# Data Curation in Interdisciplinary and Highly Collaborative Research

Inna Kouper
Indiana University

## Abstract

This paper provides a systematic analysis of publications that discuss data curation in interdisciplinary and highly collaborative research (IHCR). Using content analysis methodology, it examined 159 publications and identified patterns in definitions of interdisciplinarity, projects' participants and methodologies, and approaches to data curation. The findings suggest that data is a prominent component in interdisciplinarity. In addition to crossing disciplinary and other boundaries, IHCR is defined as curating and integrating heterogeneous data and creating new forms of knowledge from it. Using personal experiences and descriptive approaches, the publications discussed challenges that data curation in IHCR faces, including an increased overhead in coordination and management, lack of consistent metadata practices, and custom infrastructure that makes interoperability across projects, domains, and repositories difficult. The paper concludes with suggestions for future research.

International Journal of Digital Curation
2023, Vol. 17, Iss. 1, 20 pp.

1

http://dx.doi.org/10.2218/ijdc.v17i1.835
DOI: 10.2218/ijdc.v17i1.835.

# Introduction

Interdisciplinarity and collaborations are considered pivotal in addressing societal problems and advancing science. It is often argued that scientific breakthroughs, such as documenting the human genome, finding quantum particles, or understanding the COVID-19 pandemic, were possible because scientists and other professionals worked together in teams, tackling the problems from multiple perspectives and sharing their data (Fry et al., 2020; Powell, 2021; *Nature*, 2021).

The promise of interdisciplinary and highly collaborative research (IHCR), defined here broadly as research that combines resources and expertise across domains and institutions, depends on the ability of teams to effectively organize and share data and other resources. At the same time, such research requires more skills and more trust; it creates more technical and communication challenges, all of which has implications for how data is collected, managed, and preserved (Palmer, 2001). As IHCR teams grow increasingly diverse to include librarians and other professionals, it becomes important to examine who participates in IHCR research data work, understood here as all activities related to data management and curation. More specifically, it is important to understand how IHCR teams address the challenges of working with data.

This paper addresses these broad questions by conducting a systematic analysis of the literature on IHCR data and its curation. The study examined content of the papers that discuss interdisciplinarity and collaboration, focusing specifically on data challenges. It aimed to synthesize what is currently known about the practices of working with data in interdisciplinary research and to identify recommendations for IHCR data curation. Additionally, the paper aimed to identify questions and challenges that need to be addressed in the future research on IHCR data curation.

# Background

Definitions of interdisciplinarity range from selective borrowings to epistemological fusion and synthesis that leads to changes in how we produce knowledge (Klein, 1990; 2018). Grounded in the notion of disciplines, interdisciplinarity refers to a "variety of boundary transgressions" as disciplines define their relationships to each other as integrative, hierarchical, or antagonistic (Barry et al., 2008). Definitions of interdisciplinarity and collaboration are elusive because disciplinary distinctions themselves are a product of specific times and social arrangements that change (Becher, 1981; Stichweh, 1992; Weingart, 2010). Moreover, both the disciplinary knowledge and its boundary-crossing counterparts are evolving toward an increasing heterogeneity of activities, methods, theories, and forms of evidence and their hybridization (Dogan, 1996; Klein, 1996; Knorr Cetina, 1999; Jacobs, 2013).

Overall, the state of knowledge on IHCR continues to be defined by the acknowledged complexity and multiplicity, including the complexity of contexts and structures, skills and expertise, communication and leadership styles, and policies and incentives (Klein, 1985, 2008; O'Rourke, Crowley, & Gonnerman, 2016; Hirsch & Brosius, 2013). Attempts to define collaborations across disciplines and emphasize their relevance to society have even led to the adoption of new terms, such as *convergence* and *team science* (National Research Council, 2015; National Science Foundation (NSF), 2016; Bennet & Gadlin, 2012). Other terms, such as data-intensive research and open science also become part of the IHCR field (Cheruvelil & Soranno, 2018; . The discussions on interdisciplinarity and collaboration call for new forms of knowledge that bring together theory and practice, science and arts, and research and business (Barry et al., 2008; Gibbons, 1999; Moran, 2010; Palmer, 2001; Calhoun, 2017).

