Supporting Data Discovery: Comparing Perspectives of Support Specialists and Researchers

GUANGYUAN SUN ⓘD

TANJA FRIEDRICH ⓘD

KATHLEEN GREGORY ⓘD

BRIGITTE MATHIAK ⓘD

*Author affiliations can be found in the back matter of this article

]u[ubiquity press

ABSTRACT

Purpose: Much of the research in data discovery is centered on the users' viewpoint, frequently overlooking the perspective of those who develop and maintain the discovery infrastructure. Our goal is to conduct a comparative study on research data discovery, examining both support specialists' and researchers' views by merging new analysis with prior research insights.

Methods: This work summarizes the studies the authors have conducted over the last seven years investigating the data discovery practices of support specialists from different disciplines. Although support specialists were not the main target of some of these studies, data about their perspectives was collected. Our corpus comprises in-depth interviews with 6 social science support specialists, interviews with 19 researchers and 3 support specialists from multiple disciplines, a global survey with 1630 researchers and 47 support specialists, and a use case analysis of 25 support specialists. In the analysis section, we juxtapose the fresh insights on support specialists' views with the already documented perspectives of researchers for a holistic understanding. The latter is primarily discussed in the literature review, with references made in the analysis section to draw comparisons.

Results: We found that support specialists' views on data discovery are not entirely different from those of the researchers. There are, however, some differences that we have identified, most notably the interconnection of data discovery with general web search, literature search, and social networks.

Conclusion: We conclude by proposing recommendations for different types of support work to better support researchers' data discovery practices.

1. INTRODUCTION

Research data management support is an increasingly visible component in support services provided by universities and research institutions. While data librarians, archivists, and educational specialists provide support with managing and sharing data (Borgman et al., 2019; Cope and Baker, 2017; Cox et al., 2019; Lafia et al., 2021), they are also interested in supporting practices afforded by well-managed data, such as data discovery.

Many terms are used to describe research data support staff: data curator, data librarian, data manager, data steward, data specialist, and research information scientist (Tammaro et al., 2019). The field lacks standardized terminology; this inconsistency extends to both job titles and actual duties performed. For the purposes of this paper, we use the encompassing term 'research data support specialists' to represent professionals with different roles in research data management support.

Compared to studies of the behavior and practices of researchers and other data users in data discovery, studies from the perspective of data support specialists are rather sparse. Work conducted from the perspective of data support specialists focuses on typical workflows (Cox et al., 2019; Gowen and Meier, 2020; Reichmann et al., 2021), tasks (Cruz et al., 2018; Khan et al., 2021), and the details of infrastructural projects (e.g., GeRDI project), which are designed to support data discovery and reuse.[1]

With few exceptions (i.e., Borgman et al., 2019), the perspectives of support specialists and researchers exist separately from each other in the literature. This separation is problematic, as support specialists may build infrastructures and services based on perceptions of researchers' practices rather than the practices themselves. We argue that we need to bring together and analyze both the practices and perspectives of researchers and those of data support specialists in order to build effective, sustainable data infrastructures and services.

In this paper, we focus on data discovery. Drawing from a series of studies we individually conducted between 2017 and 2021 on the broader theme of research data reuse, we closely examine the nuances and intersections between researchers' data discovery practices and the perspectives and efforts of support specialists. This series encompasses data from various sources: in-depth interview studies (Friedrich, 2020; Gregory et al., 2019), a global survey (Gregory, 2020; Gregory et al., 2020), and a use case analysis by information professionals (Mathiak et al., 2023). Two of these studies emphasize the perspective of support specialists (Friedrich, 2020; Mathiak et al., 2023), and two incorporate insights from both researchers and support specialists (Gregory, 2020; Gregory et al., 2019).

Our approach in this paper is twofold. Firstly, we critically review the existing literature alongside a reflective examination of our own previous data discovery research, yielding five key themes. Simultaneously, we develop a typology of support specialists' roles and tasks.

Secondly, we delve into a targeted analysis of the data from our studies that involved support specialists' views about data discovery. This is then compared with the perspective of researchers, which draws from the findings of our prior studies. Our conclusion connects emerging themes to our typology of support roles, underscoring both best practices and areas for advancement.

2. LITERATURE ANALYSIS

We provide an overview of literature from the perspective of researchers (Section 2.1) and then focus on the perspective of support specialists (Section 2.2) to develop a typology of support specialist work for later discussion and recommendations.

2.1 PRIOR STUDIES FROM THE PERSPECTIVE OF RESEARCHERS

Many data discovery studies are motivated by studying researchers' data reuse behaviors (Curty, 2016; Gregory et al., 2019; Joo et al., 2017; Joo and Kim, 2017). They correlate concepts, such as attitude, motivation, and disciplinary norms in data reuse with discovery practices, e.g., through surveys.

---

1 GeRDI project: https://www.gerdi-project.eu/.

Qualitative studies, such as Borst and Limani (2020) investigate workflows of data search in a web-based environment and introduce patterns of data search, i.e., discover, explore, and analyze. Koesten, et al. (2017) categorize users' data-centric tasks into two types: process-oriented (e.g., using data for machine learning processes) or goal-oriented (e.g., using data to answer a question). Wu et al. (2019) identify nine requirements of users of STEM data repositories. Bishop et al. (2019) uncover a preference for known-item searches in reliable sources among environmental scientists.

Search query and log analysis are a third way to investigate data search behavior. Examples include structured data stored on the web (Koesten et al., 2017), in open government data portals (Kacprzak et al., 2018), or in research data repositories (Brickley et al., 2019; Hemphill et al., 2022; Sharifpour et al., 2022). Results show that data search queries are typically brief and that users tend to discover data repositories through search engines such as Google.

We identify five themes in existing studies, including our own, that focus on the data discovery practices of researchers, namely data discovery challenges, needs, general web search, academic literature, and social connections. For example, Krämer et al. (2021) spotlight data skill and literacy gaps and access issues in discovering data, while they and Gregory et al. (2019, 2020) highlight technical barriers like inadequate search tools and poor data-literature integration. Gregory et al. (2020) also note data dispersion as a unique challenge. Our own studies, including Sun (2019), Gregory et al. (2019, 2020), and Krämer et al. (2021), reveal a spectrum of researchers' data needs and the aiding role of academic literature in data discovery across disciplines. We also find that researchers commonly use general web search engines for various data discovery tasks, a finding consistent with earlier research (Sharifpour et al., 2022). Social connections are crucial for both data discovery and reuse, a finding that aligns with existing research (Kim, 2017; Koesten et al., 2017; Yoon, 2017; Yoon and Lee, 2019). While Friedrich (2020) found that finding data is facilitated by community involvement, Krämer et al. (2021) further emphasize the value of social connections in interpreting data, and Sun (2019) suggests their importance may vary based on data accessibility. While not exhaustive, this framework of five themes serves as a useful tool for understanding both the existing literature and the alignment between researchers and support specialists. We take these themes as a starting point for comparison in our findings.

2.2 TYPES OF WORK CONDUCTED BY SUPPORT SPECIALISTS

Studies of support specialists regarding the discoverability of data are comparably sparse. Research data support work is both social and technical, involving technical knowledge on research data management (RDM) systems and social knowledge for RDM community building (Wilson et al., 2017). Research on support roles is fragmented, with varied professional titles reflecting the lack of uniform terminology and the diverse range of duties (Tammaro et al., 2019).

