

---

# Findability and Discoverability of Qualitative Research Data in Social Sciences

## An Explorative Study Using re3data API and Systematic Evidence Searching

Master's Thesis

Master in Library and Information Science

at Faculty of Information Science and Communication Studies

TH Köln – University of Applied Sciences

Submitted by: Inji Kim

Submitted to: Prof. Dr. Mirjam Blümm

Second Reviewer: Prof. Dr. Klaus Lepsky

Porto, 30.08.2024

## Abstract

This thesis explores the findability and discoverability of qualitative research data within the social sciences, focusing on three key domains: discipline-specific repositories, scholarly journal articles, and scholarly journals. Employing an explorative and descriptive approach, the study investigates the extent to which data management practices support the finding and discovery of qualitative datasets. The research identifies that while repositories commonly use persistent identifiers (PIDs) and discipline-specific metadata standards, the availability of advanced search functionalities remains limited, which may hinder data findability. The study also highlights a concerning scarcity of data availability statements in scholarly articles, with only 17% including such statements, and reveals that data sharing is often impeded by ethical constraints. Furthermore, a significant gap was found between the datasets stored in repositories and their representation in scholarly literature, underscoring the need for better integration of repository-held data into academic discourse. The findings suggest that both repositories and journals must enhance their practices to promote the visibility of qualitative research data, with recommendations for improved search tools, stricter enforcement of data availability statement policies, and stronger linkages between datasets and scholarly publications. This research contributes to the ongoing discourse on open science and the challenges associated with qualitative data sharing in the social sciences.

*Keywords:* qualitative research; social sciences; research data repositories; data discoverability; data findability; FAIR Principles; Open Science; scholarly communication

## Table of Contents

<i>Abstract</i> .....	<i>I</i>
<i>Table of Contents</i> .....	<i>II</i>
<i>List of Tables</i> .....	<i>IV</i>
<i>List of Figures</i> .....	<i>IV</i>
<b>1 Introduction</b> .....	<b>1</b>
1.1 Background .....	1
1.2 Scope of the Study, Research Questions and Hypotheses .....	2
1.3 Structure of the Thesis .....	2
<b>2 Qualitative Research Data in Social Sciences</b> .....	<b>4</b>
2.1 Definition of Qualitative Research Data .....	4
2.2 Collection Methods and Types of Qualitative Data in Social Sciences.....	5
2.2.1 Verbal Data .....	5
2.2.2 Ethnographic Data.....	6
2.2.3 Material Data.....	8
2.3 Ethical Considerations on Qualitative Data Collection and Sharing.....	9
<b>3 Knowledge Infrastructure and Data Landscape for Qualitative Social Sciences...</b>	<b>11</b>
3.1 European, National and International Data Infrastructure .....	12
3.2 Research Data Centres and Repositories.....	15
3.3 Data Supplements in Journals and Data Journals .....	19
3.4 Summary .....	21
<b>4 Findability and Discoverability of Qualitative Research Data in Social Sciences..</b>	<b>22</b>
4.1 Findability and Discoverability in Relation to Information Needs (IN) .....	22
4.2 The FAIR Data Principles and Other Aspects for Data Findability and Discoverability.....	25
4.2.1 Persistent Identifiers (PID) .....	26
4.2.2 Metadata Standards and Data Documentation .....	28
4.2.3 Controlled Vocabularies (CV).....	30
4.2.4 Data Citation and Linkage of Scholarly Literature with Datasets.....	33
4.2.5 Inclusion in Search Services .....	34
4.2.6 Social Interactions.....	37
4.3 Summary.....	37
<b>5 Methods</b> .....	<b>39</b>

5.1	Study Design .....	39
5.2	Data Collection.....	39
5.2.1	Dataset Searching in Discipline-Specific Repositories .....	40
5.2.2	Tracing Data in Scholarly Journal Articles.....	41
5.2.3	Variable Summary.....	43
5.3	Data Analysis.....	43
6	<i>Results</i> .....	44
6.1	Sample Group 1: Discipline-Specific Repositories.....	44
6.1.1	The Role of re3data.org.....	44
6.1.2	PID Systems .....	46
6.1.3	Metadata Standards and Search Functionalities .....	47
6.1.4	Browsing with Controlled Vocabularies.....	48
6.1.5	Inclusion in Search Services through OAI-PMH .....	49
6.1.6	Data Citation and Guidance on Qualitative Data Documentation.....	50
6.2	Sample Group 2: Scholarly Journal Articles.....	50
6.2.1	Data Available Statement and Available Data .....	50
6.2.2	Publication Year and Data Available Statement .....	51
6.3	Sample Group 3: Scholarly Journals .....	51
6.3.1	Data Policy on Data Availability Statement and Data Sharing .....	52
6.3.2	Guideline on Data Citation .....	52
6.3.3	Overall Score .....	53
7	<i>Summary and Conclusion</i> .....	54
7.1	Key Findings.....	54
7.2	Unexpected Findings.....	55
7.3	Implications and Recommendations.....	55
7.4	Conclusion.....	56
	<i>References</i> .....	57
	<i>Digital Appendix</i> .....	82
	Appendix A: Repository and Metadata Retrieval Using re3data.org REST API .....	82
	Appendix B: Data Search in Discipline-Specific Repositories.....	82
	Appendix C: Tracing Data in Scholarly Journal Articles .....	83
	Appendix D: Summary of Variables and Analysis.....	83

## List of Tables

Table 3. 1 Summary of Search Services for Research Data .....	14
Table 4. 1 Aspects Influencing Data Discoverability and Findability .....	25
Table 4. 2 PID Systems Implemented by CESSDA Members .....	27
Table 4. 3 Example GET Request to a Repository's OAI-PMH Endpoint .....	35

## List of Figures

Figure 3. 1 European Qualitative Data Collections and Archives Map.....	17
Figure 6. 1 Inclusion of the Term "Qualitative" in re3data Records of Selected Repositories .....	45
Figure 6. 2 Search results in DANS Data Station Social Sciences and Humanities, AUSSDA, and ADA with the search term "qualitative" .....	46
Figure 6. 3 PID System Implemented in Social Science Repositories .....	46
Figure 6. 4 High-Level and Discipline-Specific Metadata Standards in Repositories ....	48
Figure 6. 5 OAI-PMH and Inclusion in Search Services .....	49
Figure 6. 6 Available Qualitative Data in Journal Articles .....	50
Figure 6. 7 Publication Year and Data Availability Statement .....	51
Figure 6. 8 Journal Data Policy on DAS/Data Sharing .....	52
Figure 6. 9 Distribution of Overall Score in Journals.....	53

# 1 Introduction

## 1.1 Background

Qualitative social science (QSS) generates rich and unique data, encompassing a variety of types, including verbal, ethnographic, and material data. These data are collected through a wide range of methods, such as qualitative interviews, oral histories, conversation analysis, observations, ethnography, and videography (Flick, 2018a). By capturing partial realities derived from the subjective experiences of individuals and communities, qualitative research provides evidence that facilitates our understanding and connection with the world around us (Barbour, 1999; Haase & Myers, 1988). The collection of such data requires considerable effort and can possess heritage value, serving as important resources for future use, which presents a compelling argument for sharing qualitative social research data (Berkowitz & Delacour, 2022; Corti, 2005). Furthermore, qualitative data sharing (QDS) promotes transparency and fosters synergies in research, enabling the restudy of existing material and facilitating further research endeavours, such as comparative studies, secondary analysis, text mining, and various other methodologies (Corti & Thompson, 2004; Hammersley, 1997)

In social research, digital data archiving initially focused on quantitative data, starting with opinion polls and election data in the 1960s and expanding to survey data by the late 1970s (Corti, 2012). In 1994, the first qualitative data archive, the Qualidata Resource Centre, was established at the University of Essex in the United Kingdom (Corti & Thompson, 1996). Subsequently, in 1995, the UK Economic and Social Research Council (ESRC) mandated the submission of materials produced by the qualitative research it funds as part of its research data policy (Corti, 2005).

With the growing discourse on the benefits and challenges of sharing and reusing qualitative research data (QRD), as well as the justifications for such endeavours (Antes et al., 2018; Bishop, 2009a; Branney et al., 2023; Chauvette et al., 2019; Corti, 2005; Corti & Thompson, 2004; Hammersley, 1997, 2010; Huma & Joyce, 2023; Joyce et al., 2022; V. Karhulahti, 2023; Kuula, 2011; Mannheimer et al., 2019; Mauthner et al., 1998; Mozersky et al., 2020; Parry & Mauthner, 2005; Pownall et al., 2023), there has been an increasing trend to encourage researchers, research institutions, funders, and publishers to adopt QDS practices (Chauvette et al., 2019; Crosas et al., 2018; K. Jones & Alexander, 2018; Mozersky et al., 2020).

Despite the increasing number of shared qualitative datasets and instances of their reuse (Bishop & Kuula-Luumi, 2017; Kuula, 2011), ethical and legal challenges, such as the duty of confidentiality to participants, obtaining informed consent, and de-identification (Bishop, 2009b; Hammersley, 1997; D. Thomson et al., 2005; Ursin & Bentzen, 2021), as well as methodological challenges, such as capturing the full context in shared data forms (Chauvette et al., 2019), continue to impede QDS practices. Consequently, shared qualitative data remain relatively rare compared to quantitative data.

## 1.2 Scope of the Study, Research Questions and Hypotheses

This explorative study is premised on the hypothesis that archived qualitative research data remain rare, and that searching for and discovering these data are challenging tasks. Consequently, this research focuses on the findability and discoverability of shared QRD based on various attributes. To frame the study, a comprehensive literature review and web searches were conducted to address the following foundational questions:

- What constitutes qualitative data, how is it generated in the social sciences, and what types of qualitative data exist?
- What infrastructures are available to assist in the finding and discovery of (qualitative) research data?
- Where can qualitative research data for social sciences be found?
- How can findability and discoverability be conceptualized in the context of qualitative research data?
- What are the key attributes that enhance the findability and discoverability of these data?

After this comprehensive investigation, this study seeks to answer the following primary research question:

*To what extent are data findability and discoverability practices implemented in qualitative social sciences?*

Subsequently, an empirical quantitative study is designed to answer this question. Through the literature review and the empirical study, this thesis aims to investigate the current data landscape in qualitative social research and the practices involved in finding and discovering qualitative data within the social sciences.

This study focuses on research data that are submitted to a repository for storage, curation, and access (Berg-Cross et al., 2015). Despite the multi-typed nature of qualitative data, Open Government Data (OGD), Big Data, and social media data will be excluded within the scope of this study (Borgman, 2015).

## 1.3 Structure of the Thesis

Firstly, the definition of qualitative data is discussed in the next chapter for the understanding of the objective of this thesis. Furthermore, based on the handbook edited by Uwe Flick (2018a), the varying nature of approaches to gather qualitative data, research goals pursued through these methods, and the resulting data types are briefly introduced. This is followed by brief introduction to the discourse on ethical and methodological debates on qualitative data collection and sharing.

In the third chapter, the knowledge infrastructure and data landscape for qualitative social sciences are investigated. First, European, national, and international data infrastructures for accessing research data are introduced with a focus on social research data, encompassing consortia, open clouds, metadata indexing, and search services. Second, research data centres and repositories for social research data are summarised, based on the existing scholarly literature and online resources. Through the separation between national data centres, discipline-specific repositories, institutional and generalist repositories, this section endeavours to provide an overview of the data landscape for (qualitative) social research. Lastly, data as supplements to journal articles as well as data journals are introduced as an alternative method of sharing research data.

In the fourth chapter, the findability and discoverability of qualitative research data are contextualised in relation to purposes of data reuse, searchers' information needs (IN), and information-seeking behaviours. This leads to a nuanced differentiation between finding and discovery, emphasising the importance of including both goal-directed search and browsing functions in information retrieval systems. This is followed by discussions on various formal and informal attributes for enhancing data findability and discoverability, encompassing the FAIR Data Principles as well as informal attributes such as social interaction.

In the fifth chapter, the methodological approach of the empirical study is thoroughly detailed, including the design and execution of data collection. The chapter introduces the data collection strategies employed, focusing on discipline-specific repositories and the process of tracing data in scholarly journal articles. Following this, the analysis methods used to interpret the collected data are explained. The sixth chapter presents the results obtained from the empirical study.

In the seventh chapter, the thesis concludes with a synthesis of the key findings and their implications. This chapter is structured into three sections: a summary of the research outcomes, a reflection on the limitations and strengths of the study, and suggestions for future research. The conclusion highlights how the findings contribute to the field of qualitative research data discoverability and proposes areas for further exploration.

## 2 Qualitative Research Data in Social Sciences

This chapter aims to establish a foundational understanding of the term 'qualitative data.' Qualitative data encompasses a broad spectrum, including interview transcripts, video recordings, and personal diaries, making its precise nature challenging to define. By presenting various arguments regarding the definition of qualitative research data, the following section aims to define what qualitative research entails within the context of this thesis. Consequently, it seeks to provide concrete examples of data types in qualitative social sciences by summarizing representative research methods that produce or gather such data.

### 2.1 Definition of Qualitative Research Data

Qualitative research data encompasses a diverse range of types and forms of information commonly collected and analysed in social sciences. According to the UK Data Service, the largest research data repository for social sciences and humanities in the United Kingdom, qualitative data is defined as “non-numeric information” (UK Data Service, n.d.). Jones and Alexander (2018, p. 2) describe most qualitative data as unstructured data, meaning it is “not discrete in that there are many possible measurements, characteristics, or dimensions present and/or is not organized based on a predefined framework”.

The complexity of defining qualitative data and distinguishing it from quantitative data has been emphasized by various authors (DuBois et al., 2018; Hammersley, 1996; Schoonenboom, 2023). Hammersley (1996) discusses the arbitrariness of contrasts between numeric and verbal data, structured and unstructured data, and contextualized and uncontextualized data. He argues that these distinctions often oversimplify the complex nature of research data, leading to misconceptions about their respective validity and utility. Moreover, Hammersley emphasizes that there is a range from more to less precise, structured, and contextualized types of data, suggesting that rigid categorizations fail to capture the nuanced reality of research practices.

Both Schoonenboom (2023, pp. 2–9) and DuBois et al. (2018, pp. 4–5) have attempted to identify the fundamental differences in the methods of gathering and analysing data, as well as the objectives pursued in qualitative research versus quantitative research. DuBois et al. (2018) and some other authors maintain that the primary characteristic shared by all qualitative data is its non-numeric nature (DuBois et al., 2018, p. 5; Greener, 2011; Miles & Huberman, 1994, p. 1), Schoonenboom (2023, p. 2) highlights the significance of data condensation and structure in differentiating between qualitative and quantitative data. She argues that qualitative data often require condensation to make sense of the rich, detailed information, while quantitative data are typically structured from the outset. She cautions that equating numbers solely with quantitative data and words solely with qualitative data, without considering their contexts, can lead to inappropriate choices for data analysis methods (2023, p. 4).

Moreover, qualitative evidence comprises data expressed in terms of meanings, the social contexts from which meanings are derived, or the subjective experiences of acts or events (Barbour, 1999; Haase & Myers, 1988, pp. 131–132). In this sense, qualitative data is “subjective data,” which has inherent meaning. For example, if a voice recording of an interview includes the statement, “I was extremely devastated,” one can at least assume that the speaker is expressing their experience of frustration or an equivalent negative emotion. Conversely, values of quantitative data do not include the semantic content of what was measured (Witt, 2001). To interpret such a value, one must know to which variable it belongs (Schoonenboom, 2023). If we are looking at a table and see the number 3 or the phrase “very much,” we cannot determine what this number or phrase signifies until we know the variable it is assigned to. This variable could be a survey question asking how challenging it was to filter specific search results on a barrier-free website for a visually impaired user, or how satisfying it was to do so. Furthermore, understanding where the number 3 or the phrase “very much” is situated within the full range necessitates making sense of the scale given to the survey participants.

Additionally, the interpretation of quantitative data requires external validation to be accepted, as theories are confirmed through the objectivity of information in quantitative approaches. In contrast, the meaning of subjective data, such as an individual's experience reported in an interview, is “accepted as given” without further validation. This is because any perspective “offers only partial understanding of reality,” and thus all perspectives hold validity (Haase & Myers, 1988, pp. 131–132). In this context, qualitative data are as valid as objective information, representing various views on reality, thereby contributing to and making connections between these views.

In conclusion, this thesis defines qualitative research data as any subjective data produced through research methods intended to yield data that carries semantic meaning, describing what happened or what was experienced, regardless of whether it is numeric or non-numeric.

## **2.2 Collection Methods and Types of Qualitative Data in Social Sciences**

This section enumerates several representative qualitative data collection methods as introduced in 'The SAGE Handbook of Qualitative Data Collection,' edited by Uwe Flick (2018a), and describes the types of qualitative data that are typically gathered using these methods. The methods selected from this handbook illustrate the variety of approaches used to gather qualitative data in the social sciences. As categorized by Flick, this section introduces qualitative data in three types: verbal data, ethnographic data, and material data (2018b).

### **2.2.1 Verbal Data**

#### *Qualitative Interviews*

In qualitative research, various methods are employed to collect verbal data. Qualitative Interviews, or research interviews, are prominent methods in qualitative research, typically

conducted in three forms: structured, semi-structured, and unstructured. These interviews range from highly directed questions in structured formats to open-ended discussions in semi-structured and narrative-driven unstructured formats (Roulston & Choi, 2018).

### *Focus Groups*

Focus Groups, also known as group interviews, harness the dynamics of group interaction to elicit data that might not emerge in individual interviews. This method facilitates the exploration of collective viewpoints within a group setting, employing various structuring approaches to influence participant interaction (Morgan, 1996; Morgan & Hoffman, 2018).

Such individual or group interviews for qualitative research can be conducted in person, over the phone, or through digital platforms such as internet video or text chat platforms (Padgett, 2017, pp. 130–133).

### *Narrative Data, Conversation Analysis and Discourse Analysis*

Narrative Data Collection emphasizes the storytelling aspect of human communication. This method is used to capture life stories or significant events through various means such as interviews, diaries, or letters. These narratives may reflect personal experiences or collective, representing shared cultural or community stories (Murray, 2018). *Oral histories*, as a form of narrative research, exemplify how these methods are applied within qualitative social research (Stratford, 2019).

*Conversation Analysis (CA)* involves the detailed study of naturally occurring conversations, which are audio- or video-recorded and transcribed using a notation system that captures verbal and paralinguistic elements, such as pauses and intonations, providing insights into the structure and patterns of interaction (C. Jackson, 2018).

*Discourse Analysis (DA)* examines how language is used across various contexts, from political speeches to everyday conversations. This analysis often focuses on the construction of social identities and power relations through language (Rau et al., 2018). It involves detailed textual analysis to reveal the underlying social ideologies and assumptions. Data for discourse analysis are diverse, ranging from written documents to transcripts of spoken interactions (Rau et al., 2018, p. 305).

Typically, verbal data produced from the methods mentioned above include questionnaires or interview protocols, audio or video recordings, transcripts, notes, and summaries. Additionally, conversation analysis often involves personal documents such as letters, diaries, correspondence, and personal papers. Discourse analysis extends further, incorporating a broad array of texts, from official documents and legislative proceedings to testimony in public hearings and press conferences (Qualitative Data Repository, 2014).

## **2.2.2 Ethnographic Data**

Another type of qualitative data is that which is produced in ethnographic approaches. Observation and ethnography being the general and classical approaches in this area, this

handbook introduces two more complementary methods: go-along methods and focused ethnography based on videography (Flick, 2018b, p. 231).

### *Observations*

Qualitative researchers often aim to capture data on people in their everyday lives through observations. This method involves documenting ongoing interactions, behaviours, and ambient environments, capturing the nuanced details of social practices. Observations provide a comprehensive view of the subjects' everyday lives and typically generate detailed field notes, photographs, and video recordings. These data capture interactions, behaviours, environments, and events as they naturally occur, offering deep insight into the context being studied (Wästerfors, 2018).

### *Ethnography*

Ethnography is typically defined as "long-term observation," involving participation in people's everyday lives for an extended period to collect all data relevant to the research focus (Buscatto, 2018, pp. 327–328; Hammersley & Atkinson, 2007, pp. 1–2). Primarily, ethnographic data are derived from the researcher's systematic, personal observations made in natural environments over a period, resulting in field notes and diaries (Knoblauch, 2005). Additionally, this data encompasses various other elements gathered in the field while observing, including casual discussions, structured interviews, and any documents or objects relevant to these observations (Buscatto, 2018, p. 328).

### *Go-Alongs*

These are one of mobile research methods that owe their roots in ethnography and interviewing (Adey et al., 2017; Kusenbach, 2018, p. 348). Researchers accompany individuals through their daily activities, engaging in "walking and talking (Stals et al., 2014)", which offers insights into their experiences and interactions with their environments. Data from go-alongs often consist of descriptive notes and audio- and video-recorded movements, observations and conversations, which narratively capture the physical and emotional aspects of participants' experiences (Kusenbach, 2018). Materials and technologies such as maps, photographs, Geographic Information Systems (GIS), and Global Positioning Systems (GPS) also support the research, potentially producing both qualitative data (e.g., observations) and quantitative data (e.g., GPS tracking) (Pawlowski et al., 2016).

### *Videography*

Videography integrates ethnographic fieldwork with detailed video analysis of natural interactions. It not only documents but also thoroughly analyses the dynamics of communication and interaction within various settings (Knoblauch et al., 2018, p. 362). The method employs a methodological framework known as focused ethnography, which concentrates on communicative activities. Data collection is intensely focused on the interactions in specific situations, captured by recording rather than through prolonged observation (Knoblauch, 2005). Videography utilizes advanced recording equipment to ensure high-quality audio and visual capture. The primary data produced include video recordings,

which are complemented by preliminary field notes and, potentially, interviews (Knoblauch et al., 2018).

### **2.2.3 Material Data**

Various research methods, including conversation analysis, discourse analysis, visual sociology, visual anthropology, film and television analysis, cultural studies, and sound anthropology, utilize a wide range of material data (Bowen, 2009; Bull, 2018; Collier & Collier, 1990; Eberle, 2018, p. 394; Murray, 2018; Rapley & Rees, 2018, p. 382). In this section, a closer look at different types of material data will be taken, dividing it into four categories—documents, images, audiovisual data, and sounds as data. Many researchers collect more than one type of material data for their research purposes, depending on their specific study focus and the methods they are using.

#### *Documents as Data*

Qualitative researchers utilize a variety of paper-based and digital documents, as well as archival materials, for their research purposes (Bowen, 2009). In the preceding sections on conversation and discourse analysis, examples of documents used as data were introduced, including letters, diaries, correspondence, personal papers, official documents, legislative proceedings, testimony in public hearings, and press conferences (Murray, 2018; Qualitative Data Repository, 2014). Primarily, national, regional, or local policy documents, organizational documents, newspapers, academic articles, and textbooks are the types of documents open to qualitative analysis (Rapley & Rees, 2018, p. 389).

Collecting documents can also be a precondition for the secondary analysis of qualitative data (casually known as SAQD), especially when these documents are the primary data produced by qualitative research methods such as interviews, observations, or anthropology (Bishop, 2021; Corti, 2018).

In addition to examining the static content of docile documents used in classic synthesis of qualitative research, collecting documents can accompany fieldwork research to assist analytic work. Various social interactions involving documents can be analysed, a process referred to as "documents in action" (Rapley & Rees, 2018, p. 378). Not only the written elements of texts but also the extra-textual elements such as photographs, graphs, and diagrams that are routinely embedded in documents are common objectives of analytic research on and with documents (Rapley & Rees, 2018, p. 378).

Some types of documents that are not produced for research purposes, such as blogs, diaries, letters, and magazines, can be inaccurate, uneven, or incomplete, which can be a disadvantage. However, the rawness and unfiltered nature of such documents can be highly valued depending on the research models, such as constructionism or interpretivism (Padgett, 2017, p. 133; Rapley & Rees, 2018, p. 389).

#### *Images as Data*

Visual data, used in various research methods in visual methodology, such as visual sociology, visual culture studies and visual anthropology, further expands the range of material data (Collier & Collier, 1990; Eberle, 2018, p. 394; Pink, 2012, p. 6). Eberle (2018) uses the term "image" for a comprehensive inclusion of visual data, defining it as "human-made artifacts that are perceived by our visual sense," which broadly includes still images like photographs and moving images like film or video. These images can serve various analytical purposes, such as enhancing the understanding of social environments or illustrating specific points about cultural practices (ibid.). Visual data can be both generated by the researchers through methods like photography during fieldwork or collected from existing sources, such as archives and digital platforms (ibid.).

#### *Audiovisual Data*

Audiovisual data, or media data, are typically collected for textual analysis in film and television studies (Winter, 2014). Different types of audiovisual data include documentary films, narrative films, advertising films, video clips, television entertainment shows, series, news reports, sports broadcasts, and amateur videos (Mikos, 2018). Sources of audiovisual materials vary based on the research type and include conventional public archives, private archives, online platforms like YouTube, and commercial streaming services such as Netflix (Mikos, 2018). Digitalization has significantly eased the archiving process, allowing for easier access and preservation, though older formats such as 8mm and 16mm still require specific equipment for viewing (ibid.). Licensing restrictions and the curated nature of archives present limitations in collecting audiovisual data, highlighting both legal aspects and the important role of archives in curating materials with proper metadata (Library of Congress, 2024; Mikos, 2018).

