Secondary Publication



Lachner, Sonja; Henrich, Andreas

What Makes Research Data Visible? : Factors and Implications for Humanities Research Data Infrastructures

Date of secondary publication: 26.10.2026

Version of Record (Published Version), Conferenceobject

Persistent identifier: urn:nbn:de:bvb:473-irb-117763x

Primary publication

Lachner, Sonja; Henrich, Andreas (2026): What Makes Research Data Visible?: Factors and Implications for Humanities Research Data Infrastructures, in: JCDL '26: Proceedings of the 26th ACM/IEEE Joint Conference on Digital Libraries, New York, NY; United States: ACM, pp. 1–13, doi: 10.1145/3805696.3846022

Legal Notice

This work is protected by copyright and/or the indication of a licence. You are free to use this work in any way permitted by the copyright and/or the licence that applies to your usage. For other uses, you must obtain permission from the rights-holders.

This document is made available under a Creative Commons license.



The license information is available online:
https://creativecommons.org/licenses/by/4.0/legalcode

What Makes Research Data Visible? Factors and Implications for Humanities Research Data Infrastructures

Sonja Lachner*
Media Informatics Group
University of Bamberg
Bamberg, Germany
sonja.lachner@uni-bamberg.de

Andreas Henrich
Media Informatics Group
University of Bamberg
Bamberg, Germany
andreas.henrich@uni-bamberg.de

Abstract

Research data infrastructures are designed, among other objectives, to improve the discoverability and accessibility of research data. However, this is often accompanied by an implicit assumption that data become perceptible to scholars simply by existing and being made accessible. Building on the premise that visibility is not a stable property of data itself, we argue that the visibility of research data is instead shaped by several conditions that determine whether and how data become perceptible to scholars.

Using the humanities as an illustrative domain, we focus on the modes of access to humanities research data, understood as the ways through which researchers become aware of research data and argue that each access mode is associated with a distinct *Dimension of Visibility*, shaped by different factors that influence how and whether research data become visible within these Modes of Access. To capture these dynamics, we introduce the Humanities Research Data Visibility Model (HRDVM), a literature-driven model that identifies four central Modes of Access: Social Interaction, Information Systems, Physical Access and Scholarly Literature. For each Mode of Access, the model introduces a corresponding Dimension of Visibility: Social Visibility, Digital Visibility, Physical Visibility, and Literature Visibility. Building on this model, we derive implications for the design of humanities research data infrastructures. These include, among others, supporting multiple modes of access to research data, improving metadata to better represent reuse-relevant information, and accounting for the social and institutional practices through which researchers discover and access data.

Although we focus on the humanities, the proposed framework and the implications can be adapted to other research domains and non-academic contexts, as well as to other information resources.

CCS Concepts

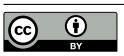
• **Applied computing** → *Arts and humanities*; • **General and reference** → *Surveys and overviews*; • **Information systems** → *Digital libraries and archives*; *Expert search.*

Keywords

Humanities Research Data, Research Data Visibility, Research Data Infrastructures, Information Behavior

ACM Reference Format:
Sonja Lachner and Andreas Henrich. 2026. What Makes Research Data Visible? Factors and Implications for Humanities Research Data Infrastructures. In *The 2026 ACM/IEEE Joint Conference on Digital Libraries (JCDL '26), October 13–16, 2026, Frisco, TX, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3805696.3846022

*Corresponding author.



This work is licensed under a Creative Commons Attribution 4.0 International License.
JCDL '26, Frisco, TX, USA
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2597-5/2026/10
https://doi.org/10.1145/3805696.3846022

1 Introduction

The sharing of information, ideas, and research materials is a fundamental feature of academic collaboration and scientific discovery [51]. Among these resources, research data play a central role, enabling scholars to explore questions, derive insights from evidence, and advance human understanding [66]. In response, funding agencies and publishers have introduced open research data policies that encourage researchers to make their data accessible with few or no restrictions. This development has been accompanied by growing efforts to establish research data infrastructures, with the aim, in addition to other objectives, to support data discovery and reuse. Prominent examples, among many others, include European-wide initiatives such as Gaia-X [12], DARIAH-EU [7], CLARIN [2], ELIXIR [6], and CESSDA [8], as well as national infrastructures such as ARDC [9], DANS [20], Recherche Data Gouv [11], and the NFDI [10].

Despite these efforts, it is often implicitly assumed that once research data exist and are available, they are simply *there* and therefore readily *visible* to potential users. We argue that this assumption is misleading. In Web Science and Search Engine Optimization (SEO), visibility is often narrowly understood as "property of being visible and locatable on the web" [107], or operationalized as a metric indicating how prominently a resource appears in search results [62]. However, search engine optimization goes well beyond purely technical aspects and also addresses issues such as general brand awareness, content quality, and backlinks.

To sharpen our use of the term *visibility*, we distinguish visibility from three closely related concepts: findability, discoverability and accessibility. In particular, the distinction between visibility, findability, and discoverability remains often conceptually unclear, as noted by Reyes-Lillo et al. [107]. We therefore delineate the boundaries between these concepts to clarify what we mean by visibility in this work. According to Morville [100], findability denotes "the quality of being locatable or navigable" and "the degree to which a particular object is easy to discover or locate". It also refers to "the degree to which a system or environment supports navigation

and retrieval" [100]. Thus, findability refers to how easily users can locate information or resources within a website or information system [107]. This aspect is also reflected in the *F* of the FAIR Principles [130]. In contrast, discoverability refers to the extent to which a resource can be encountered serendipitously, without being explicitly sought by the user [44, 92]. In this sense, it focuses on the ability of a website or system to facilitate the discovery of otherwise unknown content [107]. Accessibility, by comparison, concerns whether a resource, once located, can actually be retrieved and used. In this sense, it is related to the *A* in the FAIR Principles [130].

These concepts describe distinct and more specific conditions through which a resource may become available or perceptible to a user. We understand visibility as a broader concept that encompasses and cuts across these different dimensions, while also extending beyond them. In this sense, visibility is not reducible to any single condition, but encompasses a wider range of socio-technical factors that shape whether and how research data become perceptible, discoverable, and relevant to researchers in practice. To illustrate this, we can look at metrics that could demonstrate high visibility as we define it. These would include, for example, access statistics, figures on reuse, or the level of awareness of specific research data within a professional community—which could be determined, for instance, through surveys.

In this paper, we therefore aim to capture a broader range of visibility aspects through which research data may become perceptible to scholars. Drawing on Neumayer et al. [102], who conceptualize (in)visibility as a socio-technical process rather than a stable attribute of data, we understand *research data visibility* as emerging from the interplay of multiple technical, social, and contextual factors. Visibility is thus not solely determined by the establishment of technical foundations, but also by several other factors that determine whether and how researchers encounter research data in practice.

Of course, the "FAIR Guiding Principles for Scientific Data Management and Stewardship" [130] represent a very important and necessary step toward the visibility of research data. The FAIR principles have become a widely adopted framework for designing (research) data and metadata such that they are findable, accessible, interoperable, and reusable. FAIR establishes essential technical preconditions for making data findable, but it does not guarantee that data become visible to researchers in practice. As such, it constitutes a necessary, but not sufficient, condition for visibility. This becomes particularly evident in web-based information environments, where visibility is not only a function of technical accessibility, but is actively shaped. Approaches such as SEO and, more recently, Generative Engine Optimization (GEO) [26] aim to increase the reach of digital resources by aligning their representation with the ranking and generation mechanisms of search engines and generative systems. These practices illustrate that being indexed or technically retrievable does not automatically translate into actual visibility. Instead, whether resources are encountered depends on how they are surfaced and prioritized within these systems. As a result, research data that fulfill FAIR criteria may remain effectively invisible if they are not aligned with these mechanisms. Taken together, this also suggests that visibility cannot be reduced to technical properties alone, but must be understood as the outcome of socio-technical processes operating across different modes of access.

In this paper, we focus on the humanities and their research data as an illustrative case, characterized by heterogeneous data types, distributed storage practices, and diverse scholarly workflows. In particular, we focus on central access modes to humanities research data, understood as the ways through which researchers become aware of research data and argue that each access mode involves a Dimension of Visibility that is shaped by different factors influencing how and whether research data become visible within these modes of access. To capture these dynamics, we introduce the *Humanities Research Data Visibility Model* (HRDVM), a literature-driven model that brings together multiple strands of literature. The model integrates a typology of Humanities Data Types, Data-Holding Entities, and Modes of Access with their corresponding Visibility Dimensions.

Current research data infrastructures predominantly support access through digital search and retrieval, focusing on optimizing system-side components such as indexing and ranking. While these are essential, they reflect a narrow understanding of access that overlooks other pathways through which research data become visible. We argue that research data infrastructures should not be limited to retrieval-based access, but should support multiple modes of access that reflect the diverse ways in which researchers encounter and recognize data. This view resonates with Burrows [39], who argues that humanities infrastructures often remain search-oriented systems with sophisticated retrieval functions, while overlooking the research processes needed to exploit these sources.

Against this background, we state the following research questions:

RQ1 In what ways is the visibility of humanities research data conceptually constituted through different modes of access, and which factors shape this process?

RQ2 What implications do these conceptualizations of visibility and access have for the design of humanities research data infrastructures?

To answer these research questions, we first conducted an exploratory literature review that focused on research dealing with the search for and use of research data, particularly in the humanities. We systematized the results of this review in the HRDVM. Finally, based on this model, we drew implications regarding humanities research data infrastructures.

While our analysis focuses on the humanities, the underlying concept, that each access mode is associated with a distinct *Dimension of Visibility*, can be transferred to other domains and non-academic contexts, such as industry or cultural institutions, where it can be applied to their respective access logics and practices, as well as to their information resources.

