (Labadie et al., 2020).

By consolidating metadata from diverse origins, general-purpose data catalogues enhance the visibility and governance of research data assets across institutional and disciplinary boundaries. They play a pivotal role in supporting the FAIR principles—particularly Findability and Accessibility—and are often maintained by national infrastructures, research consortia, or domain-specific services committed to advancing open science and responsible data stewardship.

In contrast, institutional data catalogues are developed and maintained by individual research organisations, such as universities or research institutes, to showcase and manage the datasets produced by their own researchers. These catalogues contribute to institutional research profiling by highlighting ongoing research activities and outputs. They may also reveal interdisciplinary connections by identifying thematic overlaps across departments or faculties. Improved awareness of existing datasets within the institution can promote internal data reuse and foster collaboration.

Institutional data catalogues may be implemented either by integrating dataset metadata into existing Current Research Information Systems (CRIS) or by developing dedicated platforms specifically designed for managing research data. However, building a data catalogue is not necessarily straightforward, and initial versions may fall short of expectations. As Rumsey and Jefferies (2013) observed in their account of developing a data catalogue for the University of Oxford, the process can be complex and iterative.

If research data is to be recognised as a valuable scholarly output, it must be made more visible within both institutional and broader research ecosystems. Establishing and maintaining a data catalogue—whether general or institutional—is a critical step towards achieving this goal. Moreover, once metadata is systematically collected, tools such as altmetrics can be employed to monitor and demonstrate the broader impact and reach of research data.

6.5. Supporting Data Journals

As this type of service is still largely in the conceptual stage in most libraries, there is currently little to no literature available on the subject. Nevertheless, we chose to include it in our study in order to explore the extent to which libraries are beginning to engage with data journal related services.

In principle, libraries can support researchers in publishing data articles in dedicated data journals. This support may include helping researchers identify trusted repositories, ensuring appropriate metadata and persistent identifiers, which can be linked to the data article. Libraries can also advise researchers on selecting suitable data journals and crafting data articles that meet both scholarly and technical standards. Additionally, they may offer guidance on licensing, ethical considerations, and compliance with open science policies.

Beyond their advisory role, many academic libraries are also active participants in scholarly publishing. Some operate as university presses or collaborate with institutional publishing platforms, and in some cases, they may even host or publish data journals themselves.

6.6. Supporting for Domain-Specific Data Repositories

Libraries are increasingly recognised as key actors in supporting the use, development, and sustainability of domain-specific data repositories. Their contributions may include guiding researchers in selecting appropriate repositories, assisting with metadata creation and compliance with disciplinary standards, and providing training on data deposition practices. Libraries may also collaborate with repository infrastructures to ensure interoperability with institutional systems, support compliance with funder and institutional policies, and contribute to data curation and quality assurance.

In addition to these operational roles, libraries can advocate for the value of domain-specific repositories and help sustain them through institutional partnerships or funding mechanisms. By engaging in these activities, libraries enhance the visibility, accessibility, and long-term value of disciplinary research data, thereby contributing to the broader goals of open science and responsible data stewardship.

Experiences from the EOSC-Nordic project have shown that such repositories do exist within institutions and that libraries have supported efforts to improve their FAIRness and to pursue CoreTrustSeal certification (Alaterä et al., 2022; CoreTrustSeal Standards & Certification Board, 2022; Meerman et al., 2021). Libraries are well positioned to promote the use of persistent identifiers such as DOIs, ensure that repositories are registered in international directories such as re3data and FAIRsharing, and raise awareness of certification schemes. Where needed, libraries can also facilitate access to expertise to support certification applications.

Moreover, libraries can act as ambassadors for the FAIR principles, offering practical guidance on their implementation. Tools such as the FAIR Evaluation Service (F-UJI)¹⁴, tested in the EOSC-Nordic project, allow users to assess the FAIRness of datasets by entering a URL or persistent identifier (Alaterä et al., 2022). The tool generates a FAIRness score and provides targeted feedback, highlighting both strengths and areas for improvement. This feedback can serve as a valuable starting point for discussions on data quality and machine-actionability—an aspect often overlooked when assessments are made solely from a human perspective.