We identify three primary types of support work from the literature (Figure 1): people-oriented, e.g., providing consultations (Section 2.2.1); metadata-oriented, e.g., providing data documentation services that are non-technical but focus on enhancing the completeness of research data description and discoverability (Section 2.2.2); and infrastructure-oriented, e.g., providing IT-focused technical tasks (Section 2.2.3). It is important to note that these types of work are not mutually exclusive, as support specialists may be required to fulfill multiple roles. Nonetheless, the proposed categorization serves as a heuristic device for structuring and discussing the complexities of RDM support work.
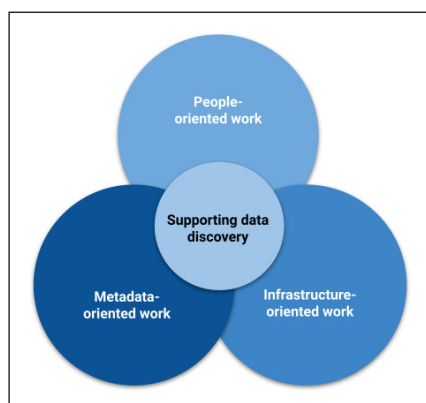

Figure 1 Types of work involved in supporting data discovery.

2.2.1 People-oriented support work

People-oriented research support varies depending on organizational infrastructure, staff expertise, and values (Darch et al., 2020). At large university libraries, services have expanded to encompass research data management (Si et al., 2019). Cox, et al. (2017) show libraries are leading in RDM, but there are also alternative organizational models. Research support may include library staff, but also professionals who are not always situated within libraries: data stewards (Tammaro et al., 2019) who engage in RDM discussions (Cruz et al., 2019) and data champions who advocate for RDM (Savage and Cadwallader, 2019). While North America and Europe lead in people-oriented research data services (Cox et al., 2019), other regions focus on policy, awareness, and RDM infrastructure development (Huang et al., 2021; Marlina and Purwandari, 2019).

As people-oriented support work evolves, data discovery has increasingly become a focus of advisory services. Cox et al. (2019) show that since 2014, large university libraries have consistently included data discovery as a core aspect of RDM services, alongside data management planning (DMP) and copyright support. In this context, Cox (2018) posits a 'data role spectrum', suggesting that facilitating data search or access is the most 'familiar task' for academic librarians, essentially serving as a natural progression from traditional resource discovery services.

A nuanced picture emerges when considering the role of data curators. Tammaro et al. (2019) analyze data curator job postings and report that, although their primary competencies focus on data management, sharing, and preservation, skills regarding data discovery are commonly sought. However, they are frequently viewed as supplementary.

Ashiq et al.'s (2020) systematic review hints at a divergence between support specialists and researchers on the value of data discovery services. While support specialists provide such services, researchers identify data discovery as a significant barrier, advocating for more specialized services to facilitate data reuse. Tang and Hu (2019) stand out in the review for finding a consensus among librarians on the need to integrate data discovery within library systems. However, there's a scarcity of research on the interactions between support specialists and data seekers in aiding data discovery.

The transformative impact of artificial intelligence (AI) on library services is gaining recognition. Lund et al. (2020) report that librarians expect AI to play a significant role in enhancing resource and data discovery within the next three decades. More recently, Cox (2023) outlines AI-driven methods poised to revolutionize knowledge discovery, suggesting that libraries might increasingly focus on advising on data discovery as part of their peripheral services to users' core data science activities. The consensus from the literature suggests an emergent recognition of AI's potential to notably advance data discovery services.

2.2.2 Metadata-oriented support work

Rich metadata enhances discoverability for datasets. Research data specialists actively participate in working groups like the eXtended Knowledge Organization System (XKOS)[2] under Data Documentation Initiative (DDI),[3] the Research Data Alliance,[4] and DataCite[5] to develop both domain-specific and generic metadata standards for research data. Techniques like Linked Open Data and knowledge graphs promise better connectivity of data with research papers, as seen in initiatives like OpenAIRE Research Graph (Manghi et al., 2019) and Open Knowledge Research Graph (Jaradeh et al., 2019). However, the transition to these techniques presents challenges for traditional libraries, which often lack the specialized knowledge required for advanced metadata transformation (Cox et al., 2017). This highlights the evolving role of research data support specialists, who not only must understand existing metadata standards but also adapt to emerging technologies to drive effective data discovery.

---

[2] https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/826408976/XKOS+-+Extended+Knowledge+Organization+System.

[3] https://ddialliance.org/.

[4] https://www.rd-alliance.org/.

[5] https://datacite.org/.

2.2.3 Infrastructure-oriented support work

Research data management demands robust digital infrastructure throughout the data lifecycle; infrastructural work is primarily managed by IT-focused support staff. Mulongo et al. (2022) evaluated research data services at leading universities in developed nations, highlighting advanced technological infrastructures for their provision of long-term data storage and improvement of data discoverability and accessibility. Other infrastructure-oriented support work frequently includes federated data repository development (Garnett et al., 2017), creation of discovery tools like DataCite Commons (Arora and Chakravarty, 2021), and data catalogs for sensitive data management (Read et al., 2015; Sheridan et al., 2021).

While various studies have outlined the key characteristics of effective RDM infrastructure (Bugaje and Chowdhury, 2018a; Bugaje and Chowdhury, 2018b; Khan et al., 2021; Liu et al., 2021; Mannheimer et al. 2021), less emphasis has been traditionally placed on data discoverability as a primary objective. For instance, Khan et al. (2021) and Smit (2019) emphasize the long-term maintenance of repositories and user engagement, noting data discoverability as a future, not immediate, priority.

Increasingly, there is a recognized need to better address data-specific requirements and enhance data discoverability. Bugaje and Chowdhury (2018a) suggest that current discovery systems are mere adaptations of traditional retrieval systems and do not adequately address the unique challenge of research data. Similarly, Hedeland (2020) highlights the infrastructure shortcomings for specialized data types, like audio-visual linguistic data. To bridge such gaps, information professionals are building expansive infrastructures like the European Open Science Cloud (Almeida et al., 2017), aiming, among other ambitions, to simplify data discovery and access.

These types of infrastructure-oriented support work tend to lean heavily on solving technological issues (e.g., data searchability and storability), potentially overlooking user-centric challenges.

3. METHOD

In this section, we briefly discuss the methodologies used in each individual study in our synthesis (see Table 1). We then discuss our process for identifying common themes across the studies.

STUDY NUMBER	METHODOLOGY	PARTICIPANTS	DISCIPLINARY FOCUS	YEAR OF DATA COLLECTION	PUBLISHED WORK
Study 1	Survey	Researchers (n = 1630); Support professionals (n = 47)	Multiple disciplines	2018	(Gregory 2020; Gregory et al. 2020)
Study 2	Interview	Researchers (n = 19); Support professionals (n = 3)	Multiple disciplines	2017	(Gregory et al. 2019)
Study 3	Interview	Support professionals (n = 6)	Social sciences	2016	(Friedrich 2020)
Study 4	Use case analysis by support professionals	Use cases (n = 100 use cases collected; n = 25 support professionals)	Multiple disciplines	2020	(Mathiak et al. 2023)

Table 1 Methodologies and study descriptions for data used in our synthesis.

3.1. METHODOLOGIES FOR INDIVIDUAL STUDIES

3.1.1. Study 1: Survey with researchers and support specialists (Gregory et al. 2020)

A global survey was conducted to investigate practices of data discovery and reuse. Respondents were recruited by emails sent to a random sample of 150,000 authors who published an article indexed in the Scopus database. An additional 40 participants were recruited by posting to library and RDM discussion lists. Responses were recorded anonymously and were analyzed using R (R Core Team n.d.); survey data are openly available (Gregory, 2020). This survey was descriptive in nature; responses only describe the practices of respondents. Both the majority of researchers (68.4%) and support specialists (74.47%) worked at universities, followed by those employed at research institutions (17.18%, 14.89%). Mid-career professionals were the largest category of both researchers (40.18%) and support specialists (46.81%). The second largest group of researchers had 16–30 years of experience (30.25%), while the second largest group

of support specialists had 0–5 years of experience (27.66%). The most selected disciplinary domains of researchers were engineering and technology, biological sciences, medicine, and social sciences. Among support specialists, the most selected domains were information science, environmental science, multidisciplinary research, and social science. Respondents were employed in 105 countries; the United States (13%), Italy (7%), and Brazil (7%) were the most represented. Here, we analyze the survey data collected from support specialists and compare it to the data from researchers. New visualizations comparing both the perspectives of support specialists and researchers were created using Tableau. This allows a comparison that was not yet performed in previous research.