The contributions of this paper can be summarized as follows. We introduce the HRDVM, which brings Humanities Data Types, Data-Holding Entities, Modes of Access, and their Visibility Dimensions together. Based on that, we derive implications for humanities research data infrastructures based on the access modes of humanities scholars.

The remainder of the paper is organized as follows. After presenting related work on research data in the humanities in Section

2, we introduce the HRDVM in Section 3. Based on that, we then outline implications for a humanities research data infrastructure in Section 4. Section 5 concludes the paper and presents directions for further research.

2 Related Work

This section provides an overview of related work on humanities research data practices. It includes definitions of research data in the humanities, as well as studies on data seeking and sharing. The overview provides insights into how humanities scholars engage with research data, while also revealing a gap in the literature with respect to connecting access modes to the visibility of humanities research data.

2.1 Humanities and the Nature of Humanities Research Data

The humanities comprise a set of disciplines concerned with the study of human society, culture [125] and human record [36]. This scope is e.g. reflected in re3data's subject hierarchy [24] or the Subject Committee of the Deutsche Forschungsgemeinschaft classification of disciplines belonging to the humanities [13]. These classifications include fields such as ancient cultures, linguistics or history. Against this disciplinary background, conceptual definitions of humanities research data vary in their emphasis. DARIAH-DE defines humanities research data as sources, materials, and research results that are collected, written, described, or evaluated in relation to a research question within the human and cultural sciences, and that are available in machine-readable form for purposes such as archiving, citation and further scholarly processing [21]. The European Federation of Academies of Sciences and Humanities ALLEA Working Group E-Humanities has a similar definition. They define humanities research data as all materials and resources that scholars collect, produce, and utilize throughout every stage of the research cycle [61]. Schöch, by contrast, describes humanities data as constructed and machine-actionable representations that abstract and select aspects of humanities research objects [111]. A function-oriented perspective is offered by Borgman, who characterizes data as evidentiary entities of phenomena in academic research [36].

Together, these definitions illustrate the conceptual breadth of research data that can be considered in the humanities. Although these conceptual definitions explicitly employ the term *(research) data*, this terminology contrasts with scholarly practice. Humanities scholars rarely refer to their materials as *data* or *research data*. Instead, they commonly use more precise and discipline-specific terms such as primary and secondary sources, research materials, readings, or sources [49, 83, 88, 103].

This preference for discipline-specific terminology reflects the practical conditions of humanities research. Unlike fields in which data are primarily collected or created for scientific purposes, the social sciences and humanities often rely on data originally collected for non-research contexts, such as business and government records, social activities, genealogical sources, or literary work [126]. In addition, humanities research relies on unique, often physical materials, and almost any evidence of human activity may be seen as a research object [36]. Such artifacts function not only as evidence but also as sources of inspiration and reinterpretation, embedded in ongoing processes of meaning-making [88]. Moreover, humanities research data are often heterogeneous and unstructured, and their study is inherently iterative and cumulative, with humanities scholars repeatedly engaging the same data over decades or even entire academic careers [101].

This diversity of research data in the humanities constitutes an initial key element for modeling in the HRDVM, which we will explore in more detail in Section 3.2 and Figure 1.

2.2 Research Data Practices in the Humanities

Studies on research data practices across disciplines have examined how data are searched [54–56, 72] and shared [70, 119]—or have addressed both [69, 73, 120, 121]. However, to the best of our knowledge, existing studies in the humanities tend to focus on individual access points rather than on the relationship between access and visibility as well as its significance for research data infrastructures. Previous work has examined individual access sources in the humanities, such as data journals [91, 96, 97], repositories [53, 91], or linked data [91] without situating them within a broader framework of access modes. Furthermore, previous literature focuses on digital access and overlooks analog forms of access. As a result, there is no perspective that links modes of access to the visibility of humanities research data. Hence, we focus our consideration of related work on data seeking and sharing in the humanities, as these represent two fundamental ways in which access to research data is created and realized.

2.2.1 Data Seeking in the Humanities: Research on information seeking in the humanities has received substantial attention [27, 28, 33, 34, 38, 43, 46, 80, 117, 128, 129]. In contrast, studies on data seeking in the humanities are limited, and those that do exist focus on different aspects of this area. Li et al. analyze characteristics of data seeking behavior and data seeking approaches adopted by humanities scholars [81], whereas Kemman et al. show that humanities scholars in the Netherlands and Belgium primarily rely on general-purpose search engines and prefer keyword-based search over other retrieval strategies such as related terms or Boolean operators [68]. Lian et al., on the other hand, investigate dataset search behavior in cultural-heritage work [85, 86] and Hauf et al. examine the criteria humanities and social science scholars apply when selecting a data source for their research [63].

2.2.2 Data Sharing in the Humanities: Research on information sharing in the humanities has also received attention [43, 124]. Some studies have also examined data sharing practices in the humanities. For example, Klungthanaboon [71] analyzes digital humanities researchers in Thailand, identifying data types, collection practices, sharing motivations, and barriers. Similarly, Ruediger and MacDougall [108] explore sharing intentions, anticipated reuse, and perceived challenges in data-intensive humanities projects. Focusing specifically on image data, Late et al. [78] show that sharing remains largely informal and is often mediated through personal networks. Li et al. [82], in contrast, examine data-related collaboration in interdisciplinary digital humanities teams, identifying core data activities and collaborative modes. From an institutional

perspective, Strange et al. [118] examine the challenges of sustaining open digital humanities data collections beyond the lifetime of research projects.

Since these aspects of data seeking and data sharing in the humanities are foundational for our considerations on the visibility of research data, we will further systematize the foundational modes of access in Section 3.4.

3 Humanities Research Data Visibility Model

Based on the related work outlined in Section 2, this section introduces the HRDVM, as shown in Figure 1, to address **RQ1**. The model comprises five core elements: Disciplinary Data Practices (a), Humanities Data (b), Data-Holding Entities (c), Modes of Access (d) and their corresponding Dimensions of Visibility (e). The Modes of Access are particularly central to addressing **RQ1**, as they capture the ways through which researchers become aware of humanities research data and the factors that shape their visibility.

The following subsections introduce each element, explain how they relate to each other within the model, and connect them to the existing literature. Rather than treating visibility as a stable property of data, the model conceptualizes it as an outcome of conditions shaped by how data are produced, stored, and accessed, with particular emphasis on the Visibility Dimensions (Section 3.5) generated by different Modes of Access (Section 3.4).

3.1 (a) Disciplinary Data Practices:

The Disciplinary Data Practices describe how humanities disciplines determine which human-made artifacts are recognized as relevant research objects. Artifacts do not become research data by their mere existence, but through processes of selection, classification, and framing guided by specific research interests. As Borgman notes, institutions organize and represent objects in accordance with their missions [36], a logic that also applies to disciplinary practices. Consequently, the same artifact may acquire different meanings depending on the disciplinary context. For example, a nineteenth-century handwritten notebook may serve as a historical source for historians or as material for authorship analysis in literary studies. These practices shape the Humanities Data used within a field and determine which materials enter Data-Holding Entities. In doing so, they establish a form of pre-access visibility by defining which materials become available for subsequent access and perception, alongside the visibility dimensions associated with different Modes of Access.

3.2 (b) Humanities Data:

Humanities research employs a wide range of data types, reflecting the inherent heterogeneity of humanities data. Existing classifications capture this diversity by focusing on observable characteristics, such as modality (e.g., textual, numerical, audio) [106] or the type of scholarly object represented (e.g., publications, primary sources, events) [57]. Complementing these classifications, Borgman proposes a creation-oriented typology that distinguishes between material objects, their digital representations, and born-digital materials [36]. Building on this perspective, we extend this typology by introducing physical representations alongside digital representations. Accordingly, the HRDVM distinguishes four categories of humanities research data: *Physical Objects*, *Physical Surrogates*, *Digital Surrogates*, and *Born Digitals*.

In the HRDVM, Physical Objects refer to tangible artifacts existing in the real world (e.g., books or paintings), Physical Surrogates to their physical reproductions (e.g., facsimiles), Digital Surrogates to digitized versions of Physical Objects (e.g., scans or 3D models), and Born Digitals to materials created in digital form (e.g., digital art works). These categories are not mutually exclusive. A single object may exist across multiple modes of this typology. For example, a Physical Object may also be available as a Physical Surrogate and as a Digital Surrogate. Whether such artifacts are recognized as research data and stored in Data-Holding Entities depends, again, on Disciplinary Data Practices.

3.3 (c) Data-Holding Entities:

The Data-Holding Entities refer to the entities in which research data are stored and maintained. We distinguish *Physical Storage* and *Digital Storage*. Physical Storage denotes geographically situated locations requiring on-site access, such as libraries, archives, museums, or private collections. Digital Storage refers to logically addressable locations within digital infrastructures, including repositories, databases, digital archives, project websites, and personal or collaborative storage environments. These examples are illustrative rather than exhaustive, as data in the humanities can be stored in a wide range of physical and digital entities. Furthermore, this distinction is not absolute but reflects a continuum of storage practices that also includes hybrid forms. For example, a museum may preserve Physical Objects on site while simultaneously maintaining Digital Surrogates within its digital infrastructure.

The Data-Holding Entities can be positioned in relation to the data pyramid of humanities research data in the institutional context proposed by Andorfer [32], which synthesizes concepts from Sahle and Kronenwett [109] and Hügi and Schneider [65] with the data pyramid model of Hey et al. [64]. The humanities data pyramid models the movement of data throughout the research workflow, distinguishing between sources (at the bottom), working data generated during the research process, and publications as forms of scholarly output (at the top). The Data-Holding Entities in our model correspond to the foundational level of Andorfer's pyramid, where he also models institutional storage next to the sources. However, HRDVM extends this focus by accounting for a broader landscape of entities in which humanities data are stored.