#### *Sounds as data*

Sound data collection focuses on capturing auditory information that provides insights into environments, social interactions, and cultural settings (Bull, 2018). This type of data is particularly valuable in disciplines like anthropology and urban studies, where the soundscape of an area can reveal underlying social dynamics and cultural practices (Bijsterveld, 2016; Bull, 2013). Researchers collect sounds using various recording technologies and often face challenges related to the quality and interpretability of audio data. Ethical considerations, especially regarding privacy and consent when recording in public or private settings, are paramount (Bull, 2018).

### **2.3 Ethical Considerations on Qualitative Data Collection and Sharing**

The collection of qualitative data necessitates careful ethical considerations, especially when involving private individuals. The ethical justification for sharing qualitative data hinges on participants being fully informed about the potential for publication and subsequent use of their data (Bishop, 2009a; Corti et al., 2000; Parry & Mauthner, 2004; VandeVusse et al., 2022).

When collecting ethnographic data, researchers must ensure that all participants are informed and consenting, particularly when methods involve close observations that potentially grant

access to the most private aspects of participants' lives (Buscatto, 2018, p. 340; Knoblauch et al., 2018, pp. 366–367; Kusenbach, 2018, p. 355; Wästerfors, 2018, p. 316). The intimate nature of ethnographic research often raises complex ethical issues related to privacy and confidentiality.

Overall, the sharing of qualitative data is complicated by these ethical considerations, requiring significant effort to make them publicly available and reusable for further research. Furthermore, de-identification of sensitive and personal data inevitably eliminates contextual information that might hold significant value for future research (D. Thomson et al., 2005). This makes archived qualitative data valuable and rare resources, highlighting the importance of enhancing their visibility for potential future needs.

### 3 Knowledge Infrastructure and Data Landscape for Qualitative Social Sciences

Sufficient data availability, coupled with appropriate analytical methodologies, enables the exploration of new questions and the acquisition of diverse evidentiary forms (Borgman, 2015, p. 31). For the right data to be released and found for future research, well-established knowledge infrastructure and fluent scholarly communication are important. Knowledge infrastructures are “robust networks of people, artefacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds and interact through social and technical processes” (adapted from Borgman, 2015 and Edwards, 2010). Within these infrastructures, formal scholarly communication such as research, publishing, and relevant activities like peer review, as well as informal communication including collaboration, personal exchanges, and presentations, can take place with the most visibility. Research data publications should be well integrated into this communication system to be properly findable and discoverable.

On one hand, data sharing has become increasingly prevalent through mandates from funding agencies, research institutions, and journal publishers (Borgman, 2015; Crosas et al., 2018; ‘Data Policies’, 2024; B. Jackson, 2021; Kratz & Strasser, 2015a; Tenopir et al., 2015). Despite the fact that data sharing and reuse practices in social sciences primarily focused on quantitative data (Corti, 2012), there have been increasing discussions about the advantages and challenges of qualitative data sharing (QDS). In accordance with this trend, national research funding bodies have recently begun supporting the sharing of qualitative data. For instance, the new data management and sharing policy of the National Institutes of Health (NIH) in the United States provides additional information on privacy considerations when sharing human research participant data, both quantitative and qualitative (DuBois et al., 2023; National Institutes of Health, 2022). However, guidelines and policies for managing and sharing qualitative data are still not sufficiently addressed in scholarly journals (Crosas et al., 2018) and by research funders, making it difficult for qualitative researchers to adopt Open Science (OS) practices (Huma & Joyce, 2023; V.-M. Karhulahti, 2022).

On the other hand, integrating shared data into the mainstream of scholarly communication to incentivize data creators has long been a challenge (Borgman, 2015; Borgman & Bourne, 2022; Meistring, 2023; Walters, 2020). A robust infrastructure is essential to facilitate seamless scholarly communication, enabling researchers to “effortlessly discover, access, and utilize relevant and trustworthy materials” (Rieger & Schonfeld, 2023).

The concept of data publication has arisen within these discussions, garnering significant attention to incentivize data documentation and sharing (Chavan & Penev, 2011; Kratz & Strasser, 2015b). Primarily, three data publication models exist that are widely accepted by the research community (1) data as a supplementary file to a traditional journal article, (2)

submitting data to a repository, and (3) publishing a data paper in a specialized data journal as a stand-alone publication describing a specific dataset (Penev et al., 2017).

The increasing amount of shared qualitative data, as well as the different models of data publication, has led to the diversification of the data landscape within qualitative social research. Consequently, this diversification makes data challenging to find for researchers (Gregory et al., 2020; Krämer et al., 2021). In this context, this chapter examines the available data infrastructures for qualitative social researchers and explore the services offered for finding and discovering qualitative research data under three categories: (1) international and national data infrastructure, (2) research data centres and repositories, and (3) data supplements in journals and data journals.

### **3.1 European, National and International Data Infrastructure**

To portray the research data landscape for social sciences in the European context, the Consortium of European Social Science Data Archives (CESSDA) of the European Research Infrastructure Consortium (ERIC) plays a significant role (Consortium of European Social Science Data Archives, 2024b). CESSDA ERIC offers metadata for over 40,000 social science data collections from more than 20 European countries through its central data catalogue (CESSDA Data Catalogue) (Consortium of European Social Science Data Archives, 2023a, 2024a; Giglia, 2019). This aligns with the mission of OPERAS, the European research infrastructure dedicated to open scholarly communication in the social sciences and humanities (Schulte, 2024).

The Social Sciences & Humanities Open Cloud (SSHOC) project aims to develop the SSH area of the European Open Science Cloud (EOSC), an e-infrastructure for publishing, storing, accessing, and reusing research data (E. J. Gray et al., 2022). SSHOC provides access to data catalogues of ERIC consortia for social sciences, including CESSDA, the European Social Survey (ESS), and the Survey of Health, Ageing, and Retirement in Europe (SHARE) (Social Sciences and Humanities Open Cloud, 2024).

Within the German National Research Data Infrastructure (NFDI), the Consortium for Social, Educational, Behavioural, and Economic Sciences (KonsortSWD) plays an important role in supporting qualitative social research (Wissenschaftszentrum Berlin für Sozialforschung, 2024). In January 2024, KonsortSWD founded QualidataNet as a “single point of entry” for qualitative research data collections and related data services from diverse data centres in Germany (Forschungsdatenzentren, FDZ) (Betancort Cabrera & Mozygamba, 2022; Miller, 2021; Universität Bremen, 2024a).

On an international level, data services and metasearch engines are crucial components of the research infrastructure. B2FIND, a metadata indexing service offered by the EUDAT Collaborative Data Infrastructure (EUDAT CDI), provides a discovery portal enabling

exploratory searches of metadata harvested from diverse data providers ([Unknown], 2023). As of July 2024, B2FIND offers metadata for about 1.4 million datasets from 100 European and international repositories and catalogues, partially including CESSDA (Deutsches Klimarechenzentrum GmbH, 2024).

OpenAIRE is an Open Science service provider for European researchers (*OpenAIRE*, 2024a). OpenAIRE EXPLORE integrates metadata from journal archives, repositories, current research information systems (CRIS), research entity registries (such as Crossref and DataCite), aggregators, and catalogues of research products, including CESSDA and B2FIND. Currently, it indexes over 61 million research data items from approximately 140,000 global data sources (OpenAIRE, 2024b). Even though EUDAT CDI and OpenAIRE are both infrastructures built for European researchers, they provide a comprehensive coverage of research data on an international level.

Crossref (Hendricks, 2021), DataCite (DataCite – International Data Citation Initiative, 2024), and OpenAlex (*OpenAlex*, 2024a) are significant open infrastructures that are committed to the Principles of Open Scholarly Infrastructure (POSI), which ensure the openness, lawfulness, and availability of research data (Bilder et al., 2020). These organizations act as significant data aggregators, indexing sources such as institutions, journals, and repositories to enable metasearch of aggregated datasets. Crossref and DataCite also function as registration agencies (RAs) of Digital Object Identifiers (DOIs), providing services to register and manage DOIs and metadata for research data publishing institutions such as repositories, archives, and universities (DataCite Metadata Working Group, 2024).

Furthermore, Elsevier Mendeley Data (Elsevier, 2024b) and Harvard Dataverse (The President and Fellows of Harvard College, 2024) are open research data repositories that facilitate metadata search through large-scale data aggregation, making it easier to access deposited datasets (Friedrich & Recker, 2021).

Additionally, Google Dataset Search, launched in September 2018, reflects the growing demand for datasets in research (Chapman et al., 2020). This tool indexes over 45 million datasets from more than 13,000 websites, including publisher websites, data repositories, and personal websites, enabling a “Google-style search” for user exploration and interaction (Chapman et al., 2020; Google, 2024). It relies on dataset providers using schema.org markup, a standardized vocabulary for describing data on the web, to aggregate, normalize, and reconcile dataset descriptions from web pages (Noy & Benjelloun, 2023). The metadata aggregated in Google Dataset Search are also linked to other Google resources, such as Google Scholar (Burgess & Noy, 2018).

The following table summarizes the search services offered by the research data infrastructures discussed above. The scale refers exclusively to research datasets and collections\*:

Search Service	Features	Scale	Data Source	Link
CESSDA Data Catalogue	Metasearch service for social research data in Europe	Over 40,000 data collections	22 European national data service providers for social sciences (Consortium of European Social Science Data Archives, n.d.-a)	<a href="https://datacatalogue.cessda.eu/">https://datacatalogue.cessda.eu/</a>
Suchportal QualidataNet	Metasearch service for qualitative research data in Germany	104 datasets	7 German research data centres and archives (Universität Bremen, 2024b)	<a href="https://www.qualidatanet.com/de/daten-suchen.html">https://www.qualidatanet.com/de/daten-suchen.html</a>
B2FIND EUDAT	Interdisciplinary metasearch service for research data	About 1.4 million datasets	100 data centres, repositories and catalogues	<a href="https://b2find.eudat.eu/">https://b2find.eudat.eu/</a>
OpenAIRE EXPLORE	Interdisciplinary metasearch service for research data	Over 61 million research data	Over 140,000 data sources including journals, repositories, aggregators etc.	<a href="https://explore.openaire.eu/">https://explore.openaire.eu/</a>
Crossref Metadata Search	Metasearch service of DOI records registered with Crossref	About 3 million datasets ( <i>Crossref Stats</i> , 2024)	DOI registering members of Crossref (Hendricks, 2021)	<a href="https://search.crossref.org/">https://search.crossref.org/</a>
DataCite Commons	Comprehensive metasearch service of DOI records	About 18 million datasets**	DOIs registered with DataCite, Crossref, or other DOI registration agencies (Fenner, 2020)	<a href="https://commons.datacite.org/">https://commons.datacite.org/</a>
OpenAlex	Interdisciplinary metasearch service for research data	About 8 million datasets (OpenAlex, 2024b)	Microsoft Academic Graph (MAG), Crossref, and others ( <i>Sources</i> , 2024)	<a href="https://openalex.org/">https://openalex.org/</a>
Mendeley Data	Data repository	About 30 million datasets	Data Monitor (Elsevier, 2022)	<a href="https://data.mendeley.com/">https://data.mendeley.com/</a>
Harvard Dataverse	Data repository	Over 180,000 datasets	Deposited datasets in its dataverses	Harvard University: <a href="https://dataverse.harvard.edu/">https://dataverse.harvard.edu/</a>
Google Dataset Search	Search engine for research datasets	About 45 million datasets	Web sources using schema.org markup (Noy & Benjelloun, 2023)	Google: <a href="https://datasetsearch.research.google.com/">https://datasetsearch.research.google.com/</a>

Table 3. 1 Summary of Search Services for Research Data

\*All the information in the table was accessed through the links and citations provided. Last access on July 21, 2024.

\*\* Data retrieved using REST API query: <https://api.datacite.org/doi/resource-type-id=dataset>. Retrieved July 21, 2024, 17:09.

The integration of datasets into these search services, whether through search engines, data aggregators, or metasearch engines, is essential for ensuring the findability and discoverability of research data (Chapman et al., 2020; Gregory et al., 2020; Krämer et al., 2021; Kratz & Strasser, 2015a; Wu et al., 2019). Further discussions on related studies and technical terms, including schema.org markup, can be found in subchapter 2.2, specifically section 4.2.5 Inclusion in Search Services.

## 3.2 Research Data Centres and Repositories

A data repository is an infrastructure component designed to store, manage and curate, and provide access to data object (Baker & Duerr, 2017; Berg-Cross et al., 2015). Submitting data to a research data repository (RDR) is recognized as one of the primary methods of data publication (Penev et al., 2017).

Horizon Europe, the European Union's research and innovation funding programme, mandates the deposition of research data underlying funded research in a trusted repository (OpenAIRE, 2022b). This requirement is echoed by other funding bodies, including the UK Research and Innovation (UKRI) (UK Research and Innovation, 2023), the National Institutes of Health (NIH) (National Institutes of Health, 2020), and Swiss National Science Foundation (SNSF) (Swiss National Science Foundation, 2016).

Moreover, Crosas et al. (2018) conducted an analysis of data policies in social science journals, revealing that over half of the journals with a data policy either recommend or require the deposition of underlying datasets in a repository. This underscores the significant role of research data repositories in the field of social sciences.

OpenAIRE identifies three major types of research data repositories where researchers may deposit their data: domain-specific repositories, general-purpose repositories, and institutional repositories (OpenAIRE, 2022b). In addition to these categories, this thesis adopts Baker and Duerr's classification of repositories, which is originally based on the work of Armbruster and Romary (2009) (Baker & Duerr, 2017, p. 142). Consequently, five repository types are considered for investigating the research data landscape in social sciences, with a focus on qualitative data: national data centres, discipline-specific repositories, thematic repositories, institutional repositories, and generalist repositories.

National data centres, or national repository systems, are intended to broadly capture scholarly output. These infrastructures typically aim not only to preserve a country's academic records but also to support activities such as teaching and learning in higher education (Armbruster & Romary, 2009). In Europe, the 22 CESSDA members serve as representative national data centres for the social science domain (Consortium of European Social Science Data Archives, n.d.-a). The Australian Data Archive (ADA) also portrays an important national repository system for social researchers, providing as well archiving and access service of qualitative data (Australian National University, 2024b, 2024a).

Among these, the UK Data Service (UKDS) stands out as the most significant national data service for qualitative data, pioneering qualitative data archiving with the establishment of the ESRC Qualitative Data Archival Resource Centre (Qualidata) in 1994 (Corti & Thompson, 1996; UK Data Service, 2024c). In 2012, Qualidata was integrated into the UK Data Service (UK Data Service, n.d.-b). The UK Data Service has also assisted numerous European

countries in establishing qualitative data archives (Neale & Bishop, 2011; Rasmussen, 2011). Currently, the UK Data Service provides an online search and browse interface for their qualitative data collections, named QualiBank, which allows users to search the content and metadata of qualitative data objects such as interviews, essays, open-ended questions, reports, photos, and audio recordings (UK Data Service, 2024b).

Additionally, the Finnish Social Science Data Archive (FSD) is recognized for its mature service for qualitative data (Corti, 2011; Finnish Social Science Data Archive (FSD), 2024). Countries initially focused on quantitative data, such as Germany (Leibniz Institute for the Social Sciences [GESIS]) (GESIS, 2024), Switzerland (Swiss Centre of Expertise in the Social Sciences [FORS]) (FORS, 2024b), Norway (Norwegian Agency for Shared Services in Education and Research [Sikt]) (Sikt, 2024), and Slovenia (Social Science Data Archives [ADP]) (ADP, 2024), have increasingly begun accepting qualitative data in recent years (Bishop & Kuula-Luumi, 2017).

In 2011, IASSIST Quarterly published a series of fourteen papers on qualitative data sharing and archiving in European countries (Neale & Bishop, 2011; Rasmussen, 2011). Articles cover the establishment and development of qualitative data archives in Switzerland (Kleiner et al., 2011), Ireland (J. Gray & O'Carroll, 2011), the UK (Bishop & Neale, 2011), Austria (Smioski, 2011), the Czech Republic (Cizek, 2011), Denmark (Kjeldgaard, 2011), as well as efforts in Germany (Medjedovic & Witzel, 2011), Norway (Henriksen et al., 2011), Slovenia (Stebe et al., 2011), Belarus (Titarenko & Tereschenko, 2011), Hungary (Gardos & Ivacs, 2011), Lithuania (Butviliene & Butvilas, 2011), and Poland (Binder & Filipkowski, 2011).

To continue and update the overview of European practices in archiving qualitative research data, the UK Data Service maintains a list of qualitative data archives in Europe (UK Data Service, n.d.-b). Notable updates include Austria's new data deposit guideline for qualitative data (AUSSDA) (Kernecker, 2023), the Czech Social Science Data Archive's search advice for their qualitative data collections (CSDA, 2024), France's Qualitative Social Science Survey Bank (beQuali) (beQuali, 2024), Hungary's Voices of the 20th Century Archive and Research Center (Voices of the 20th Century Archive, n.d.), the Irish Qualitative Data Archive (IQDA) (National University of Ireland Maynooth, 2024) the Northern Ireland Qualitative Archive (NIQA) (ARK, 2017), the ODISSEI Portal managed by the Dutch National Centre of Expertise and Repository for Research Data (DANS) (DANS, 2024), Poland's Qualitative Data Archive (QDA) (Archiwum Danych Jakościowych, 2024), and Germany's Qualiservice (QualiService, n.d.). Furthermore, QualidataNet offers a comprehensive overview of qualitative data collections in German research data centres, including Qualiservice (Universität Bremen, 2024a).

The map below illustrates the current locations of qualitative datasets archived in Europe:



Figure 3. 1 European Qualitative Data Collections and Archives Map

Additionally, the Australian Data Archive (ADA) is another example of a national data service for diverse research disciplines including qualitative social research (Australian National University, 2024b).

Studies indicate that discipline-specific repositories are among the most preferred methods for data retrieval by researchers (Contaxis et al., 2022; Gregory et al., 2019, 2020; Kratz & Strasser, 2015a). A recent large-scale survey highlights that repository reputation and disciplinary norms play crucial roles when researchers choose repositories for data sharing, thereby enhancing the volume of discipline-relevant data collections in these repositories (Khan et al., 2023).

Discipline-specific repositories, or domain repositories, offer research data associated with a specific research field or discipline (Baker & Duerr, 2017, p. 143). National social science data archives, such as those affiliated with CESSDA, serve as domain repositories due to their focus on social science research.

While GESIS aggregates metadata from several national and international social science repositories and offers a metadata search service (GESIS-Suche), primarily for quantitative data (GESIS, n.d.-b), Qualidatanet focuses on the network of qualitative data centres to provide services regarding deposition, curation, and access to qualitative data (Universität Bremen, 2024a). Other German data centres contributing to Qualidatanet's catalogue include the Research Data Centre for Higher Education Research and Science Studies (FDZ-DZHW) (FDZ-DZHW, n.d.), the Research Data Centre for Education (FDZ Bildung) (DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation, 2022), the Research Data Centre for

Business and Organizational Data (FDZ-BO) (Deutsches Institut für Wirtschaftsforschung, 2024), eLabour (FDZ eLabor, 2017), and aviDa (Technische Universität Berlin, 2024).

Internationally, notable social science data repositories include the Interuniversity Consortium for Political and Social Research (ICPSR) at the University of Michigan, which houses the world's largest collection of digital social science data and has been acquiring qualitative data since 2011 (UK Data Service, n.d.-c). The Qualitative Data Repository (QDR) at Syracuse University, founded in 2014, is dedicated to storing and sharing digital data from qualitative or multi-method research in the social sciences (Qualitative Data Repository, 2023b). Both ICPSR and QDR have protocols for processing sensitive data and offer guidelines for managing, depositing, and reusing qualitative data, exemplifying good practice in qualitative research data services (Antes et al., 2017; K. Jones & Alexander, 2018).

In addition, the Henry A. Murray Research Archive, a Harvard Dataverse repository at the Institute for Quantitative Social Science (IQSS, Harvard University), is renowned for its collections of quantitative and qualitative data in psychology and sociology (The Murray Research Archive, n.d.).

Compared to domain repositories, thematic or themed repositories are typically confined to a specific topic within a domain or subject area (Baker & Duerr, 2017, p. 143). Weller (2023) advocates for smaller, specialized repositories that carefully consider the structure and flexibility necessary to accommodate the complex and evolving nature of qualitative data. A prime example is the ESRC Timescapes Qualitative Longitudinal Data Archive at the University of Leeds (hereafter referred to as the Timescapes Archive), which has been collecting qualitative longitudinal research (QLR) data since 2007 (Consortium of European Social Science Data Archives, 2024b). This archive is originated from the nationally funded QLR study in the UK named Timescape (ibid.). It exemplifies competence by fostering participant-led archiving practices, developing methodologies for qualitative secondary analysis, encouraging collective secondary analysis activities, and providing guides on teaching using deposited data in the archive.

Another notable example is the NOAA Voices Oral History Archives, initiated by the US National Oceanic and Atmospheric Administration (NOAA Voices) (NOAA Fisheries, n.d.). This project has been crucial in capturing and preserving oral histories detailing the experiences of individuals involved in U.S. fisheries, the National Weather Service, the Office of Oceanic and Atmospheric Research, and the National Ocean Service, developed in response to the need for long-term archiving and the protection of these recordings that account for climate change and ecological knowledge from across the U.S. As of June 2024, it includes a total of 2457 voice recordings.

Additionally, the 'Everyday Childhoods' dataset is an open access public collection with a range of multimedia data documenting the everyday lives of children and their families (R.

Thomson & Berriman, 2023). This collection distinctively involves participants in its creation, soliciting their input on potential secondary uses through interactive mediums such as short films and postcards, and is housed in a generalist repository Figshare (Everyday Childhoods, 2024). These specialized thematic repositories illustrate the unique characteristics of qualitative data collected with specific research methods or approaches, promoting active participation in archiving practices (Weller, 2023).

Institutional repositories "contain various outputs of an institution" and are "generally not differentiated by topic or domain but rather concerned solely with the outputs of members of the organization within which they reside" (Armbruster & Romary, 2009; Baker & Duerr, 2017, p. 143). The Registry of Research Data Repositories (re3data.org), an internationally recognized research data repository registry, currently lists 921 institutional research data repositories, 188 of which contain data within the scope of social and behavioural sciences<sup>1</sup>. Institutional repositories primarily focus on the research outputs of their affiliated members, making them useful for accessing local research data (Vitale et al., 2017). They are also vital for managing research data from researchers across various disciplines globally and contribute to showcasing an institution's research accomplishments and earning peer recognition (M. J. Foster & Jewell, 2022, p. 276; D. J. Lee & Stvilia, 2017).

Generalist repositories, also known as general-use or general-purpose repositories, are discipline-independent and open to any research data, providing an option for researchers without a specific disciplinary or institutional repository (Johnston, 2017, p. 138). The Generalist Repository Ecosystem Initiative (GREI) at the U.S. National Institutes of Health (NIH) aims to make NIH-funded research datasets openly accessible. GREI includes Dryad, Harvard Dataverse, Figshare, Mendeley Data, Open Science Framework (OSF), Vivli, and Zenodo. Figshare, the sole commercial service, has grown significantly, holding over 8 million digital objects and nearly 2 million datasets as of April 2024. The quality of (meta)data varies, with repositories like Dryad requiring strict data submission standards (Dryad, 2024), while Figshare and Zenodo are fully open (Zenodo, 2024).

Discipline-specific repositories generally have more rigorous review processes compared to generalist repositories (Austin et al., 2017). For instance, UK Data Archive for social sciences uses different levels of quality control depending on how much additional value is to be added to the data (UK Data Archive, 2024).

### **3.3 Data Supplements in Journals and Data Journals**

The concept of data publication has emerged within the discussion of incentivising researchers who create, document, and share their data (Chavan & Penev, 2011; Kratz & Strasser, 2015b).

---

<sup>1</sup> Search result filtered by the metadata property *Repository Types* with the value *institutional* and the metadata property *Subject* with the value *Social and Behavioural Sciences*. Last access 24. July 2024.

*Data publication*, or *data publishing*, can be understood as “the release of research data, associated metadata, accompanying documentation [...] 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” (Austin et al., 2017). Despite its general definition, data publication can be narrowed down to three primary types (1) data as supplementary materials to a traditional journal article, (2) data submission to a repository, and (3) data paper as a stand-alone publication describing a specific dataset or a specific collection of datasets.

Scholarly journals have been promoting the sharing of datasets underlying research findings through their data policies. The study by Crosas et al. (2018) indicates that most of the highly ranked social science journals with a data sharing policy (n=155) recommend or require authors to deposit data in a repository (n=80) or submit the underlying data directly to the journal as supplementary material to the published article, which is made available on the journal’s own website (n=53). However, the deposition of qualitative data is typically mentioned by only a limited subset of journals, with occasional reflection on ethical and practical issues (Crosas et al., 2018; Prosser et al., 2023, 2024).

Prosser and her colleagues reviewed open data requirements of 261 journals in social psychology with focus on qualitative data (Prosser et al., 2021, 2023). Among these journals, 74 explicitly accepted qualitative research data, and only two – the British Journal of Social Psychology (Wiley) and the Journal of Environmental Psychology (Elsevier) – had guidelines for open qualitative data. It is generally encouraged to deposit data in established repositories rather than publishing data as supplementary files on a journal’s website, as journal websites do not offer the permanence of a data repository, nor assign a DOI or other PID (see an exemplary statement by [The British Psychological Society, 2024](#) and also see Penev et al., 2017, p. 9).