3.1.2. Study 2: Interviews with researchers and support specialists (Gregory et al., 2019)

One-hour, audio-recorded, semi-structured interviews were conducted with researchers and support specialists to elicit information about participants' data needs, sources, and strategies for finding and evaluating data. Detailed summaries were made following a protocol for every interview. Summaries were thematically coded and analyzed.

The majority of participants were active researchers (n = 18); three participants worked as support specialists, and one respondent identified as being a concerned citizen. Participants came from a range of disciplines, with more from information science (n = 3) and computer science (n = 3). The majority worked in the United States (n = 6) or the Netherlands (n = 3) and were mid-career (n = 10). Support specialists were mid-career (n = 2) or early-career (n = 1) professionals; two worked in university libraries and one worked in industry.

The interview study was previously analyzed primarily from the perspective of researchers (Gregory et al., 2019). Here, we report further details about researchers' data needs that were not previously published, analyze the perspectives of support specialists, and compare these with the results from later studies.

3.1.3. Study 3: Interviews with support specialists (Friedrich 2020)

Six support specialists from the German GESIS data archive for the social sciences were interviewed regarding their interactions and experiences with users of the data archive. The goal of the study was to find out more about data users' information-seeking behavior and their challenges when looking for data. Following constructivist grounded theory methodology, the interviewees were sampled by initial and theoretical sampling. Two participants were help desk staff who received requests via the general helpdesk line. The other four were specialized experts for data from one or more complex large-scale social science surveys. Participants had been working as data support staff for different periods of time, ranging from 3 years to 25 years. Audio recordings and transcripts were made of all interviews, which were approximately one hour in length. Transcripts were analyzed according to constructivist grounded theory methodology, using the software atlas.ti. The interview data were previously used to inform the development of a survey questionnaire for data catalog users. Here, we re-analyze the interview data to explicitly reveal the perspective of support specialists' regarding the challenges, problems, and needs of data users and to identify differences and similarities with researchers' and other data users' perspectives.

3.1.4. Study 4: Use case analysis and ranking by support specialists (Mathiak et al. 2023)

The GOFair Data Discovery Implementation Network[6] collected over 100 use cases for data discovery by polling researchers and support specialists and by consolidating existing lists. These use cases were then clustered around common themes and requirements. A survey tool administered during the GOFAIR meeting (February 3rd, 2021) was used to gain consensus about the identified themes. Meeting attendees were asked to answer a demographic survey and to rank the two most and least 'relevant' use case clusters. 25 out of approximately 50 meeting attendees completed the survey. Most respondents were from mainland Europe (n = 17), with three additional participants from the UK (n = 3). Three attendees were from North America, and one was from South America. The majority of respondents identified themselves as infrastructure providers (n = 17), with eight respondents indicating that they produce data

---

6   https://www.go-fair.org/implementation-networks/overview/discovery.

themselves. Most respondents were involved in the life sciences (n = 8), general science (n = 7), or natural sciences (n = 5). Fewer worked in social science (n = 3) and computer science. Earlier published work about this study (Mathiak et al. 2023) only includes a very short analysis about how the ranking of use contrasts with the needs found in user studies. Here, we deepen the analysis by re-examining the survey data and differentiating between different aspects, such as web search, literature search, and social networks.

3.2 RESEARCH ETHICS AND INFORMED CONSENT

Studies received ethical approval according to our institutional policies. For studies 1 and 2, ethical approval was received from Maastricht University. Study 3 and 4 did not undergo formal ethical review, as the institution did not yet have an ethical approval process. Nonetheless, the authors adhered to ethical principles for research involving human subjects. The authors obtained informed consent after giving a thorough briefing of the studies' objectives and data handling. Measures of anonymization were taken during and after data collection. After audio transcription, files were deleted.

3.3 METHODOLOGY FOR SYNTHESIS

Our synthesis process spanned four analysis phases:

- Exploratory Phase: We individually reassessed our study data, focusing on aspects of data discovery and reuse, and then summarized our insights using shared documents and data visualizations.
- Theme Identification: We exchanged summaries and collaboratively identified five themes (i.e., data discovery challenges, needs, general web searches, academic literature, and social connections) reflecting researchers' perspectives. All five themes were salient in at least three of the studies and the wider literature. Alternative themes were either dismissed, i.e., organizational issues, or combined into one of the final five topics, i.e., different forms of social networking. Throughout this process, we documented our theme development in a shared document and added examples from each study.
- Support Specialist Data Re-analysis: We re-examined data from our studies involving support specialists, narrowing our focus solely on their perspective on data discovery. The analysis was framed around the aforementioned five themes.
- Comparative Analysis: In the analysis, we compared fresh insights on support specialists with previously documented perspectives of researchers to spotlight convergences and divergences. We began by searching for alignment between the perspectives, iteratively discussing differences and similarities in a series of weekly meetings over a period of two months, thoroughly documenting findings and decisions.

A major challenge during this process was the heterogeneity of the studies from both a methodological standpoint, study populations, and even timeframes. However, we found that the qualitative agreement between studies was high enough to minimize this challenge. We address these challenges more thoroughly in the Discussion.

4. FINDINGS

Our findings are arranged in five thematic clusters: challenges and problems in data discovery; data discovery needs; the role of general web search in data discovery; the role of academic literature; and the role of social connections. Within each cluster, we begin by presenting our findings from support specialists in a new analysis. We then discuss how the perspectives of support specialists relate to our own earlier analyses of the practices of researchers. We also indicate direct questions asked to participants in parentheses when needed for clarification, except for Study 3, which did not use a fixed set of questions.

4.1 CHALLENGES AND PROBLEMS IN DATA DISCOVERY

In Study 3, the interviewed support specialists, when asked what requests users typically made, reported that users requested help for both exploratory search and known-item searches, particularly in locating suitable data sources. Users also requested help when they

had problems accessing data. Access-related problems included challenges related to data location and retrieval, cost of access, and restricted access.

The interviewees also discussed problems users face when they have found relevant data but cannot make use of them. These problems fall into three categories: problems with erroneous data, when users find potential errors in the data; problems with documentation, when users need assistance in finding or understanding documentation; and problems with data literacy skills, when users lack the necessary skills to actually use data they have found.

In Study 4, research support specialists were asked to prioritize clusters of data discovery use cases to identify the most urgent areas for future work ('Which cluster do you think is (second) most relevant?'). The cluster of use cases related to metadata/metadata quality was ranked as being most important, followed by the clusters data citation and convenience, the last of which refers to improving the user experience of data search. Participants were also asked which clusters they felt were least important in data discovery ('Which cluster do you think is least relevant?/also not very relevant?'). Here, clusters related to lesser-known approaches, such as search within data, linking with persons, and machine discoverability, were seen as having low importance. What is somewhat contradictory is that while metadata quality was rated as highly important, the documentation cluster was often rated as least important.

Relation to researchers' perspective

Data discovery is perceived as challenging by researchers and support specialists alike. Both Gregory et al. (2020), which includes a survey of 1,630 researchers, and Krämer et al. (2021), an interview study with 12 social scientists, found that challenges in data seeking as reported by researchers are mostly about lack of data search skill, lack of data literacy, and lack of access to data. Study 3 reveals that support specialists who provide data services to researchers are aware of these challenges and work to support these challenges through training and consultancy efforts. This is supported in Cox et al. (2017).