7. Data Publishing in University Data Policies

The data policies of the universities that responded to the survey were examined, with particular attention given to the guidance provided on the publication of research data. These policies were articulated in a variety of ways across institutions, reflecting differences in institutional priorities, infrastructures, and the maturity of research data management services.

According to the responses, organisations appear to be at varying stages of maturity regarding the development of policies for publishing research data. Three institutions are currently in the process of formulating such principles, one organisation is at the planning stage. Meanwhile, two organisations reported having no principles in place, which may reflect either gaps in policy development or alternative strategic priorities.

Eighteen participants reported that their university had a data policy in place. However, upon closer inspection, one of the documents identified as a data policy did not, in fact, address research data at all, but rather provided general guidance on research publication. Thus, 17 different data policies were included in the final analysis. To systematically explore these variations, we employed qualitative text analysis, allowing us to identify thematic patterns, divergences, and institutional approaches to research data governance.

Two of the data policies included in the study were not available in English: one was written in Norwegian and the other in Lithuanian. While every effort was made to ensure accurate interpretation, it is important to acknowledge that the analysis of terminology may not fully capture the nuances of policies translated using automated tools such as Google Translate. However, in the case of the Scandinavian languages, the linguistic proximity to English—combined with the authors' proficiency in Swedish, a closely related language—supports the reliability of the interpretations. We therefore consider the translations to be sufficiently faithful to the original expressions for the purposes of this analysis.

Nonetheless, it is acknowledged that certain nuances may have been lost in translation. For instance, a Lithuanian-language policy may have used the term *publishing* in relation to data in a way that was not captured by the text analysis. This limitation should be kept in mind when interpreting the findings. To minimize the impact of such potential inaccuracies, all example

sentences cited in the article have been drawn exclusively from data policies originally written in English.

Out of the 17 data policies reviewed, only seven explicitly use the term “publishing” to refer to publishing research data (see examples 1–4). Some policies mention “publication” or “publishing,” but they refer to publishing articles or other research outputs, not the data itself (see examples 5–7). In some cases, research data is seen as supplementary material to published results, rather than being considered a published output on its own.

Example 1:

“Data underlying publication must as default be published, unless it would violate legal or ethical rules.”

Example 2:

“The data management plan includes the methods, processing, ownership and access rights, storage (including long-term storage), re-use, opening, publishing and planned disposal (if necessary) of the data and data collection, as well as the resources required for these measures.”

Example 3:

“The Swedish Research Council recommends that research data financed via public funds, and applicable legislation allows to be published, should be published openly on the internet within a reasonable time after the research results have been published.”

Example 4:

“Research data are published in data archives that safeguard the findability of data and enable references to them (for the definition of ‘data archive’, see the glossary).”

Example 5:

“Research data attached to published research results is principally available for shared use and open.”

Example 6:

“Data which forms the basis of **scientific publications** should be made available as early as possible, and no later than **at the time of publishing**.”

Example 7:

“*Research data related to research outputs should be opened after the outputs have been published.*”

In the discourse of data policy, the terms “*opening*” (15 out of 17 policies) and “*sharing*” or “*open sharing*” (10 out of 17) are more commonly used when referring to research data. Even in those policies where the term “*publishing*” was employed, it was used alongside expressions such as “*opening*” or “*sharing*”.