Scholarly journals also provide bidirectional data linking services between data published in a repository and the online article published by them. For example, Elsevier enables data linking to an article via ScholXplorer, a data DOI, or a link to a repository (Elsevier, 2024a). Further investigation into data linkage to journal articles can be found in subchapter 2.2, specifically section 2.2.2.4 Data Citation and Linkage of Scholarly Literature with Datasets.

As discussed above, data sharing practices of traditional journals and curation standards of repositories for data deposition are well established. However, data as supplementary materials in journals as well as deposition in repositories typically do not include peer-review of the dataset to be published. Peer-review of dataset yet has a significant meaning to researchers as potential data re-users (Kratz & Strasser, 2015b).

To push research data into existing academic reward system, a publication type called *data paper* has emerged. *Data paper*, defined by Chavan and Penev (2011) as “a scholarly publication of a searchable metadata document describing a particular online accessible

dataset, or a group of datasets, published in accordance to the standard academic practices”, is a data publication model that most closely resembles the existing scholarly literature in that its core processes consist of peer review and dissemination of datasets (Austin et al., 2017; Kratz & Strasser, 2015b).

A data paper is a stand-alone journal publication that focuses on the description of the originality of data, the data collection process, the analysis method, and the usability of data, whereas the published data described is usually hosted in a repository and linked to the paper (Austin et al., 2017; Chavan & Penev, 2011; J. Lee & Kim, 2021). It offers scholarly credit to data creators and brings the existence of the data to the attention of the scholarly community, since it is easily citable as well as peer-reviewed data enjoy much greater trust and prestige (Chavan & Penev, 2011; Kratz & Strasser, 2015b; Seo & Kim, 2020).

Data papers appear in existing journals or in specialized data journals. Founded in 2016 and funded by CESSDA from 2021, the Research Data Journal for the Humanities and Social Sciences from Brill (RDJ) is a representative example for a dedicated social science data journal (Koninklijke Brill, 2024). RDJ provides a comprehensive instruction for researchers on ethical and legal conditions, submission process, paper structure, licensing as well as choosing the right repository for deposition (Research Data Journal for the Humanities and Social Sciences, 2023). As of July 2024, it includes 60 data papers in social sciences and humanities domain, among which 10 papers describe existing qualitative datasets. These datasets are deposited in diverse repositories, including UKDS (n=3), Zenodo (n=2), figshare (n=1), OSF (n=1), DataverseNL (n=1), DANS (n=1), IISH Data Collection (n=1).

### **3.4 Summary**

The investigation, supported by a comprehensive literature review and extensive internet resources, highlights the complex data landscape for qualitative social research. The dispersed locations of archived data necessitate that potential data seekers possess precise knowledge of where to find relevant information. While search services such as metadata catalogues and metasearch platforms aid in this endeavor, their effectiveness is limited due to the uneven coverage of their data sources. Domain-specific search services, like the CESSDA Data Catalogue, serve as prominent research data aggregators for the social sciences, though they only encompass specific European national data providers and exclude smaller or non-European data sources. Despite advancements in data discovery infrastructures, the ability of data seekers to identify and navigate pertinent repositories, archives, data centers, and search platforms remains essential. In this context, re3data.org plays a crucial role, bearing the significant responsibility of providing a comprehensive list of repositories and ensuring accessibility within this intricate data environment.

## 4 Findability and Discoverability of Qualitative Research Data in Social Sciences

The motivations to seek existing research data are varied and can be understood in relation to the purpose of the data seekers' intended usage. A large-scale survey on data discovery and reuse among researchers (n=1,677) indicated that over 70 percent of researchers require secondary data as "the basis for a new study." This is followed by purposes such as "preparing for a new project or proposal," "teaching," "verifying their own data," "generating new ideas," and other uses (Gregory et al., 2020, p. 128). Specifically, in the realm of qualitative data, a user study conducted by the UK Data Service revealed that more than two-thirds of downloaded qualitative data were utilized for learning and teaching research methods and substantive topics within social science disciplines (n=7,115) (Bishop & Kuula-Luumi, 2017).

In qualitative research, these modes of data reuse are linked to further research activities such as comparative research, follow-up studies, secondary analysis, text mining, natural language processing, replication of published studies, and advancements in research design and methodology (Corti, 2020; Corti & Thompson, 2004). Additionally, while some researchers browse for potentially relevant data to address their research questions, others may have specific data in mind and require only an effective retrieval method (Henrich et al., 2021, pp. 429–430). Depending on seekers' different needs for research data, searches can be exploratory—either as completely free browsing or somewhat directed exploration—or targeted, aiming to retrieve precise data needed.

This subchapter aims to connect data findability and discoverability with people's needs for information, which determine their information seeking behaviours (Allen, 1991; Frants et al., 1997, p. 27). It builds upon the two distinct types of information needs—Concrete Information Need (CIN) and Problem-Oriented Need (POIN)—introduced by Frants and Brush (1989). These needs involve different approaches to information seeking and consequently result in a nuanced differentiation between findability and discoverability. The discussion will delve into each aspect of data findability and discoverability, incorporating attributes introduced in the FAIR Data Principles along with other relevant factors such as social contact and linkage with scholarly literature (Friedrich & Recker, 2021; Shigapov, 2024; Wilkinson et al., 2016).

### 4.1 Findability and Discoverability in Relation to Information Needs (IN)

Finding a piece of information inevitably involves a user's known need for specific information and consequently the act of searching (Sutcliffe & Ennis, 1998, p. 333), since "to find something" implies that someone is looking for it, hence necessitating its acquisition. Conversely, "discovery" implies the unexpected encounter with information resulting from serendipitous exploration, without necessarily knowing what exactly one will find in the end (Makri et al., 2019). Such separation of these two terms is evident in the context of information science, as several studies have discussed the notion of serendipity utilized in information

seeking behaviour (A. Foster & Ford, 2003; Makri et al., 2015; McKenzie, 2003) and the tension between findability and discoverability (Makri et al., 2019).

Information seeking activities which result in finding or discovering desired information vary. Sutcliffe and Ennis (1998) considered information searching as a range of behaviours from goal directed information searching to more exploratory information browsing, also expressing it as “serendipitous or exploratory information seeking.” The former is chosen when the user has a specific information need, whereas the latter is required when “no specific goal is present besides the intention to explore an information repository” (ibid.). In this sense, finding desired information while browsing is unlikely because there is no clear desire. However, it is common to discover unexpected but interesting information during goal-directed searching.

The incidents of finding and discovering can complement each other’s inability to wholly satisfy one’s information need. Loosely directed exploration, which can be seen as a search activity situated between goal directed information searching and exploratory information browsing, can involve both finding desired information and discovering unexpected but unknowingly desired information. This is evident in the study of Makri and his colleagues (2019) on how design features of a physical bookshop designed for serendipity affect the findability and discoverability of information based on users’ information behaviour. They found that while participants rarely found desired information by accident, they all unexpectedly encountered books they found interesting. In another study, it was found that two different search strategies (using generic search terms and specific search terms) retrieved different information relevant to a specific research purpose in electronic databases, with only 7 percent of identified information retrieved by both search strategies (Egan et al., 2012). This aligns with Frants et al.’s (1997, p. 36) observation on information needs (IN):

*“In fact, in satisfying an IN we always strive for clarity. But clarity ensues only when we can make a decision using obtained information [...] However, in many cases we cannot know beforehand exactly what information leads to the achievement of clarity. We want information, we want it very much [...], but unfortunately we do not always know exactly what we want.”*

Through this statement, we can assume that the discovery of unanticipated information can contribute to the clarity, the satisfaction of an IN, by revealing the hidden desire, or to the realisation of a new IN. Furthermore, Frants and Brush (1989) recognize two distinct types of information needs. A concrete information need (CIN) has well-defined thematic boundaries, and satisfying a CIN involves meeting these specific boundaries. For instance, a question like “What was the inflation rate in the United Kingdom in May 2024?” represents a CIN that can be answered with just one document that contains the right data. Once the desired information is found, the CIN is resolved (Frants et al., 1997, p. 37).

Problem-oriented information needs (POIN), on the contrary, require more complex process of information acquisition to be satisfied, since a POIN lacks clear thematic boundaries, making it difficult to formulate a precise request. For example, the information need expressed as "How has the inflation in the United Kingdom been influencing the living situation of the citizens since 2020?" cannot be fully satisfied with a single document or a certain number of documents. Even if multiple individuals share the same POIN and access the same information, their levels of satisfaction will vary, and no one will be completely satisfied. While a CIN requires searching and finding specific information to conclude the search, a POIN may lead to the expansion, alteration or refinement of its boundaries and persist over time (Frants et al., 1997, pp. 36–38).

Search strategies are determined by the type of the user's information need (Allen, 1991). A qualitative research question, such as the example provided in the preceding paragraph, can often be seen as a problem-oriented information need. The search to answer such a question requires a loosely directed exploration to not only find the initially desired information but also to discover additional findings that enhance the searcher's understanding of the topic, which allows iterative refinements of queries (Henrich et al., 2021, p. 430). This iterative process enables the refinement of the research question and reveals information needs that were not initially anticipated.

Searching for research data is one of many information seeking activities to satisfy such information needs. In social sciences, such exploration can lead to the development of new connections and insights, originating new ideas and hypotheses (Gregory et al., 2020, p. 128; Merton, 1968, p. 158). In this context, finding and discovering existing qualitative research data can be understood as part of the process of answering, expanding, or refining a problem-oriented question in social sciences, as these retrieved data can be reused in learning research methods and discipline-relevant topics, comparative research, follow-up studies, secondary analysis, text mining and natural language processing, replication of published studies, and methodological advancement.

While it may seem intuitive that searches related to research data primarily fall into the category of exploratory search, as they involve seeking potentially relevant data for a specific research question, Henrich et al. (2021, pp. 429–430) argue that searchers often know precisely which data they are looking for. Consequently, goal-directed information search should also be considered in the design of a research data search solution. Therefore, an effective search strategy for research data should incorporate both exploratory and goal-directed search aspects. Several social science data repositories exemplify this practice, including the Inter-university Consortium for Political and Social Research (ICPSR), the UK Data Service and GESIS Data Archive. Through enabling advance searches with filter options as well as generic search with keywords and thematic and topical browsing faceted through pre-defined classifications and controlled vocabularies, they support both exploratory and goal-directed search strategies for data seekers.

The next chapter will inspect various aspects related to data findability and discoverability which enhance goal-directed search and loosely directed exploration to satisfy diverse information needs in research.

## 4.2 The FAIR Data Principles and Other Aspects for Data Findability and Discoverability

The FAIR Data Principles, introduced and developed by The Future of Research Communications and e-Scholarship (FORCE11) (*The FAIR Data Principles – FORCE11*, 2014; Wilkinson et al., 2016), are widely referenced by data curators and researchers for managing research data across various disciplines, including social sciences (Berkowitz & Delacour, 2022; Mannheimer, 2024; Zondergeld et al., 2020). To enable the discovery of research data on the Internet, the FAIR Data Principles suggest the following:

- 1) (meta)data are assigned a globally unique and persistent identifier.
- 2) Data are described with rich metadata.
- 3) Metadata clearly and explicitly include the identifier of the data it describes.
- 4) (meta)data are registered or indexed in a searchable resource.

While the FAIR Principles primarily focus on enhancing machines' ability to automatically find and use data at a large scale by defining essential standard protocols (Wilkinson et al., 2016), Friedrich and Recker (2021) further discuss data findability in three different perspectives: (1) system perspective, (2) user perspective, and (3) infrastructure perspective. Adopting Friedrich and Recker's classification, different aspects which influence research data's findability and discoverability can be summarized as in the following table:

System perspective	User perspective	Infrastructure perspective
Research data repositories	Research communities	Research data communities
Data journals with peer-review process	Knowledge from academic training periods and research practices	Metadata standards for data documentation
Inclusion in search services for aggregated datasets from different sources	Social contact	Controlled vocabularies
	Relevant scholarly literature	Linkage of scholarly literature with datasets
	Keyword search, filter options, thematic browsing and advanced search in search services and web search	
	Thematic relevance	

Table 4. 1 Aspects Influencing Data Discoverability and Findability

Reflecting the FAIR Data Principles on findability and Friedrich and Recker's classification, the following sections will discuss each essential aspect that possibly influences the discoverability and findability of qualitative research data in social science domain. The

aspects to be discussed are: (1) Persistent identifiers; (2) Metadata standards and data documentation; (3) Controlled vocabularies; (4) Data citation and linkage of scholarly literature with datasets; (5) Inclusion in search services; (6) Social interaction.

#### 4.2.1 Persistent Identifiers (PID)

Persistent identifier (PID) is a unique name given to a digital object that remains stable over time, even if the object's location changes (Hilse & Kothe, 2006). This stability ensures access to the object whenever a searcher wants to revisit the source. Data findability and discovery can only be achieved if the data are uniquely specifiable, and permanently and stably retrievable through a protocol associated with such a persistent identifier (Juty et al., 2020). Common types of PIDs include Digital Object Identifiers (DOIs), Handles, Archival Resource Keys (ARKs), Persistent Uniform Resource Locators (PURLs), Uniform Resource Names (URNs), National Bibliography Numbers (NBNs), and OpenURL (Hilse & Kothe, 2006).

The Handle System is a general-purpose identifier system that provides a broad framework for naming and resolving digital objects over the Internet (ibid.). DOIs are a specialized implementation of the Handle System, supported by a standardized metadata schema and governed by the DOI Foundation (DOI Foundation, 2023). Widely used in academic publishing, DOIs uniquely identify and provide persistent access to digital documents such as journal articles, books, and datasets (Juty et al., 2020). URNs, on the other hand, are part of the URI (Uniform Resource Identifier) scheme, providing unique, persistent, location-independent identifiers relying on namespace-specific resolution mechanisms. But they do not have the same level of standardized metadata or centralized governance as DOIs (Hilse & Kothe, 2006).

A Digital Object Identifier (DOI) is a unique alphanumeric string used to identify content and provide a persistent link to its location on the internet. Each DOI consists of a prefix and a suffix, separated by a forward slash. The DOI Foundation allocates prefixes to organizations known as Registration Agencies (RAs) that register DOIs. DataCite and Crossref are prominent examples of such RAs. A registrant, typically a researcher or an institution, assigns a DOI to their data publications through an RA. The suffixes, provided by the registrant, are unique within their allocated prefix (DOI Foundation, 2023).

Supported by data centres, libraries, archives, universities, publishers, and research institutes that host repositories, DataCite has registered about 64 million DOIs as of July 2024 (DataCite, 2024a). A DOI is resolvable using the HTTPS Proxy Server of the DOI System (<https://doi.org/>), where it can be displayed as a link (DOI Foundation, 2023):

<https://doi.org/10.1000/182>  
proxy server   prefix   suffix

This link resolves to a resource, such as a web or internet resource, metadata describing the entity, or a landing page with access to further resources (DOI Foundation, 2023). Data

discovery can be achieved without making the data fully accessible, since "data discovery is the process of learning that a dataset exists" (Contaxis et al., 2022). Therefore, providing long-term access to a set of metadata describing the data or a landing page hinting at further access through a persistent identifier enhances the discoverability of data, even when the data itself is not openly available.

As of July 2024, among 3,231 repositories registered in re3data.org, 1,074 utilize the DOI system, 275 use the Handle system, and 63 employ the URN system, whereas 1,322 repositories implement no PID system for their digital objects (re3data.org, 2013). Following is a table summarizing the PID systems implemented by CESSDA members, the representative national data repositories for social science domains in Europe:

<b>Data Centre/Archive*</b>	<b>Country</b>	<b>PID System**</b>
Austrian Social Science Data Archive (AUSSDA)	Austria	DOI
Croatian Social Science Data Archive (CROSSDA)	Croatia	DOI
Czech Social Science Data Archive (CSDA)	Czech Republic	DOI
Data Archive Social Sciences Italy (DASSI)	Italy	None
Data Archiving and Networked Services (DANS)	Netherlands	DOI
Finnish Social Science Data Archive (FSD)	Finland	DOI, URN
GESIS Data Archive	Germany	DOI
Icelandic Social Science Data Archive (DATICE)	Iceland	DOI
Irish Social Science Data Archive (ISSDA)	Ireland	None
Portuguese Archive of Social Information (APIS)	Portugal	Handle
Sikt Research Data Archive	Norway	DOI
Social Science Data Archive of North Macedonia (MK DASS)	North Macedonia	None
Social Sciences and Digital Humanities Archive (SODHA)	Belgium	DOI
Swedish National Data Service (SND)	Sweden	DOI
Swiss Centre of Expertise in the Social Sciences (FORS)	Switzerland	DOI
TARKI Data Archive	Hungary	None
Social Data Network (SoDaNet)	Greece	DOI
UK Data Service (UKDS)	United Kingdom	DOI

*Table 4. 2 PID Systems Implemented by CESSDA Members*

\*Source: re3data.org and respective archive websites (last visit 11. July 2024)

\*\*DASSI, CROSSDA, MK DASS, DCS, SASD are not registered in re3data

The implementation of persistent identifiers like DOIs, Handles, and URNs is essential for ensuring the long-term accessibility and findability of research data in the scholarly community (Juty et al., 2020). Furthermore, PIDs enhance the stability and reliability of digital resources, thereby supporting robust scholarly communication as well as data sharing and reuse

practices (Data Citation Synthesis Group, 2014; Fenner et al., 2019; Juty et al., 2020; Khan et al., 2023). As evidenced by the majority of CESSDA member data centres and archives, DOIs are the most widely adopted PID system, demonstrating their importance in the social sciences. However, a significant number of repositories do not utilize any PID system, highlighting an area for potential improvement.

#### **4.2.2 Metadata Standards and Data Documentation**

Creating metadata that is both human-comprehensible and machine-readable enhances data findability and discoverability (Contaxis et al., 2022; Wilkinson et al., 2016). Rich metadata provides detailed and clear information about the dataset, making it easier to find and identify relevant data (Bensmann et al., 2020). During web searches, users scan text returned by information retrieval systems for relevant information. Properly described metadata makes the content of the data more understandable, facilitating data discovery during such explorations (Lin et al., 2020).

Furthermore, using structured metadata helps machines to parse and link to other resources (Contaxis et al., 2022). For instance, when retrieving data via an application programming interface (API), a structured metadata scheme facilitates the interaction with the resources and the metadata describing them, enabling retrieval of the correct data (Pomerantz, 2015, p. 193). DataCite, a DOI registration agency, exemplifies such practice. It offers the ability to retrieve DOI metadata records using various metadata properties listed in the DataCite Metadata Schema via an API. The DataCite API returns DOI records in JSON format (DataCite, 2024b; DataCite Metadata Working Group, 2024).

Registering metadata using the DataCite metadata schema enhances the findability of data by making it accessible through metadata discovery tools. When a registrant assigns a DOI to their dataset and conforms to the DataCite metadata format, the dataset becomes searchable via DataCite Commons (DataCite, n.d.). DataCite Commons serves as an interface for the PID Graph, which connects diverse PIDs and metadata records from DataCite, Crossref, ORCID, re3data, and other sources, aggregating their information (Fenner, 2020; Fenner & Aryani, 2019). Utilizing the DataCite metadata schema improves the findability and discoverability of datasets by enabling searches in DataCite Commons. OpenAIRE also expects repositories' metadata to be in the DataCite format so they can be harvested and included in OpenAIRE EXPLORE, the comprehensive metasearch service for research data (OpenAIRE, 2022a). Details about DataCite and other metadata schemas for inclusion in dataset search services will be further discussed in section 2.2.2.5: Inclusion in Search Services.

In addition to the DataCite Metadata Schema, several other metadata schemas are widely used across various disciplines, enhancing their broad applicability. Notable examples include the Dublin Core Metadata Initiative Schema (DCMI Schema, often referred as the Dublin Core

or DC) and the Metadata Object Description Schema (MODS) of the Library of Congress (LOC) (D. J. Lee & Stvilia, 2017). While the DataCite Metadata Schema is specifically designed to describe datasets, Dublin Core and MODS are utilized in a broader context, describing any digital resources (DCMI Usage Board, 2012). DataCite collaborates with Dublin Core to provide crosswalk mappings from its metadata schema to the DCMI schema, promoting openness and extensibility (DataCite Metadata Working Group, 2024). These metadata schemas provide high-level elements essential for digital research objects across disciplines.

The Data Documentation Initiative (DDI) is designed to structure metadata for describing social, behavioural, economic, and health data and is widely used in social science domains (Zondergeld et al., 2020, p. 5). The DDI is a metadata standard with a set of XML rules, ensuring automatic discovery of datasets through machine-readable metadata (*Why Use DDI? | Data Documentation Initiative*, n.d.). Numerous social science repositories have adopted the DDI, including AUSSDA, CESSDA Data Catalogue, ICPSR, QDR, and GESIS (*Wall of Adopters | Data Documentation Initiative*, n.d.). ICPSR documents their datasets with DDI standards, combined with DC whereas the Qualitative Data Repository at Syracuse University (QDR) curates their data collections both with DataCite and DDI metadata schemas to enhance their discoverability and findability (Qualitative Data Repository, 2023a). Additionally, the Text Encoding Initiative (TEI) is widely used for describing textual documents (TEI Consortium, n.d.). The UK Data Service (UKDS) utilizes TEI schema to develop a metadata standard for describing their qualitative data collections (UK Data Service, 2024a).

ICPSR identifies key metadata elements crucial for social science data in their guidelines. These elements include project description, sample and sampling procedures, unit of analysis/observation, variables, related publications, data collection instruments, among others (Inter-university Consortium for Political and Social Research, 2020). Although the Data Documentation Initiative (DDI) was initially designed to standardise metadata for quantitative social science data, it can also be applied at the study level to describe qualitative research (Mannheimer, 2024, p. 36; Vardigan, 2013). Additionally, the QualidataNet of KonsortSWD in Germany defines a metadata core set for qualitative data using DDI standard by consenting which DDI metadata are suitable as minimum criterion for creating an index of qualitative datasets. These are Title; Creator; Description/Abstract; Keywords; Temporal coverage; Geographic coverage; Resource type; PID; Access type; and Archive organization (Betancort Cabrera & Mozygemba, 2022).

However, the application of DDI metadata to qualitative data presents several challenges, such as accommodating complex study designs, managing relationships between files, preserving the hierarchical structure of codes, and attaching comments or memos to specific text segments or codes (Mannheimer et al., 2019). To address these challenges, Mannheimer et al. advocate for individual qualitative researchers to provide detailed narrative documentation (2019). This approach enables data curators to create structured descriptions

that are both machine- and human-readable, thereby enhancing data discoverability and ensuring long-term preservation.

Comprehensive documentation of qualitative datasets should include details on research methods and practices, project plans, informed consent forms, interview or observation settings and guidelines, participant details, and data collection instruments (Inter-university Consortium for Political and Social Research, 2020; Van Den Eynden & Chatsiou, 2011). This information can be provided as extensive metadata, assuming the metadata schema allows for such values, or in the form of documentation files or a README file (a plain-text document stored alongside the data). Such extensive documentation supplements DDI metadata, addressing its limitations in comprehensively describing qualitative research data (Inter-university Consortium for Political and Social Research, 2020; Mannheimer et al., 2016).

Furthermore, UKDS provides templates for interview transcription and data listing to describe contextual information, participant details, and observations for qualitative data documentation (UK Data Service, n.d.-a). To enable machine-actionability and rich descriptive metadata for qualitative data, the UK Data Archive also maintains the Qualitative Data Exchange Schema (QuDEx) as a complement to DDI, along with the usage of TEI metadata standards (Van Den Eynden & Chatsiou, 2011). QuDEx is utilized in QualiBank, the UK Data Service's discovery tool for qualitative datasets (UK Data Service, 2024b). Furthermore, QDR has developed the Annotation for Transparent Inquiry (ATI) tool to support contextual information and cross-linking of qualitative datasets (Karcher & Weber, 2019). ATI allows scholars to annotate specific passages to provide extended commentary or link to additional sources, enhancing dataset findability through persistent identifiers (*ibid.*)

### **4.2.3 Controlled Vocabularies (CV)**

Controlled vocabulary (CV) traditionally refers to a standardized set of terms and phrases used in information retrieval systems to ensure consistency and accuracy in the indexing and retrieval of information. It involves the selection of preferred terms and the exclusion of synonyms or near-synonyms, enabling more effective and precise searching (Buckland, 1999; Svenonius, 2001). Types of CV include subject headings, thesauri, ontologies, and taxonomies (Svenonius, 2001). The Library of Congress Subject Headings (LCSH) and the Thesaurus of Psychological Index Terms by the American Psychological Association are examples of controlled vocabularies that contain complex hierarchical relationships and cover a broad range of terms in subjects.

Scholarly databases typically index their items under terms from these subject languages (Fidel, 1992; Svenonius, 2001). The application of controlled vocabularies during the information retrieval process increases the recall of the search—this means retrieving a larger number of related items out of the total number of related items in the system—by recognizing all the possible terms under which a database would index the relevant items a searcher is

looking for (McGowan et al., 2016). This practice also enables advanced or targeted searches using these terms, as well as thematic browsing and search filtering using facets derived from them (Fowler & Jewell, 2022; Wu et al., 2019). Supporting both targeted search and explorative browsing in a search service is essential for data discoverability and findability, conforming to users' different information needs and consequent search behaviours (Allen, 1991; Henrich et al., 2021; Wu et al., 2019).