In Gregory et al. (2019; 2020) and Krämer et al. (2021) researchers mention technical challenges to data discovery, such as a lack of suitable search tools and links between research data and other relevant sources of information, in particular the academic literature. Support staff in our studies don't see priorities in this area. Instead, they observed (Study 3) or self-assessed (Study 4) that the quality of data or data documentations is an important factor supporting data seeking behaviors. This reflects the fact that support specialist work includes tasks related to metadata curation and improving metadata and data standardization (Arora and Chakravarty, 2021; Lafia et al., 2021; Vardigan et al., 2008).

Though in our findings, data being distributed across different locations is reported as a challenge by researchers (Gregory et al., 2020) but neglected by support specialists, it is a challenge recognized by support specialists and information professionals in other studies, which describe the development of metadata standards and tools for federated dataset search to support data discoverability (Garnett et al., 2017).

4.2 DATA DISCOVERY NEEDS

The data needs of support specialists are often the data needs of their clients. One of the interviewees in Study 2 searched for data in the literature about particular technologies, such as intravascular ultrasounds, to pass on to her clients. Finding data and associated literature is just one way in which support specialists work to meet the data needs of their clients. Support specialists also report engaging in educational activities such as teaching people how to discover and evaluate data or how to curate and manage their own data (Study 4; 'What is your role in data discovery?'). Study 3 shows that support specialists get requests from users who need data for a broad array of purposes. The purposes include not only various research-related tasks but also using data for professional qualification and in journalism.

Support specialists also need data for their own research and projects. Forty percent of support respondents in Study 1 (n = 19) reported needing data both for their own research and to support others, while fewer individuals (n = 2) need data solely for their own research.

There are some differences between how researchers and support specialists use, or view the use of, secondary data by researchers (Figure 2). In Study 1, respondents were asked about their purposes for seeking data. Researchers were asked, 'Why do you use or need secondary data?' while support specialists were asked, 'Why do you or the people whom you support use secondary data?'
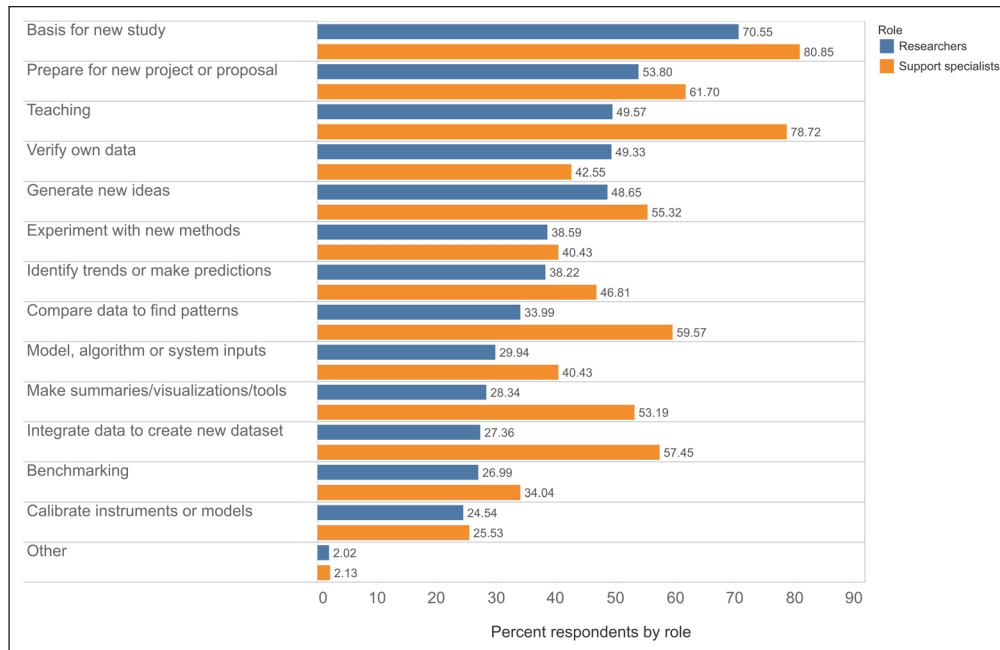

Figure 2 Purposes for using secondary data by role: researchers (n = 1630) and support specialists (n = 47). Multiple responses possible.

Approximately 50% of researcher respondents use data for teaching, while nearly 80% of support specialist respondents selected this use. In a separate question, support specialists indicated that they help others through teaching; the high frequency of using data for teaching indicated in Figure 2 mirrors this practice. Similarly, support specialists use data to make summarizations or visualizations at a higher percentage than researchers; this could be attributable to their own teaching practices also.

Support specialists also indicated that they integrate data (or believe that researchers integrate data) at higher percentages than did researchers. This could reflect the fact that data integration can be a key step in curatorial workflows for certain data, e.g., long-running survey data in the social sciences. It could also indicate a potential mismatch between how support specialists view the activities of researchers, as data integration is often discussed in both the literature and efforts around FAIR data, which is perhaps reflected in the answers of support specialists.

Support specialists in Study 3 described the data needs of users seeking survey data, as presented in support requests. This data was used to identify six categories of data needs, as presented in Figure 3 below, where categories are presented by increasing levels of specificity. Interviewees in Study 3 reported that requests with only a few specifics or very general requests require more support than very specific requests.

Figure 3 Data needs of users seeking survey data.

	Data Need	Example Requests
General	General need for any kind of data	"What data do you have?"
	Need for new research ideas	"Do you have data on any exciting topics?"
	Need for data on a specific topic	"Which new or novel data do you have on this topic?"
	Need for known data	"Someone recommended this data to me; do you have it?"
		"Do you have an updated version of this data?"
	Need for specific variables in known data	"I know the data I need, but I can't find the variable I am looking for."
Specific	Need for data with specific features	"I need longitudinal data / linked data / comparative data."

Relation to researchers' perspective

Study 3, in line with Sun (2019), Gregory et al. (2019; 2020), and Krämer et al. (2021), provides evidence that researchers have multiple evolving data needs ranging from general to specific. Not all of these uses may be cited in a research publication (Sun, 2019; Gregory et al., 2019); some data use happens outside academic research (Study 3). The data needs of support specialists commonly reflect the data needs of researchers in both social sciences and multidisciplinary studies (Gregory et al., 2019; Gregory et al., 2020; Krämer et al., 2021). The interviewed support specialists from Study 3 proved to be aware of both very general as well as very specific needs and purposes for data use. Support specialists sometimes need data for their own research (Study 1), albeit to a lesser extent than researchers. Studies 1–3, as well as Sun (2019), also highlight that both support specialists and researchers need data to use in their own teaching.

Many support specialists' own data needs are service-oriented and are tied to the support work that they do (i.e., teaching, summarizing, or visualizing data). Study 1 suggests that support specialists' perceptions and researchers' actual data needs may not always be in alignment and that support specialists are perhaps more invested in pooling or integrating datasets. This finding may reflect the curatorial nature of some support specialists' work, i.e., integrating and maintaining longitudinal survey data, as found in prior studies, for example (Lafia et al., 2021).

4.3 THE ROLE OF GENERAL WEB SEARCH IN DATA DISCOVERY

In our studies as well as in the available literature overall, there is a lack of information on how support specialists locate data by searching the web. Study 1 suggests that researchers may rely on general web search engines more heavily than support specialists, as more respondents who are researchers (59%) reported often making use of web search engines than did support specialists (40%) ('How frequently do you use the following (sources) to find data?'). This difference becomes much smaller, however, when looking at the percentages for respondents who selected never using search engines to find data (11% of researchers and 13% of support specialists).