3.4 (d) Modes of Access:

The Modes of Access refer to the ways through which researchers become aware of and encounter sources preserved by the Data-Holding Entities. In this context, access is understood as awareness and discovery.

To derive these Modes of Access, we reviewed literature on information-seeking and research data-seeking practices in the humanities. Relevant publications were identified through a manual search conducted with Google Scholar by using the terms "*humanities information seeking*", "*humanities research data seeking*", and "*humanities information behavior*". The search was not restricted by publication date and was complemented by backward and forward

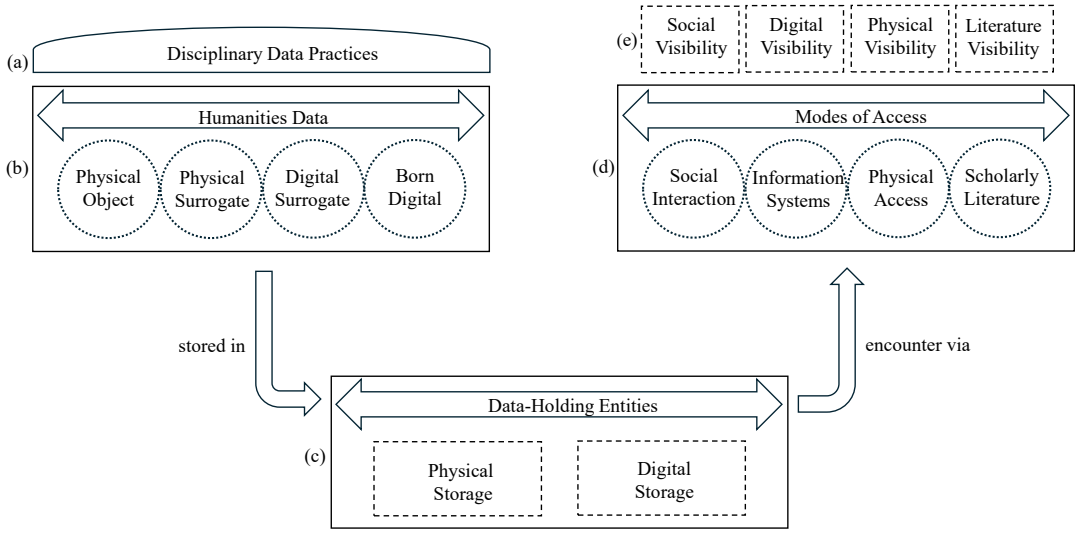


Figure 1: Architecture of the Humanities Research Data Visibility Model

citation searches. Citation chaining was continued until no additional relevant publications emerged. Only publications written in English were included.

From this literature, we extracted the resources and channels through which humanities researchers are reported to become aware of and locate relevant sources or research data. Table 1 summarizes these literature-derived resources and shows how they can be grouped into broader access patterns. Rather than treating each resource as a separate category, we abstracted them according to the underlying way in which access is enabled. This resulted in four elementary Modes of Access: *Social Interaction*, *Information Systems*, *Physical Access*, and *Scholarly Literature*.

Social Interaction refers to access mediated through interpersonal practices, including peer-to-peer communication (e.g. conversations or email), as well as interactions embedded in broader social settings (e.g. workshops, conferences, and seminars). *Information Systems* denote access via local and web-based infrastructures, including institutional catalogs, repository search interfaces, recommender systems, search engines, and generative search systems (see Kelly et al. [67] for an example of a RAG-System applied to cultural heritage materials). *Physical Access* captures access through on-site engagement with materials in libraries, archives, or museums. *Scholarly Literature* refers to access mediated through publications, where research data are encountered via citations, references, or descriptions.

It should be noted that the resources and channels mentioned in Table 1 do not constitute an exhaustive list of all possible access routes. Additional or emerging resources and channels, such as generative search, may exist and can still be assigned to the proposed Modes of Access, even if they are not yet prominently addressed in the reviewed literature.

Table 1: Derivation of Modes of Access from literature-derived resources and channels

Mode of Access	Resources/Channels
Social Interaction	Archivists, librarians [31, 129]
	Colleagues [28, 31, 33, 34, 129]
	Conferences [28, 34, 80]; seminars [28]
	Supervisors [28, 34]; lecturers [33]
	Social media [80]
Information Systems	Databases [28, 33, 34, 68, 84, 86, 129]
	Online finding aids of libraries/archives [84]
	Online catalogs [84]
	Search engines [28, 68, 86, 123]
	Video platforms [68, 86]
	Social media [86]
	Open data platforms [86]
	Repositories [31, 123]
	OPACs [28, 31, 34, 123]
Physical Access	GLAM institutions [28, 33, 34, 129]
	Personal collections [28, 129]
	Bookstores [28, 34]; book exhibitions [28]
Scholarly Literature	Journal articles/literature [33, 34, 123]
	Reference materials [33]
	Bibliographies [34, 123]

3.5 (e) Dimensions of Visibility:

Within the model, research data do not become visible merely by being stored or made available in Data-Holding Entities. Instead, visibility emerges through the Modes of Access by which researchers become aware of them. Accordingly, each mode is associated with a distinct *Dimension of Visibility*, which shapes how data become perceptible in practice. These dimensions are structured by a set of factors that condition this perceptibility. The model distinguishes four such dimensions: *Social Visibility*, associated with Social Interaction; *Digital Visibility*, associated with Information Systems;

Physical Visibility, associated with Physical Access; and *Literature Visibility*, associated with Scholarly Literature.

This perspective aligns with Neumayer et al. [102], who conceptualize visibility as a socio-technical process rather than a stable property of data. While their framework, developed for social media data, emphasizes dimensions such as actors, technologies, and meaning, our model adapts this core premise to the context of research data by linking visibility to distinct Modes of Access. In the following, each Dimension of Visibility is introduced by outlining central and initial factors that shape how research data become perceptible within the respective access modes.

Social Visibility. Social Visibility shapes the perceptibility of research data through Social Interaction. In peer-to-peer contexts, factors contributing to the Social Visibility arise from the data seeker and the information holder.

On the side of the data seeker, one factor is the articulation of information needs. Research data may remain socially invisible when researchers do not ask others about their existence. Such restraint can be driven by concerns about social exposure such as appearing incompetent [37]. More generally, seeking help can signal incompetence, dependence, or inferiority to others [79]. In academic environments where competence is highly valued, revealing such uncertainty may be perceived as risky. As a result, researchers may refrain from asking even when relevant knowledge about research data is available within their social network. For example, a researcher searching for archival data may hesitate to ask a junior colleague for assistance, despite that colleague's direct knowledge of the data, because seeking advice from someone perceived as less established may be interpreted as signaling a lack of expertise. In this way, research data may remain socially invisible not because they are unknown within a community, but because articulating information needs entails a perceived risk of negative social reaction. This mechanism mirrors Chatman's theory of information poverty, which claims that individuals often avoid revealing their problems because they believe the potential negative consequences outweigh the benefits [41].

On the side of the information holder, one factor is the willingness to share information about research data. Research data may remain socially invisible when researchers who are aware of the data choose not to disclose its existence to others. Such restraint may be shaped by strategic or normative considerations. Researchers may withhold information due to concerns about competition, expectations of reciprocity, and prevailing disciplinary norms [58, 59]. As a result, information about research data may circulate selectively within social networks. For example, a scholar may share knowledge about a dataset with close collaborators while withholding it from others when no reciprocal relationship is anticipated. In this way, research data may remain socially invisible because information holders deliberately limit disclosure based on social, strategic, or normative considerations.

Beyond direct peer-to-peer interactions, Social Visibility is also shaped through disciplinary norms in broader social communication settings such as conferences, workshops, seminars, and social media platforms. These communication settings provide opportunities for research data to be mentioned or informally circulated, thereby increasing their perceptibility within a scholarly community. Disciplines differ in their data sharing activities [35]. In communities characterized by a strong culture of data sharing, these forms can increase the perceptibility of research data, whereas in more strict disciplines, they may contribute little to data perceptibility.

Digital Visibility. Digital Visibility shapes the perceptibility of research data through Information Systems. Factors contributing to Digital Visibility arise both from the research data itself and from the design of the Information Systems through which they are accessed.

On the resource side, one factor is the quality and scope of metadata associated with research data, also addressed by the FAIR Principles [130]. Because many information systems rely on metadata to represent, organize, and retrieve information objects, metadata determine how resources are matched to user queries and thus whether and how they become perceptible to researchers. Consequently, metadata quality and completeness constitute critical components for effective access and retrieval [52, 130]. If metadata are incomplete, inaccurate, or inconsistent, this negatively affects the discoverability of the resources [104]. As a result, suitable research data may remain digitally invisible even when they are technically available within an information system. For example, a digitized archival collection may record the place of origin inconsistently across metadata entries, using modern names in some cases and historical or abbreviated variants in others. As a result, region-based queries may retrieve only a subset of relevant materials, since the system cannot reconcile these variants as referring to the same location.