“*The researcher shall make research data openly available for further use to all relevant users, providing there are no legal, ethical, security or commercial reasons for not doing so.*”

“*The University is committed to promoting good research data management and the responsible open accessibility of research data, infrastructures, and methods.*”

“*Thus, data sharing should be in accordance with the principle of ‘as open as possible, as closed as necessary’. If the research data cannot be made available, sharing the metadata associated with the research data must be considered.*”

“*The University provides a research data infrastructure that supports management, sharing and reproducibility of data.*”

8. Libraries Offer Services that Support Data Publishing

Data repositories or data archives are offered quite often by libraries. According to our survey institutional data repositories or archives are offered or at least planned to be offered in most libraries (see Table 3). Half of the respondents said they already had such a service. For existing services, the service provider was equally likely to be a library (3 cases), a university (3 cases), a national (4 cases) or international actor (2 cases).

Table 3: Does your university provide institutional data repository or data archive services?

Does your university provide institutional data repository or data archive services?	Count
No	3
At planning stage	2
In process	7
Yes	12
Total	24

Table 4: Does your university provide long-term preservation services?

Does your university provide long-term preservation services?	Count
I don't know	1
No	3
At planning stage	2
In process	3
Yes	15
Total	24

A slightly larger proportion of libraries reported offering long-term preservation services for research data (see Table 4). These services were provided at various organisational levels, including the library (4 cases), university (5 cases), and national level (7 cases). A cross-analysis revealed that nine libraries offered both types of services—data repositories and long-term preservation—while six libraries provided only one or the other. Notably, three of the institutions offering long-term preservation services were also in the process of establishing, or planning to establish, an institutional repository.

The survey also explored whether libraries provide services related to the assessment of the value of research data (see Table 5). Only three respondents indicated that such services are currently in place. One service is under development, and four others are in the planning stage. Of the existing services, two are offered by libraries and one by a university. These findings suggest that data evaluation services are beginning to emerge, although their broader implementation remains limited. This finding appears to contrast with the more frequent reporting of long-term preservation services. As Whyte and Wilson (2010) have noted, effective long-term preservation necessarily entails the appraisal and selection of datasets. Given that such preservation is both resource-intensive and costly, ensuring the long-term accessibility and usability of digital information requires careful and systematic evaluation.

Table 5: Does your library offer services related to assessment of value of research data?

Does your library offer services related to assessment of the value of research data?	Count
No	16
At planning stage	4
In process	1
Yes	3
Total	24

Data catalogue services are currently available in significantly fewer universities compared to long-term data preservation services, although efforts to develop such services are ongoing. Eight respondents indicated that no data catalogue for research data is maintained by the library, nor is one currently planned. A nearly equal number of institutions, however, reported that a data catalogue is already in use (7 cases) or under development (8 cases) (see Table 6). In three cases, data catalogue services were provided by the library; two were offered at the national level, and a further two were delivered through international services.

Promoting data journals could be one of the roles of the library. Services related to data journals were already provided by three libraries or information services (see Table 7). Two of these libraries were in Finland and one in Sweden. Support services for data journals are therefore the rarest among data publishing services—at least for now.

Support related to domain-specific data repositories is also provided by university libraries (see Table 8). The question concerning the support offered by universities for such repositories was interpreted in various ways, which is beneficial, as support can take many forms. In the open-ended responses to this question, participants described the types of services available. For example, one respondent noted, “The library can assist in finding domain-specific data repositories if the researcher requests support.” Some responses indicated that support is available at least at the university level. For instance, the Swedigarch service—the Swedish National Infrastructure for Digital Archaeology—is hosted by a university. As the role of libraries was not explicitly addressed in the question, it remains somewhat unclear what specific services they provide in relation to domain-specific repositories. Nonetheless, it is evident that at least some level of support is already established within universities.

Table 6: Does your university provide data catalogue services?

Does your university provide a data catalogue service?	Count
No	8
At planning stage	1
In process	8
Yes	7
Total	24

Table 7: Does your university provide data journal services?

Does your university provide data journal services?	Does your university provide data journal services? (Count All)
I don't know	6
No	15
Yes	3
Total	24

Table 8: Does your university provide support for domain specific data repositories?