There was also less variety in how support specialists reported their success using web search engines to find data in Study 1, with 80% of responses from support specialists indicating that they are 'sometimes successful/sometimes not successful' in their data searches with general web search engines.

Support specialists interviewed in Study 3 indicated that researchers use web searches in particular for finding data from studies that they already know. They also indicated that users expect data repositories to be as easy to use as general web search engines, which they often are not.

In contrast, support specialists in Study 4 did not recognize or prioritize the role of web search in searching for data. None of the over 200 use cases in this study refer to web search and therefore had no priority assigned to them. While absence of proof is not proof of absence, web search does not seem to be the first thing that came to the minds of the many support specialists who compiled the use cases in the study. This is different from what we found for our next two topics: using literature and social networks, which are both well-represented in Study 4.

Relation to researchers' perspective

As shown in Sun (2019), Gregory et al. (2019; 2020), and Krämer et al. (2021), in both multidisciplinary and social science studies, researchers use general web search engines extensively to discover data, as they search the web to find data repositories, conduct known-item searches, and find data directly. This is also found in prior studies (Sharifpour et al., 2022).

Study 3 shows that support specialists are aware of this practice. When asked how or where they believe that users find data, they also concede that the retrieval tools at data repositories do not meet users' expectations regarding searchability. There also appear to be differences in perspectives. Study 1 suggests that support specialists may rely less on web search but rather spread their efforts across sources. Study 4 likewise suggests that support specialists may not consider web search when thinking about supporting researchers' data discovery practices, as it was not mentioned in the use case ranking.

4.4 THE ROLE OF ACADEMIC LITERATURE IN DATA DISCOVERY

Study 1 suggests that support specialists make more use of a diversity of sources to locate data than do researchers, with support specialists selecting that they often use the literature (42%), search engines, domain repositories, and governmental sources at roughly the same percentage. More researchers reported turning to the literature with the explicit goal of finding data than did support staff.

Support specialists in Study 1 tend to view data discovery and literature discovery as distinct practices (Figure 4). In Study 1, 49% of support specialists (n = 23) stated that their processes for finding data and literature are always different, while less than 20% of researchers responded similarly to this question (Figure 4).
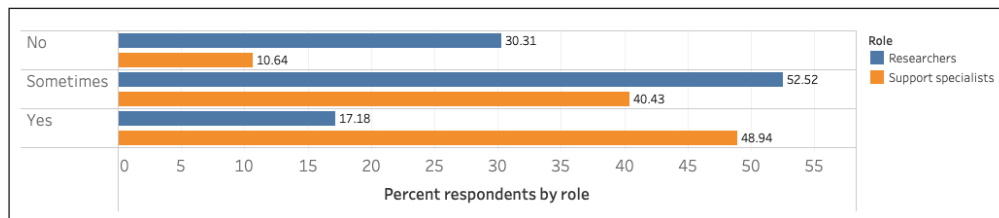
Figure 4 Responses to the question 'Do you discover data differently than the literature?' by role: researchers (n = 1630) and support specialists (n = 47).

In Study 3, the interviewed support specialists reported that research papers are one of the main sources for learning about existing data for data archive users. One support specialist also reported observing a frequent use of DOI references to enter records in the data catalog, which suggests that users are citation chaining from papers. In particular for students or young researchers, textbooks were seen to be another important source for discovering data.

Of all the data discovery use cases collected by support specialists in Study 4, data citation had the greatest number of use cases. However, when it came to prioritization, some ranked it as high priority, while others explicitly ranked it as low priority. (Participants could only pick two of each).

Relation to researchers' perspective

A point of alignment of perspectives is the use of academic literature. All studies show in various ways that academic literature is important for researchers to discover data. The degree of importance, however, seems to be less pronounced among data support specialists (Study 1 and 4), although they do seem to recognize that this is common practice among researchers (Study 3). Our findings also indicate that support specialists use literature less as an avenue for finding data, perhaps seeing data discovery as distinct from literature discovery. Nevertheless, there are increasing efforts in linking research data to papers to enhance data discoverability, such as Scholix (Burton et al., 2017) and knowledge graphs (Aryani et al., 2018; Färber and Lamprecht, 2021; Jaradeh et al., 2019; Manghi et al., 2019).

4.5 THE ROLE OF SOCIAL CONNECTIONS IN DATA DISCOVERY

Personal networks are also important for support specialists. Study 3 revealed that community involvement plays an important role in finding data. According to data support specialists, researchers may be directed to use certain data, for example, by supervisors. Others exchange datasets with peers, sometimes with access restriction (*data dealing*, Friedrich, 2020). The interviews also revealed that support specialists make use of a larger data-related community when catering to users' needs. These data communities include, for instance, principal investigators who have collected the data and work closely with data professionals to prepare the data for archiving, publication, or reuse. The interviews also suggest that data reusers themselves are part of these communities, for example, when they are invited to detect and report errors in datasets or make other suggestions for improvements.

In Study 4, the use case cluster regarding social connections was one of the smallest and was not seen as relevant by the support specialists.

Relation to researchers' perspective

With exception of Study 4, all studies show that social connections are viewed as important by both researchers and support specialists, though the ways of connecting may be different (Study 1). For researchers across disciplines, social connections are an important means to find data (Sun, 2019; Gregory et al., 2019; Gregory et al., 2020; Krämer et al., 2021) and are of value in reusing and making sense of data (Krämer et al., 2021). This reliance is in line with many other studies on data discovery practices and data reuse (Kim, 2017; Koesten et al., 2017; Yoon, 2017; Yoon and Lee, 2019).

Sun (2019) suggests that the relevance of social connections may depend on the accessibility of the data; finding open datasets may not require social interactions as much as finding access-restricted datasets. Study 3 shows support specialists confirm this finding and also describes the composition of these social networks, which include data producers, primary researchers, and support specialists. For support specialists, these connections are important to prepare data as well as to assist researchers in finding and reusing data.

Despite this awareness, social connections are not defined as a formal way of data discovery that deserves attention from the support specialist community (Study 4). This contrasts with the fact that support work related to data discovery is itself a collaborative process that involves many personal exchanges (Gregory et al., 2019; Joo and Schmidt, 2021).

5. DISCUSSION

5.1 MAIN INSIGHTS FROM THE SYNTHESIS

First, an interesting point of distinction from the perspective of support specialists is the separation between data discovery and literature discovery. Support specialists across disciplines may tend to view data discovery as a distinct practice in comparison to literature search, while researchers tend to see them as more interconnected. This may be explained by the different roles of the groups. Researchers tend to be highly specialized within their domains. Literature is important to them to gauge the impact of their work, including working with data, within their community. Research support specialists, on the other hand, often search data as part of a delegated task without the need to understand the bigger picture of the research. They have more hands-on experience with data management, curation, and preservation, which could lead them to view data discovery as a separate practice that requires specific skills and expertise.

Secondly, our studies consistently show a blurring of lines between 'research' and 'support' roles, a theme that resonates with findings from Teperek et al. (2022). In Study 1, we found that support specialists, while aiding others in data discovery, may also require data for their own projects, signaling a shift from traditional support functions to more multifaceted roles. Reciprocally, researchers might be involved in data preparation and curation alongside their primary role of using data for investigations. These dual roles challenge the traditional delineation between 'support' and 'research' roles.

Study 3 explored this dynamic further, illuminating the role of support specialists within, what we choose to call, data communities—networks comprising not only data collectors, such as principal investigators, but also data managers, curators, archivists, and librarians. Our findings highlight the active involvement of support specialists in research and data communities, leveraging their social connections to assist researchers in finding data. This not only emphasizes their crucial role in fostering data discovery but also indicates that 'research' and 'support' roles may become less clear as these two professions are brought together in the 'data communities'. Study 3 further reveals that researchers and curators often collaboratively enhance the value of data, with each bringing unique expertise to the table. In certain specialized cases, such as within social science data archives, researchers themselves exclusively take on data curation tasks, applying their expertise to maximize data utility and integrity.