This practice can also be applied in the context of research data, where metadata standards for research data are often accompanied by controlled vocabularies to define their contents (Dierkes, 2021, p. 317). Domain-independent controlled values for data description (data terms) can be represented by the controlled list values provided by the DataCite Metadata Schema (DataCite Metadata Working Group, 2024). Several metadata properties in the schema, such as `identifierType`, `nameType`, `titleType`, `contributorType`, `dateType`, `resourceTypeGeneral`, `relationType`, `descriptionType`, `funderIdentifierType`, and `numberType`, must be described with predefined terms (*ibid.*). Similarly, DCMI Metadata Terms specifies all metadata terms maintained by the DCMI, including properties, vocabulary encoding schemes, resource types and others (DCMI Usage Board, 2020).

DataCite strongly recommends the inclusion of different types of relations between resources (e.g., `IsCitedBy`, `IsSupplementTo`, `IsReferencedBy`) as well as relevant types of data description (Abstract and Methods) to enhance data discoverability (*ibid.*). The Data Citation Implementation Pilot (DCIP) Expert Group appoints important discovery metadata for data repositories with examples of Dublin Core, Schema.org and DataCite metadata schema. These are Description, Keywords, License, Related Dataset and Related Publication (Fenner et al., 2019). More details on Schema.org are provided in section 2.2.2.5 Cross-indexing.

In addition to the pre-defined metadata element types for general information, domain-specific controlled vocabularies exist to further aid the finding and discovery of social research datasets. The CESSDA Vocabulary Service currently lists 35 CVs developed by GESIS, DDI and CESSDA, that are tailored to suit the distinctive research practices in the social sciences. These encompass various sets of terms for data formats, concept groups, data collection sampling methods, types of notes, instruments used and data types produced in social sciences, and many others (Consortium of European Social Science Data Archives, 2023b). Many of these vocabularies are developed by the DDI Controlled Vocabularies Group (DDI-CVG), of which usage is mandatory or recommended in certain DDI elements for consistency and efficiency of data description. This practice is implemented in the CESSDA Data Catalogue (Jaaskelainen et al., 2010).

Furthermore, several social science subject terms exist for more specific content information of the data such as an abstract, data description, topic classifications and keywords. The UK Data Archive, for example, developed and uses the Humanities and Social Science Electronic Thesaurus (HASSET) to index their data collections, allowing the retrieval of data and related

documentation using hierarchies of keywords (UK Data Archive, n.d.). The European Language Social Science Thesaurus (ELSST) is a multilingual thesaurus for the social sciences, published and maintained by CESSDA and its national service provider (CESSDA And Service Providers, 2023). ELSST is originally based on HASSET and consists of over 3,300 concepts covering diverse social science disciplines and is available in 16 languages, enabling broader discovery and access of data resources across Europe (ibid.). Along with the CESSDA Topic Classification for topic search, CESSDA Data Catalogue utilizes ELSST terms provided by its metadata providers for keyword search (CESSDA, 2022).

In the United States, the ICPSR has developed a controlled vocabulary system called the ICPSR Thesaurus. It includes Personal Names Authority List, Organization Names Authority List, Subject Thesaurus and Geographic Names Thesaurus, facilitating filtered search as well as browsing using these terms within the repository (ICPSR, n.d.). The Thesaurus for the Social Sciences owned by GESIS (THESOZ Thesaurus or Thesaurus Sozialwissenschaften) is a representative instrument for originally indexing content of documents in social science databases in German, English, French, and Russian languages (GESIS, n.d.-a; Zapilko & Sure, 2009). Given that keyword search in Google, data portals, and library catalogues is a popular search strategy for research data among social scientists (Krämer et al., 2021), the development and usage of controlled vocabularies in data portals is a significant advancement for enabling the data discovery in social science domain.

Furthermore, recent endeavours to improve search of social science data on the web, such as Google, are evident. The Linked Open Research Data for Social Science Pilot Study (LORDpilot), led by Kern, acknowledges the challenges posed by inconsistent or missing semantic indexing in social science research data (Kern & Momeni, 2022). The study aims to address these issues by integrating concepts from THESOZ Thesaurus into the Linked Open Data Cloud (LOD Cloud) (McCrae, 2024). This integration will be achieved by linking between concepts using RDF triples (RDF Working Group, 2014), thereby enabling comprehensive and thematic searches on the web through semantic indexing of research data. Similarly, the DDI-RDF Discovery Vocabulary (Disco) is designed to facilitate the discovery of social survey research data on the web. It is based on DDI XML formats and employs RDF technologies within the Web of Linked Data (Bensmann et al., 2020). Given the significant role of web searching in finding and discovering relevant data among researchers (Gregory et al., 2020), these advancements are crucial for enhancing discoverability and findability of research data.

On November 29, 2023, at the European DDI User Conference (EDDI) in Ljubljana, a new DDI subgroup “Non-numeric data/qualitative data” was founded. Currently, QualidataNet, part of KonsortSWD in Germany, is collaborating with this DDI working group to develop standardized and comprehensive terminologies for qualitative data (Betancort Cabrera & Mozygamba, 2022, 2023). In conjunction with the DDI-CVG, KonsortSWD is actively involved in developing the QualiTerm controlled vocabulary. This initiative aims to enhance the

accurate description of qualitative data by expanding existing DDI controlled vocabularies and introducing new concepts (Betancort Cabrera & Mozygema, 2023; Vardigan, 2013).

#### **4.2.4 Data Citation and Linkage of Scholarly Literature with Datasets**

The relevance of scholarly literature in data search and discovery practices is evident in recent studies (Gregory et al., 2019, 2020; Krämer et al., 2021; Kratz & Strasser, 2015a). A qualitative study of social scientists' dataset search behaviours (n=12) highlighted the critical role of literature search in intended data finding (Krämer et al., 2021). Additionally, a series of interviews on general data search practices (n=22) underscored the importance of reading scholarly literature for the serendipitous discovery of data. A large-scale survey (n=1677) further confirmed that academic literature is a key source for discovering data, with one-third of respondents identifying it as a primary method, alongside general search engines (e.g., Google) and disciplinary data repositories (Gregory et al., 2020).

Various strategies for utilizing scholarly literature to find data include following citations to data, literature searching with the goal of finding data, discovering data serendipitously while reading literature, extracting data directly from literature, tables, or graphs, and "footnote chasing" (Gregory et al., 2020; Krämer et al., 2021).

As seen above, effective data citation practices and the linkage of scholarly literature with datasets are essential components for enhancing data discoverability and findability. Various initiatives and frameworks have been developed to facilitate these practices. These include the Joint Declaration of Data Citation Principles (JDDCP) (Data Citation Synthesis Group, 2014), the data citation roadmap for scholarly data repositories by the Data Citation Implementation Pilot (DCIP) Expert Group (Fenner et al., 2019), and the Scholarly Link eXchange (Scholix) framework (Burton et al., 2017; Cousijn et al., 2019; Khan et al., 2020).

The JDDCP, released in 2014, outlines guidelines concerning the purpose, function, and attributes of data citations (Data Citation Synthesis Group, 2014). The data citation roadmap for scholarly data repositories further focuses on the practical implementation of JDDCP principles for data discovery (Fenner et al., 2019). The identified metadata elements for data citation available in representative metadata schemas (Dublin Core, Schema.org, DataCite) include identifier, title, creator, data repository or archive, publication date, version, and type (Fenner et al., 2019). The DCIP project found that essential elements for data citation, such as machine-actionable PIDs resolving to a specific landing page for the dataset or dataset collection, PID granularity to support the citation of a specific or unspecified version of a dataset, and metadata and PID embedded in the landing page, are well implemented in research data repositories. However, recommended practices such as providing metadata using schema.org markup still need further implementation (ibid.).

Scholix is a framework developed to create a universal mechanism for linking scholarly literature with datasets (Burton et al., 2017). It uses metadata aggregators such as DataCite, Crossref, and OpenAIRE as community hubs to collect and exchange information about links

between scholarly literature and data. These hubs then exchange information using standardized methods and protocols, such as JSON, XML, and RDF formats, as well as RESTful APIs, OAI-PMH, and SPARQL protocols. Further details on metadata exchange will be provided in the following section 2.2.2.5: Metadata Exposure, Harvesting, and Inclusion in Search Services.

Scholix specifies the necessary information about the elements involved in data-literature links, such as PIDs, object types, and relationship types. Potentially applicable metadata standards for this information include the XML Schema `dateTime` property, DataCite Metadata Schema relationship properties (e.g., `IsCitedBy`, `IsSupplementTo`, `IsReferencedBy`), DataCite resource type property, ORCID for author information, and many others (*ibid.*). This linkage enables the identification of data citations, thus enhancing the visibility of datasets (Cousijn et al., 2019).

Scholixplorer is an API based on the Scholix framework (OpenAIRE, 2024c). A quantitative study on the usability and quality of data retrieved from the Scholixplorer API identified a significant number of datasets associated with research outputs at a particular research organization, proving the effectiveness of Scholix implementation in data discovery (Khan et al., 2020). The study found evidence of data reuse and improved metadata quality by linking author information to research outputs. Additionally, the study recommends enriching Scholixplorer metadata with controlled vocabularies and PIDs (*ibid.*).

According to the existing guidelines and frameworks aimed at improving data citation practices, it is essential to assign the necessary metadata elements, including a persistent identifier, title, creator, data publisher, publication date, version, resource type, and relationship type. Data publishers should also ensure that metadata is provided to Scholix hubs, such as DataCite or OpenAIRE, to facilitate linkage to scholarly literature. Implementing these metadata elements ensures that datasets are easily discoverable, properly cited, and effectively linked to scholarly literature, thereby enhancing the overall findability and reusability of qualitative research data in social sciences.

#### **4.2.5 Inclusion in Search Services**

Numerous metasearch services for research datasets are available, including the CESSDA Data Catalogue, Suchportal of QualidataNet, B2FIND, DataCite Commons, Crossref Metadata Search, OpenAIRE EXPLORE, and OpenAlex. Metasearch refers to “the process of simultaneously searching multiple remote databases from a single user interface” (Dorman, 2008). These services aggregate metadata from various data providers, such as national data centres, repositories, archives, and journals, offering users a unified platform to access a comprehensive list of research data. By collecting and integrating metadata from diverse sources, metadata indexing enables these services to create a cohesive and comprehensive index. This practice is crucial for improving the findability and discoverability of qualitative research data, especially in the social sciences where data can be dispersed across multiple

locations, as discussed in subchapter 1.1, “Knowledge Infrastructure and Data Landscape for Qualitative Social Sciences”.

Data repositories and dataset metasearch service providers typically use the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) to expose and harvest metadata. OAI-PMH provides a framework for retrieving metadata from multiple repositories, utilizing Hypertext Transfer Protocol (HTTP) as the transport layer and XML for encoding metadata (Lagoze & Van de Sompel, 2015). Conventionally, data repositories expose their metadata in general formats such as Dublin Core. However, the protocol also supports the dissemination of records in multiple metadata schemas (ibid.).

When a metasearch service provider sends an HTTP GET request to the repository’s OAI-PMH endpoint, the repository (data provider) responds with an XML document containing the requested metadata records. For example, the following query requests all records in the “oai\_dc” (Dublin Core) metadata format from a repository:

```
GET http://example-repository.org/oai?verb=ListRecords&metadataPrefix=oai_dc
HTTP/2.0
Host: example-repository.org
```

*Table 4.3 Example GET Request to a Repository's OAI-PMH Endpoint*

This process allows metasearch services to aggregate metadata efficiently from various data providers, thereby enhancing the findability and discoverability of research data. OAI-PMH supports the use of other metadata schemas, such as the DataCite Metadata Schema (metadataPrefix=oai\_datacite), through which DataCite exposes its metadata of registered DOI records. Similarly, OpenAIRE uses OAI-PMH to harvest dataset metadata encoded in the DataCite format (DataCite, 2023; OpenAIRE, 2022).

In the field of social sciences, the Data Documentation Initiative (DDI) metadata schema (e.g., metadataPrefix=oai\_ddi25 for DDI-Codebook v2.5) is predominantly employed for exposing and harvesting metadata associated with social science research datasets. This practice is exemplified by CESSDA, which provides comprehensive procedural documentation in the CESSDA Data Catalogue User Guide (Consortium of European Social Science Data Archives, n.d.-b). Additionally, ICPSR (ICPSR, 2024a) and the UK Data Service (UK Data Service, 2024a) also expose their dataset metadata to aggregators. The successful implementation of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) necessitates that data are described using interoperable metadata schemas (Lagoze & Van de Sompel, 2015). This underscores the critical role of metadata schemas in enhancing data findability and discoverability.

Following the implementation of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) for metadata exchange among repositories and dataset search services, another critical strategy for enhancing the discoverability of research datasets is the use of Schema.org

markup for indexing the content of dataset landing pages in Google. Numerous studies have demonstrated that searching for research data in search engines such as Google is a well-established and common practice among researchers (Chapman et al., 2020; Gregory, 2021; Gregory et al., 2019, 2020; Krämer et al., 2021; Kratz & Strasser, 2015b; Wu et al., 2019)

Schema.org is a collaborative initiative among major search engines like Google, Bing, and Yahoo, aimed at creating a common vocabulary for structured data on the web (Schema.org, 2024). Schema.org markup for datasets involves embedding structured data (Schema.org vocabulary) in the HTML of dataset landing pages, usually hosted by data repositories. This structured data, added to the landing page of a dataset, can be recognized by Google, enhancing the visibility of the web content in search results (Stathis et al., 2022). This structured data provides search engines with detailed information about the dataset, such as the title, authors, publication date, and download links, typically using the JavaScript Object Notation for Linked Data (JSON-LD) format (Google, n.d.). By adding the value "Dataset" in Schema.org property @type, the landing page of the dataset can also be found in Google Dataset Search (Google, n.d.; Stathis et al., 2022).

While DataCite itself offer metadata for each DOI through platforms like DataCite Commons and exposes its metadata via OAI-PMH for inclusion in more comprehensive metadata indexing services such as OpenAIRE, the organization further encourages individual repositories to implement Schema.org on their own websites (DataCite, 2023). This decentralized approach ensures that datasets are not only visible within research data ecosystems but also across popular search engines like Google. The use of Schema.org complements existing metadata exposure methods like OAI-PMH, providing a robust framework for maximizing the discoverability of research data.

In addition to including a dataset's metadata in a metasearch service, indexing a repository in a widely recognized registry can further enhance the dataset's findability. The Registry of Research Data Repositories (re3data.org) is a comprehensive registry that encompasses 3,210 research data repositories across various academic disciplines (re3data.org, 2013). It is widely acknowledged as a valuable tool for identifying repositories suitable for data deposition, with endorsements from journals and funding bodies, including the European Commission (European Commission, 2017). Other scholarly registries, such as FAIRsharing (FAIRsharing, 2009) and the Directory of Open Access Repositories (OpenDOAR) (Jisc, 2005), also exist. FAIRsharing and re3data complement each other in terms of domain coverage and curation. Specifically, "the FAIRsharing registry is manually curated and has a historical focus on the Life Sciences, while re3data is not curated and is less updated, but has a larger corpus of institutional repositories" (Jahn, Najko et al., 2023, p. 8).

#### **4.2.6 Social Interactions**

In discovering and finding data, social contact plays an essential role (Gregory et al., 2019). Through a series of 22 interviews conducted in 2019, Gregory and her colleagues identified the importance of social interactions in acquiring data among researchers (ibid.). The routes to accessing data in these findings are diverse, including “conference presentations,” “informal conversations with colleagues,” “networking and training events,” and “professional networks and connections.” In some communities, data access is granted only to a small group of trusted colleagues. The study also provides an example of historians, who often need to develop intimate relationships with the individuals they study to obtain the necessary documents (ibid.). Researchers also share their know-hows for better data retrieval in databases and consult support professionals like librarians. Combining data retrieval techniques such as using APIs, human-created metadata schemas and documentation, and social interactions, the study team concludes that data search is a complex socio-technical practice (ibid.).

These findings are supported by further studies (Gregory et al., 2020; Krämer et al., 2021). A qualitative study on data search behaviour in the social sciences (n=12) highlighted the importance of personal contact with “relevant researchers who publish in portals,” “acquaintances at the university dealing with datasets previously at work or during education,” and “presentations of datasets at conferences” in data search (Krämer et al., 2021). One participant even reported hearing about a dataset on the radio, illustrating the diverse ways data can be discovered (ibid.). Participants also relied on study series and portals they learned about during their education when beginning their data search. Additionally, they reported receiving help from colleagues and making contacts for data acquisition when initial searches were unsuccessful (ibid.). A large-scale survey with researchers (n=1,630) further proves this, showing that 85% of the survey participants answered that they often or occasionally use personal networks to find data as well as 58% who use support professionals (Gregory et al., 2020).

The role of social interactions extends to digital platforms as well. Academic social networks (ASNs) are a new interest in terms of data discovery. For example, further research by Raffaghelli and Manca (2023) explores the social activity of open research data (ORD) on ResearchGate, revealing that there is still underdeveloped social activity around ORDs in terms of reads and citations.

### **4.3 Summary**

Ensuring that a dataset is findable and discoverable is a complex practice involving various elements. Formal metadata schemas, such as Dublin Core or DataCite Metadata Schema, enhance data interoperability, facilitating inclusion in various aggregators and search engines. Discipline-specific metadata schemas, like DDI, provide additional description and context, which is crucial for data seekers. Comprehensive documentation is also essential for

qualitative data, as it supplements the information and context not fully captured by formal metadata schemas.

To maximize discoverability, metadata and descriptions should be written using community-accepted vocabularies, ensuring that they are searchable with relevant keywords. Assigning appropriate controlled vocabularies as keywords or topics facilitates faceted search or browsing by topic/keywords. Additionally, using standardized protocols and formats, relationship metadata, and proper data citation with correct identifiers streamlines the linkage between scholarly literature and datasets, thereby fostering data discovery within well-established literature reviews among researchers.

The efficacy of these practices within the domain of Qualitative Social Science (QSS) remains uncertain, motivating this study to investigate their implementation. The following chapter details the design of the empirical study and analysis framework, followed by the results of data collection and analysis.

## 5 Methods

### 5.1 Study Design

As outlined in the introduction, this study seeks to answer the primary research question: *To what extent are findability and discoverability practices implemented in qualitative social sciences for research data?* To explore this question, the study employs two prevalent data retrieval strategies: (1) searching within domain-specific repositories and (2) tracing data through scholarly literature.

These methods form the basis for the following sub-questions:

1. To what extent are findability and discoverability practices implemented in discipline-specific repositories containing social qualitative data?
2. To what extent are research data findable and discoverable in scholarly journal articles on qualitative social research?
3. To what extent do scholarly journals that publish papers on qualitative social research implement findability and discoverability practices for research data?

These sub-questions categorize the inquiry into specific areas, focusing on the practices within repositories, scholarly literature, and scholarly journals. This structure allows for a targeted analysis of key components related to data findability and discoverability in qualitative social sciences (QSS).

The study employs a specific research inquiry within the field of qualitative sociology as a use case. This approach reflects the relevance of the research to real-world contexts and enables a detailed exploration of qualitative data findability and discoverability on a practical scale. The use case examines the life-worlds of refugee families with a disabled child, a topic identified in collaboration with a master's student in sociology at the University of Bern, Switzerland (Personal communication, S. Melattukunnel, 24 April 2024).

Within the scope of this thesis, the use case focuses on identifying and discovering relevant qualitative research data in both English and German, without geographical limitations. The two data retrieval strategies selected for the study guide this process.

### 5.2 Data Collection

As outlined earlier, this study investigates data findability and discoverability within the qualitative social sciences using two distinct approaches: dataset searches in discipline-specific repositories and tracing data within scholarly journal articles. These approaches focus on three sample groups—discipline-specific repositories, scholarly journal articles, and

scholarly journals—and are designed to assess the implementation of key practices, as defined in Subchapter 4.2, “The FAIR Data Principles and Other Aspects for Data Findability and Discoverability.” Specifically, the study examines the following practices: (1) Persistent Identifiers; (2) Metadata standards and data documentation; (3) Controlled vocabularies; (4) Data citation and the linkage of scholarly literature with datasets; and (5) Inclusion in search services. The aspect of social interactions was excluded from this study due to its irrelevance to the selected approaches. The methodologies and variables employed in each approach are summarized below, with detailed information available in Appendices B, C, and D.

### **5.2.1 Dataset Searching in Discipline-Specific Repositories**

This study employs a systematic approach to investigate how qualitative data can be identified and discovered within discipline-specific repositories, focusing on a specific use case involving refugee families with a disabled child. Data centres and repositories identified through the landscape investigation (Subchapter 3.2, “Research Data Centres and Repositories”) as well as those indexed in re3data.org were selected for dataset searches and metadata extraction.

The repository selection process involved utilizing the re3data.org REST API to retrieve relevant repositories by applying specific filters such as the query “qualitative,” the subject area (111 Social Sciences), and the repository type (disciplinary). The final query, executed within a Jupyter Notebook environment using Python, retrieved 18 repositories. Details of the GET requests and the repositories retrieved via the API are provided in Appendix A. These repositories were combined with those identified during the landscape investigation, resulting in a final selection of 21 repositories. The detailed criteria and process for the final repository selection are documented in Appendix B, Worksheet B.2.

Within each repository, exploratory searches were conducted using the primary search terms “refugee” and “disability,” supplemented by Boolean operators, synonyms, and thematic browsing. These searches were refined based on the repository’s size and search capabilities. The search protocols are documented in Appendix B, Worksheet B.3. The datasets identified as pertinent to the use case were recorded for potential comparison with datasets associated with scholarly journal articles, as outlined in the subsequent approach.

In addition to identifying relevant datasets, further information about each repository was manually collected on variables related to advanced search and browsing functionalities, controlled vocabularies, data citation practices, and guidelines on qualitative data documentation. This information is detailed in Appendix B, Worksheet B.4.

Relevant metadata from these repositories were retrieved via API requests, including PIDs, metadata standards, as well as keywords and descriptions in each re3data record. These metadata are documented in Worksheet B.5 and subsequently merged with the data in

Worksheet B.4. The combined data is presented in Worksheet B.6, offering an overview of the data collection process through this approach.

The complete process of data retrieval from repositories, including criteria for dataset relevance and repository information collected, is thoroughly outlined in Appendix B.

### **5.2.2 Tracing Data in Scholarly Journal Articles**

The second approach employed in this study involves a comprehensive search of scholarly journal articles, specifically focusing on tracing cited, linked, or associated datasets within these articles. To ensure an unbiased identification of relevant scholarly articles, a systematic search strategy, hereafter referred to as systematic searching, was implemented. This approach explores the potential for serendipitous discovery of research data through conventional research synthesis and the examination of research evidence. Unlike the structured search of repositories, this method illustrates how research data can be uncovered through broader academic discourse, such as the reading and analysis of scholarly articles.

This approach evaluates the effectiveness of current practices in linking scholarly literature to relevant research data within the field of qualitative social science (QSS). The assessment involves not only the examination of data availability statements and supplementary data within each identified research paper but also an investigation of the journals' efforts toward data visibility. This includes a thorough review of each journal's data policy on data sharing and data availability statements, submission guidelines for data as supplementary material, and data citation guidelines.

#### *Systematic Evidence Searching*

This approach involves employing a systematic evidence search strategy to locate scholarly literature pertinent to the use case, with the ultimate goal of tracing the underlying datasets. Systematic evidence synthesis methods, such as systematic reviews, are designed to mitigate the subjectivity and potential biases associated with traditional literature reviews by utilizing exhaustive search strategies and predefined selection criteria (M. J. Foster & Jewell, 2022; Grant & Booth, 2009; Mulrow, 1987, 1994).

#### *Step 1: Locating Relevant Journal Articles to the Use Case*

To conduct systematic searches for this thesis, the guidelines of the Peer Review of Electronic Search Strategies (PRESS) (McGowan et al., 2016) were adapted for the use case:

1. Formulation of Review Question and Definition of Search Criteria:

The review question was formulated using the population concepts defined in the use case. To maintain comparability between the two approaches used in the study, the search strategy focused solely on the key concepts considered during data retrieval in the repositories, namely "refugee" and "disability." Thus, the review question centred on the population group "refugee families with at least one child with a disability." The

criteria for determining document relevance were flexible to allow a broader search scope, ensuring alignment with the generic search strategy used in repository dataset searches (See Appendix C, Worksheet C.6).

## 2. Preliminary Search:

Initial searches were conducted in Google Scholar using generic terms. Search phrases provided by a sociology student conducting the use case research were translated into English and formulated into search queries using Boolean operators. This preliminary search yielded 123,000 results, from which 150 titles were screened, leading to the identification of 13 relevant journal articles. Of these, 4 articles were selected for their high thematic relevance and methodological appropriateness (qualitative interviews and narrative inquiries) (See Appendix C, Worksheet C.2).

## 3. Database Selection:

Based on where the example articles were indexed, disciplinary relevance, and availability to the author, the following were selected: one interdisciplinary database (Scopus), one academic search engine (Google Scholar), and one disciplinary database (PsycINFO) (See Appendix C, Worksheet C.3).

## 4. Term Harvesting:

Keywords and phrases were extracted from the titles and abstracts of the preliminary articles, and synonyms were added. Additionally, controlled vocabulary terms from the APA Thesaurus of Psychological Index Terms were used for searching within PsycINFO (See Appendix C, Worksheet C.4).