Does your university provide support for domain specific data repositories?	Count
I don't know	1
No	9
At planning stage	3
Yes	11
Total	24

9. Discussion

This article examines the evolving role of research libraries in supporting the formal publication of research data as a recognised scholarly output. A qualitative, multi-method approach was employed, incorporating literature review, policy analysis, and a survey of university libraries in the Nordic and Baltic regions. A combination of data and methodological triangulation, alongside theoretical sampling, was used to provide a holistic perspective. The various types and origins of research material were not intended to serve as direct comparison pairs but were selected to enrich the understanding of the phenomenon.

9.1. Reframing Research Data as a Scholarly Output

The repositioning of research data from a supplementary resource to a primary scholarly output has been underscored. This shift necessitates a redefinition of institutional responsibilities and a broader understanding of academic publishing. If data were more widely regarded as publishable content—subject to peer review, persistent identification, and quality control—its role in research evaluation and scholarly communication could be significantly enhanced. Equally important is the citability of research data, which ensures that datasets receive appropriate academic recognition and can be reliably traced and reused. Enhancing data citation practices not only supports transparency and reproducibility but also incentivises data sharing by formally acknowledging the intellectual contribution of data creators.

The influence of national and international data policies on the discourse of service providers, as compared to researchers' own conceptualisations, warrants further investigation. When seeking to position research data alongside other scholarly outputs, the terminology employed becomes particularly significant. If data policies were to explicitly refer to the publication of data, this could help raise awareness that research data is not solely to be opened and shared—it can also be formally published.

In this context, scholarly publishers are uniquely positioned to advance the recognition of research data as a legitimate academic output. By mandating data availability statements (DAS) and requiring authors to cite underlying datasets, publishers can help embed data sharing into the fabric of scholarly communication. This, however, presupposes that data are made available in a citable form, with persistent identifiers and adequate metadata. Even in cases where access to the data must be restricted, the publication of metadata remains essential to ensure discoverability and transparency. Furthermore, peer reviewers should be encouraged—and guided—to consider the quality, accessibility, and relevance of the supporting data as part of the review process. Such measures would not only reinforce the scholarly value of data but also foster a more robust and accountable research ecosystem.

9.2. The Role of Research Libraries

Research libraries have been identified as key actors in facilitating the publication of research data through services such as institutional repositories, long-term preservation infrastructures, data catalogues, and support for data

journals and domain-specific repositories. Their role is both strategic and operational, requiring coordination of services and alignment with institutional research goals.

However, notable gaps remain in areas such as data appraisal, quality assurance, and the development of sustainable service models. Collaborative service models that align with disciplinary practices and institutional capacities are essential. Libraries must support researchers in navigating data publication pathways, promote the use of data journals, and ensure that metadata standards are upheld to maximise discoverability and reuse.

The integration of data publishing into research assessment frameworks remains limited. Without robust metrics and cataloguing systems, the scholarly impact and reuse of datasets cannot be effectively measured or recognised. Libraries must therefore advocate for the inclusion of data outputs in evaluation systems and support the development of appropriate indicators.

While libraries already play a pivotal role in research data management—through guidance, training, and coordination of multiprofessional support—an exclusive focus on data management risks overlooking the broader benefits of FAIR data and the importance of quality assurance.

A data catalogue serves to record the metadata of research data produced within an institution. In the absence of such a catalogue, the establishment of effective assessment practices becomes significantly more challenging. Beyond supporting the development of metrics, catalogues are essential for identifying datasets containing personal information (in accordance with GDPR requirements) and those requiring long-term preservation.

At present, many CRIS systems do not support the construction of high-quality data catalogues, as they often lack compatibility with established metadata standards. It is the responsibility of libraries to ensure that metadata adhere to international standards and schemas, thereby enhancing interoperability and supporting broader discoverability and reuse.

9.3. Building Competencies for Data Publishing Services

The emergence of research data publishing as a formal scholarly activity presents new challenges to existing skill sets, service models, and

business frameworks within academic libraries. While libraries already possess expertise in knowledge management, long-term accessibility, and open access publishing, further development is required in areas such as data quality, documentation, and metadata. These competencies are critical for effective data appraisal and selection and extend beyond traditional librarianship.