It is in this context that we draw Figure 5 to graphically illustrate the interconnectivity and overlap of roles in the data discovery process, showing the fluidity with which professionals in 'research' and 'support' now operate.
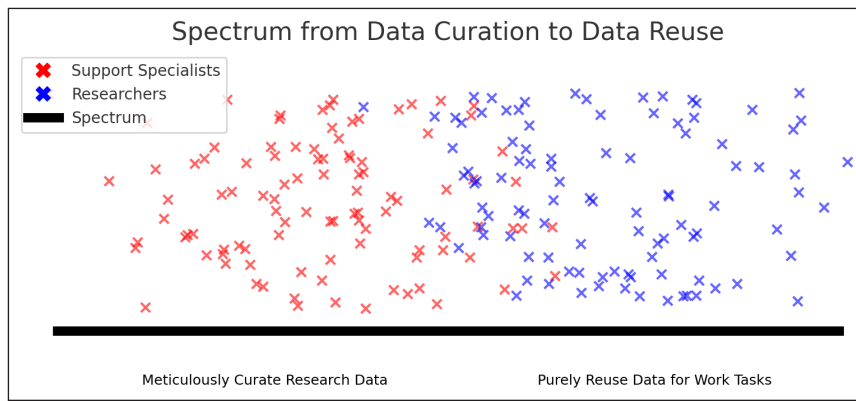
Spectrum from Data Curation to Data Reuse

Meticulously Curate Research Data   Purely Reuse Data for Work Tasks

Figure 5 Conceptual illustration of role distribution in RDM. *This abstract representation, not based on actual data, contrasts support specialists (red) favoring meticulous data curation with researchers (blue) inclined towards data reuse.*

5.2 RECOMMENDATIONS FOR DIFFERENT AREAS OF SUPPORT WORK

We conclude by thinking about how the different types of support work identified in Section 2.2 could better support the data discovery practices of researchers by identifying existing best practices and by making suggestions for improvement.

People-oriented support work, especially in university libraries, may consider embedding data literacy skills into the fabric of research support services. Data literacy encompasses a broad skill set, from locating relevant datasets to mastering sophisticated data analysis techniques. Training should not only encompass fundamental data search techniques and the evaluation of results but also consider the implications of employing AI tools in these searches. This recommendation emerges from observed challenges that support professionals encounter in data discovery (Section 4.1 above). Recognizing the varied needs of researchers (Section 4.2), we advocate for bespoke training programs that align with researchers' diverse requirements at various research stages.

Our findings, particularly in Section 4.4, indicate that researchers often rely on literature as a pathway to data. Given that data literacy training is mostly an optional service in university libraries, we suggest a strategic shift towards a unified approach to the teaching of data and literature discovery.

We suggest that support specialists co-develop curricula with professors and other members of the research community and contribute to the nurturing of data communities. Social connections are invaluable in this context, as highlighted in Section 4.5. Fostering interdisciplinary cooperation can enhance data quality and promote its reuse.

Metadata-oriented support work may need improvement to meet the evolving needs of researchers. Metadata quality is a problem, as seen in Study 3, where data seekers were not satisfied with documentation. Another problem involves determining the optimal granularity for contextual information in data documentation, as this significantly influences data's reusability.

Another promising strain of metadata-oriented support work is to continue expanding the scope of existing initiatives to build links between research data and other research materials and entities (e.g., literature, researchers, research institutions). Our studies reveal that researchers depend on such interlinkages to discover and evaluate data for reuse, whereas support specialists have yet to prioritize these link-building efforts, signaling a critical area for development.

Infrastructure-oriented support work may fork into the development of search tools customized for different data reuse tasks (Koesten et al., 2017). To do this, an in-depth understanding of data-centric reuse tasks is needed. As we have found, researchers appear to have multiple, evolving data needs. We call for more user studies into researchers' data reuse behaviors, needs, and requirements.

We also conclude that more user studies are needed to improve the findability of research data via general web search engines and to increase the usability of data repositories, which may be a topic for both metadata-oriented and infrastructure-oriented support work. It is not uncommon that researchers start with web search engines for searching and end up in repositories. While existing efforts focus on how to make research data more discoverable, i.e., through using Linked Open Data or the use of schema.org vocabularies, we also observe an equally relevant issue: retaining researchers after they land in a data repository.

5.3 LIMITATIONS

Putting the findings from these different studies together is challenging given their different methods, subject demographics, and levels of abstraction. The studies thus do not support representativity but provide in-depth data from a variety of perspectives.

RDM is a quickly evolving topic, and the studies we analyze cover multiple years, in which much has changed. The FAIR data principles (Wilkinson et al., 2016) evolved from an idea in 2016 to a widespread movement. The field of data discovery has evolved from being a specialized subcategory within data reuse and information retrieval research to becoming a distinct topic with dedicated support from associations such as the Research Data Alliance. Despite these changes, we still see consistency in our results across time.

6. CONCLUSION

In this synthesis, we have compared the practices and perspectives of support specialists and researchers. We have found differences in the perception of challenges and problems of finding data; of data needs; of the role of web search to find data; of the role of literature as a source of data discovery; and of the role of social connections.

Our results sharpen awareness of the relations between the two perspectives, but further studies are needed to yield additional insight and more detailed answers to our research questions. Furthermore, we want to emphasize that while support specialists and researchers have different perspectives, they also share a community space regarding practices of data discovery and data use. As we have found, both groups are part of larger 'data communities' where people with different roles rely on each other to find and use, but also to prepare and enhance data. Therefore, it could be very insightful to conduct studies that bring together researchers and support specialists in co-design activities.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

All authors contributed to the study conception and design. All authors contributed the methods and findings of their respective studies. The first draft and consecutive reviews were done collaboratively by all authors. All authors contributed equally to the discussion and conclusion. Literature review was primarily done by Guangyuan Sun. Guangyuan Sun also developed the typology of support work.

AUTHOR AFFILIATIONS

Guangyuan Sun orcid.org/0000-0001-7352-2158
National Institute of Education, Nanyang Technological University, Singapore

Tanja Friedrich orcid.org/0000-0003-1557-3728
German Aerospace Center, Scientific Information, Cologne, Germany

Kathleen Gregory orcid.org/0000-0001-5475-8632
Centre for Science and Technology Studies, Leiden University, Netherlands

Brigitte Mathiak orcid.org/0000-0003-1793-9615
GESIS – Leibniz Institute for the Social Sciences, Cologne, Germany

REFERENCES

Almeida, A.V. de, Borges, M.M. and **Roque, L.** (2017) 'The European open science cloud: A new challenge for Europe', in *Proceedings of the 5th International Conference on Technological Ecosystems for Enhancing Multiculturality*. New York, NY: Association for Computing Machinery, pp. 1–4. Available at: https://doi.org/10.1145/3144826.3145382

Arora, S. and **Chakravarty, R.** (2021) 'Making research data discoverable: an outreach activity of Datacite', *Library Philosophy and Practice (e-journal)*, p. 5199.