## 5. Search Strategy Development and Execution of the Search:

Combining harvested terms with Boolean operators and database-specific syntax created a comprehensive search strategy aimed at retrieving all relevant items in each database and search engine (See Appendix C Worksheet C.5).

## 6. Deduplication and Item Selection

Retrieved results were exported to a library within the reference management software Zotero for deduplication and item selection. After title and abstract screening, 41 relevant items were selected for further data collection (See Appendix C, Worksheet C.6).

### *Step 2: Identification of Associated Research Data and Manual Information Collection*

The landing page of each selected journal article was reviewed to locate any associated research data. Additionally, information was collected on variables related to data availability statements within each article, the presence of underlying data linked to the study, the journal's data policy on data availability statements and data sharing, and data citation guidelines. This detailed information collection process is documented in Appendix C, Worksheet C.7.

### **5.2.3 Variable Summary**

As outlined in the preceding sections, this study evaluates the implementation of data findability and discoverability practices in qualitative social research across three sample groups: discipline-specific repositories (S1), scholarly journal articles (S2), and scholarly journals (S3). The assessment focuses on five key indicators aligned with the FAIR Data Principles: (1) persistent identifiers, (2) metadata standards and data documentation, (3) controlled vocabularies, (4) data citation and the linkage of scholarly literature with datasets, and (5) inclusion in search services.

These indicators are operationalized through a set of variables, each associated with one of the sample groups. The specific variables collected through the approaches detailed in sections 5.2.1 and 5.2.2 include those related to repositories' PID systems, metadata standards, search and browsing functionalities, controlled vocabularies, utilization of OAI-PMH, indexing in search services, data citation practices, data availability statements in journal articles, and journals' data policies.

A comprehensive summary of all variables, along with their associated sample groups, labels, definitions, types, and categories, is provided in Appendix D, Worksheet D.2.

## **5.3 Data Analysis**

This thesis adopts an explorative and descriptive approach, emphasizing the analysis of frequency and percentage distributions within each variable category (J. S. Jones & Goldring, 2021). The study focuses on three small sample groups: discipline-specific repositories (S1), scholarly journal articles (S2), and scholarly journals (S3). By examining the quantitative evidence derived from the defined variables and the overall dataset collected during the exploratory phase, outlined in Appendices A, B, and C, this analysis seeks to identify patterns and narratives that emerge from the data.

## 6 Results

The data collected through the two explorative approaches, as detailed in Appendices A, B, and C, were meticulously cleaned and integrated to ensure accurate analysis and clear interpretation. For clarity and coherence, the data are presented with a distinct separation between the three sample groups, with variables systematically organized under clearly defined column headings (Appendix D). This chapter presents the final results of the data collection, structured according to the three sample groups, and is followed by a comprehensive conclusion in Chapter 7.

### 6.1 Sample Group 1: Discipline-Specific Repositories

Sample Group 1 comprises the 21 selected social science repositories that house qualitative data, which are detailed in Appendix B, Worksheet B.2. These repositories were chosen based on their relevance to the study's focus on data findability and discoverability within the field of qualitative social sciences.

#### 6.1.1 The Role of re3data.org

The variables RERE3DATA (indicating whether a repository is indexed in re3data.org) and RERE3DATAQUAL (indicating whether a repository indexed in re3data.org can be found using the search term “qualitative”) were analysed to assess the accessibility of social science repositories for qualitative data within the re3data.org registry. Although 95.2% (n = 21) of the selected repositories are indexed in re3data.org, only 60% (n = 20) could be retrieved using the predefined search parameters (“types[]”: “disciplinary”, “subjects[]”: “111 Social Sciences”, “query”: “qualitative”) (see Appendix D, Worksheet D.3).

This discrepancy reveals critical insights into the documentation practices within re3data.org. Among the eight repositories that were indexed but not retrieved with the specified parameters (Appendix B, Worksheet B.2), seven did not include the term “qualitative” in either their descriptions or keywords. The sole exception, FDZ Bildung, did include the term in its keywords but was not retrieved due to an incompatibility with the subject parameter values (re3data.org, 2023). This finding is visually summarized in Figure 6.1, which illustrates how the term “qualitative” is incorporated across different metadata fields in the re3data records of the selected repositories.

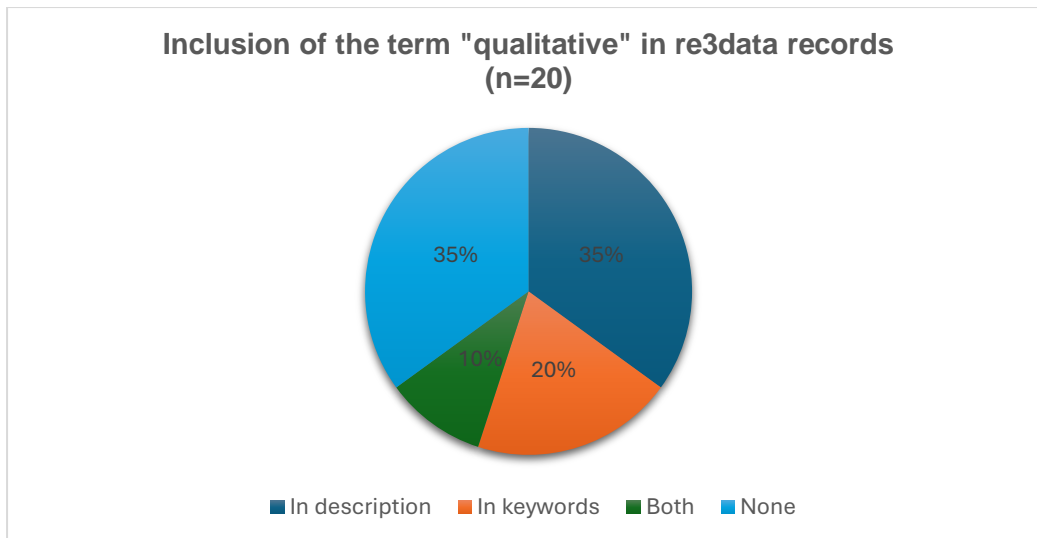


Figure 6. 1 Inclusion of the Term "Qualitative" in re3data Records of Selected Repositories

The seven repositories that were not retrieved with the search term “qualitative” are predominantly large-scale repositories, each containing over 1,000 datasets with a potentially significant number of qualitative datasets (see Appendix B, Worksheet B.4, last updated 2024-08-01). For instance, ICPSR currently holds 202 qualitative datasets (ICPSR, 2024b), which represent a substantial portion of the qualitative social data landscape. Similarly, the DANS Data Station Social Sciences and Humanities, with 301 results for the search term “qualitative,” and repositories like AUSSDA and ADA, which retrieve 32 and 23 results respectively, underscore the importance of these repositories for qualitative social science research (Figure 6.2).

About User Guide Support Log In

DANS Data Station Social Sciences and Humanities This Data Station allows you to deposit and search for data within the field of SSH.

Metrics 292,852 Downloads

Contact [Share](#)

qualitative  [Advanced Search](#) + Add Data

Datasets (63)

Files (238)

Publication Year

2024 (38)

2023 (21)

2022 (33)

2021 (16)

2020 (17)

[More...](#)

1 to 10 of 301 Results Sort

Data annex for the qualitative analysis

Nov 15, 2019

A Berti Suman, 2019, "Data annex for the qualitative analysis", <https://doi.org/10.17026/dans-262-3jwv>, DANS Data Station Social Sciences and Humanities, V1

... Empirical qualitative analysis for PhD Project "Sensing the risk - In search of the factors influencing the policy uptake of citizen sensing" ... Relation or Related Material Text: Data annex for the fuzzy-set Qualitative Comparative Analysis

Alternative Title: Qualitative Analysis

Online Child Sexual Abuse Qualitative Exploration Guidelines

Jan 18, 2024

---

AUSTRIAN SOCIAL SCIENCE DATA ARCHIVE
Search About User Guide Support Sign Up Log In

AUSSDA We make social science data accessible and reusable.

Metrics 126,003 Downloads

Contact [Share](#)

qualitative  [Advanced Search](#)

Datasets (11)

Files (19)

Dataverse Category

Organization or Institution (1)

Publication Year

2024 (2)

1 to 10 of 32 Results Sort

Qualitative Data

Mar 11, 2024

Explore scientific depths with **qualitative** data from AUSSDA. In this dataverse you will find data collected using **qualitative** research methods.

A Quantitative and **Qualitative** Analysis of the Austrian's Subjective Social Position from 1993 to 2016

Jul 13, 2018 - SWS-Rundschau

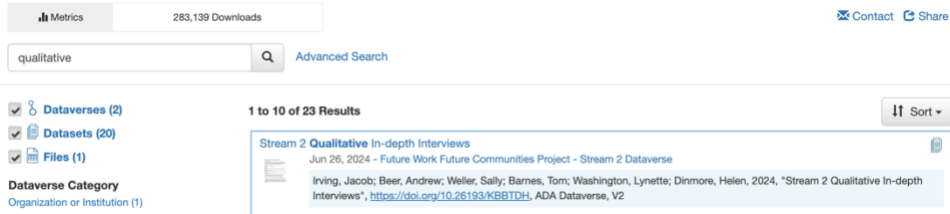


Figure 6. 2 Search results in DANS Data Station Social Sciences and Humanities, AUSSDA, and ADA with the search term "qualitative"\*

\*Last accessed 2024-08-27

This analysis highlights that while re3data.org offers extensive coverage for repositories containing qualitative research data in the social sciences, the insufficiently detailed metadata descriptions impede the effective identification of these repositories. Consequently, the discoverability of qualitative datasets within these repositories remains limited, emphasizing the need for more comprehensive and consistent documentation practices.

### 6.1.2 PID Systems

Among the 21 selected repositories, 90.5% employ DOI as the primary persistent identifier (PID) system for their data objects. Notably, two of these repositories (9.5%) combine DOI with other PID systems, such as URN or the Handle System. No repositories rely solely on these alternative PID systems. However, it is important to highlight that two repositories, Eurobarometer and Northern Ireland Qualitative Archive, do not implement any PID system at all.

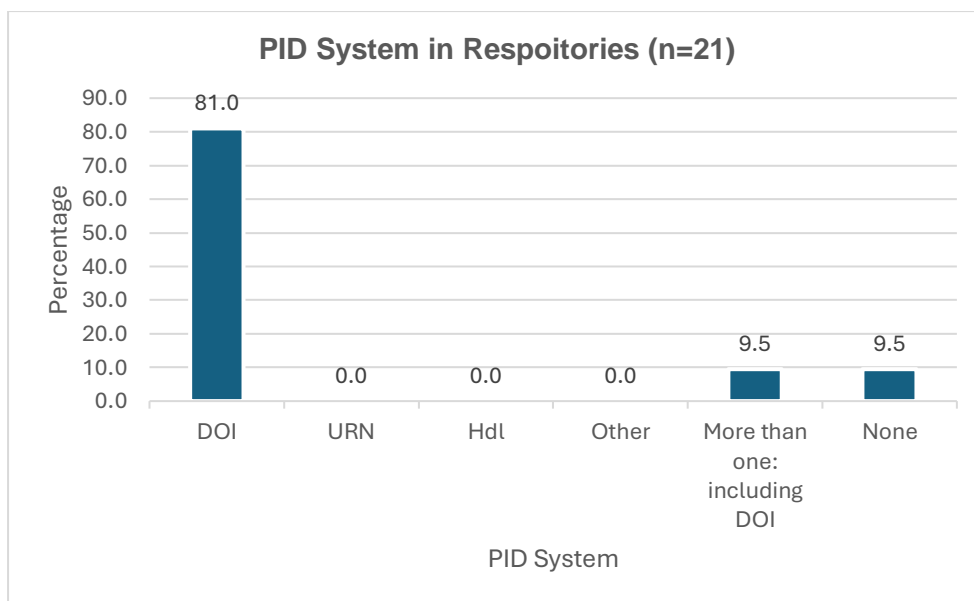


Figure 6. 3 PID System Implemented in Social Science Repositories

This distribution indicates a strong preference for DOI among social science repositories, reflecting its widespread adoption as a standard for ensuring the persistence and accessibility of digital objects. The absence of PID systems in two repositories suggests potential challenges in data citation and retrieval, which could hinder the discoverability and long-term preservation of the datasets they house.

### **6.1.3 Metadata Standards and Search Functionalities**

The variables REMDHL (indicating whether a repository describes its data objects using high-level metadata standards such as Dublin Core or DataCite) and REMDDS (indicating the use of discipline-specific metadata standards like DDI) provide insights into the adoption and utilization of metadata standards in social science repositories that house qualitative data.

In terms of discipline-independent (high-level) metadata standards, more than one-third of the repositories (33.3%) utilize Dublin Core, while only one repository (4.7%) exclusively employs the DataCite Metadata Standard. Additionally, 23.8% of the repositories use both Dublin Core and DataCite standards, reflecting a hybrid approach to metadata management. However, it is concerning that 38.1% of the repositories do not use any high-level metadata standards for their data holdings, which may limit the interoperability and discoverability of their datasets.

In contrast, the adoption of discipline-specific metadata standards is more pronounced, with 81% of the repositories employing DDI to assign metadata to their data objects. Nevertheless, 19% of the repositories do not utilize any discipline-specific metadata standards, potentially impacting the detailed description and accessibility of their qualitative datasets. Among the repositories that employ DDI, 28.6% also implement other specific standards, such as repository-developed metadata schemas, OAI-ORE (Open Archives Initiative Object Reuse and Exchange), and QuDEx (Qualitative Data Exchange Format) (see Appendix B, Worksheet B.5).

The cross-tabulation of these two variables (REMDHL and REMDDS) is visualized in Figure 6.4, highlighting the mixed usage of high-level and discipline-specific metadata standards across the selected repositories.

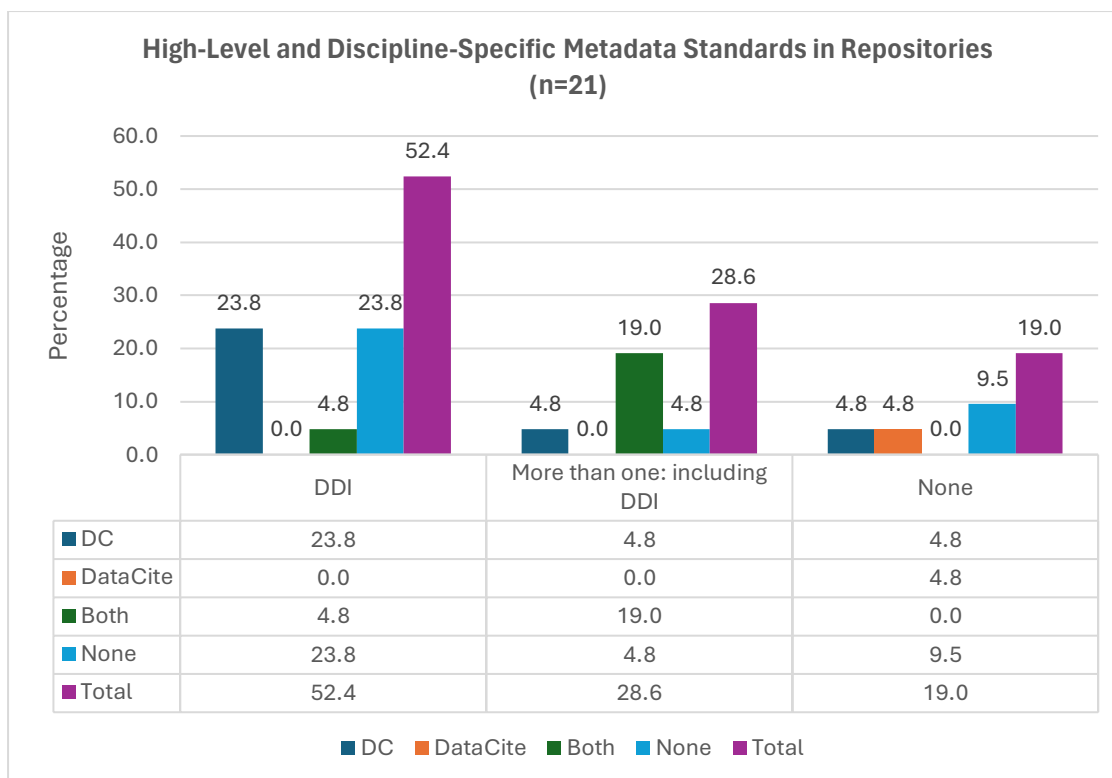


Figure 6. 4 High-Level and Discipline-Specific Metadata Standards in Repositories

As depicted in Figure 6.4, 23.8% of the selected repositories rely solely on DDI, while another 23.8% combine Dublin Core with DDI. Furthermore, four repositories employ a combination of four or more metadata standards, including both high-level and discipline-specific standards. In contrast, one repository exclusively uses Dublin Core, another solely uses DataCite, and two repositories (9.5%) do not implement any commonly recognized metadata schema.

The variable RESEARCH, which captures the availability of advanced search functionalities and filter options, reveals that while most repositories employ DDI, only 66.7% (n=14) offer advanced search or faceted filtering based on data type, data collection instruments, or detailed project/study descriptions. This finding suggests that even well-documented qualitative datasets may remain difficult to discover if repositories do not provide sufficient search functionalities, underscoring a critical gap in the accessibility of qualitative data within social science repositories.

#### 6.1.4 Browsing with Controlled Vocabularies

A significant proportion of the selected repositories (85.7%, n = 21) provide users with the ability to browse datasets through controlled vocabularies. However, three repositories (FDZ eLabour, FORS Data Service, and Qualiservice) do not offer this browsing functionality. The case of FORS Data Service is particularly noteworthy; although it indexes its datasets using the CESSDA Topic Classification (FORS, 2024a), it does not provide a browsing option for users. This inconsistency in indexing and the subsequent lack of a browsing function may be

due to concerns about the reliability of search results when using the classification, potentially leading FORS Data Service to omit this feature from their search capabilities.

Further analysis of the controlled vocabularies used by other repositories reveals a diverse landscape. For instance, CSDA utilizes the European Language Social Science Thesaurus (ELSST), while the UK Data Service indexes its data using the Humanities and Social Science Electronic Thesaurus (HASSET), which served as the foundation for ELSST. ICPSR employs controlled terms from its proprietary thesaurus (ICPSR Thesaurus), and ADA uses an Australian thesaurus known as the National Libraries APAIS thesaurus (see Appendix B, Worksheet B.6). Unlike these repositories, which rely on a single controlled vocabulary, 23.8% of the repositories (n = 5) use more than one controlled vocabulary to index their datasets. However, for 52.4% of the selected repositories, the specific controlled vocabularies employed remain unidentified, highlighting a gap in the transparency and documentation of indexing practices.

### 6.1.5 Inclusion in Search Services through OAI-PMH

A substantial majority of the repositories (76.2%, n = 21) expose their metadata through the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), facilitating the inclusion of their datasets in various search services. Conversely, five repositories either do not expose their metadata or relevant information on their OAI-PMH status could not be determined. The variable REINDEX categorizes repositories based on the number of search services in which they are indexed. The results indicate that 61.9% of the repositories are indexed across all three major metasearch services: DataCite Commons, OpenAIRE EXPLORE, and Google Dataset Search. However, 9.5% of the repositories are not indexed in any of these services. The bar graph presented in Figure 6.5 illustrates the relationship between the implementation of OAI-PMH and the extent of repository indexing, demonstrating that repositories exposing their metadata through OAI-PMH are more likely to be indexed in multiple search services.

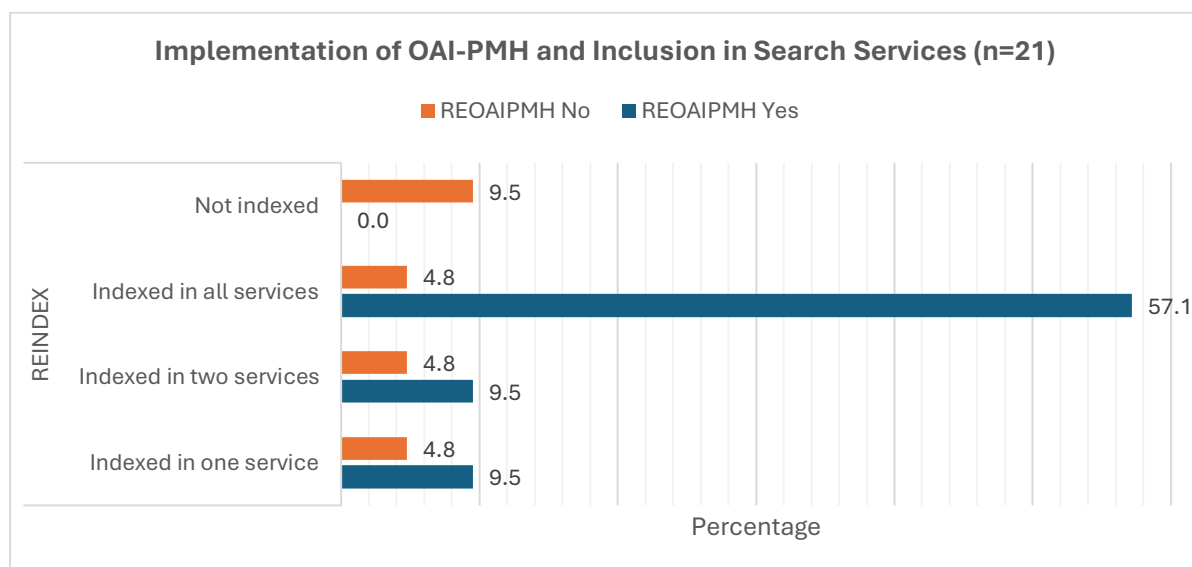


Figure 6. 5 OAI-PMH and Inclusion in Search Services

### 6.1.6 Data Citation and Guidance on Qualitative Data Documentation

A notable 81% of the repositories offer data citation options directly on the dataset’s landing page, facilitating easy citation by users. However, a significant gap remains in the provision of guidance on qualitative data documentation. Over half of the repositories (52.4%) do not offer separate guidelines detailing how qualitative datasets should be documented or which metadata properties are particularly important for qualitative data submissions. This lack of guidance may result in inconsistent or incomplete metadata, potentially hindering the discoverability and usability of the datasets.

## 6.2 Sample Group 2: Scholarly Journal Articles

Sample Group 2 comprises scholarly journal articles relevant to the use case, identified through a systematic search process. A total of 41 articles, published between 2003 and 2024, were selected for their pertinence to qualitative studies within the specified research context (see Appendix C, Worksheet C.6).

### 6.2.1 Data Available Statement and Available Data

Among the 41 journal articles identified from 2003 to 2024, only 7 (17%) included a data availability statement (DAS). Of these, datasets from three articles could be accessed by requesting them from the author(s), while two articles were unable to provide underlying data due to ethical restrictions. For one article, data sharing was not applicable because no new data were created or analyzed in the study (variable ARTDA, category “Other”). Only one article made the underlying data directly available as supplementary material to the publication (see Appendix C, Worksheet C.7). Notably, in one case, the underlying data were directly accessible within the article itself, although no formal data availability statement was provided.

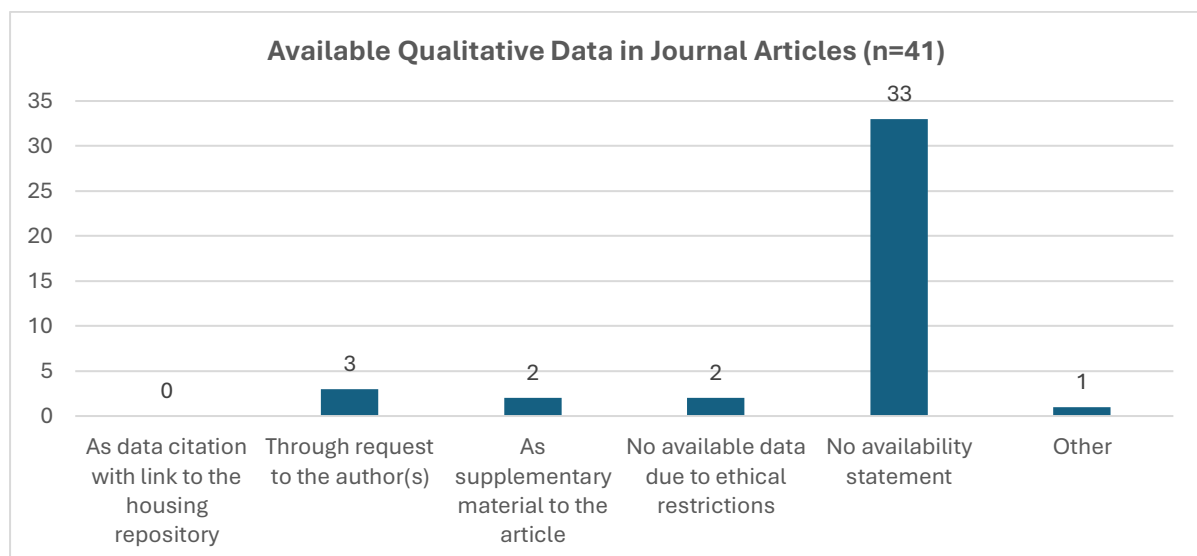


Figure 6. 6 Available Qualitative Data in Journal Articles

## 6.2.2 Publication Year and Data Available Statement

The following graph (Figure 6.7) illustrates the emergence of data availability statements over time, beginning in 2018.