To support the development of these new competencies, sustained and deep collaboration is essential. Initiatives such as the *Advancing RDM Careers: A Framework for Expert Education in Finland* (2024) provide valuable frameworks for guiding this evolution.

The implementation of services such as long-term data preservation necessitates close cooperation with institutional stakeholders, including IT services (for technical infrastructure and data security), university leadership (for policy alignment), data protection officers (for GDPR compliance), and legal experts (for data rights management). Within this collaborative environment, libraries are expected to act as active expert partners.

Proactive engagement requires that library professionals are afforded opportunities to familiarise themselves with academic practices from the perspective of research data. The ability to specialise as data management experts enables the delivery of high-quality, targeted services that directly benefit the research community. Investment in such specialisation is not only beneficial but essential, as it facilitates the acquisition of the necessary expertise to meet evolving institutional and disciplinary needs.

9.4. Limitations

While the sample of the survey is neither large nor representative of all university libraries in the region, the responses appear to come predominantly from institutions that have been early adopters in the provision of research data management (RDM) services. Given that the survey was mainly distributed through mailing lists specifically aimed at RDM professionals—such as the DataCite list in Estonia, the Dataverse network in Norway, and the DataSupport chat in Finland—it is likely that the survey primarily reached libraries actively engaged in developing such services.

Had the survey been disseminated more broadly via general university library mailing lists, it might have reached institutions where RDM services are less developed or still emerging. It is also plausible that libraries with dedicated RDM personnel were more inclined to respond to a survey on this topic. This may help explain the absence of responses from libraries where such services are currently limited—stitutions which, in our view, do exist within the region.

The primary aim of this study was to explore the nature and scope of data publishing services being offered, rather than to assess their overall distribution or uptake. The data nonetheless offer valuable insights into the types of services currently being provided in support of research data publishing, although the number of responses is modest. However, the limited sample size precludes any definitive conclusions regarding the prevalence of these services across the broader library landscape.

10. Conclusions

In conclusion, the formal publication of research data presents both a challenge and an opportunity for research libraries. By embracing this role, libraries can contribute meaningfully to the advancement of open science, the integrity of research, and the recognition of diverse scholarly contributions. To fully realise this potential, a cultural shift is required—one that positions data publishing as an integral part of the research lifecycle. Continued investment in infrastructure, policy alignment, and professional development will be essential to ensure that libraries remain at the forefront of this transformation.

Acknowledgements

We warmly thank Dr. Mikael Niku, Dr. Manna Satama and MA. Soile Manninen, for their invaluable comments on the draft of this article and for the encouraging discussions throughout various stages of our work. We also extend our gratitude to the peer reviewers for their valuable contributions, which have significantly shaped the final outcome. Finally, we thank all the respondents who took the time to participate in our survey, as well as those who helped disseminate it.

Note: This article reflects the independent work of the authors. Generative AI tools (ChatGPT versions 3, 4, and 4o) were used solely to support language editing. The tools assisted in rephrasing selected portions of the author's own writing to improve clarity and conciseness. At no point were AI tools used to generate or summarise content from external sources. All AI-assisted outputs were reviewed and revised by the authors prior to inclusion in the manuscript.

Data Availability

Survey template and survey results are published in Zenodo.