Aryani, A. *et al.* (2018) 'A research graph dataset for connecting research data repositories using RD-Switchboard', *Scientific Data*, 5(1), p. 180099. Available at: https://doi.org/10.1038/sdata.2018.99

Ashiq, M., Usmani, M.H. and **Naeem, M.** (2020) 'A systematic literature review on research data management practices and services', *Global Knowledge, Memory and Communication*, 71(8/9), pp. 649–671. Available at: https://doi.org/10.1108/GKMC-07-2020-0103

Bishop, B.W. *et al.* (2019) 'Scientists' data discovery and reuse behavior: (Meta)data fitness for use and the FAIR data principles', *Proceedings of the Association for Information Science and Technology*, 56(1), pp. 21–31. Available at: https://doi.org/10.1002/pra2.4

Borgman, C.L., Scharnhorst, A. and **Golshan, M.S.** (2019) 'Digital data archives as knowledge infrastructures: Mediating data sharing and reuse', *Journal of the Association for Information Science and Technology*, 70(8), pp. 888–904. Available at: https://doi.org/10.1002/asi.24172

Borst, T. and **Limani, F.** (2020) 'Patterns for searching data on the web across different research communities', *LIBER Quarterly: The Journal of the Association of European Research Libraries*, 30(1), pp. 1–21. Available at: https://doi.org/10.18352/lq.10317

Brickley, D., Burgess, M. and **Noy, N.** (2019) 'Google dataset search: Building a search engine for datasets in an open web ecosystem', in *The World Wide Web Conference*. New York, NY: Association for Computing Machinery, pp. 1365–1375. Available at: https://doi.org/10.1145/3308558.3313685

Bugaje, M. and **Chowdhury, G.** (2018a) 'Data retrieval = text retrieval?', in G. Chowdhury, J. McLeod, V. Gillet, and P. Willett (eds.), *Transforming Digital Worlds*. Presented at the iConference 2018. Cham: Springer International Publishing, pp. 253–262. Available at: https://doi.org/10.1007/978-3-319-78105-1_29

Bugaje, M. and **Chowdhury, G.** (2018b) 'Identifying design requirements of a user-centered research data management system', in M. Dobreva, A. Hinze, and M. Žumer (eds.), *Maturity and Innovation in Digital Libraries*. Presented at the ICADL 2018. Cham: Springer International Publishing, pp. 335–347. Available at: https://doi.org/10.1007/978-3-030-04257-8_35

Burton, A. *et al.* (2017) 'The Scholix framework for interoperability in data-literature information exchange', *D-Lib Magazine*, 23(1/2). Available at: https://doi.org/10.1045/january2017-burton

Cope, J. and **Baker, J.** (2017) 'Library carpentry: Software skills training for library professionals', *International Journal of Digital Curation*, 12(2), pp. 266–273. Available at: https://doi.org/10.2218/ijdc.v12i2.576

Cox, A. (2018) 'Academic librarianship as a data profession: The familiar and unfamiliar in the data role spectrum', *eLucidate*, 15(1–2). Available at: https://elucidate-ukeig.org.uk/index.php/elucidate/article/view/251 (Accessed: 3 November 2023).

Cox, A. (2023) 'How artificial intelligence might change academic library work: Applying the competencies literature and the theory of the professions', *Journal of the Association for Information Science and Technology*, 74(3), pp. 367–380. Available at: https://doi.org/10.1002/asi.24635

Cox, A.M. *et al.* (2017) 'Developments in research data management in academic libraries: Towards an understanding of research data service maturity', *Journal of the Association for Information Science and Technology*, 68(9), pp. 2182–2200. Available at: https://doi.org/10.1002/asi.23781

Cox, A.M. *et al.* (2019) 'Maturing research data services and the transformation of academic libraries', *Journal of Documentation*, 75(6), pp. 1432–1462. Available at: https://doi.org/10.1108/JD-12-2018-0211

Cruz, M. *et al.* (2019) 'Policy needs to go hand in hand with practice: The learning and listening approach to data management', *Data Science Journal*, 18(1), p. 45. Available at: https://doi.org/10.5334/dsj-2019-045

Cruz, M.J. *et al.* (2018) 'From passive to active, from generic to focussed: How can an institutional data archive remain relevant in a rapidly evolving landscape?', *International Journal of Digital Curation*, 13(1), pp. 385–391. Available at: https://doi.org/10.2218/ijdc.v13i1.613

Curty, R.G. (2016) 'Factors influencing research data reuse in the social sciences: An exploratory study', *International Journal of Digital Curation*, 11(1), pp. 96–117. Available at: https://doi.org/10.2218/ijdc.v11i1.401

Darch, P.T. *et al.* (2020) 'Library cultures of data curation: Adventures in astronomy', *Journal of the Association for Information Science and Technology*, 71(12), pp. 1470–1483. Available at: https://doi.org/10.1002/asi.24345

Färber, M. and **Lamprecht, D.** (2021) 'The data set knowledge graph: Creating a linked open data source for data sets', *Quantitative Science Studies*, 2(4), pp. 1324–1355. Available at: https://doi.org/10.1162/qss_a_00161

Friedrich, T. (2020) *Looking for data: Information seeking behaviour of survey data users*. PhD thesis., Humboldt-Universität zu Berlin. Available at: https://edoc.hu-berlin.de/handle/18452/22813 (Accessed: 10 March 2022).

Garnett, A. *et al.* (2017) 'Open metadata for research data discovery in Canada', *Journal of Library Metadata*, 17(3–4), pp. 201–217. Available at: https://doi.org/10.1080/19386389.2018.1443698

Gowen, E. and **Meier, J.J.** (2020) 'Research data management services and strategic planning in libraries today: A longitudinal study', *Journal of Librarianship and Scholarly Communication*, 8(1), p. eP2336. Available at: https://doi.org/10.7710/2162-3309.2336

Gregory, K. (2020) 'A dataset describing data discovery and reuse practices in research', *Scientific Data*, 7(1), p. 232. Available at: https://doi.org/10.1038/s41597-020-0569-5

Gregory, K. *et al.* (2020) 'Lost or found? Discovering data needed for research', *Harvard Data Science Review*, 2(2). Available at: https://doi.org/10.1162/99608f92.e38165eb

Gregory, K.M. *et al.* (2019) 'Understanding data search as a socio-technical practice', *Journal of Information Science*, 46(4), pp. 459–475. Available at: https://doi.org/10.1177/0165551519837182

Hedeland, H. (2020) 'Providing digital infrastructure for audio-visual linguistic research data with diverse usage scenarios: Lessons learnt', *Publications*, 8(2), p. 33. Available at: https://doi.org/10.3390/publications8020033

Hemphill, L. *et al.* (2022) 'How do properties of data, their curation, and their funding relate to reuse?', *Journal of the Association for Information Science and Technology*, 73(10), pp. 1432–1444. Available at: https://doi.org/10.1002/asi.24646

Huang, Y., Cox, A.M. and **Sbaffi, L.** (2021) 'Research data management policy and practice in Chinese university libraries', *Journal of the Association for Information Science and Technology*, 72(4), pp. 493–506. Available at: https://doi.org/10.1002/asi.24413

Jaradeh, M.Y. *et al.* (2019) 'Open research knowledge graph: A system walkthrough', in A. Doucet, A. Isaac, K. Golub, T. Aalberg, and A. Jatowt (eds.), *Digital Libraries for Open Knowledge*. Cham: Springer International Publishing, pp. 348–351. Available at: https://doi.org/10.1007/978-3-030-30760-8_31

Joo, S., Kim, S. and **Kim, Y.** (2017) 'An exploratory study of health scientists' data reuse behaviors: Examining attitudinal, social, and resource factors', *Aslib Journal of Information Managagement*, 69, pp. 389–407. Available at: https://doi.org/10.1108/AJIM-12-2016-0201

Joo, S. and **Schmidt, G.M.** (2021) 'Research data services from the perspective of academic librarians', *Digital Library Perspectives*, 37(3), pp. 242–256. Available at: https://doi.org/10.1108/DLP-10-2020-0106