Attempts to develop empirically grounded typologies of IHCR have resulted in the creation of typologies that differentiate between what is integrated, how, and why (Huutoniemi et al., 2010). Within this framework, data may fit within each category of interdisciplinarity depending on how interactions around data are organized. A wide range of factors affects the success or failure of interdisciplinary work, including funding, training, institutional support and organizational structures, team dynamics and leadership, problem scoping and complexity, and infrastructure and continuity (Porter et al., 2006). Data sharing and integration are among the prerequisites of successful interactions between domains.

IHCR data management and curation pose additional challenges as data in such projects is collected by individuals and groups with a variety of backgrounds. Such data forms research collections with particular challenges in standardization or alignment of methods, documentation, storage, and sharing of such data (Parsons et al., 2011). Interdisciplinary and highly collaborative data require more work in integration and discovery (Esteva et al., 2020). If data is stored in multiple repositories with separate catalogs and systems, a lot of work goes into aggregating heterogeneous data and metadata through coordination of standards, infrastructure, and other resources (Yarmey & Khalsa, 2014). While some issues are solvable from the technical point of view, the issues of long-term maintenance and governance as well as building trust are harder to address (Norris & Mader, 2019).

Several large initiatives focused on IHCR data management and discussed related challenges (Borgman et al, 2015; Karasti, Baker, & Halkola, 2006; Baron et al., 2017). The International Polar Year, for example, had a designated Data Committee and an ambitious data management plan, and found that building "data infrastructure to enable international sharing and reuse of multidisciplinary datasets is a complex and fraught sociotechnical exercise" (Mokrane & Parsons, 2014). The World Data System proposed a data curation framework that distinguished between curating digital and analog data, while focusing on the archival component of data management and ingest into a repository (Laughton & du Plessis, 2013). Support for interdisciplinarity in this model was addressed mostly through the development of unified standards for data description. The Group on Earth Observations (GEO), a global network of data providers and researchers, developed a system that tried to "better integrate observing systems and share data by connecting existing infrastructures using common standards" (*GEO*, 2021).

Building upon these previous efforts and the broader literature, this paper synthesized approaches from many project-oriented efforts in IHCR data curation. It created a sample that bridges literature on interdisciplinarity and data curation. The study focuses on the role data plays in IHCR and ways data curation efforts can contribute to the development of open, interconnected, and transparent research infrastructures. The findings of the study contribute to a better understanding of the role of data and interdisciplinarity in knowledge production and help to establish better curation and management practices within the increasingly complex informational environments.

# Methodology

This study addressed the following questions:

RQ1. How are interdisciplinarity and collaboration defined in the context of data work?

RQ2. Who is involved in IHCR data practices and how?

RQ3. What approaches and recommendations are provided for IHCR data curation?

The methodology included the following: 1) develop a search protocol and selection criteria, 2) search for publications and compile a sample, 3) develop a coding protocol, 4) analyze and code publications, 5) synthesize results, and 6) report the findings.

The search protocol used the database keyword approach to balance comprehensiveness and time spent compiling the database. The following search statements that cover IHCR data curation broadly were used:

1) (interdisciplinary OR collaborative OR distributed) AND (data curation OR data management OR case study OR data infrastructure OR data communit*) AND (research OR team)
2) interdisciplinary AND collaboration AND (data curation OR data management OR data infrastructure OR data stewardship).

The search was conducted in two rounds. Round 1 involved searching the following databases: Google Scholar, Web of Science Core Collection, and EBSCO Academic Search and Library Literature & Information Science Full Text. Round 2 involved searching specific journals to ensure that the domains of library science and data curation are sufficiently covered. The papers that were retrieved via both searches were screened for relevance to make sure that they 1) directly address interdisciplinary and/or collaborative research activities, 2) directly address data practices or other practices that may affect data practices (e.g., academic library resources, interdisciplinary communication and so on), and 3) describe an approach or system that was used in an interdisciplinary project. Many papers, such as bibliometric papers that measured collaboration or education-oriented papers, were excluded. The number of papers retrieved via search and downloaded after the screening are provided in Table 1 below.