The types of metadata used to describe a resource are a further factor contributing to Digital Visibility. Different metadata types, such as administrative, descriptive, preservation, technical, and use metadata [52], capture different aspects of research data and enable different forms of discovery and access. As a consequence, only those dimensions of a resource for which corresponding metadata exist can be taken into account during retrieval. If a resource lacks the metadata types corresponding to the criteria used in a search query, the information system will not surface it, even when the resource itself would be highly relevant to the user's research interest. For example, a digitized correspondence collection can be described thematically through descriptive metadata and therefore be discoverable via topic-based searches. However, if contextual metadata are not recorded, the same collection may not be retrieved by researchers searching for sources from a specific region. In this context, one can try to improve the quality and completeness of metadata using LLMs. Currently, many projects are reporting results in this direction (see, e.g., [29, 76, 99, 127]). While some of these results are promising, in many cases they are still far from perfect. For the time being, well-maintained metadata will remain a crucial foundation in many fields.

On the information system side, Digital Visibility is shaped by several interrelated mechanisms that impact how research data are retrieved for users.

One factor is indexing. Indexing determines which elements of research data are represented in the system's internal structures and are therefore available for search and retrieval. Depending on

the indexing strategy employed (see, e.g., [45, 60]), certain aspects of a resource are foregrounded while others remain invisible, even when they are documented in the underlying record. For example, an archival dataset may include metadata describing the geographic origin of its source, but if the system indexes only full text and not the place-related metadata fields, location-based search queries will not retrieve the data.

Building on indexing, retrieval and especially ranking algorithms constitute a further factor, as they determine which indexed resources are selected and how they are ordered in response to a query. A variety of retrieval and ranking techniques can be applied to an information system (see Hambarde and Proenca [60] for an overview of prominent categories and techniques for information retrieval and ranking). Consequently, even when research data are indexed and technically retrievable, their visibility also depends on the specific retrieval and ranking techniques and their combination.

Searcher actions, tactics, and strategies are shaped by the capabilities provided by a search system [74]. As these capabilities are made accessible through the user interface, the interface itself constitutes a further factor contributing to Digital Visibility. By defining the available interaction options, such as search fields, filters, and navigation structures, interfaces shape how users formulate queries and explore research data, enabling certain search strategies while constraining others. Consequently, which search options are provided determines which research data become easily accessible and which remain difficult to explore, regardless of their presence and indexing within the system. For example, Kules and Shneiderman show that faceted search interfaces shape searchers' tactics and influence which results users choose to examine [75]. This is corroborated by a systematic literature review of the challenges humanities scholars encounter when adopting software tools in digital humanities projects, which identifies insufficient search support as one of the interface-related practice barriers [89]. This is further supported by usability testing on archival retrieval systems, where unclear interface elements prevented users from effectively using integrated digital humanities features [87]. This underscores the importance of search systems as gatekeepers to research data: both the way results are prioritized and ranked, and the clarity with which functionality is presented determine whether relevant research data actually become visible to users.

Physical Visibility. Physical Visibility shapes how research data become perceptible through Physical Access. One factor contributing to Physical Visibility is the ordering and shelving systems of physical Data-Holding Entities, which determine how materials are arranged and grouped. In libraries, archives, and museums, objects are typically organized according to thematic, chronological, or classificatory schemes (see Simmons et al. [113] for an overview of different classification principles used for museum objects). As a result, related materials are often located in close physical proximity, enabling scholars to discover additional relevant resources, even when these materials are not part of the original search intent. Such forms of exploratory discovery align with findings that serendipitous encounters play a significant role in humanities research [46, 48, 50, 80, 110].

Another factor is staff expertise and institutional knowledge practices. Personnel in libraries, archives, and museums often possess detailed, experience-based knowledge about collections, their organization, and the context of materials, much of which remains tacit and not fully captured in catalogs or finding aids [48, 90, 132, 133]. As a result, staff actively mediates access and shapes visibility by directing researchers to relevant materials, suggesting related sources, or highlighting resources that might otherwise remain unnoticed. In this sense, they act as knowledge gatekeepers whose expertise significantly influences what becomes perceptible during Physical Access. Accordingly, Physical Visibility depends not only on organizational structures but also on the availability and willingness of personnel to share their knowledge. Recent efforts aim to externalize and formalize such tacit knowledge, e.g., through consultation services like the NFDI4Culture Helpdesk [94].

A further factor concerns institutional rules and access conditions. Whether research materials can be viewed and used is shaped by usage regulations, data protection requirements, and legal constraints such as copyright [47, 133]. In some cases, access is conditional on meeting specific requirements, e.g., demonstrating a legitimate research purpose. These rules and conditions directly affect the perceptibility by determining which materials are made available to humanities researchers and under what circumstances. Materials subject to strict restrictions may remain unnoticed during Physical Access, whereas less restricted materials are more readily encountered.

Literature Visibility. Literature Visibility shapes the perceptibility of research data through Scholarly Literature. The first group of factors contributing to Literature Visibility can be described as *contextual visibility factors*, as they influence whether a publication, and the research data associated with it, are encountered at all.

One contextual factor is the publication outlet. Data reported in highly visible journals or widely read venues are more likely to attract attention than those published in peripheral outlets. This pattern reflects cumulative advantage processes in science, commonly referred to as the Matthew Effect, whereby recognition concentrates on already prominent scholars [98]. Empirical evidence further shows that the publication venue itself shapes visibility. Larivière and Gingras demonstrate that the journal in which a paper is published has a significant influence on its citation rate [77].

Another contextual factor is author reputation, which shapes whether a publication, and the data associated with it, receives attention. Reputation acts as a signal of credibility and quality, helping researchers cope with information overload and directing attention toward particular works [105]. It also serves as a proxy for trustworthiness, allowing scholars to rely on findings without incurring the costs of verification [115]. Empirical evidence shows that reputation significantly influences citation dynamics, especially early in a publication's life cycle [105]. In practice, author names function as heuristics in the evaluation of newly encountered work, reinforcing cumulative visibility advantages [105]. Consequently, publications by well-established authors are more likely to be noticed, read, and cited.

The second group of factors contributing to Literature Visibility are *content-based visibility factors*. These factors concern how data are represented within the publication, including the ways data are described, referenced, and contextualized. The type, depth, and quality of research data descriptions in scholarly publications

Table 2: Dimensions of Visibility and their associated factors shaping perceptibility

Dimension of Visibility	Factors shaping perceptibility
Social Visibility	Articulation of information needs
	Willingness to share information
	Disciplinary data-sharing norms
Digital Visibility	Metadata quality and scope
	Metadata types
	Indexing
	Retrieval and ranking techniques
	Interface affordances
Physical Visibility	Ordering and shelving systems
	Staff expertise and institutional knowledge practices
	Institutional rules and access conditions
Literature Visibility	Publication outlet
	Author reputation
	Representation and contextualization of data in publications

shape how such data are perceived. When researchers rely on the scholarly literature to identify relevant research data, publications typically serve as the primary interface through which data are first encountered. Moreover, they constitute the primary form of documentation through which data are introduced, described, and interpreted according to Borgman [35]. Consequently, the information provided in publications plays a central role in determining whether data are recognized as relevant. If data are only briefly mentioned or insufficiently contextualized, they may be overlooked or judged as irrelevant, even when they would be suitable for a given research purpose. This is consistent with prior work showing that data reuse depends, among other factors, on the availability of contextual information and documentation [35, 134].

Table 2 summarizes the Dimensions of Visibility and the influencing factors discussed in the section. Several of the factors identified above have corresponding implications for the design of research data infrastructures. We will discuss these in more detail in the following Section 4, focusing on the scenario of the humanities.

4 Implications for Humanities Research Data Infrastructures

Building on the proposed HRDVM and addressing **RQ2**, this section derives implications for humanities research data infrastructures by translating the identified Modes of Access, their corresponding Dimensions of Visibility, and the initial factors contributing to these dimensions into basic considerations for infrastructure design and implementation.

For the Digital Visibility factors summarized in Table 2, one implication concerns the role of metadata for research data. Metadata are commonly understood as data about data, primarily used to describe and represent research objects by documenting their characteristics. From this perspective, metadata serve a representational function by specifying what a resource is. However, this perspective limits the potential of metadata for reuse of research data. Rather than being understood solely as descriptive, metadata should also be conceived as carriers of potential use. This perspective shifts attention from only documenting the existing properties of a research object to also capturing information about how, in which contexts, and for which research interests the research data can also be meaningfully used. This (re)use-oriented perspective is not directly reflected in currently used metadata schemata. For example, an examination of the Dataset schema [22] of schema.org shows it does not include an explicit field for documenting the purposes for which a dataset is (or might be) suitable or the contexts in which it could be (re)used. Instead, such information can only be expressed indirectly through the CreativeWork properties, such as *exampleOfWork* or *workExample*. Incorporating such information into the metadata schema through explicit fields (e.g., *potentialUse*, *reusePotential* or *usageContext*) would enable research data infrastructures to account for reuse contexts during discovery and retrieval processes. A similar situation can be observed in the humanities. Although prior work has proposed metadata models for humanities research data [30, 114], these approaches primarily focus on descriptive, structural, and contextual aspects rather than explicitly representing potential reuse contexts. One obvious problem here is the question of whether possible usage scenarios can be comprehensively anticipated when curating research data. This naturally has its limits. On the other hand, the concept of SEO is well known in the creation of websites, demonstrating that the findability of digital content can be actively shaped by anticipating typical search behaviors and query formulations for which the provided content could be relevant, even in the absence of clearly defined use cases. Against this background, it becomes clear that when creating metadata, it is ultimately a matter of changing perspective. Instead of simply describing the data, metadata should be written clearly with the aim of being found by relevant search queries.

Literature Visibility is modeled as a mode of access in HRDVM emphasizing the importance of scholarly literature as a way to raise awareness and interest in research data. Clearly, this does not mean that data infrastructures need to manage publications in the same way as research data themselves. Nevertheless, one implication of this perspective is that research data infrastructures should integrate literature-based access mechanisms for discovering research data. In consequence, this requires publication metadata models that explicitly and unambiguously represent the research data used in scholarly work, particularly for research data infrastructures that primarily rely on metadata.