References

Alaterä, T., Kleemola, M., Ala-Lahti, H., & Jerlehag, B. (2022). *D4.5 Report on completed FAIR data standard adoption and certifications of data repositories in the region* [Project delivery]. Zenodo. <https://doi.org/10.5281/ZENODO.7303538>

Albani, M., Maggio, I., & Ceos Data Stewardship Interest Group. (2020). Long-term data preservation data lifecycle, standardisation process, implementation and lessons learned. *International Journal of Digital Curation*, 15(1), 1–10. <https://doi.org/10.2218/ijdc.v15i1.715>

Austin, C. C., Bloom, T., Dallmeier-Tiessen, S., Khodiyar, V. K., Murphy, F., Nurnberger, A., Raymond, L., Stockhause, M., Tedds, J., Vardigan, M., & Whyte, A. (2017). Key components of data publishing: Using current best practices to develop a reference model for data publishing. *International Journal on Digital Libraries*, 18(2), 77–92. <https://doi.org/10.1007/s00799-016-0178-2>

Ayris, P., Bernal, I., Cavalli, V., Dorch, B., Frey, J., Hallik, M., Hormia-Poutanen, K., Labastida, I., MacColl, J., Ponsati Obiols, A., Sacchi, S., Scholze, F., Schmidt, B., Smit, A., Sofronijevic, A., Stojanovski, J., Svoboda, M., Tsakonas, G., van Otegem, M., ... Horstmann, W. (2018). *Liber open science roadmap*. Zenodo. <https://doi.org/10.5281/ZENODO.1303002>

Callaghan, S., Donegan, S., Pepler, S., Thorley, M., Cunningham, N., Kirsch, P., Ault, L., Bell, P., Bowie, R., Leadbetter, A., Lowry, R., Moncoiffé, G., Harrison, K., Smith-Haddon, B., Weatherby, A., & Wright, D. (2012). Making data a first class scientific output: Data citation and publication by NERC's Environmental Data Centres. *International Journal of Digital Curation*, 7(1), 107–113. <https://doi.org/10.2218/ijdc.v7i1.218>

CODATA-ICSTI Task Group on Data Citation Standards and Practices. (2013). Out of cite, out of mind: The current state of practice, policy, and technology for the citation of data. *Data Science Journal*, 12, CIDCR1–CIDCR75. <https://doi.org/10.2481/dsj-OSOM13-043>

CoreTrustSeal Standards and Certification Board. (2022). *CoreTrustSeal requirements 2023-2025 (Version V01.00)*. Zenodo. <https://doi.org/10.5281/ZENODO.7051012>

Janesick, V. J. (2000). The choreography of qualitative research design: Minuets, improvisations, and crystallization. In N. K. Denzin, K. Norman & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 379–399). Sage.

Johnston, L. R., Curty, R., Braxton, S. M., Carlson, J., Hadley, H., Lafferty-Hess, S., Luong, H., Petters, J. L., & Kozlowski, W. A. (2024). Understanding the value of curation: A survey of US data repository curation practices and perceptions. *PLoS One*, 19(6), Article e0301171. <https://doi.org/10.1371/journal.pone.0301171>

Klump, J., Bertelmann, R., Bräse, J., Diepenbroek, M., Grobe, H., Höck, H., Lautenschlager, M., Schindler, U., Sens, I., & Wächter, J. (2006). Data publication in the open access initiative. *Data Science Journal*, 5, 79–83. <https://doi.org/10.2481/dsj.5.79>

Kuusniemi, M. E., & Nykyri, S. (2022, July 6–8). *Data as a new research publication type: What could be the role of research libraries as service providers?* [Conference presentation] 55th LIBER Annual Conference (Odense, Denmark. Zenodo. <https://doi.org/10.5281/zenodo.6812223>

Kuusniemi, M. E., & Moisio, M. (2024). *Advancing RDM careers: A framework for expert education in Finland*. National Open. Roadmap for professional qualification of data management experts working group. <https://doi.org/10.5281/ZENODO.10815480>

Kuusniemi, M. E., & Nykyri, S. (2025a). *Dataset for the study research data publication services provided by University Libraries in the Nordic and Baltic Countries* [Data set]. University of Helsinki & Tampere University. <https://doi.org/10.5281/zenodo.15434405>