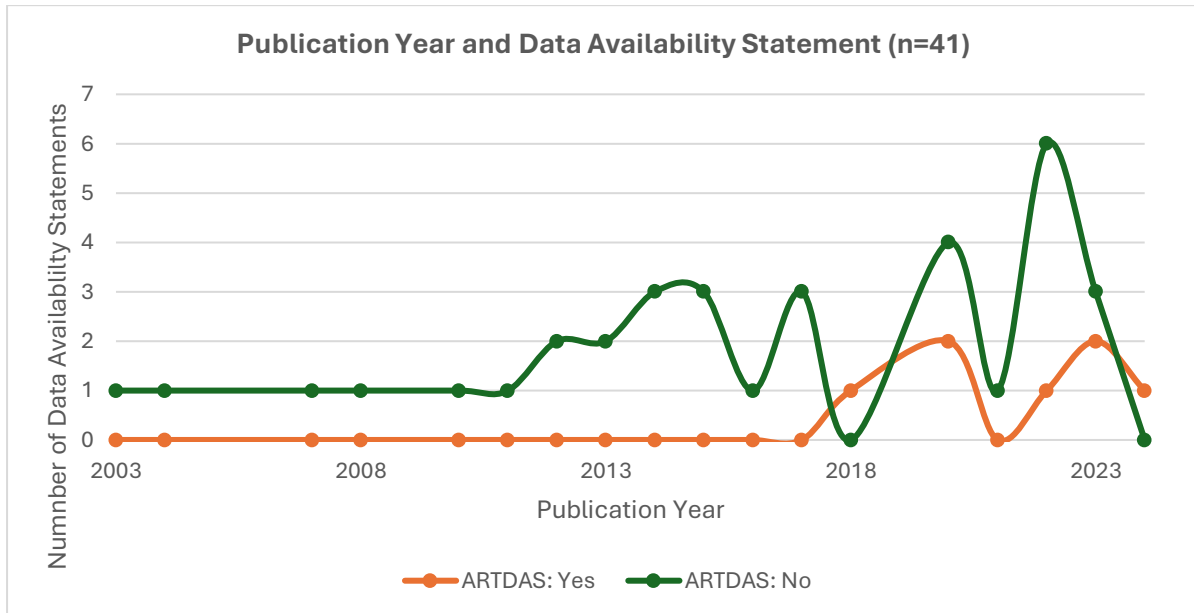


Figure 6.7 Publication Year and Data Availability Statement

Although the 41 journal articles selected for this use case are not representative of all qualitative studies, they provide a useful example of how data availability in scholarly articles of qualitative studies in a particular topic has evolved. The data suggest that while data availability statements have become more common in recent years, their adoption remains inconsistent. For example, the year 2022 shows that even recent articles from qualitative studies may still lack a data availability statement, with 6 articles not providing a DAS compared to only 1 article that did.

## 6.3 Sample Group 3: Scholarly Journals

From the 41 journal articles identified, 33 unique scholarly journals were selected for further analysis. One journal was excluded due to its discontinuation in 2020 (see Appendix C, Worksheet C.7). The selected journals share two key characteristics: (1) they publish qualitative research findings, and (2) they focus on research related to refugees, immigrants, and/or disabilities. These journals were evaluated based on three variables, each assigned ordinal values (0, 1, or 2) to rank their practices regarding data findability and discoverability. Higher values indicate a stronger commitment to promoting the visibility and accessibility of underlying research data. The values were summed to provide an overall score for each journal, ranging from 0 to 6.

### 6.3.1 Data Policy on Data Availability Statement and Data Sharing

Among the 32 journals analyzed, 62.5% require or encourage authors to provide a data availability statement (DAS) when submitting a manuscript. Similarly, 68.8% require or encourage authors to publicly share the underlying data associated with their studies. However, a notable difference emerges when examining whether these practices are mandatory. While at least seven journals require a DAS (score 2), only two journals mandate data sharing (score 2). This discrepancy may reflect the less prevalent culture of data sharing in qualitative research or the recognition by journals of the challenges associated with sharing sensitive qualitative data, particularly when it involves human subjects. Nevertheless, 13 journals encourage the inclusion of a DAS, and 20 journals promote data sharing (score 1), indicating a growing awareness among journals of the importance of data findability and discoverability, as well as a commitment to open research data practices. Conversely, 12 journals do not mention DAS in their policies or submission guidelines, and 10 journals do not address data sharing (score 0).

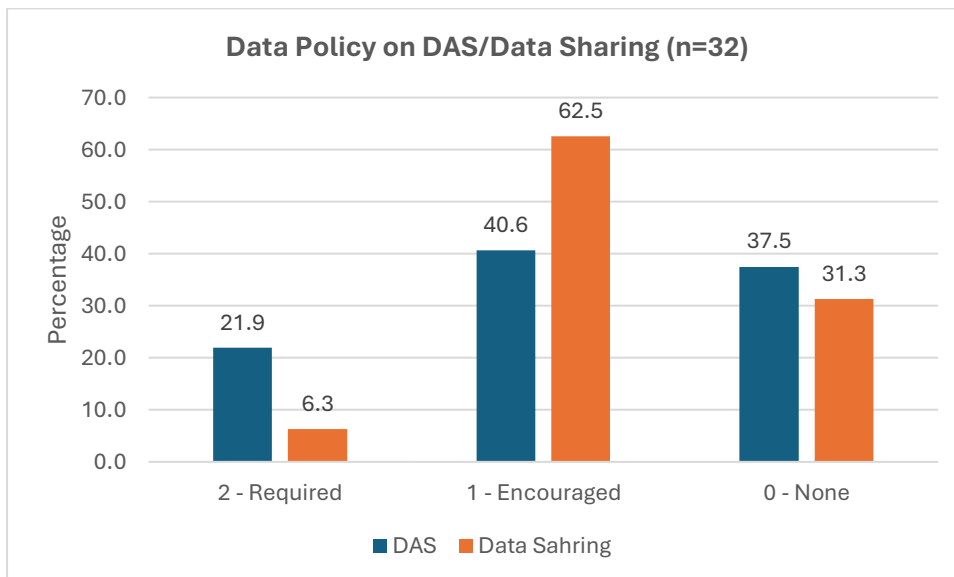


Figure 6. 8 Journal Data Policy on DAS/Data Sharing

### 6.3.2 Guideline on Data Citation

Regarding data citation, more than half of the journals (57.6%) provide detailed guidelines, including examples of how to cite datasets (score 2). Only one journal encourages data citation but does not offer specific guidance on how to do so (score 1). In contrast, 12 journals (37.5%) do not mention data citation in their data policies or author submission guidelines (score 0). This lack of guidance in a significant portion of journals may hinder the proper acknowledgment and discoverability of datasets used in qualitative research.

### 6.3.3 Overall Score

By summing the values of the three variables, an overall score for each journal's data findability and discoverability practices was calculated. A score of 0 indicates no implementation of these practices, 1 signifies very limited implementation, 2 denotes limited implementation, 3 indicates sufficient implementation, 4 represents good implementation, 5 reflects very good implementation, and 6 exemplifies ideal implementation. The latter implies that the journal requires a DAS and data sharing (when applicable) and provides detailed guidance on data citation for datasets generated, used, or analysed in a study.

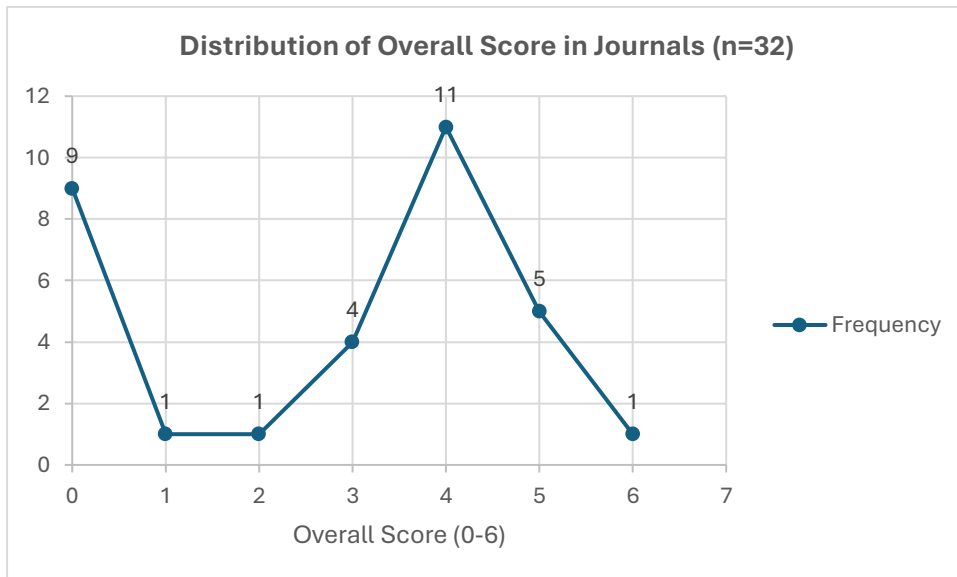


Figure 6. 9 Distribution of Overall Score in Journals

As illustrated in the accompanying line graph, 11 journals (34.4%) exhibit either no or very limited implementation of practices that enhance the visibility of research data. In contrast, only one journal (3.1%) demonstrates exemplary practices in data findability and discoverability. However, it is encouraging to note that 62.5% of the journals show sufficient to very good implementation of these practices, indicating a positive trend toward improving the discoverability of research data in qualitative studies.

## 7 Summary and Conclusion

This study aimed to explore the extent to which findability and discoverability practices are implemented in the qualitative social sciences, focusing on discipline-specific repositories, scholarly journal articles, and scholarly journals. Through an explorative and descriptive approach, the research examined three distinct sample groups, identifying patterns, trends, and gaps within each group to provide a comprehensive understanding of current data practices in this field.

### 7.1 Key Findings

#### Sample Group 1: Discipline-Specific Repositories

The analysis of discipline-specific repositories revealed critical insights into the management and accessibility of qualitative data. The majority of repositories utilize persistent identifiers (PIDs) such as DOIs and employ discipline-specific metadata standards like DDI. However, the availability of advanced search functionalities is limited, which potentially hinders the discoverability of even well-documented qualitative datasets. This limitation underscores the need for enhanced search tools within repositories to support researchers in locating relevant data more efficiently.

#### Sample Group 2: Scholarly Journal Articles

In contrast, the analysis of scholarly journal articles highlighted a significant deficiency in data availability practices. Only 17% of the articles included a data availability statement (DAS), and the actual availability of underlying data was often restricted. Ethical considerations were cited as barriers to data sharing, which is particularly concerning in qualitative research. This finding emphasizes the challenges in promoting open data practices within qualitative research, especially when dealing with sensitive human data, represented by the use case.

#### Sample Group 3: Scholarly Journals

The investigation of scholarly journals revealed that while many journals encourage or require the inclusion of DAS and data sharing, the enforcement of these practices remains inconsistent. Only a small proportion of journals fully implement detailed guidelines for data citation, which is crucial for enhancing the discoverability of research data. This inconsistency suggests that while there is an increasing recognition of the importance of data availability and citation, significant gaps remain in the actual implementation of these practices.

#### Comparison Across Sample Groups

When comparing the findings across the three sample groups, several overarching themes emerge. First, there is a clear disconnect between the repositories' potential to store and provide access to qualitative data and the practices within scholarly literature and journals that would make these datasets more discoverable. While repositories often follow robust metadata and PID practices, these efforts are not mirrored in the scholarly articles derived

from or linked to these datasets. Furthermore, the lack of a strong culture of data sharing and citation within qualitative research, particularly in published articles, further exacerbates this disconnect, suggesting that the role of repositories in enhancing data visibility is underutilized.

## 7.2 Unexpected Findings

A particularly unexpected finding was the significant gap between datasets stored in repositories and their representation in scholarly literature. The comparison revealed that although repositories often house a wealth of qualitative datasets relevant to the research use case—as demonstrated by the dataset search detailed in Appendix B, Worksheet B.3, which identified 10 uniquely relevant datasets—these datasets are frequently underrepresented or entirely absent in the corresponding scholarly literature. Notably, of the eight publications associated with the identified datasets, only one journal article provided a direct link to the available dataset, and another publication, a book chapter, correctly cited the dataset in its references. The remaining publications, which consisted of various forms of "grey literature," such as research reports, theses, and additional book chapters, did not offer any evidence, links, or citations for the datasets (see Appendix B, Worksheet B.3).

This discrepancy highlights a critical barrier to the findability of qualitative research data. Despite the availability of relevant datasets within repositories, their discoverability is significantly hindered when they are not adequately cited or linked within peer-reviewed journal articles. This disconnect suggests that the potential of repositories to enhance data visibility remains underutilized, underscoring the necessity for more robust efforts to integrate repository-held datasets into the academic discourse.

## 7.3 Implications and Recommendations

The findings of this study carry several important implications for the future of data management and sharing practices in qualitative social sciences:

1. **Enhanced Search Functionalities in Repositories:** The limited availability of advanced search functionalities in repositories suggests a need for improved tools that allow researchers to locate qualitative data more efficiently. Repositories should invest in developing and implementing more sophisticated search capabilities, including faceted search options that can filter by data type, collection methods, and other relevant criteria.
2. **Advocacy for Data Availability Statements:** The low prevalence of DAS in scholarly journal articles indicates a need for greater advocacy and implementation of open data practices. Scholarly journals, particularly those in the qualitative research domain, should play a more active role in enforcing data availability and citation guidelines. This could involve making DAS a mandatory component of the publication process and providing clear instructions on how to cite data.

3. **Strengthening Linkages Between Repositories and Scholarly Publications:** The unexpected finding regarding the disconnect between datasets in repositories and their representation in scholarly literature underscores the need for better integration of these resources. Repositories and journals should collaborate to develop clearer guidelines for data citation in academic writing and enhance the metadata standards used by repositories to ensure that datasets are more easily integrated into the scholarly ecosystem.
4. **Cultural Shift Towards Open Data in Qualitative Research:** Given the unique challenges associated with sharing qualitative data, particularly sensitive data, there is a need for a cultural shift towards more openness in qualitative research. This could be facilitated by developing best practices for sharing qualitative data in a way that respects ethical considerations while promoting greater transparency and data reuse.

## **7.4 Conclusion**

In conclusion, this study provides valuable insights into the current practices of data findability and discoverability in the qualitative social sciences. While significant progress has been made in the use of PIDs, metadata standards, and controlled vocabularies, challenges remain in ensuring that qualitative data are both findable and accessible to researchers. The unexpected finding of the gap between datasets in repositories and their representation in scholarly literature highlights an area in need of further research and development. Addressing these challenges will require ongoing collaboration between researchers, repositories, and journals to promote best practices in data management and sharing. Only through such collaborative efforts can the full potential of qualitative research data be realized, ensuring that these valuable resources are both accessible and utilized to their fullest extent.

## References

- Adey, P., Bissell, D., Hannam, K., Merriman, P., & Sheller, M. (Eds.). (2017). *The Routledge handbook of mobilities* (First issued in paperback). Routledge.
- ADP. (2024). *ADP - Social Science Data Archives*. ADP – Social Science Data Archives. Alalyze Data! Deposit Study! Promote Science! <https://www.adp.fdv.uni-lj.si/eng/>
- Allen, B. L. (1991). Cognitive Research in Information Science: Implications for Design. *Annual Review of Information Science and Technology*, 26, 3–37.
- Antes, A. L., Walsh, H. A., Strait, M., Hudson-Vitale, C. R., & DuBois, J. M. (2017). Examining Data Repository Guidelines for Qualitative Data Sharing. *Journal of Empirical Research on Human Research Ethics*, 13(1), 61–73.
- Antes, A. L., Walsh, H. A., Strait, M., Hudson-Vitale, C. R., & DuBois, J. M. (2018). Examining Data Repository Guidelines for Qualitative Data Sharing. *Journal of Empirical Research on Human Research Ethics*, 13(1), 61–73. <https://doi.org/10.1177/1556264617744121>
- Archiwum Danych Jakościowych. (2024). *Welcome*. Archiwum Danych Jakościowych. Qualitative Data Archive (QDA). <https://adj.ifispan.pl/index.php/en/welcome/>
- ARK. (2017). *ARK Qualitative Archives*. NIQA - Northern Ireland Qualitative Archive. <https://www.maynoothuniversity.ie/iqda>
- Armbruster, C., & Romary, L. (2009). Comparing Repository Types: Challenges and Barriers for Subject-Based Repositories, Research Repositories, National Repository Systems and Institutional Repositories in Serving Scholarly Communication. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1506905>
- 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>

- Australian National University. (2024a). *Depositing Data*. ADA – The Australian Data Archive. <https://ada.edu.au/depositing-data/>
- Australian National University. (2024b). *The Australian Data Archive*. ADA – The Australian Data Archive. <https://ada.edu.au/>
- Baker, K. S., & Duerr, R. E. (2017). Data and a Diversity of Repositories. In *Curating research data* (Vol. 2, pp. 139–144). Association of College and Research Libraries, a division of the American Library Association.
- Barbour, R. S. (1999). The Case for Combining Qualitative and Quantitative Approaches in Health Services Research. *Journal of Health Services Research & Policy*, 4(1), 39–43. <https://doi.org/10.1177/135581969900400110>
- Bensmann, F., Papenmeier, A., Kern, D., Zapilko, B., & Dietze, S. (2020). Semantic Annotation, Representation and Linking of Survey Data. In *Semantic Systems. In the Era of Knowledge Graphs: 16th International Conference on Semantic Systems, SEMANTiCS 2020, Amsterdam, The Netherlands, September 7–10, 2020, Proceedings* (Vol. 12378, pp. 53–69). Springer International Publishing. <https://doi.org/10.1007/978-3-030-59833-4>
- beQuali. (2024). *Qualitative social science survey bank*. beQuali – Qualitative Social Science Survey Bank. <https://bequali.fr/en/>
- Berg-Cross, G., Ritz, R., & Wittenburg, P. (2015). *RDA Data Foundation and Terminology—DFT: Results RFC*.
- Berkowitz, H., & Delacour, H. (2022). Opening Research Data: What Does It Mean for Social Sciences? *M@n@gement*, 1–15. <https://doi.org/10.37725/mgmt.v25.9123>
- Betancort Cabrera, N., & Mozygemba, K. (2022). *QualidataNet—A single point of entry to qualitative research data*. <https://doi.org/10.5281/ZENODO.7406055>
- Betancort Cabrera, N., & Mozygemba, K. (2023). *Qualitative Data and DDI: Chances to move forward*. <https://doi.org/10.5281/ZENODO.10225400>
- Bijsterveld, K. (2016). Ethnography and Archival Research in Studying Cultures of Sound. In J. G. Papenburg & H. Schulze (Eds.), *Sound as Popular Culture* (pp. 99–110). The MIT Press. <https://doi.org/10.7551/mitpress/9975.003.0015>

- Bilder, G., Lin, J., & Neylon, C. (2020). *The Principles of Open Scholarly Infrastructure*.  
<https://doi.org/10.24343/C34W2H>
- Binder, P., & Filipkowski, P. (2011). Data Sharing and Archiving Qualitative and QL Data in Poland. *IASSIST Quarterly*, 34(3–4), 70. <https://doi.org/10.29173/iq467>
- Bishop, L. (2009a). Ethical sharing and reuse of qualitative data. *Australian Journal of Social Issues*, 44(3), 255–272.
- Bishop, L. (2009b). Ethical Sharing and Reuse of Qualitative Data. *Australian Journal of Social Issues*, 44(3), 255–272. <https://doi.org/10.1002/j.1839-4655.2009.tb00145.x>
- Bishop, L. (2021). Secondary Analysis of Qualitative Data. In *Qualitative research* (pp. 429–446). SAGE.
- Bishop, L., & Kuula-Luumi, A. (2017). Revisiting Qualitative Data Reuse: A Decade On. *SAGE Open*, 7(1), 215824401668513. <https://doi.org/10.1177/2158244016685136>
- Bishop, L., & Neale, B. (2011). Sharing Qualitative and Qualitative Longitudinal Data in the UK. *IASSIST Quarterly*, 34(3–4), 23. <https://doi.org/10.29173/iq457>
- Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. The MIT Press.
- Borgman, C. L., & Bourne, P. E. (2022). Why It Takes a Village to Manage and Share Data. *Harvard Data Science Review*. <https://doi.org/10.1162/99608f92.42eec111>
- Bowen, G. A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, 9(2), 27–40. <https://doi.org/10.3316/QRJ0902027>
- Branney, P. E., Brooks, J., Kilby, L., Newman, K., Norris, E., Pownall, M., Talbot, C. V., Treharne, G. J., & Whitaker, C. M. (2023). Three steps to open science for qualitative research in psychology. *Social and Personality Psychology Compass*, 17(4), e12728. <https://doi.org/10.1111/spc3.12728>
- Buckland, M. (1999). Vocabulary as a Central Concept in Library and Information Science. *Digital Libraries: Interdisciplinary Concepts, Challenges, and Opportunities. Proceedings of the Third International Conference on Conceptions of Library and Information Science*, 3–12. <https://people.ischool.berkeley.edu/~buckland/colisvoc.htm>

- Bull, M. (2013). Sound Mix: The Framing of Multi-Sensory Connections in Urban Culture. *SoundEffects - An Interdisciplinary Journal of Sound and Sound Experience*, 3(3), 25–45. <https://doi.org/10.7146/se.v3i3.15730>
- Bull, M. (2018). Sounds as Data. In *The SAGE Handbook of Qualitative Data Collection* (pp. 426–438). SAGE Publications Ltd.
- Burgess, M., & Noy, N. (2018, September 26). *Building Google Dataset Search and Fostering an Open Data Ecosystem*. Google Research. <http://research.google/blog/building-google-dataset-search-and-fostering-an-open-data-ecosystem/>
- Burton, A., Aryani, A., Koers, H., Manghi, P., La Bruzzo, S., Stocker, M., Diepenbroek, M., Schindler, U., & Fenner, M. (2017). The Scholix Framework for Interoperability in Data-Literature Information Exchange. *D-Lib Magazine*, 23(1/2). <https://doi.org/10.1045/january2017-burton>
- Buscatto, M. (2018). Doing Ethnography: Ways and Reasons. In *The SAGE Handbook of Qualitative Data Collection* (pp. 327–342). SAGE Publications Ltd.
- Butviliene, J., & Butvilas, T. (2011). Exploring Qualitative Longitudinal Research and Qualitative Resources: The Lithuanian Case. *IASSIST Quarterly*, 34(3–4), 66. <https://doi.org/10.29173/iq466>
- CESSDA. (2022). *CESSDA Topic Classification (Version 4.2.2) [Controlled vocabulary]*. *Urn:ddi:int.cessda.cv:TopicClassification:4.2.2*. <https://vocabularies.cessda.eu/urn/urn:ddi:int.cessda.cv:TopicClassification:4.2.2>
- CESSDA And Service Providers. (2023). *The European Language Social Science Thesaurus (ELSST) (Version 4) [Dataset]*. Zenodo. <https://doi.org/10.5281/ZENODO.8364268>
- Chapman, A., Simperl, E., Koesten, L., Konstantinidis, G., Ibáñez, L.-D., Kacprzak, E., & Groth, P. (2020). Dataset search: A survey. *The VLDB Journal*, 29(1), 251–272. <https://doi.org/10.1007/s00778-019-00564-x>
- Chauvette, A., Schick-Makaroff, K., & Molzahn, A. E. (2019). Open Data in Qualitative Research. *International Journal of Qualitative Methods*, 18, 160940691882386. <https://doi.org/10.1177/1609406918823863>

- Chavan, V., & Penev, L. (2011). The data paper: A mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics*, *12*(S15), S2. <https://doi.org/10.1186/1471-2105-12-S15-S2>
- Cizek, T. (2011). Archiving Qualitative and Quantitative Longitudinal Social Sciences Data In the Czech Republic. *IASSIST Quarterly*, *34*(3–4), 36. <https://doi.org/10.29173/iq459>
- Collier, J., & Collier, M. (1990). *Visual anthropology: Photography as a research method* (Rev. and expanded ed). University of New Mexico Press.
- Consortium of European Social Science Data Archives. (n.d.-a). *National Service Providers*. CESSDA – Consortium of European Social Science Data Archives. Retrieved 21 July 2024, from <https://www.CESSDA.eu/About/Consortium-and-Partners/List-of-Service-Providers>
- Consortium of European Social Science Data Archives. (n.d.-b). *Providing metadata*. CESSDA Data Catalogue User Guide. Retrieved 28 July 2024, from <https://www.CESSDA.eu/CESSDA-Documents/Annual-reports/CESSDA-Annual-Report-2022.pdf>
- Consortium of European Social Science Data Archives. (2023a). *CESSDA Annual Report 2022*. CESSDA – Consortium of European Social Science Data Archives. <https://www.CESSDA.eu/CESSDA-Documents/Annual-reports/CESSDA-Annual-Report-2022.pdf>
- Consortium of European Social Science Data Archives. (2023b, November 21). *CESSDA Vocabulary Service version 3.3.1*. CESSDA Vocabulary Service. <https://vocabularies.CESSDA.eu/>
- Consortium of European Social Science Data Archives. (2024a). *CESSDA Data Catalogue*. CESSDA – Consortium of European Social Science Data Archives. <https://datacatalogue.CESSDA.eu/>
- Consortium of European Social Science Data Archives. (2024b). *CESSDA ERIC in a Nutshell*. CESSDA – Consortium of European Social Science Data Archives. <https://www.CESSDA.eu/About>
- Contaxis, N., Clark, J., Dellureficio, A., Gonzales, S., Mannheimer, S., Oxley, P. R., Ratajeski, M. A., Surkis, A., Yarnell, A. M., Yee, M., & Holmes, K. (2022). Ten simple rules for improving research data discovery. *PLOS Computational Biology*, *18*(2), e1009768. <https://doi.org/10.1371/journal.pcbi.1009768>