Modeling research data as identifiable objects within publication metadata can enable such infrastructures to more systematically leverage publications for data discovery and reuse. Currently, publication metadata schemata rarely provide properties that are explicitly designed to represent research data as part of scholarly work. For example, this limitation is evident in the ScholarlyArticle schema [25] of schema.org. Although this schema allows scholarly works to be linked to other resources through properties such as *isBasedOn*, *isPartOf* or *workIsDerivedFrom*, it does not include a property that explicitly denotes the use of research data such as *usedDataset*. As a result, research data can only be referenced indirectly, through properties that are not designed to distinguish research data from other scholarly resources in an easy way.

A second, complementary approach to supporting literature-based access may be the integration of literature-derived knowledge graphs into a research data infrastructure. Scholarly literature, including conventional publications and data papers, often contains valuable information about research data, expressed in narrative descriptions of datasets, their research context and reuse potential, as well as in figures and tables. Extracting such information from the literature and linking it to the corresponding research data through a knowledge graph can enhance the perceptibility, the discoverability and the reuse of research data (see, e.g., Stocker et al. for a related approach towards an Open Research Knowledge Graph [116], Martin and Henrich for a more technical idea to link publications and related resources [93] and McGillivray et al. for linking research papers to data papers [95]).

Following the Social Visibility factors summarized in Table 2 and Social Interaction as a Mode of Access, another implication is that research data infrastructures should support the visibility and accessibility of implicit knowledge that is generated, carried, and exchanged through social interaction. The following approaches outline initial possibilities for integrating this aspect into humanities research data infrastructures.

One possible approach is the explicit modeling of responsibilities and points of contact for research data that are held by the infrastructure. Such modeling can help integrate socially distributed expertise into the infrastructure by indicating whom to consult regarding the research data.

Another approach is the integration of forum-like spaces that support scholarly exchange about research data. An illustrative example of this approach is Forum4MICA [23], a forum dedicated to topics related to research data and its management in the social, educational, behavioral, and economic sciences. It supports direct communication between data users and data providers via a forum-based interface. In contrast to Forum4MICA, which is organized as a general discussion platform, a potential forum for humanities research data could be designed as a data-anchored space in which scholarly exchange is explicitly linked to specific research data, thereby supporting the articulation and preservation of implicit knowledge. Regardless of its nature as a general discussion forum, Forum4MICA can serve as a starting point and reference model, in combination with empirical work, for developing comparable forum-based components for a humanities research data infrastructure tailored to the needs of humanities scholars.

A further implication derived from access through Social Interaction is that research data infrastructures should disseminate the research data they hold in order to increase awareness within scholarly communities and society of this data, rather than primarily promoting the infrastructure itself. Promoting a research data infrastructure as a platform provides limited informational value for potential users. Although such promotion signals the existence of the infrastructure, it does not necessarily convey what kinds of research data it stores or how they can be used. Data-centered dissemination addresses this limitation by making the data and their scholarly potential visible. Existing research on how (research) data infrastructures disseminate research data is limited, with only a few contributions such as Butzlaff et al. [40]. An illustrative example of data-centered dissemination, however, is provided by bavarikon [3]. Bavarikon is the digital portal of the Free State of Bavaria for presenting art, culture, and knowledge from museums, archives, and libraries across the region. Although bavarikon does not function as a research data infrastructure in the first place, it demonstrates how active, data-focused promotion can contribute to the visibility of data and cultural objects and can serve as an inspiration for data-centered promotion strategies in research data infrastructures. Bavarikon uses a range of communication channels to create awareness of cultural and scientific objects, including its website [3], a podcast [17], and multiple social media platforms [14–16, 18, 19]. Through social media promotion, these objects gain visibility and circulate via interaction mechanisms such as sharing, reposting, linking, and commenting. Beyond this, bavarikon further amplifies visibility through curated virtual exhibitions [5] and active dissemination via press releases, posters, and other printed materials [4].

Of course, bavarikon's target audience differs from that of typical research data infrastructures, as it is explicitly designed for both the general public and scientific users [112], rather than being primarily researcher-oriented. Therefore, the prioritization of channels must be optimized specifically in each case. However, the clear focus on the target audience and the promotion of the infrastructure via its content rather than as an infrastructure in itself remain as a model for research data infrastructures.

As a side note, cultural heritage portals such as bavarikon are sometimes assumed to primarily address a non-academic public. This assumption, however, does not fully hold. According to its mission statement, bavarikon targets both scientific users and the general interested public [112]. Findings by the cultural heritage portal Europeana [1] further support this picture. A study by Clough et al. [42] on search motivation and search targets within Europeana shows that, among 240 survey responses, 30.4% of users—the largest group identified in the study—described themselves as academics. This indicates that researchers constitute a substantial and active user group of cultural heritage portals and suggests that such platforms are already situated closer to the researcher side than is often assumed. This raises the question of whether such platforms may, in terms of their user base, already approximate or increasingly converge with research data infrastructures, even if their institutional tasks and design priorities continue to differ.

The Physical Visibility factors summarized in Table 2 point to the importance of ordering and shelving systems. In physical Data-Holding Entities, ordering and shelving practices allow researchers to discover research data through spatial proximity. Research data infrastructures can draw on this principle by translating physical ordering logics into the digital environment, thereby creating new possibilities for making research data visible by enabling contextual browsing and the perception of relationships between research data (see, e.g., the bohemian bookshelf as an illustration of how different ordering principles are translated into digital systems to support serendipitous discovery [122] or Windhager et al. for a survey on visualization of cultural heritage collection data to enhance access to cultural collections [131]).

Even the factor of staff expertise and institutional knowledge practices can be adapted to the digital world. Infrastructures can provide access to experts through various channels (email, phone,

chat, etc.) [94]. This often benefits not only those asking questions but also the operators of the infrastructure, as it allows them to gain a better understanding of the target audience's needs over time.

A last implication derived from Physical Access is that research data infrastructures should also support the perceptibility of research data that exist only in physical form but are not digitally available. By referencing descriptive records, metadata or finding aids that represent these physical research data, infrastructures can signal their existence and make researchers aware of relevant physical materials.

The implications outlined in this section are intended merely as illustrative examples. In our view, the HRDVM (Figure 1) and the factors summarized in Tables 1 and 2 can effectively support the targeted optimization of research data infrastructures with the goal of achieving high visibility for research data.

5 Conclusion

This paper challenged the implicit assumption that research data are visible to potential users simply by existing and being made accessible. Building on the premise that visibility is not a stable property of the data itself [102], we argued that visibility is instead the outcome of interacting conditions that shape whether and how research data become perceptible to humanities scholars. We focused in particular on four central Modes of Access to humanities research data and their corresponding Dimensions of Visibility, each shaped by individual factors. To conceptualize these relationships, we introduced the HRDVM, distinguishing Information Systems, Physical Access, Scholarly Literature, and Social Interaction as central modes through which data are encountered. For Information Systems, Digital Visibility is shaped by metadata quality, indexing and ranking mechanisms, and interface design. Physical Visibility depends on spatial organization, institutional regulations, and human mediation. Literature Visibility is influenced by publication outlets, authorship, and the representation of research data within texts. Social Visibility emerges from the articulation of information needs, willingness to share, and disciplinary norms.

These factors should be understood as initial factors illustrating how visibility is structured within each Mode of Access. Based on this framework, we derived implications for research data infrastructures. Rather than viewing them solely as retrieval systems, we argue for a broader perspective that supports multiple modes of access and their visibility conditions. This includes rethinking metadata as carriers of reuse potential, integrating literature- and physical-based access, incorporating social expertise and tacit knowledge, and actively promoting research data. Improving visibility, therefore, cannot be achieved through technical optimization alone but requires accounting for the different modes through which data become perceptible. The HRDVM provides a conceptual basis for guiding future research and infrastructure design. Future work should test and refine the factors proposed as contributing to the different Visibility Dimensions and translate the resulting implications into concrete infrastructural practices. In the context of literature-based access, a next step is to examine how and to what extent humanities publications document research data in their content and metadata.

Acknowledgments

This paper is based on work and experiences in the Text+ project, funded by the Deutsche Forschungsgemeinschaft (DFG, project number 460033370) as part of the German National Research Data Infrastructure (NFDI e. V.). The authors sincerely thank their colleagues in Text+ and the broader NFDI context for their invaluable contributions and insights.