Kuusniemi, M. E., & Nykyri, S. (2025b). *Survey form for the study Research Data Publication Services Provided by University Libraries in the Nordic and Baltic Countries (1.0)* [Data set]. University of Helsinki & Tampere University. <https://doi.org/10.5281/zenodo.15433857>

Labadie, C., Legner, C., Eurich, M., & Fadler, M. (2020). FAIR enough? Enhancing the usage of enterprise data with data catalogs. In 2020 IEEE 22nd Conference on Business Informatics (CBI), 201–210. <https://doi.org/10.1109/CBI49978.2020.00029>

Lawrence, B., Jones, C., Matthews, B., Pepler, S., & Callaghan, S. (2011). Citation and peer review of data: Moving towards formal data publication. *International Journal of Digital Curation*, 6(2), 4–37. <https://doi.org/10.2218/ijdc.v6i2.205>

Lee, D. J., & Stvilia, B. (2017). Practices of research data curation in institutional repositories: A qualitative view from repository staff. *PLoS One*, 12(3), Article e0173987. <https://doi.org/10.1371/journal.pone.0173987>

Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., De Giusti, M., L'Hours, H., Hugo, W., Jenkyns, R., Khodiyar, V., Martone, M. E., Mokrane, M., Navale, V., Petters, J., Sierman, B., Sokolova, D. V., Stockhouse, M., & Westbrook, J. (2020). The TRUST Principles for digital repositories. *Scientific Data*, 7(1), Article 144. <https://doi.org/10.1038/s41597-020-0486-7>

Mayernik, M. S., Callaghan, S., Leigh, R., Tedds, J., & Worley, S. (2015). Peer review of datasets: When, why, and how. *Bulletin of the American Meteorological Society*, 96(2), 191–201. <https://doi.org/10.1175/BAMS-D-13-00083.1>

McGillivray, B., Marongiu, P., Pedrazzini, N., Ribary, M., Wigdorowitz, M., & Zordan, E. (2022). Deep impact: A study on the impact of data papers and datasets in the humanities and social sciences. *Publications*, 10(4), Article 39. <https://doi.org/10.3390/publications10040039>

Meerman, B., Gaiarin, S. P., & Gallas, S. M. (2021). *EOSC-NORDIC FAIRification study testing F-UJI* (Version 1.0). Zenodo. <https://doi.org/10.5281/ZENODO.5226082>

National Information Standards Organization. (2023). *ANSI/NISO Z39.106-2023, Standard Terminology for Peer Review*. <https://doi.org/10.3789/ansi.niso.z39.106-2023>

OECD & Science Europe. (2020). *Optimising the operation and use of national research infrastructures. OECD Science, Technology and Industry Policy Papers 91*. <https://doi.org/10.1787/7cc876f7-en>

Open Science Coordination in Finland, Federation of Finnish Learned Societies. (2023). *Open research data and methods. National policy and executive plan by the higher education and research community for 2021–2025: Policy component 1 (Open access to research data) and 2 (Open access to research methods and infrastructures)*. The Committee for Public Information (TJNK) and Federation of Finnish Learned Societies (TSV). <https://doi.org/10.23847/tsv.669>

Parsons, M. A., & Fox, P. A. (2013). Is data publication the right metaphor? *Data Science Journal*, 12, WDS32–WDS46. <https://doi.org/10.2481/dsj.WDS-042>

Pop, M., & Salzberg, S. L. (2015). Use and mis-use of supplementary material in science publications. *BMC Bioinformatics*, 16(1), Article 237. <https://doi.org/10.1186/s12859-015-0668-z>

Publications Office of the European Union. (2022). *Data citation: A guide to best practice*. European Union. <https://data.europa.eu/doi/10.2830/59387>

Publications Office of the European Union. (2021). *Data.europa.eu data quality guidelines*. European Union. <https://data.europa.eu/doi/10.2830/79367>

PwC EU Services. (2018). *Cost-benefit analysis for FAIR research data: Cost of not having FAIR research data*. European Commission. <https://data.europa.eu/doi/10.2777/02999>