- Corti, L. (2005). Qualitative Archiving and Data Sharing: Extending the reach and impact of qualitative data. *IASSIST Quarterly*, 29(3). <https://iassistquarterly.com/public/pdfs/iqvol293corti.pdf>
- Corti, L. (2011). The European Landscape of Qualitative Social Research Archives: Methodological and Practical Issues. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, Vol 12, No 3 (2011): Qualitative Archives and Biographical Research Methods. <https://doi.org/10.17169/FQS-12.3.1746>
- Corti, L. (2012). Recent developments in archiving social research. *International Journal of Social Research Methodology*, 15(4), 281–290. <https://doi.org/10.1080/13645579.2012.688310>
- Corti, L. (2018). Data Collection in Secondary Analysis. In U. Flick, *The SAGE Handbook of Qualitative Data Collection* (pp. 164–181). SAGE Publications Ltd. <https://doi.org/10.4135/9781526416070.n11>
- Corti, L. (2020). Archiving Qualitative Data. In *SAGE Research Methods Foundations*. SAGE Publications Ltd. <https://doi.org/10.4135/9781526421036813114>
- Corti, L., Day, A., & Backhouse, G. (2000). Confidentiality and Informed Consent: Issues for Consideration in the Preservation of and Provision of Access to Qualitative Data Archives. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, Vol 1, No 3 (2000): Text . Archive . ReAnalysis. <https://doi.org/10.17169/FQS-1.3.1024>
- Corti, L., & Thompson, P. (1996). *ESRC Qualitative Data Archival Resource Centre (QUALIDATA)*. 1(3). <https://www.socresonline.org.uk/1/3/qualidata.html#top>
- Corti, L., & Thompson, P. (2004). Secondary analysis of archive data. In *Qualitative research practice* (pp. 327–343). SAGE Publications Ltd.
- Cousijn, H., Feeney, P., Lowenberg, D., Presani, E., & Simons, N. (2019). Bringing Citations and Usage Metrics Together to Make Data Count. *Data Science Journal*, 18, 9. <https://doi.org/10.5334/dsj-2019-009>
- Crosas, M., Gautier, J., Karcher, S., Kirilova, D., Otalora, G., & Schwartz, A. (2018). *Data policies of highly-ranked social science journals* [Preprint]. SocArXiv. <https://doi.org/10.31235/osf.io/9h7ay>
- Crossref stats*. (2024). Crossref. <https://www.crossref.org/06members/53status.html>

CSDA. (2024). *Our data*. CSDA – Czech Social Science Data Archive. <https://archiv.soc.cas.cz/en/pristup-k-datum-2/nase-data>

DANS. (2024). *Data Station Social Sciences and Humanities*. DANS - the Dutch National Centre of Expertise and Repository for Research Data. <https://dans.knaw.nl/en/social-sciences-and-humanities/>

Data Citation Synthesis Group. (2014). *Joint Declaration of Data Citation Principles*. Force11. <https://doi.org/10.25490/A97F-EGYK>

Data Policies. (2024). In *Forschungsdaten.org*. [https://www.forschungsdaten.org/index.php/Data\\_Policies](https://www.forschungsdaten.org/index.php/Data_Policies)

DataCite. (n.d.). *About*. DataCite Commons. Retrieved 28 July 2024, from <https://commons.datacite.org/about>

DataCite. (2023, December). *DataCite OAI-PMH Guide*. DataCite Support. <https://support.datacite.org/docs/datacite-oai-pmh>

DataCite. (2024a). *DataCite Statistics*. DataCite Statistics. <https://stats.datacite.org>

DataCite. (2024b, March). *Introduction to the DataCite REST API*. DataCite Support. <https://support.datacite.org/docs/api>

DataCite – International Data Citation Initiative. (2024). *DataCite*. DataCite. <https://datacite.org/>

DataCite Metadata Working Group. (2024). *DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs v4.5*. <https://doi.org/10.14454/G8E5-6293>

DCMI Usage Board. (2012, June 14). *Dublin Core™ Metadata Element Set, Version 1.1: Reference Description*. DublinCore. <https://www.dublincore.org/specifications/dublin-core/dces/>

DCMI Usage Board. (2020, January 20). *DCMI Metadata Terms*. DublinCore. <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

Deutsches Institut für Wirtschaftsforschung. (2024). *Forschungsdatenzentrum Betriebs- und Organisationsdaten (FDZ-BO)*. DIW Berlin. [https://www.diw.de/de/diw\\_01.c.670982.de/seiten/forschungsdatenzentrum\\_betriebs\\_und\\_organisationsdaten\\_fdz\\_bo.html](https://www.diw.de/de/diw_01.c.670982.de/seiten/forschungsdatenzentrum_betriebs_und_organisationsdaten_fdz_bo.html)

- Deutsches Klimarechenzentrum GmbH. (2024). *B2FIND*. B2FIND EUDAT Collaborative Data Infrastructure. <https://b2find.eudat.eu/>
- Dierkes, J. (2021). Planung, Beschreibung und Dokumentation von Forschungsdaten. In *Praxishandbuch Forschungsdatenmanagement* (pp. 303–326). De Gruyter.
- DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation. (2022). *Forschungsdatenzentrum (FDZ) Bildung*. FDZ Bildung. <https://www.fdz-bildung.de/home>
- DOI Foundation. (2023, April). *DOI Handbook*. [https://www.doi.org/doi-handbook/DOI\\_Handbook\\_Final.pdf](https://www.doi.org/doi-handbook/DOI_Handbook_Final.pdf)
- Dryad. (2024). *Submission requirements*. Dryad. <https://datadryad.org/stash/requirements>
- DuBois, J. M., Mozersky, J., Parsons, M., Walsh, H. A., Friedrich, A., & Pienta, A. (2023). Exchanging words: Engaging the challenges of sharing qualitative research data. *Proceedings of the National Academy of Sciences*, *120*(43), e2206981120. <https://doi.org/10.1073/pnas.2206981120>
- Eberle, T. S. (2018). Collecting Images as Data. In *The SAGE Handbook of Qualitative Data Collection* (pp. 392–411). SAGE Publications Ltd.
- Edwards, P. N. (2010). *A Vast Machine*. The MIT Press; JSTOR. <http://www.jstor.org/stable/j.ctt5hhds1>
- Egan, M., MacLean, A., Sweeting, H., & Hunt, K. (2012). Comparing the effectiveness of using generic and specific search terms in electronic databases to identify health outcomes for a systematic review: A prospective comparative study of literature search methods. *BMJ Open*, *2*(3), e001043. <https://doi.org/10.1136/bmjopen-2012-001043>
- Elsevier. (2022). *Data Monitor Content and Data Policy*. Elsevier. [https://assets.ctfassets.net/o78em1y1w4i4/1onRJtxpieRzfVCgieapJo/c040109498026fe0f42f574db31f4825/Data\\_Monitor\\_Content\\_Policy.pdf](https://assets.ctfassets.net/o78em1y1w4i4/1onRJtxpieRzfVCgieapJo/c040109498026fe0f42f574db31f4825/Data_Monitor_Content_Policy.pdf)
- Elsevier. (2024a). *Database linking*. Elsevier. <https://www.elsevier.com/researcher/author/tools-and-resources/research-data/data-base-linking>
- Elsevier. (2024b). *Mendeley Data*. Mendeley Data. <https://data.mendeley.com/>

- European Commission. (2017). *Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020*.  
[https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)
- Everyday Childhoods. (2024). *Accessing the 'Everyday Childhoods' Archive*. Everyday Childhoods.  
<https://blogs.sussex.ac.uk/everydaychildhoods/the-archive/>
- FAIRsharing. (2009). *FAIRsharing*. FAIRsharing.Org. Standards, Databases, Policies.  
<https://fairsharing.org/>
- FDZ eLabor. (2017). *Forschungsdatenzentrum für die Archivierung und Bereitstellung qualitativer Forschungsdaten*. eLabour. <http://elabour.de/>
- FDZ-DZHW. (n.d.). *Research Data Centre for Higher Education Research and Science Studies (FDZ-DZHW)*. Deutsches Zentrum Für Hochschul- Und Wissenschaftsforschung (DZHW). Retrieved 21 July 2024, from <https://fdz.dzhw.eu/en>
- Fenner, M. (2020). *DataCite Commons—Exploiting the Power of PIDs and the PID Graph*.  
<https://doi.org/10.5438/F4DF-4817>
- Fenner, M., & Aryani, A. (2019). *Introducing the PID Graph*. <https://doi.org/10.5438/JWVF-8A66>
- Fenner, M., Crosas, M., Grethe, J. S., Kennedy, D., Hermjakob, H., Rocca-Serra, P., Durand, G., Berjon, R., Karcher, S., Martone, M., & Clark, T. (2019). A data citation roadmap for scholarly data repositories. *Scientific Data*, 6(1), 28. <https://doi.org/10.1038/s41597-019-0031-8>
- Fidel, R. (1992). Who needs controlled vocabulary? *Special Libraries*, 83(1), 1–9.
- Finnish Social Science Data Archive (FSD). (2024). *Finnish Social Science Data Archive*. Finnish Social Science Data Archive (FSD). <https://www.fsd.tuni.fi/en/>
- Flick, U. (2018a). *The SAGE Handbook of Qualitative Data Collection*. SAGE Publications Ltd.  
<https://doi.org/10.4135/9781526416070>
- Flick, U. (2018b). Types of Data and How to Collect Them. In *The SAGE Handbook of Qualitative Data Collection* (pp. 231–232). SAGE Publications Ltd.
- FORS. (2024a). *2026-09-04—FORS - CoreTrustSeal Requirements 2020-2022* [Dataset]. DataverseNL. <https://doi.org/10.34894/IJ3CNP>

- FORS. (2024b). *FORS - Swiss Centre of Expertise in the Social Sciences*. FORS Explore.Understand.Share. <https://forscenter.ch/>
- Foster, A., & Ford, N. (2003). Serendipity and information seeking: An empirical study. *Journal of Documentation*, 59(3), 321–340. <https://doi.org/10.1108/00220410310472518>
- Foster, M. J., & Jewell, S. T. (2022). *Piecing Together Systematic Reviews and Other Evidence Syntheses*. Rowman & Littlefield.
- Fowler, S. A., & Jewell, S. T. (2022). Database Searching. In *Piecing Together Systematic Reviews and Other Evidence Syntheses* (pp. 93–110). Rowman & Littlefield.
- Frants, V. I., & Brush, C. B. (1989). The Need for Information and some Aspects of Information Retrieval Systems Construction. *Journal of the American Society for Information Science*, 39, 86–91.
- Frants, V. I., Shapiro, J., & Voiskunskii, V. G. (1997). *Automated information retrieval: Theory and methods*. Academic Press.
- Friedrich, T., & Recker, J. (2021). Auffindbarkeit und Nutzbarkeit von Daten. In *Praxishandbuch Forschungsdatenmanagement* (pp. 405–426). De Gruyter.
- Gardos, J., & Ivacs, G. (2011). Qualitative Longitudinal Research and Qualitative Resources: The Hungarian Case. *IASSIST Quarterly*, 34(3–4), 64. <https://doi.org/10.29173/iq465>
- GESIS. (n.d.-a). *GESIS Controlled Vocabulary Service*. GESIS – Leibniz Institute for the Social Sciences. Retrieved 23 August 2024, from <https://lod.gesis.org/thesoz/en/>
- GESIS. (n.d.-b). *Welche Informationen finde ich in der Kategorie „Forschungsdaten“?* GESIS – Leibniz Institute for the Social Sciences. Retrieved 21 July 2024, from [https://search.gesis.org/faq#explain\\_research\\_data](https://search.gesis.org/faq#explain_research_data)
- GESIS. (2024). *Finding and accessing data*. GESIS – Leibniz Institute for the Social Sciences. <https://www.gesis.org/en/services/finding-and-accessing-data>
- Giglia, E. (2019). OPERAS: Bringing the long tail of Social Sciences and Humanities into Open Science. *JLIS*, 1. <https://doi.org/10.4403/jlis.it-12523>

- Google. (n.d.). *Dataset (Dataset, DataCatalog, DataDownload) structured data*. Google Search Central. Retrieved 28 July 2024, from <https://developers.google.com/search/docs/appearance/structured-data/dataset>
- Google. (2024). *Dataset Search*. Google Dataset Search. <https://datasetsearch.research.google.com/>
- Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies: A typology of reviews, *Maria J. Grant & Andrew Booth. Health Information & Libraries Journal*, 26(2), 91–108. <https://doi.org/10.1111/j.1471-1842.2009.00848.x>
- Gray, E. J., Broeder, D., Kleemola, M., Thiel, C., Chambers, S., & Fischer, F. (2022). *The European Open Science Cloud (EOSC) and Its Implications for the Digital Humanities and Social Sciences*. <https://doi.org/10.5281/ZENODO.6913361>
- Gray, J., & O’Carroll, A. (2011). Qualitative Research In Ireland: Archiving Strategies and Development. *IASSIST Quarterly*, 34(3–4), 18. <https://doi.org/10.29173/iq456>
- Gregory, K. (2021). *Findable and reusable?: Data discovery practices in research* [maastricht university]. <https://doi.org/10.26481/dis.20210302kg>
- Gregory, K., Cousijn, H., Groth, P., Scharnhorst, A., & Wyatt, S. (2019). Understanding data search as a socio-technical practice. *Journal of Information Science*, 46(4), 459–475. <https://doi.org/10.1177/0165551519837182>
- Gregory, K., Groth, P., Scharnhorst, A., & Wyatt, S. (2020). Lost or Found? Discovering Data Needed for Research. *Harvard Data Science Review*. <https://doi.org/10.1162/99608f92.e38165eb>
- Haase, J. E., & Myers, S. T. (1988). Reconciling Paradigm Assumptions of Qualitative and Quantitative Research. *Western Journal of Nursing Research*, 10(2), 128–137. <https://doi.org/10.1177/019394598801000202>
- Hammersley, M. (1997). QUALITATIVE DATA ARCHIVING: SOME REFLECTIONS ON ITS PROSPECTS AND PROBLEMS. *Sociology*, 31(1), 131–142. JSTOR.
- Hammersley, M. (2010). Can We Re-Use Qualitative Data via Secondary Analysis? Notes on Some Terminological and Substantive Issues. *Sociological Research Online*, 15(1), 47–53. <https://doi.org/10.5153/sro.2076>
- Hammersley, M., & Atkinson, P. (2007). *Ethnography: Principles in practice* (3rd ed). Routledge.

- Hendricks, G. (2021, February 26). *About us*. Crossref. <https://www.crossref.org/about/>
- Henrich, A., Jegan, R., & Gradl, T. (2021). Data Retrieval. In *Praxishandbuch Forschungsdatenmanagement* (pp. 427–449).
- Henriksen, G.-H., Orvik, M. B., & Pedersen, T. (2011). Mapping Out Qualitative Data Resources in Norway. *IASSIST Quarterly*, 34(3–4), 46. <https://doi.org/10.29173/iq462>
- Hilse, H.-W., & Kothe, J. (2006). *Implementing persistent identifiers: Overview of concepts, guidelines and recommendations*. Consortium of European Research Libraries.
- Huma, B., & Joyce, J. B. (2023). ‘One size doesn’t fit all’: Lessons from interaction analysis on tailoring Open Science practices to qualitative research. *British Journal of Social Psychology*, 62(4), 1590–1604. <https://doi.org/10.1111/bjso.12568>
- ICPSR. (n.d.). *Thesaurus*. ICPSR. Retrieved 22 June 2024, from <https://www.icpsr.umich.edu/web/ICPSR/thesaurus#:~:text=The%20ICPSR%20Thesaurus%20is%20a,controlled%20name%20forms%20and%20spellings.>
- ICPSR. (2024a). *Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)*. ICPSR Inter-University Consortium for Political and Social Research. <https://www.icpsr.umich.edu/web/pages/membership/or/metadata/oai.html>
- ICPSR. (2024b, August 27). *Search Results*. ICPSR Inter-University Consortium for Political and Social Research. [https://www.icpsr.umich.edu/web/ICPSR/search/studies?start=0&ARCHIVE=ICPSR&PUBLISHED\\_STATUS=PUBLISHED&DATAKIND\\_FACET=qualitative&sort=score%20desc%20CTITLE\\_SORT%20asc&rows=50&q=](https://www.icpsr.umich.edu/web/ICPSR/search/studies?start=0&ARCHIVE=ICPSR&PUBLISHED_STATUS=PUBLISHED&DATAKIND_FACET=qualitative&sort=score%20desc%20CTITLE_SORT%20asc&rows=50&q=)
- Inter-university Consortium for Political and Social Research. (2020). *Guide to Social Science Data Preparation and Archiving. Best Practice Throughout the Data Life Cycle: 6th Edition*. ICPSR. <https://www.icpsr.umich.edu/files/deposit/dataprep.pdf>
- Jaaskelainen, T., Moschner, M., & Wackerow, J. (2010). Controlled Vocabularies for DDI 3: Enhancing Machine-Actionability. *IASSIST Quarterly*, 33(1), 34. <https://doi.org/10.29173/iq649>
- Jackson, B. (2021). Open Data Policies among Library and Information Science Journals. *Publications*, 9(2), 25. <https://doi.org/10.3390/publications9020025>

- Jackson, C. (2018). Data Collection in Conversation Analysis. In *The SAGE Handbook of Qualitative Data Collection* (pp. 280–296). SAGE Publications Ltd.
- Jahn, Najko, Laakso, Mikael, Lazzeri, Emma, & McQuilton, Peter. (2023). *Study on the readiness of research data and literature repositories to facilitate compliance with the Open Science Horizon Europe MGA requirements* (Version Version 1.0). Zenodo. <https://doi.org/10.5281/ZENODO.7728016>
- Jisc. (2005). *Welcome to OpenDOAR - Sherpa Services*. OpenDOAR. <https://v2.sherpa.ac.uk/pendoar/>
- Johnston, L. R. (2017). *Curating research data*. Association of College and Research Libraries, a division of the American Library Association.
- Jones, J. S., & Goldring, J. (2021). *Exploratory and Descriptive Statistics*. SAGE Publications Ltd. <https://doi.org/10.4135/9781529682786>
- Jones, K., & Alexander, S. M. (2018). *Qualitative data sharing and re-use for socio-environmental systems research: A synthesis of opportunities, challenges, resources and approaches*. <https://doi.org/10.13016/M2WH2DG59>
- Joyce, J. B., Douglass, T., Benwell, B., Rhys, C. S., Parry, R., Simmons, R., & Kerrison, A. (2022). Should we share qualitative data? Epistemological and practical insights from conversation analysis. *International Journal of Social Research Methodology*, 1–15. <https://doi.org/10.1080/13645579.2022.2087851>
- Juty, N., Wimalaratne, S. M., Soiland-Reyes, S., Kunze, J., Goble, C. A., & Clark, T. (2020). Unique, Persistent, Resolvable: Identifiers as the Foundation of FAIR. *Data Intelligence*, 2(1–2), 30–39. [https://doi.org/10.1162/dint\\_a\\_00025](https://doi.org/10.1162/dint_a_00025)
- Karcher, S., & Weber, N. (2019). Annotation for transparent inquiry: Transparent data and analysis for qualitative research. *IASSIST Quarterly*, 43(2), 1–9. <https://doi.org/10.29173/iq959>
- Karhulahti, V. (2023). Reasons for qualitative psychologists to share human data. *British Journal of Social Psychology*, 62(4), 1621–1634. <https://doi.org/10.1111/bjso.12573>
- Karhulahti, V.-M. (2022). Registered reports for qualitative research. *Nature Human Behaviour*, 6(1), 4–5. <https://doi.org/10.1038/s41562-021-01265-8>

- Kern, D., & Momeni, F. (2022). Linked Open Research Data for Social Science Pilot Study (LORDpilot\_KE). *GESIS – Leibniz Institute for the Social Sciences*.  
<https://www.gesis.org/en/research/external-funding-projects/details/project/156/linked-open-research-data-for-social-science-pilot-study>
- Kernecker, T. (2023, April 3). *Data Deposit Guideline for qualitative data v1.0*. Vienna: The Austrian Social Science Data Archive.  
[https://aussda.at/fileadmin/user\\_upload/p\\_aussda/Documents/Data-Deposit-Guideline\\_Qualitative\\_V1.pdf](https://aussda.at/fileadmin/user_upload/p_aussda/Documents/Data-Deposit-Guideline_Qualitative_V1.pdf)
- Khan, N., Pink, C. J., & Thelwall, M. (2020). Identifying Data Sharing and Reuse with Scholix: Potentials and Limitations. *Patterns*, *1*(1), 100007.  
<https://doi.org/10.1016/j.patter.2020.100007>
- Khan, N., Thelwall, M., & Kousha, K. (2023). Data sharing and reuse practices: Disciplinary differences and improvements needed. *Online Information Review*, *47*(6), 1036–1064.  
<https://doi.org/10.1108/OIR-08-2021-0423>
- Kjeldgaard, A. S. F. (2011). Archiving and Disseminating Qualitative Data in Denmark. *IASSIST Quarterly*, *34*(3–4), 38. <https://doi.org/10.29173/iq460>
- Kleiner, B., Heinzmann, C., Eberle, T. S., & Bergman, M. M. (2011). Qualitative Data Archiving in Switzerland. *IASSIST Quarterly*, *34*(3–4), 76. <https://doi.org/10.29173/iq203>
- Knoblauch, H. (2005). Focused Ethnography. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, Vol 6, No 3 (2005): The State of the Art of Qualitative Research in Europe.  
<https://doi.org/10.17169/FQS-6.3.20>
- Knoblauch, H., Schnettler, B., & Tuma, R. (2018). Videography. In *The SAGE Handbook of Qualitative Data Collection* (pp. 362–377). SAGE Publications Ltd.
- Koninklijke Brill. (2024). *Research Data Journal for the Humanities and Social Sciences*. Brill.  
<https://brill.com/view/journals/rdj/rdj-overview.xml?language=en>
- Krämer, T., Papenmeier, A., Carevic, Z., Kern, D., & Mathiak, B. (2021). Data-Seeking Behaviour in the Social Sciences. *International Journal on Digital Libraries*, *22*(2), 175–195.  
<https://doi.org/10.1007/s00799-021-00303-0>

- Kratz, J. E., & Strasser, C. (2015a). Making data count. *Scientific Data*, 2(1), 150039. <https://doi.org/10.1038/sdata.2015.39>
- Kratz, J. E., & Strasser, C. (2015b). Researcher Perspectives on Publication and Peer Review of Data. *PLOS ONE*, 10(2), e0117619. <https://doi.org/10.1371/journal.pone.0117619>
- Kusenbach, D. (2018). Go-Alongs. In *The SAGE Handbook of Qualitative Data Collection* (pp. 344–360). SAGE Publications Ltd.
- Kuula, A. (2011). Methodological and Ethical Dilemmas of Archiving Qualitative Data. *IASSIST Quarterly*, 34(3–4), 12. <https://doi.org/10.29173/iq455>
- Lagoze, C., & Van de Sompel, H. (2015, January 8). *The Open Archives Initiative Protocol for Metadata Harvesting. Protocol Version 2.0*. Open Archives Initiative. <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>
- Lee, D. J., & Stvilia, B. (2017). Practices of research data curation in institutional repositories: A qualitative view from repository staff. *PLOS ONE*, 12(3), e0173987. <https://doi.org/10.1371/journal.pone.0173987>
- Lee, J., & Kim, J. (2021). Korean researchers' motivations for publishing in data journals and the usefulness of their data: A qualitative study. *Science Editing*, 8(2), 145–152. <https://doi.org/10.6087/kcse.246>
- Library of Congress. (2024). *Library of Congress Recommended Formats Statement 2024-2025*. Library of Congress. <https://www.loc.gov/preservation/resources/rfs/index.html>
- Lin, S., Ali, I., & Wilson, G. (2020). Ten quick tips for making things findable. *PLOS Computational Biology*, 16(12), e1008469. <https://doi.org/10.1371/journal.pcbi.1008469>
- Makri, S., Bhuiya, J., Carthy, J., & Owusu-Bonsu, J. (2015). Observing serendipity in digital information environments. *Proceedings of the Association for Information Science and Technology*, 52(1), 1–10. <https://doi.org/10.1002/pr2.2015.145052010019>
- Makri, S., Chen, Y.-C., McKay, D., Buchanan, G., & Ocepek, M. (2019). Discovering the Unfindable: The Tension Between Findability and Discoverability in a Bookshop Designed for Serendipity. In D. Lamas, F. Loizides, L. Nacke, H. Petrie, M. Winckler, & P. Zaphiris (Eds.), *Human-*

- Computer Interaction – INTERACT 2019* (Vol. 11747, pp. 3–23). Springer International Publishing. [https://doi.org/10.1007/978-3-030-29384-0\\_1](https://doi.org/10.1007/978-3-030-29384-0_1)
- Mannheimer, S. (2024). *Scaling Up: How Data Curation Can Help Address Key Issues in Qualitative Data Reuse and Big Social Research*. Springer International Publishing. <https://doi.org/10.1007/978-3-031-49222-8>
- Mannheimer, S., Pienta, A., Kirilova, D., Elman, C., & Wutich, A. (2019). Qualitative Data Sharing: Data Repositories and Academic Libraries as Key Partners in Addressing Challenges. *American Behavioral Scientist*, 63(5), 643–664. <https://doi.org/10.1177/0002764218784991>
- Mannheimer, S., Serman, L., & Borda, S. (2016). Discovery and Reuse of Open Datasets: An Exploratory Study. *Journal of eScience Librarianship*, 5(1), e1091. <https://doi.org/10.7191/jeslib.2016.1091>
- Mauthner, N. S., Parry, O., & Backett-Milburn, K. (1998). The Data are Out there, or are They? Implications for Archiving and Revisiting Qualitative Data. *Sociology*, 32(4), 733–745. <https://doi.org/10.1177/0038038598032004006>
- McCrae, J. P. (2024). *The Linked Open Data Cloud*. The Linked Open Data Cloud. <https://lod-cloud.net/>
- McGowan, J., Sampson, M., Salzwedel, D. M., Cogo, E., Foerster, V., & Lefebvre, C. (2016). PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *Journal of Clinical Epidemiology*, 75, 40–46. <https://doi.org/10.1016/j.jclinepi.2016.01.021>
- McKenzie, P. J. (2003). A model of information practices in accounts of everyday-life information seeking. *Journal of Documentation*, 59(1), 19–40. <https://doi.org/10.1108/00220410310457993>
- Medjedovic, I., & Witzel, A. (2011). Sharing and Archiving Qualitative and Qualitative Longitudinal Research Data in Germany. *IASSIST Quarterly*, 34(3–4), 42. <https://doi.org/10.29173/iq461>
- Meistring, M. (2023). Überlegungen zu einem Bewertungssystem für Forschungsdatenpublikationen unter Einbezug der FAIR-Prinzipien: Am Beispiel der Geowissenschaften. *Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin*, 502.
- Merton, R. K. (1968). *Social theory and social structure* (1968 Edition). Free Press.