Declaration on Generative AI

During the preparation of this work, the authors used DeepL and ChatGPT to check grammar and spelling, and to improve the writing style. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] 2008. Europeana. Accessed: 20.04.2026. https://www.europeana.eu/en
[2] 2012. CLARIN. Accessed: 20.04.2026. https://www.clarin.eu/
[3] 2013. bavarikon. Accessed: 20.04.2026. https://www.bavarikon.de/?lang=en
[4] 2013. bavarikon - Press and Public relations. Accessed: 20.04.2026. https://www.bavarikon.de/press?lang=en
[5] 2013. bavarikon - Virtual Exhibitions. Accessed: 20.04.2026. https://www.bavarikon.de/exhibitions?lang=en
[6] 2013. ELIXIR. Accessed: 20.04.2026. https://www.elixir-europe.org/
[7] 2014. DARIAH-EU. Accessed: 20.04.2026. https://www.dariah.eu/
[8] 2015. CESSDA. Accessed: 20.04.2026. https://www.cessda.eu/
[9] 2018. ARDC. Accessed: 20.04.2026. https://www.ardc.edu.au/
[10] 2020. NFDI. Accessed: 20.04.2026. https://www.nfdi.de/
[11] 2022. Recherche Data Gouv. Accessed: 20.04.2026. https://recherche.data.gouv.fr/en
[12] 2023. Gaia-X. Accessed: 20.04.2026. https://www.gaia-x.eu/
[13] 2024. DFG - Systematik der Fachkollegien Amtsperiode 2024-2028. Accessed: 20.04.2026. https://www.dfg.de/resource/blob/175690/774fa9bbe276519d52bda4a1395adfb5/2023-systematik-faecher-fachkollegien-data.pdf
[14] 2026. bavarikon Facebook. Accessed: 20.04.2026. https://www.facebook.com/bavarikon
[15] 2026. bavarikon Instagram. Accessed: 20.04.2026. https://www.instagram.com/bavarikon.de/
[16] 2026. bavarikon Pinterest. Accessed: 20.04.2026. https://de.pinterest.com/bavarikon/
[17] 2026. bavarikon Podcast - Crime History Talk. Accessed: 20.04.2026. https://bavarikon.blogs.julephosting.de/
[18] 2026. bavarikon TikTok. Accessed: 20.04.2026. https://www.tiktok.com/@bavarikon
[19] 2026. bavarikon Youtube. Accessed: 20.04.2026. https://www.youtube.com/@bavarikon4U
[20] 2026. DANS. Accessed: 20.04.2026. https://www.dans.knaw.nl/en/
[21] 2026. DARIAH - Research Data. Accessed: 20.04.2026. https://de.dariah.eu/en/forschungsdaten
[22] 2026. Dataset - Schema.org Type. Accessed: 20.04.2026. https://schema.org/Dataset
[23] 2026. Forum4MICA - Discussion forum about social sciences data. Accessed: 20.04.2026. https://forum.lifbi.de/
[24] 2026. re3data - Browse by subject. Accessed: 20.04.2026. https://www.re3data.org/browse/by-subject/
[25] 2026. ScholarlyArticle - Schema.org Type. Accessed: 20.04.2026. https://schema.org/ScholarlyArticle
[26] Pranjal Aggarwal, Vishvak Murahari, Tanmay Rajpurohit, Ashwin Kalyan, Karthik Narasimhan, and Ameet Deshpande. 2024. Geo: Generative Engine Optimization. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 5–16. doi:10.1145/3637528.3671900
[27] Mohammad Khaled Al Shboul and Abdullah Abrizah. 2014. Information Needs: Developing Personas of Humanities Scholars. *The Journal of Academic Librarianship* 40, 5 (2014), 500–509. doi:10.1016/j.acalib.2014.05.016
[28] Mohammad Khaled Al Shboul and Abdullah Abrizah. 2016. Modes of information seeking: Developing personas of humanities scholars. *Information Development* 32, 5 (2016), 1786–1805. doi:10.1177/0266666915627673
[29] Zaid Alyafeai, Maged S Al-Shaibani, and Bernard Ghanem. 2025. MOLE: Metadata Extraction and Validation in Scientific Papers Using LLMs. *Findings of*

the Association for Computational Linguistics: EMNLP 2025 (2025), 12236–12264. doi:10.18653/v1/2025.findings-emnlp.655

[30] Byoung-Goon An. 2025. A metadata design for humanities research data based on semantic structure of research literature. *Journal of Information Science* (2025). doi:10.1177/01655515241305345

[31] Ian G. Anderson. 2004. Are you being Served? Historians and the Search for Primary Sources. *Archivaria* 58 (2004).

[32] Peter Andorfer. 2015. Forschungsdaten in den (digitalen) Geisteswissenschaften: Versuch einer Konkretisierung. *DARIAH-DE Working Papers* 14 (2015). https://publications.goettingen-research-online.de/handle/2/108308

[33] Emmanuel E Baro, George O Onyenania, and Oni Osaheni. 2010. Information seeking behaviour of undergraduate students in the humanities in three universities in Nigeria. *South African Journal of Libraries and Information Science* 76, 2 (2010), 109–117. https://journals.co.za/doi/abs/10.10520/EJC61346

[34] Andy Barrett. 2005. The Information-Seeking Habits of Graduate Student Researchers in the Humanities. *The Journal of Academic Librarianship* 31, 4 (2005), 324–331. doi:10.1016/j.acalib.2005.04.005

[35] Christine L Borgman. 2012. The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology* 63, 6 (2012), 1059–1078. doi:10.1002/asi.22634

[36] Christine L Borgman. 2015. *Big Data, Little Data, No Data: Scholarship in the Networked World*. The MIT Press. doi:10.7551/mitpress/9963.001.0001

[37] Alison Wood Brooks, Francesca Gino, and Maurice E Schweitzer. 2015. Smart People Ask for (My) Advice: Seeking Advice Boosts Perceptions of Competence. *Management Science* 61, 6 (2015), 1421–1435. doi:10.1287/mnsc.2014.2054

[38] George Buchanan, Sally Jo Cunningham, Ann Blandford, Jon Rimmer, and Claire Warwick. 2005. Information Seeking by Humanities Scholars. *International Conference on Theory and Practice of Digital Libraries* (2005), 218–229. doi:10.1007/11551362_20

[39] Toby Burrows. 2011. Sharing humanities data for e-research: conceptual and technical issues. *Sustainable data from digital research: Humanities perspectives on digital scholarship*. (2011). http://hdl.handle.net/2123/7938

[40] Iris Butzlaff, Lisa Hönegger, and Lars Kaczmirek. 2020. Dissemination Strategy. *Vienna: The Austrian Social Science Data Archive* (2020).

[41] Elfreda A Chatman. 1996. The impoverished life-world of outsiders. *Journal of the American Society for Information Science* 47, 3 (1996), 193–206. doi:10.1002/(SICI)1097-4571(199603)47:3<193::AID-ASI3>3.0.CO;2-T

[42] Paul Clough, Timothy Hill, Monica Lestari Paramita, and Paula Goodale. 2017. Europeana: What Users Search for and Why. In *International Conference on Theory and Practice of Digital Libraries*. Springer, 207–219. doi:10.1007/978-3-319-67008-9_17

[43] Ellen Collins and Michael Jubb. 2012. How do Researchers in the Humanities Use Information Resources? *LIBER Quarterly: The Journal of the Association of European Research Libraries* 21, 2 (2012), 176–187. doi:10.18352/lq.8017

[44] Renato Cortinovis, Alexander Mikroyannidis, John Domingue, Paul Mulholland, and Robert Farrow. 2019. Supporting the discoverability of open educational resources. *Education and Information Technologies* 24 (2019), 3129–3161. doi:10.1007/s10639-019-09921-3

[45] Bruce Croft, Donald Metzler, and Trevor Strohman. 2009. *Search Engines: Information Retrieval in Practice* (1st ed.). Addison-Wesley Publishing Company, USA.

[46] Roberto Delgadillo and Beverly P Lynch. 1999. Future Historians: Their Quest for Information. *College & Research Libraries* 60, 3 (1999), 245–259. doi:10.5860/crl.60.3.245

[47] Virginia Dressler, Kelley Rowan, and Rebecca Bakker. 2021. Managing privacy: A survey of practices in digital archives and libraries. *Journal of Digital Media Management* 9, 2 (2021), 177–190.

[48] Wendy M Duff and Catherine A Johnson. 2002. Accidentally Found on Purpose: Information-Seeking Behavior of Historians in Archives. *The Library Quarterly* 72, 4 (2002), 472–496. doi:10.1086/lq.72.4.40039793

[49] Jennifer Edmond and Erzsébet Tóth-Czifra. 2018. Open Data for Humanists, A Pragmatic Guide. (2018). https://shs.hal.science/halshs-02115443v1

[50] Allen Foster and Nigel Ford. 2003. Serendipity and information seeking: an empirical study. *Journal of Documentation* 59, 3 (2003), 321–340. doi:10.1108/00220410310472518

[51] Massimo Franceschet and Antonio Costantini. 2010. The effect of scholar collaboration on impact and quality of academic papers. *Journal of Informetrics* 4, 4 (2010), 540–553. doi:10.1016/j.joi.2010.06.003

[52] Anne J Gilliland. 2016. Setting the Stage. In *Introduction to Metadata*. Getty Publications, 1–20.

[53] Nancy-Diana Gómez, Eva Méndez, and Tony Hernández-Pérez. 2016. Social sciences and humanities research data and metadata: A perspective from thematic data repositories. *Profesional de la información* 25, 4 (2016), 545–556. doi:10.3145/epi.2016.jul.04

[54] Kathleen Gregory, Paul Groth, Helena Cousijn, Andrea Scharnhorst, and Sally Wyatt. 2019. Searching Data: A Review of Observational Data Retrieval Practices in Selected Disciplines. *Journal of the Association for Information Science and Technology* 70, 5 (2019), 419–432. doi:10.1002/asi.24165

[55] Kathleen Gregory, Paul Groth, Andrea Scharnhorst, and Sally Wyatt. 2020. Lost or Found? Discovering Data Needed for Research. *Harvard Data Science Review* (2020). doi:10.1162/99608f92.e38165eb

[56] Kathleen M Gregory, Helena Cousijn, Paul Groth, Andrea Scharnhorst, and Sally Wyatt. 2019. Understanding data search as a socio-technical practice. *Journal of Information Science* 46, 4 (2019), 459–475. doi:10.1177/0165551519837182

[57] Bianca Gualandi, Luca Pareschi, and Silvio Peroni. 2023. What do we mean by "data"? A proposed classification of data types in the arts and humanities. *Journal of Documentation* 79, 7 (2023), 51–71. doi:10.1108/JD-07-2022-0146

[58] Carolin Haeussler. 2011. Information-sharing in academia and the industry: A comparative study. *Research Policy* 40, 1 (2011), 105–122. doi:10.1016/j.respol.2010.08.007