Joo, Y.K. and **Kim, Y.** (2017) 'Engineering researchers' data reuse behaviours: a structural equation modelling approach', *The Electronic Library*, 35, pp. 1141–1161. Available at: https://doi.org/10.1108/EL-08-2016-0163

Kacprzak, E. *et al.* (2018) 'Characterising dataset search queries', in *Companion Proceedings of the The Web Conference 2018*. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, pp. 1485–1488. Available at: https://doi.org/10.1145/3184558.3191597

Khan, N., Thelwall, M. and **Kousha, K.** (2021) 'Are data repositories fettered? A survey of current practices, challenges and future technologies', *Online Information Review*, ahead-of-print(ahead-of-print). Available at: https://doi.org/10.1108/OIR-04-2021-0204

Kim, Y. (2017) 'Fostering scientists' data sharing behaviors via data repositories, journal supplements, and personal communication methods', *Information Processing & Managagement*, 53, pp. 871–885. Available at: https://doi.org/10.1016/j.ipm.2017.03.003

Koesten, L.M. *et al.* (2017) 'The trials and tribulations of working with structured data: A study on information seeking behaviour', in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. New York, NY: Association for Computing Machinery, pp. 1277–1289. Available at: https://doi.org/10.1145/3025453.3025838

Krämer, T., Papenmeier, A., Carevic, Z., Kern, D. and **Mathiak, B.** (2021) 'Data-Seeking Behaviour in the Social Sciences', *International Journal on Digital Libraries*, 22(2), pp. 175–195. Available at: https://doi.org/10.1007/s00799-021-00303-0

Lafia, S. *et al.* (2021) 'Leveraging machine learning to detect data curation activities', in *2021 IEEE 17th International Conference on eScience (eScience)*. Presented at the 2021 IEEE 17th International Conference on eScience (eScience). Innsbruck, Austria: IEEE, pp. 149–158. Available at: https://doi.org/10.1109/eScience51609.2021.00025

Liu, S. *et al.* (2021) 'An infrastructure with user-centered presentation data model for integrated management of materials data and services', *npj Computational Materials*, 7(1), pp. 1–8. Available at: https://doi.org/10.1038/s41524-021-00557-x

Lund, B.D. *et al.* (2020) 'Perceptions toward artificial intelligence among academic library employees and alignment with the diffusion of innovations' adopter categories', *College & Research Libraries*, 81(5), pp. 865–882. Available at: https://doi.org/10.5860/crl.81.5.865

Manghi, P. *et al.* (2019) *The OpenAIRE research graph data model*. Available at: https://doi.org/10.5281/zenodo.2643199

Mannheimer, S. *et al.* (2021) 'Dataset search: A lightweight, community-built tool to support research data discovery', *Journal of eScience Librarianship*, 10(1), p. e1189. Available at: https://doi.org/10.7191/jeslib.2021.1189

Marlina, E. and **Purwandari, B.** (2019) 'Strategy for research data management services in Indonesia', *Procedia Computer Science*, 161, pp. 788–796. Available at: https://doi.org/10.1016/j.procs.2019.11.184

Mathiak, B. *et al.* (2023) 'What are researchers' needs in data discovery? Analysis and ranking of a large-scale collection of crowdsourced use cases', *Data Science Journal*, 22(1), p. 3. Available at: https://doi.org/10.5334/dsj-2023-003

Mulongo, J. *et al.* (2022) 'Investigation of research data maturity in academic libraries of developed countries', in *Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance*. New York, NY: Association for Computing Machinery, pp. 288–300. Available at: https://doi.org/10.1145/3560107.3560153

R Core Team. (n.d.) *R: a language and environment for statistical computing*. Available at https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing (Accessed 13 July 2022).

Read, K. *et al.* (2015) 'Promoting data reuse and collaboration at an academic medical center', *International Journal of Digital Curation*, 10(1), pp. 260–267. Available at: https://doi.org/10.2218/ijdc.v10i1.366

Reichmann, S. *et al.* (2021) 'Between administration and research: Understanding data management practices in an institutional context', *Journal of the Association for Information Science and Technology*, 72(11), pp. 1415–1431. Available at: https://doi.org/10.1002/asi.24492

Savage, J.L. and **Cadwallader, L.** (2019) 'Establishing, developing, and sustaining a community of data champions', *Data Science Journal*, 18(1), p. 23. Available at: https://doi.org/10.5334/dsj-2019-023

Sharifpour, R., Wu, M. and **Zhang, X.** (2022) 'Large-scale analysis of query logs to profile users for dataset search', *Journal of Documentation*, ahead-of-print(ahead-of-print). Available at: https://doi.org/10.1108/JD-12-2021-0245

Sheridan, H. *et al.* (2021) 'Data curation through catalogs: A repository-independent model for data discovery', *Journal of eScience Librarianship*, 10(3). Available at: https://doi.org/10.7191/jeslib.2021.1203

Si, L. *et al.* (2019) 'Investigation and analysis of research support services in academic libraries', *The Electronic Library*, 37(2), pp. 281–301. Available at: https://doi.org/10.1108/EL-06-2018-0125

Smit, M. (2019) 'Code convention adherence in research data infrastructure software: An exploratory study', in *2019 IEEE International Conference on Big Data (Big Data)*. Presented at the 2019 IEEE International Conference on Big Data (Big Data). pp. 4691–4700. Available at: https://doi.org/10.1109/BigData47090.2019.9006130

Sun, G. (2019) *Knowledge representation of social science quantitative research data for data curation and reuse*. PhD thesis. Nanyang Technological University. Available at: https://dr.ntu.edu.sg/handle/10356/107575 (Accessed 12 July 2022).

Tammaro, A.M. *et al.* (2019) 'Data curator's roles and responsibilities: An international perspective', *Libri*, 69(2), pp. 89–104. Available at: https://doi.org/10.1515/libri-2018-0090

Tang, R. and **Hu, Z.** (2019) 'Providing research data management (RDM) services in libraries: Preparedness, roles, challenges, and training for RDM practice', *Data and Information Management*, 3(2), pp. 84–101. Available at: https://doi.org/10.2478/dim-2019-0009

Teperek, M., Cruz, M. and **Kingsley, D.** (2022) 'Time to re-think the divide between academic and support staff', *Nature*. Available at: https://doi.org/10.1038/d41586-022-01081-8

Vardigan, M., Heus, P. and **Thomas, W.** (2008) 'Data documentation initiative: Toward a standard for the social sciences', *International Journal of Digital Curation*, 3(1), pp. 107–113. Available at: https://doi.org/10.2218/ijdc.v3i1.45

Wilkinson, M.D. *et al.* (2016) 'The FAIR guiding principles for scientific data management and stewardship', *Scientific Data*, 3(1), p. 160018. Available at: https://doi.org/10.1038/sdata.2016.18

Wilson, L., Colborne, A. and **Smit, M.** (2017) 'Preparing data managers to support open ocean science: Required competencies, assessed gaps, and the role of experiential learning', *Presented at the 2017 IEEE International Conference on Big Data (Big Data)*. IEEE Computer Society, pp. 3984–3993. Available at: https://doi.org/10.1109/BigData.2017.8258412

Wu, M. *et al.* (2019) 'Data discovery paradigms: User requirements and recommendations for data repositories', *Data Science Journal*, 18(1), p. 3. Available at: https://doi.org/10.5334/dsj-2019-003

Yoon, A. (2017) 'Role of communication in data reuse', *Proceedings of the Association for Information Science and Technology*, 54, pp. 463–471. Available at: https://doi.org/10.1002/pra2.2017.14505401050

Yoon, A. and **Lee, Y.Y.** (2019) 'Factors of trust in data reuse', *Online Information Review*, 43, pp. 1245–1262. Available at: https://doi.org/10.1108/oir-01-2019-0014