- Mikos, L. (2018). Collecting Media Data: TV and Film Studies. In *The SAGE Handbook of Qualitative Data Collection* (pp. 412–425). SAGE Publications Ltd.
- Miller, B. (2021). *Bausteine Forschungsdatenmanagement: 2021, 2KonsortSWD: Vom Netzwerk zur integrierten Dateninfrastruktur der Gesellschaftsforschung* [Application/pdf].  
<https://doi.org/10.17192/BFDM.2021.2.8330>
- Morgan, D. L. (1996). Focus Groups. *Annual Review of Sociology*, 22(1), 129–152.  
<https://doi.org/10.1146/annurev.soc.22.1.129>
- Morgan, D. L., & Hoffman, K. (2018). Focus Groups. In *The SAGE Handbook of Qualitative Data Collection* (pp. 250–263). SAGE Publications Ltd.
- Mozersky, J., Walsh, H., Parsons, M., McIntosh, T., Baldwin, K., & DuBois, J. M. (2020). Are we ready to share qualitative research data? Knowledge and preparedness among qualitative researchers, IRB members, and data repository curators. *IASSIST Quarterly*, 43(4), 1–23.  
<https://doi.org/10.29173/iq952>
- Mulrow, C. D. (1987). The Medical Review Article: State of the Science. *Annals of Internal Medicine*, 106(3), 485. <https://doi.org/10.7326/0003-4819-106-3-485>
- Mulrow, C. D. (1994). Systematic Reviews: Rationale for systematic reviews. *BMJ*, 309(6954), 597–599. <https://doi.org/10.1136/bmj.309.6954.597>
- Murray, M. (2018). Narrative Data. In *The SAGE Handbook of Qualitative Data Collection* (pp. 264–279). SAGE Publications Ltd.
- National Institutes of Health, O. of the D. (2020, October 29). *Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research*. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-016.html>
- National Institutes of Health, O. of the D. (2022, September 21). *Supplemental Information to the NIH Policy for Data Management and Sharing: Protecting Privacy When Sharing Human Research Participant Data*. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-22-213.html>
- National University of Ireland Maynooth. (2024). *Irish Qualitative Data Archive*. National University of Ireland Maynooth. <https://www.maynoothuniversity.ie/iqda>

- Neale, B., & Bishop, L. (2011). Qualitative and Quantitative Longitudinal Resources in Europe. *IASSIST Quarterly*, 34(3–4), 6. <https://doi.org/10.29173/iq189>
- NOAA Fisheries. (n.d.). *Voices. Oral history archives*. NOAA Voices. Retrieved 21 July 2024, from <https://voices.nmfs.noaa.gov/>
- Noy, N., & Benjelloun, O. (2023, February 28). *Datasets at your fingertips in Google Search*. Google Research. <http://research.google/blog/datasets-at-your-fingertips-in-google-search/>
- OpenAIRE. (2022a). *Use of OAI-PMH*. OpenAIRE Guidelines. [https://guidelines.openaire.eu/en/latest/data/use\\_of\\_oai\\_pmh.html](https://guidelines.openaire.eu/en/latest/data/use_of_oai_pmh.html)
- OpenAIRE. (2022b, August 8). *Guides for Researchers. How to comply with Horizon Europe mandate for Research Data Management*. OpenAIRE. <https://www.openaire.eu/how-to-comply-with-horizon-europe-mandate-for-rdm>
- OpenAIRE. (2024a). OpenAIRE. <https://www.openaire.eu/>
- OpenAIRE. (2024b). *OpenAIRE EXPLORE*. OpenAIRE EXPLORE. <https://explore.openaire.eu/>
- OpenAIRE. (2024c). *Scholarly Literature & Data interlinking*. ScholeXplorer. The Daa Literature Interlinking Service. <https://scholexplorer.openaire.eu>
- OpenAlex. (2024a). OpenAlex. <https://openalex.org/>
- OpenAlex. (2024b, July 21). *Current stats*. OpenAlex. <https://openalex.org/stats>
- Padgett, D. (2017). *Qualitative methods in social work research* (Third edition). SAGE.
- Parry, O., & Mauthner, N. (2005). Back to Basics: Who Re-uses Qualitative Data and Why? *Sociology*, 39(2), 337–342. JSTOR. <https://doi.org/10.1177/0038038505050543>
- Parry, O., & Mauthner, N. S. (2004). Whose Data are They Anyway?: Practical, Legal and Ethical Issues in Archiving Qualitative Research Data. *Sociology*, 38(1), 139–152. <https://doi.org/10.1177/0038038504039366>
- Pawlowski, C. S., Andersen, H. B., Troelsen, J., & Schipperijn, J. (2016). Children’s Physical Activity Behavior during School Recess: A Pilot Study Using GPS, Accelerometer, Participant Observation, and Go-Along Interview. *PLOS ONE*, 11(2), e0148786. <https://doi.org/10.1371/journal.pone.0148786>

- Penev, L., Mietchen, D., Chavan, V., Hagedorn, G., Smith, V., Shotton, D., Ó Tuama, É., Senderov, V., Georgiev, T., Stoev, P., Groom, Q., Remsen, D., & Edmunds, S. (2017). Strategies and guidelines for scholarly publishing of biodiversity data. *Research Ideas and Outcomes*, 3, e12431. <https://doi.org/10.3897/rio.3.e12431>
- Pink, S. (2012). *Advances in Visual Methodology*. SAGE Publications Ltd. <https://doi.org/10.4135/9781446250921>
- Pomerantz, J. (2015). *Metadata*. The MIT Press.
- Pownall, M., Talbot, C. V., Kilby, L., & Branney, P. (2023). Opportunities, challenges and tensions: Open science through a lens of qualitative social psychology. *British Journal of Social Psychology*, 62(4), 1581–1589. <https://doi.org/10.1111/bjso.12628>
- Prosser, A. M. B., Bagnall, R., & Higson-Sweeney, N. (2024). Reflection over compliance: Critiquing mandatory data sharing policies for qualitative research. *Journal of Health Psychology*, 13591053231225903. <https://doi.org/10.1177/13591053231225903>
- Prosser, A. M. B., Hamshaw, R. J. T., Meyer, J., Bagnall, R., Blackwood, L., Huysamen, M., Jordan, A., Vasileiou, K., & Walter, Z. (2021, October 6). *A Content Analysis of Qualitative Open Data Guidelines in Social Psychology Journals*. [osf.io/zhrpn](https://osf.io/zhrpn)
- Prosser, A. M. B., Hamshaw, R. J. T., Meyer, J., Bagnall, R., Blackwood, L., Huysamen, M., Jordan, A., Vasileiou, K., & Walter, Z. (2023). When open data closes the door: A critical examination of the past, present and the potential future for open data guidelines in journals. *British Journal of Social Psychology*, 62(4), 1635–1653. <https://doi.org/10.1111/bjso.12576>
- QualiService. (n.d.). *Profile*. QualiService Data Sharing. Retrieved 21 July 2024, from <https://bequali.fr/en/>
- Qualitative Data Repository. (2014, January 13). *Types of Qualitative Data* [Text]. QDR; Qualitative Data Repository. <https://qdr.syr.edu/node/20648>
- Qualitative Data Repository. (2023a). *QDR Curation Policy*. QDR. The Qualitative Data Repository. <https://qdr.syr.edu/policies/curation>
- Qualitative Data Repository. (2023b). *The Qualitative Data Repository*. QDR. The Qualitative Data Repository. <https://qdr.syr.edu/>

- Raffaghelli, J. E., & Manca, S. (2023). Exploring the social activity of open research data on ResearchGate: Implications for the data literacy of researchers. *Online Information Review*, 47(1), 197–217. <https://doi.org/10.1108/OIR-05-2021-0255>
- Rapley, T., & Rees, G. (2018). Collecting Documents as Data. In *The SAGE Handbook of Qualitative Data Collection* (pp. 378–390). SAGE Publications Ltd.
- Rasmussen, K. B. (2011). Editor's Notes. *IASSIST Quarterly*, 34(3–4), 5. <https://doi.org/10.29173/iq188>
- Rau, A., Elliker, F., & Coetzee, J. K. (2018). Collecting Data for Analyzing Discourses. In *The SAGE Handbook of Qualitative Data Collection* (pp. 298–312). SAGE Publications Ltd.
- RDF Working Group. (2014). *RDF - Semantic Web Standards*. W3.Org. <https://www.w3.org/RDF/>
- re3data.org. (2013). *re3data.org—Registry of Research Data Repositories*. <https://www.re3data.org/>
- re3data.org. (2023). *re3data.org: Forschungsdatenzentrum Bildung*. re3data.org - Registry of Research Data Repositories. <https://doi.org/10.17616/R3R010>
- Research Data Journal for the Humanities and Social Sciences. (2023, September 28). *Instructions for authors*. Brill. [https://brill.com/fileasset/downloads\\_products/Author\\_Instructions/RDJ.pdf](https://brill.com/fileasset/downloads_products/Author_Instructions/RDJ.pdf)
- Rieger, O. Y., & Schonfeld, R. (2023). *Common Scholarly Communication Infrastructure Landscape Review*. Ithaka S+R. <https://doi.org/10.18665/sr.318775>
- Roulston, K., & Choi, M. (2018). Qualitative Interviews. In *The SAGE Handbook of Qualitative Data Collection* (pp. 233–249). SAGE Publications Ltd.
- Schema.org. (2024, July 1). *Schema.org*. Schema.Org. <https://schema.org>
- Schulte, J. (2024). *OPERAS. Open Scholarly Communication in the European Research Area for Social Sciences and Humanities*. OPERAS AISBL. <https://operas-eu.org/>
- Seo, S., & Kim, J. (2020). Data journals: Types of peer review, review criteria, and editorial committee members' positions. *Science Editing*, 7(2), 130–135. <https://doi.org/10.6087/kcse.207>
- Shigapov, R. (2024). *ChatGPT for FAIR Research Data*. <https://doi.org/10.5281/ZENODO.10664553>
- Sikt. (2024). *Sikt. Norwegian Agency for Shared Services in Education and Research*. Sikt – Norwegian Agency for Shared Services in Education and Research. <https://sikt.no/en/home>

- Smioski, A. (2011). Establishing a Qualitative Data Archive in Austria. *IASSIST Quarterly*, 34(3–4), 30. <https://doi.org/10.29173/iq458>
- Social Sciences and Humanities Open Cloud. (2024). *About European Research Infrastructures*. SSHOC – Social Sciences and Humanities Open Cloud. <https://sshopencloud.eu/about-european-research-infrastructures>
- Sources. (2024, January). OpenAlex Technical Documentation. <https://docs.openalex.org/api-entities/sources>
- Stals, S., Smyth, M., & Ijsselsteijn, W. (2014). Walking & talking: Probing the urban lived experience. *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*, 737–746. <https://doi.org/10.1145/2639189.2641215>
- Stathis, K., Ross, C., Dreyer, B., & Vierkant, P. (2022). *DataCite Metadata Schema 4.4 to Schema.org Mapping* (Version 1.0). Zenodo. <https://doi.org/10.5281/ZENODO.7661399>
- Stebe, J., Hudales, J., & Kragelj, B. (2011). Archiving and Re-using Qualitative and Qualitative Longitudinal Data in Slovenia. *IASSIST Quarterly*, 34(3–4), 50. <https://doi.org/10.29173/iq463>
- Stratford, E. (2019). Oral History and Narrative. In D. Richardson, N. Castree, M. F. Goodchild, A. Kobayashi, W. Liu, & R. A. Marston (Eds.), *International Encyclopedia of Geography* (1st ed., pp. 1–7). Wiley. <https://doi.org/10.1002/9781118786352.wbieg1038.pub2>
- Sutcliffe, A., & Ennis, M. (1998). Towards a cognitive theory of information retrieval. *Interacting with Computers*, 10(3), 321–351. [https://doi.org/10.1016/S0953-5438\(98\)00013-7](https://doi.org/10.1016/S0953-5438(98)00013-7)
- Svenonius, E. (2001). *The intellectual foundation of information organization* (Third printing). MIT Press.
- Swiss National Science Foundation, E. C. of the F. C. (2016, January 1). *Funding Regulations. Regulations of the Swiss National Science Foundation on research grants*. [https://www.snf.ch/media/en/ICCrvpOHZ38Hbg5Y/allg\\_reglement\\_16\\_e.pdf](https://www.snf.ch/media/en/ICCrvpOHZ38Hbg5Y/allg_reglement_16_e.pdf)
- Technische Universität Berlin. (2024). *aviDa. Forschungsdatenzentrum für audio-visuelle Daten der empirischen qualitativen Sozialforschung*. Technische Universität Berlin. <https://depositonce.tu-berlin.de/handle/11303/9911>

TEI Consortium. (n.d.). *Text Encoding Initiative*. TEI. Text Encoding Initiative. Retrieved 28 July 2024, from <https://tei-c.org/>

Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., & Dorsett, K. (2015). Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLOS ONE*, *10*(8), e0134826. <https://doi.org/10.1371/journal.pone.0134826>

The British Psychological Society. (2024). *FAQ*. The British Psychological Society. <https://bpspsychub.onlinelibrary.wiley.com/hub/datasharingfaqs>

*The FAIR Data Principles – FORCE11*. (2014, September 3). <https://force11.org/info/the-fair-data-principles/>

The Murray Research Archive. (n.d.). *Henry A. Murray Research Archive*. Henry A. Murray Research Archive. Retrieved 21 July 2024, from <https://murray.harvard.edu/dataverse>

The President and Fellows of Harvard College. (2024). *Harvard Dataverse*. Harvard Dataverse. <https://dataverse.harvard.edu/>

Thomson, D., Bzdel, L., Golden-Biddle, K., Reay, T., & Estabrooks, C. A. (2005). Central Questions of Anonymization: A Case Study of Secondary Use of Qualitative Data [33 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *6*(1), Art. 29.

Thomson, R., & Berriman, L. (2023). Starting with the archive: Principles for prospective collaborative research. *Qualitative Research*, *23*(2), 234–251. <https://doi.org/10.1177/14687941211023037>

Titarenko, L., & Tereschenko, O. (2011). Qualitative and Qualitative Longitudinal Research and Resources in Belarus. *IASSIST Quarterly*, *34*(3–4), 60. <https://doi.org/10.29173/iq464>

UK Data Archive. (n.d.). *Controlled vocabularies*. UK Data Archive. Retrieved 23 August 2024, from <https://www.data-archive.ac.uk/managing-data/standards-and-procedures/controlled-vocabularies/>

UK Data Archive. (2024). *Quality control*. UK Data Archive. <https://www.data-archive.ac.uk/managing-data/digital-curation-and-data-publishing/quality-control/>

UK Data Service. (n.d.-a). *Data documentation: Qualitative data*. UK Data Service. Retrieved 23 August 2024, from <https://ukdataservice.ac.uk/learning-hub/research-data-management/document-your-data/data-level/data-documentation-qualitative-data/>

- UK Data Service. (n.d.-b). *European qualitative data archives*. UK Data Service. Retrieved 21 July 2024, from <https://ukdataservice.ac.uk/help/other-data-providers/other-data-providers-qualitative-data/european-archives/>
- UK Data Service. (n.d.-c). *International qualitative archives*. UK Data Service. Retrieved 21 July 2024, from <https://ukdataservice.ac.uk/help/other-data-providers/other-data-providers-qualitative-data/international-qualitative-archives/>
- UK Data Service. (2024a). *Metadata standards*. UK Data Service. <https://www.data-archive.ac.uk/managing-data/standards-and-procedures/metadata-standards/>
- UK Data Service. (2024b). *QualiBank*. UK Data Service. <https://discover.ukdataservice.ac.uk/qualibank>
- UK Data Service. (2024c). *UK Data Service*. UK Data Service. <https://ukdataservice.ac.uk/>
- UK Research and Innovation. (2023, November 27). *Publishing your research findings*. UK Research and Innovation. <https://www.ukri.org/manage-your-award/publishing-your-research-findings/making-your-research-data-open/>
- Universität Bremen. (2024a). *Qualidata Network*. QualidataNet – Qualidata Network. <https://www.qualidatanet.com/de/>
- Universität Bremen. (2024b). *Über uns*. QualidataNet – Qualidata Network. <https://www.qualidatanet.com/de/ueber-uns.html>
- Ursin, G., & Bentzen, H. B. (2021). Open science and sharing personal data widely – legally impossible for Europeans? *Acta Oncologica*, 60(12), 1555–1556. <https://doi.org/10.1080/0284186X.2021.1995894>
- Van Den Eynden, V., & Chatsiou, K. (2011). *Data management for qualitative data using NVivo9*. UK Data Archive. <https://ukdataservice.ac.uk//app/uploads/ukda-datamanagement-nvivo.pdf>
- Vandevusse, A., Mueller, J., & Karcher, S. (2022). Qualitative Data Sharing: Participant Understanding, Motivation, and Consent. *Qualitative Health Research*, 32(1), 182–191. <https://doi.org/10.1177/10497323211054058>
- Vardigan, M. (2013). *A Qualitative Data Model for DDI*. <https://doi.org/10.3886/DDIWORKINGPAPER05>

- Vitale, C. R. H., Carlson, J., Hodge, A. E., Hswe, P., Johns, E., Johnston, L. R., Kozlowski, W., Nurnberger, A., Petters, J., Rolando, E., Shorish, Y., Schneider, J., & Zilinski, L. (2017). Making the Case for Institutional Data Repositories. In *Curating research data* (Vol. 2, pp. 158–161). Association of College and Research Libraries, a division of the American Library Association.
- Voices of the 20th Century Archive, C. for S. S. (n.d.). *About us*. Voices of the 20th Century Archive and Research Group. Retrieved 28 July 2024, from <https://20szazadhangja.tk.hun-ren.hu/en/about-us>
- Wall of Adopters | Data Documentation Initiative*. (n.d.). Retrieved 13 July 2024, from <https://ddialliance.org/ddi-adopters>
- Walters, W. H. (2020). Data journals: Incentivizing data access and documentation within the scholarly communication system. *Insights the UKSG Journal*, 33, 18. <https://doi.org/10.1629/uksg.510>
- Wästerfors, D. (2018). Observations. In *The SAGE Handbook of Qualitative Data Collection* (pp. 314–325). SAGE Publications Ltd.
- Weller, S. (2023). Fostering habits of care: Reframing qualitative data sharing policies and practices. *Qualitative Research*, 23(4), 1022–1041. <https://doi.org/10.1177/14687941211061054>
- Why Use DDI? | Data Documentation Initiative*. (n.d.). Retrieved 13 July 2024, from <https://ddialliance.org/learn/why-use-ddi>
- 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), 160018. <https://doi.org/10.1038/sdata.2016.18>
- Winter, R. (Ed.). (2014). *Die Zukunft der Cultural Studies: Theorie, Kultur und Gesellschaft im 21. Jahrhundert* (H. Thies, Trans.). transcript.
- Wissenschaftszentrum Berlin für Sozialforschung. (2024). *KonsortSWD*. KonsortSWD. <https://www.konsortswd.de/>

- Wu, M., Psomopoulos, F., Khalsa, S. J., & De Waard, A. (2019). Data Discovery Paradigms: User Requirements and Recommendations for Data Repositories. *Data Science Journal*, 18, 3. <https://doi.org/10.5334/dsj-2019-003>
- Zapilko, B., & Sure, Y. (2009). Converting the TheSoz to SKOS. *GESIS-Technical Reports*, 2009(7). [https://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/gesis\\_methodenberichte/2009/TechnicalReport\\_09\\_07.pdf](https://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2009/TechnicalReport_09_07.pdf)
- Zenodo. (2024). *General Policies v1.0*. Zenodo. <https://about.zenodo.org/policies/#:~:text=All%20users%20are%20allowed%20to,the%20parties%20prior%20to%20submission>
- Zondergeld, J. J., Scholten, R. H. H., Vreede, B. M. I., Hessels, R. S., Pijl, A. G., Buizer-Voskamp, J. E., Rasch, M., Lange, O. A., & Veldkamp, C. L. S. (2020). FAIR, safe and high-quality data: The data infrastructure and accessibility of the YOUth cohort study. *Developmental Cognitive Neuroscience*, 45, 100834. <https://doi.org/10.1016/j.dcn.2020.100834>

## Digital Appendix

**Overview:** This digital appendix provides detailed documentation of the methodologies, data collection processes, and analyses performed as part of the study on data findability and discoverability in qualitative social sciences. The appendix is divided into multiple sections, each focusing on different aspects of the research. The content includes technical scripts, data retrieval processes, and summaries of the variables and findings across different datasets.

### **Appendix A: Repository and Metadata Retrieval Using re3data.org REST API**

**Content:** This section includes detailed Python scripts used to retrieve repositories and their metadata via re3data.org REST API. It includes steps for repository retrieval, processing of repository list according to the selection process detailed in Appendix B, Worksheet B.2, and steps for metadata extraction.

**File included:**

APPENDIX A.pdf

### **Appendix B: Data Search in Discipline-Specific Repositories**

**Content:** Appendix B documents the process of identifying and collecting data from discipline-specific repositories in the social sciences that host qualitative research data. The goal of this process is to gather the necessary data for further analysis, particularly to evaluate how effectively these repositories implement findability and discoverability practices. Metadata properties of the identified repositories were retrieved via the re3data REST API, with additional information collected manually due to the API's limitations. Furthermore, these repositories were explored for datasets relevant to the use case using generic search terms such as "refugee" and "disability," supplemented by AND/OR Boolean operators, synonyms and thematic/topical browsing, depending on the size of each repository's holdings and search capabilities. Datasets identified during this process are documented for related publication and link back to the datasets.

**Content description:**

- Worksheet B.1: Overview and Content
- Worksheet B.2: Repository Selection
- Worksheet B.3: Dataset Search Protocol
- Worksheet B.4: Manually Gathered Repository Information
- Worksheet B.5: Repository Metadata Retrieved via re3data REST API
- Worksheet B.6: Cleaned and Merged Data

- Worksheet B.7: References

**File included:**

APPENDIX B.xlsx

## **Appendix C: Tracing Data in Scholarly Journal Articles**

**Content:** Appendix C documents the systematic evidence searching process used to identify linked datasets and data availability statements. The file's structure follows the steps outlined in Section 5.2.2, "Tracing Data in Scholarly Literature," from Chapter 5, "Methods", covering the entire process from the preliminary search to the identification of associated research data.

**Content description:**

- Worksheet C.1: Overview and Content
- Worksheet C.2: Preliminary Search
- Worksheet C.3: Database Selection
- Worksheet C.4: Term Harvesting
- Worksheet C.5: Search Strategy Development
- Worksheet C.6: Deduplication and Item Selection
- Worksheet C.7: Final Data Collection - Information on Journals, Data Availability Statements and Journals' Data Policies
- Worksheet C.8: References

**File included:**

APPENDIX C.xlsx

## **Appendix D: Summary of Variables and Analysis**

**Content:** Appendix D provides a comprehensive overview of the final datasets and analyses used in this study. It consolidates the data collected through the exploratory approaches detailed in the Methods chapter and presents the analysis results for the variables defined in the study.

**Content description:**

- Worksheet D.1: Overview and Content
- Worksheet D.2: Variable Summary
- Worksheet D.3: Frequency Tables for Sample Group 1 (Discipline-Specific Repositories)
- Worksheet D.4: Frequency Tables for Sample Group 2 (Scholarly Journal Articles)
- Worksheet D.5: Frequency Tables for Sample Group 3 (Scholarly Journals)

**File included:**

APPENDIX D.xlsx