[59] Carolin Haeussler, Lin Jiang, Jerry Thursby, and Marie Thursby. 2014. Specific and general information sharing among competing academic researchers. *Research policy* 43, 3 (2014), 465–475. doi:10.1016/j.respol.2013.08.017

[60] Kailash A Hambarde and Hugo Proenca. 2023. Information Retrieval: Recent Advances and Beyond. *IEEE Access* 11 (2023), 76581–76604. doi:10.1109/ACCESS.2023.3295776

[61] Natalie Harrower, Beat Immenhauser, Gehrard Lauer, Maciej Maryl, Tito Orlandi, Bernard Rentier, and Eveline Wandl-Vogt. 2020. Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities. (2020). doi:10.7486/DRI.tq582c863

[62] Ralf-Christian Härting, Maik Mohl, Philipp Steinhauser, and Michael Möhring. 2016. Search Engine Visibility Indices Versus Visitor Traffic on Eebsites. In *International Conference on Business Information Systems*. Springer, 91–101. doi:10.1007/978-3-319-39426-8_8

[63] Nicolai Hauf, Andreas Fürholz, Vanessa Christina Klaas, Jennifer Morger, Elena Šimukovič, and Martin Jaekel. 2021. Data reuse in the social sciences and humanities: project report of the SWITCH Innovation Lab "Repositories & Data Quality". (2021). doi:10.21256/zhaw-2404

[64] Tony Hey, Stewart Tansley, and Kristin M Tolle. 2009. Jim Gray on eScience: a transformed scientific method. *The Fourth Paradigm: Data-intensive Scientific Discovery* (2009). ISBN: 978-0982544204.

[65] Jasmin Hügi and René Schneider. 2013. Digitale Forschungsinfrastrukturen für die Geistes-und Geschichtswissenschaften. *Genève: Haute école de gestion de Genève, 2013. 89 p.* (2013).

[66] Wei Jeng, Daqing He, and Jung Sun Oh. 2016. Toward a conceptual framework for data sharing practices in social sciences: A profile approach. *Proceedings of the Association for Information Science and Technology* 53, 1 (2016), 1–10. doi:10.1002/pra2.2016.14505301037

[67] Paul Kelly, Jonathan Schild, and Amir Jafari. 2025. FolkRAG: a retrieval-augmented generation system for cultural heritage materials. *Neural Computing and Applications* 37, 24 (2025), 20281–20297. doi:10.1007/s00521-025-11455-4

[68] Max Kemman, Martijn Kleppe, and Stef Scagliola. 2013. Just Google It - Digital Research Practices of Humanities Scholars. *arXiv preprint* (2013). doi:10.48550/arXiv.1309.2434

[69] Nushrat Khan, Mike Thelwall, and Kayvan Kousha. 2023. Data sharing and reuse practices: disciplinary differences and improvements needed. *Online Information Review* 47, 6 (2023), 1036–1064. doi:10.1108/OIR-08-2021-0423

[70] Youngseek Kim and Jeffrey M Stanton. 2012. Institutional and Individual Influences on Scientists' Data Sharing Practices. *Journal of Computational Science Education* 3, 1 (2012), 47–56. doi:10.22369/issn.2153-4136/3/1/6

[71] Wachiraporn Klungthanaboon. 2026. To share or not to share research data: Perspectives and practices of researchers in the digital humanities community in Thailand. *Journal of Librarianship and Information Science* 58, 1 (2026), 525–543. doi:10.1177/09610006251319029

[72] Thomas Krämer, Andrea Papenmeier, Zeljko Carevic, Dagmar Kern, and Brigitte Mathiak. 2021. Data-Seeking Behaviour in the Social Sciences. *International Journal on Digital Libraries* 22 (2021), 175–195. doi:10.1007/s00799-021-00303-0

[73] John E Kratz and Carly Strasser. 2015. Making data count. *Scientific Data* 2 (2015). doi:10.1038/sdata.2015.39

[74] Bill Kules, Robert Capra, Matthew Banta, and Tito Sierra. 2009. What do exploratory searchers look at in a faceted search interface?. In *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries*. 313–322. doi:10.1145/1555400.1555452

[75] Bill Kules and Ben Shneiderman. 2008. Users can change their web search tactics: Design guidelines for categorized overviews. *Information Processing & Management* 44, 2 (2008), 463–484. doi:10.1016/j.ipm.2007.07.014

[76] Manika Lamba, You Peng, Sophie Nikolov, Glen Layne-Worthey, and J. Stephen Downie. 2025. Metadata Enrichment of Long Text Documents using Large Language Models. *Proceedings of the Association for Information Science and Technology* 62, 1 (Oct. 2025), 990–994. doi:10.1002/pra2.1327

[77] Vincent Larivière and Yves Gingras. 2010. The impact factor's Matthew Effect: A natural experiment in bibliometrics. *Journal of the American Society for Information Science and Technology* 61, 2 (2010), 424–427. doi:10.1002/asi.21232

[78] Elina Late, Mette Skov, and Sanna Kumpulainen. 2024. To share or not to share? Image data sharing in the social sciences and humanities. *Information Research*