Rantasaari, J. (2022). Multi-stakeholder research data management training as a tool to improve the quality, integrity, reliability and reproducibility of research. *LIBER Quarterly*, 32(1), 1–54. <https://doi.org/10.53377/lq.11726>

Rieger, O. Y. (2007). Select for success: Key principles in assessing repository models. *D-Lib Magazine*, 13(7/8). <https://doi.org/10.1045/july2007-rieger>

Rumsey, S., & Jefferies, N. (2013). Challenges in building an institutional research data catalogue. *International Journal of Digital Curation*, 8(2), 205–214. <https://doi.org/10.2218/ijdc.v8i2.284>

Schmidt, B., Chiarelli, A., Loffreda, L., & Sondervan, J. (2023). Emerging roles and responsibilities of libraries in support of reproducible research. *LIBER Quarterly*, 33(1), 1–21. <https://doi.org/10.53377/lq.14947>

Seale Clive (ed.) (1998). *Researching society and culture*. SAGE Publications.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research. Grounded theory procedures and techniques*. SAGE Publications.

UNESCO. (2021). *UNESCO Recommendation on open science*. <https://doi.org/10.54677/MNMH8546>

Whyte, A. (2016, January 22). Where to keep research data: Version 1.1 of the DCC checklist for evaluating data repositories. *DCC How-to Guides*. <https://www.dcc.ac.uk/guidance/how-guides/where-keep-research-data>

Whyte, A. & Wilson, A. (2010). How to Appraise and Select Research Data for Curation. *DCC How-to Guides*. <https://www.dcc.ac.uk/guidance/how-guides/appraise-select-data>

Wilkinson, M. D., Dumontier, M., Aalbersberg, IJ. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), Article 160018. <https://doi.org/10.1038/sdata.2016.18>

Notes

¹ *San Francisco Declaration on Research Assessment (DORA)* (2012). The Annual Meeting of the American Society for Cell Biology in San Francisco, <https://sfdora.org/read/>.

² Data Publication, <https://www.ncbi.nlm.nih.gov/guides/data-glossary/data-publication>.

³ Scientific Data, <https://www.nature.com/sdata/>.

⁴ Data in Brief, <https://www.sciencedirect.com/journal/data-in-brief>.

⁵ Matthews, T. 2021, November 22). *Data articles: What are they and how can they benefit me?* Research Communities by Springer Nature. <http://researchdata.springernature.com/posts/data-articles-what-are-they-and-how-can-they-benefit-me>.

⁶ Carpenter, T. A. (2017, April 11). *What Constitutes Peer Review of Data? A Survey of Peer Review Guidelines.* The Scholarly Kitchen. <https://scholarlykitchen.sspnet.org/2017/04/11/what-constitutes-peer-review-research-data/>.

⁷ FSD Operational guidelines, <https://www.fsd.tuni.fi/en/data-archive/documents/records-management-and-archives-formation-plan/operational-guidelines/>.

⁸ Research data, <https://www.springernature.com/gp/authors/research-data>.

⁹ Why Publish Data? <https://edirepository.org/resources/why-publish-data>.

¹⁰ Coalition for Advancing Research Assessment. (2022, July 20). *Agreement on reforming research assessment.* CoARA. <https://coara.eu/agreement/the-agreement-full-text/>.

¹¹ DataWiz – Data Archives and Repositories, <https://datawizkb.leibniz-psychology.org/index.php/after-collection/what-should-i-know-about-archives-and-repositories/>.

¹² Little, J. Code Repository vs Archival Repository. You need both. (2022), <https://blogs.library.duke.edu/data/author/jrlduke-edu/>.

¹³ OAIS Reference Model (ISO 14721), <http://www.oais.info/>.

¹⁴ F-UJI is a web service to assess FAIRness of research data objects at the dataset level based on the FAIRsFAIR Data Object Assessment Metrics, <https://www.f-iji.net/>.