29, 2 (2024), 386–400. doi:10.47989/ir292834
[79] Fiona Lee. 2002. The Social Costs of Seeking Help. *The Journal of Applied Behavioral Science* 38, 1 (2002), 17–35. doi:10.1177/0021886302381002
[80] Carl Lehnen and Glenda M Insua. 2021. Browsing, Networking, Contextualizing: Research Practices of Humanists and Implications for Library Instruction. *portal: Libraries and the Academy* 21, 2 (2021), 275–297. doi:10.1353/pla.2021.0016
[81] Kai Li, Rongqian Ma, and Zhichao Fang. 2024. Explicit or implicit digital humanities? An examination of search strategies to retrieve digital humanities publications from large-scale scholarly databases. *Quantitative Science Studies* 5, 3 (2024), 718–735. doi:10.1162/qss_a_00321
[82] Wenqi Li, Zhenyi Tang, Pengyi Zhang, and Jun Wang. 2024. Collaborative Data Behaviors in Digital Humanities Research Teams. *Proceedings of the 24th ACM/IEEE Joint Conference on Digital Libraries* (2024), 1–5. doi:/10.1145/3677389.3702537
[83] Wenqi Li, Pengyi Zhang, and Jun Wang. 2023. Humanities Scholars' Understanding of Data and the Implications for Humanities Data Curation. *Proceedings of the Association for Information Science and Technology* 60, 1 (2023), 1034–1036. doi:10.1002/pra2.936
[84] Wenqi Li, Pengyi Zhang, and Jun Wang. 2024. Analysing humanities scholars' data seeking behaviour patterns using Ellis' model. *Information Research an international electronic journal* 29, 2 (2024), 401–418.
[85] Jingwen Lian. 2025. User-centered Research on Cultural Heritage Open Data Access and Reuse. In *Proceedings of the 24th ACM/IEEE Joint Conference on Digital Libraries*. 1–3. doi:10.1145/3677389.3702614
[86] Jingwen Lian, Yuxiang Zhao, Xinyue Li, and Qinghua Zhu. 2025. An Investigation of Searching Behavior for Open Datasets: Insights from Cultural Heritage Data Curation Competitions. *Proceedings of the Association for Information Science and Technology* 62, 1 (2025), 1007–1012. doi:10.1002/pra2.1330
[87] Chiao-Min Lin and Yu-Ting Huang. 2026. Evaluating usability in digital humanities features for archival retrieval systems. *Archival Science* 26 (2026). doi:10.1007/s10502-026-09536-x
[88] Chi-Shiou Lin. 2023. Exploring Humanities Researchers' Perceptions of "Data": A Phenomenological Approach. *Proceedings of the Association for Information Science and Technology* 60, 1 (2023), 1052–1054. doi:10.1002/pra2.942
[89] Rui Liu, Dana McKay, and George Buchanan. 2021. Humanities Scholars and Digital Humanities Projects: Practice Barriers in Tools Usage. *International Conference on Theory and Practice of Digital Libraries* (2021), 215–226. doi:10.1007/978-3-030-86324-1_25
[90] Rui Liu, Dana McKay, and George Buchanan. 2024. Sociotechnical Factors Affecting Interoperability of Digital Humanities. *Proceedings of the Association for Information Science and Technology* 61, 1 (2024), 218–231. doi:10.1002/pra2.1022
[91] Rongqian Ma. 2024. Toward an open humanities data: Current states, challenges, and cases. In *Intelligent Computing for Cultural Heritage*. Routledge, 3–24. doi:10.4324/9781032707211-2
[92] Gary Marchionini. 2006. Exploratory search: from finding to understanding. *Commun. ACM* 49, 4 (2006), 41–46. doi:10.1145/1121949.1121979
[93] Leon Martin and Andreas Henrich. 2024. RDFtex in-depth: knowledge exchange between LATEX-based research publications and Scientific Knowledge Graphs. *International Journal on Digital Libraries* 25 (2024), 517–535. doi:10.1007/S00799-023-00370-5
[94] Desiree Mayer and Angela Kailus. 2023. Der NFDI4Culture Helpdesk - ein Beratungsangebot für die Kulturwissenschaften. In *9. Tagung des Verbands Digital Humanities im deutschsprachigen Raum, DHd 2023, Belval, Luxembourg and Trier, Germany, March 13 - 17, 2022*, Anna Busch and Peer Trilcke (Eds.). 372. doi:10.5281/ZENODO.7715428
[95] Barbara McGillivray, Kaveh Aryan, Viola Harperath, Marton Ribary, and Mandy Wigdorowitz. 2025. Finding the Paper Behind the Data: Automatic Identification of Research Articles related to Data Publications. In *Proceedings of the Third Workshop for Artificial Intelligence for Scientific Publications*. 34–43. doi:10.18653/v1/2025.wasp-main.5
[96] Barbara McGillivray, Daniele Borkowski, Andrea Farina, Simon Mahony, and Xiaoguang Wang. 2025. Open Data Adoption Across the Humanities: Insights from the Journal of Open Humanities Data. In *AI and Smart Data for Cultural Heritage*. Routledge.
[97] Barbara McGillivray, Paola Marongiu, Nilo Pedrazzini, Marton Ribary, Mandy Wigdorowitz, and Eleonora Zordan. 2022. Deep Impact: A Study on the Impact of Data Papers and Datasets in the Humanities and Social Sciences. *Publications* 10 (2022), 39. doi:10.3390/publications10040039
[98] Robert K Merton. 1968. The Matthew Effect in Science: The reward and communication systems of science are considered. *Science* 159, 3810 (1968), 56–63. doi:10.1126/science.159.3810.56
[99] Adriana Morejón, Borja Navarro-Colorado, Carmen García-Barceló, Alberto Berenguer, David Tomás, and Jose-Norberto Mazón. 2025. Automatic Metadata Extraction Leveraging Large Language Models in Digital Humanities. *Electronics* 14, 24 (2025). doi:10.3390/electronics14244962
[100] Peter Morville. 2005. *Ambient Findability: What We Find Changes Who We Become*. O'Reilly Media, Inc.
[101] Heather Moulaison-Sandy and André G Wenzel. 2023. The Records Data Ecosystem in Humanities and Human Sciences Scholarship. *portal: Libraries and the Academy* 23, 1 (2023), 67–88. doi:10.1353/pla.2023.0000
[102] Christina Neumayer, Luca Rossi, and David M Struthers. 2021. Invisible Data: A Framework for Understanding Visibility Processes in Social Media Data. *Social Media + Society* 7, 1 (2021). doi:10.1177/2056305120984472
[103] D O'Donnell. 2018. The Bird in Hand: Humanities Research in the Age of Open Data. *Zenodo* (2018). doi:10.5281/zenodo.1470822
[104] Jung-Ran Park. 2009. Metadata Quality in Digital Repositories: A Survey of the Current State of the Art. *Cataloging & Classification Quarterly* 47, 3-4 (2009), 213–228. doi:10.1080/01639370902737240
[105] Alexander Michael Petersen, Santo Fortunato, Raj K Pan, Kimmo Kaski, Orion Penner, Armando Rungi, Massimo Riccaboni, H Eugene Stanley, and Fabio Pammolli. 2014. Reputation and impact in academic careers. *Proceedings of the National Academy of Sciences* 111, 43 (2014), 15316–15321. doi:10.1073/pnas.1323111111
[106] Ljiljana Poljak Bilić and Kristina Posavec. 2024. FAIRness of Research Data in the European Humanities Landscape. *Publications* 12, 1 (2024), 6. doi:10.3390/publications12010006
[107] Danilo Reyes-Lillo, Alejandro Morales-Vargas, and Cristòfol Rovira. 2025. Visibility, Discoverability, Findability, Search Engine Optimization (SEO) and Academic SEO in Digital Repositories: A scoping review. *BiD* 54 (2025). doi:10.1344/bid2025.54.06
[108] Dylan Ruediger and Ruby MacDougall. 2023. Are the Humanities Ready for Data Sharing? *ITHAKA S+R* (2023). https://www.jstor.org/stable/resrep49500
[109] Patrick Sahle and Simone Kronenwett. 2013. Jenseits der Daten: Überlegungen zu Datenzentren für die Geisteswissenschaften am Beispiel des Kölner "Data Center for the Humanities". *LIBREAS. Library Ideas* 23 (2013). https://libreas.eu/ausgabe23/09sahle/.
[110] Anna Schlander. 2026. Zwischen Struktur und Überraschung – Gestaltung explorativer Suchfunktionen für digitale Korpora. In *DHd 2026 Nicht nur Text, nicht nur Daten*. Zenodo. doi:10.5281/zenodo.18696161
[111] Christof Schöch. 2013. Big? Smart? Clean? Messy? Data in the Humanities. *Journal of Digital Humanities* 2, 3 (2013), 2–13. https://hal.science/hal-00920254v1
[112] Florian Sepp. 2017. Virtuelle Ausstellungen in bavarikon. Vortrag auf dem Workshop "Virtuelle Ausstellungen in Bibliotheken: Konzepte, Praxis und Perspektiven", 8 September 2017.
[113] John E Simmons and Toni M Kiser. 2020. *Museum Registration Methods*. Bloomsbury Publishing. ISBN: 9781538113110.
[114] Jane Smith. 2025. CREATING A SEMANTIC METADATA MODEL FOR HUMANITIES RESEARCH DATA. *International Research Journal of Library and Information Sciences* 2, 01 (2025), 1–6. doi:10.55640/irjlis-v02i01-01
[115] Paula E Stephan. 1996. The Economics of Science. *Journal of Economic Literature* 34, 3 (1996), 1199–1235. http://www.jstor.org/stable/2729500
[116] Markus Stocker, Allard Oelen, Mohamad Yaser Jaradeh, Muhammad Haris, Omar Arab Oghli, Golsa Heidari, Hassan Hussein, Anna-Lena Lorenz, Salomon Kabenamualu, Kheir Eddine Farfar, Manuel Prinz, Oliver Karras, Jennifer D'Souza, Lars Vogt, and Sören Auer. 2023. FAIR scientific information with the Open Research Knowledge Graph. *FAIR Connect* 1, 1 (2023), 19–21. doi:10.3233/FC-221513
[117] Sue Stone. 1982. HUMANITIES SCHOLARS: INFORMATION NEEDS AND USES. *Journal of Documentation* 38, 4 (1982), 292–313.
[118] Damon Strange, Megan Gooch, and Alwyn Collinson. 2023. Equality, findability, sustainability: the challenges and rewards of open digital humanities data. *International Journal of Performance Arts and Digital Media* 19, 3 (2023), 348–368. doi:10.1080/14794713.2023.2206286
[119] Carol Tenopir, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame. 2011. Data Sharing by Scientists: Practices and Perceptions. *PLOS ONE* 6, 6 (2011), e21101. doi:10.1371/journal.pone.0021101
[120] Carol Tenopir, Elizabeth D Dalton, Suzie Allard, Mike Frame, Ivanka Pjesivac, Ben Birch, Danielle Pollock, and Kristina Dorsett. 2015. Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLOS ONE* 10, 8 (2015), e0134826. doi:10.1371/journal.pone.0134826
[121] Carol Tenopir, Natalie M Rice, Suzie Allard, Lynn Baird, Josh Borycz, Lisa Christian, Bruce Grant, Robert Olendorf, and Robert J Sandusky. 2020. Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLOS ONE* 15, 3 (2020), e0229003. doi:10.1371/journal.pone.0229003
[122] Alice Thudt, Uta Hinrichs, and Sheelagh Carpendale. 2012. The bohemian bookshelf: supporting serendipitous book discoveries through information visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1461–1470. doi:10.1145/2207676.2208607
[123] Helen R. Tibbo. 2002. Primarily history: historians and the search for primary source materials. *Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries* (2002), 1–10. doi:10.1145/544220.544222
[124] Daniel Tracy and Graham Jensen. 2025. Humanities scholars' needs for open social scholarship platforms as online scholarly information sharing infrastructure. *First Monday* 30, 2 (2025). doi:10.5210/fm.v30i2.13742

[125] René van Horik. 2019. The Research Data Alliance and the Humanities. *Zenodo* (2019). doi:10.5281/zenodo.3355145

[126] Jillian C Wallis, Elizabeth Rolando, and Christine L Borgman. 2013. If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. *PLOS ONE* 8, 7 (2013), e67332. doi:10.1371/journal.pone.0067332

[127] Yu Watanabe, Koichiro Ito, and Shigeki Matsubara. 2024. Capabilities and Challenges of LLMs in Metadata Extraction from Scholarly Papers. In *International Conference on Asian Digital Libraries*. Springer, 280–287. doi:10.1007/978-981-96-0865-2_23

[128] Stephen Wiberley and William Goodrich Jones. 1989. Patterns of Information Seeking in the Humanities. (1989).

[129] Pradeepa Wijetunge. 2019. Information Seeking Behavior of the Humanities and Social Science Undergraduates: A Case of University of Colombo. *Journal of the University Librarians Association of Sri Lanka* 22, 1 (2019), 1–22. doi:10.4038/jula.v22i1.7927

[130] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3 (2016). doi:10.1038/sdata.2016.18

[131] Florian Windhager, Paolo Federico, Günther Schreder, Katrin Glinka, Marian Dörk, Silvia Miksch, and Eva Mayr. 2018. Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges. *IEEE Transactions on Visualization and Computer Graphics* 25, 6 (2018), 2311–2330. doi:10.1109/TVCG.2018.2830759

[132] Elizabeth Yakel. 2005. Hidden collections in archives and libraries. *OCLC Systems & Services: International digital library perspectives* 21, 2 (2005), 95–99. doi:10.1108/10650750510598675

[133] Elizabeth Yakel and Deborah Torres. 2003. AI: Archival Intelligence and User Expertise. *The American Archivist* 66, 1 (2003), 51–78. doi:10.17723/aarc.66.1.q022h85pn51n5800

[134] Ayoung Yoon. 2017. Data reusers' trust development. *Journal of the Association for Information Science and Technology* 68, 4 (2017), 946–956. doi:10.1002/asi.23730