Table 1. Search results and downloads (Zero download in some cells was due to the overlap in search results, i.e., papers have already been downloaded in another search).

| Source | Statement 1 | | Statement 2 | |
|---|---|---|---|---|
| | Search Results | Papers Downloaded | Search Results | Papers Downloaded |
| Google Scholar | 17,800 | 10 | 17,600 | 28 |
| Web of Science | 17,049 | 47 | 1,034 | 22 |
| EBSCO | 2,587 | 8 | 21 | 1 |
| International Journal of Digital Curation (IJDC) | 15 | 3 | 1 | 0 |
| Journal of Information Science (JIS) | 100 | 7 | 5 | 0 |
| Journal of the Association for Information Science and Technology (JASIST) | 237 | 15 | 23 | 0 |
| Big Data and Society (BDaS) | 92 | 4 | 14 | 0 |
| Scientific Data (SD) | 88 | 2 | 64 | 2 |
| Library & Information Science Research (LISR) | 242 | 4 | 22 | 2 |
| Data Science Journal | 96 | 11 | 102 | 5 |
| **Total** | **38,306** | **111** | **18,886** | **60** |

After additional screening for relevance and representation and duplicate removal the resulting dataset consisted of 159 papers. The earliest publication year for the search was set to 2011, because this was the year when many funding agencies started implementing data management plan mandates, which spurred activities around data management and data sharing and, possibly, increased the number of publications (Akers et al., 2014; Cox & Pinfield, 2014; Kowalczyk & Shankar, 2011; Maienschein et al., 2018). The resulting dataset consisted of

159 papers that were relatively evenly distributed across years, with a slight increase in the years 2017, 2018, and 2020 (see Figure 1). The methodology inevitably has limitations as it is based on a set of keywords and a time range. As such it may miss papers from less interdisciplinary or collaborative domains or from prior years. At the same time, it enabled a more focused qualitative analysis.
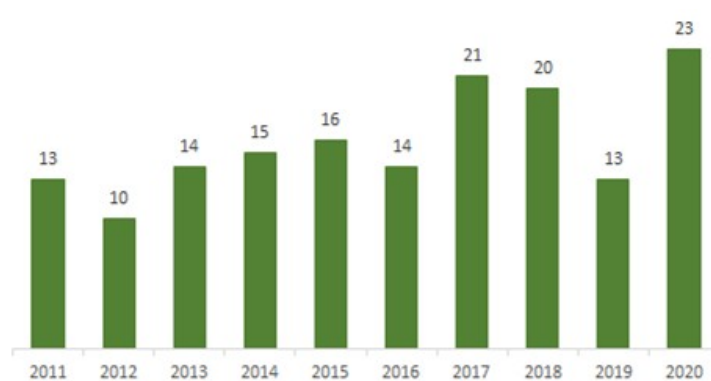


**Figure 1**. Number of papers in the sample by year (N = 159).

After the papers were downloaded and saved, metadata for each of them was extracted into a spreadsheet. The metadata included the names of the first three authors, their affiliations, countries, year of the publication, publication venue, and paper availability. For subsequent data analysis the disciplinary areas mentioned in the papers were standardized into broader categories using a broad classification of the branches of science, namely, the division into formal, natural, social, and applied sciences[1]. Librarianship was included into the applied sciences category.

Each paper was read closely, and summary statements were extracted based on the set of categories developed in advance (see Appendix B). The statements that were extracted and coded included IHCR definitions (explicitly stated or inferred), methods and research subjects / participants, findings and recommendations, conclusions, and whether any specific IHCR projects were discussed. A small sample of papers (10%) was coded by two coders separately; the results were then discussed, and the differences were reconciled. After developing a consensus on how to interpret the categories and statements that belonged to them, the coders proceeded to code papers independently, with occasional spot-checking for mutual agreement.

# Findings

## Defining Interdisciplinarity and Collaboration in Data Work

While many papers in the sample relied on the existing definitions of interdisciplinarity and emphasized crossing the existing boundaries that delineate research, many of them also highlighted the role of data in blurring those boundaries (see Table 2 below).

The majority of the papers (64%) mentioned crossing boundaries as part of the IHCR. "Spanning", "coordinating", "working across", "interacting", "integrating", and "transcending" were the verbs that were used to describe such crossings. Knowledge integration, for example, often referred to working together to not only generate new knowledge, but also to align the existing understandings and approaches through joint discussions and to develop a common language to present results.

---

[1] https://en.wikipedia.org/wiki/Branches_of_science

**Table 2**. Definitions of IHCR (Npapers = 159, each paper could contain more than one definitional statement).

| Definition emphasis | Number of statements | Percent of papers |
|---|---|---|
| Crossing boundaries (disciplines, institutions, sectors, and regions) | 102 | 64% |
| Integrating, sharing, or reusing heterogeneous data, methods, or results | 74 | 47% |
| Understanding complex systems or societal problems and creating new forms of knowledge | 39 | 25% |
| Creating and curating data for multiple uses and building infrastructure to facilitate data lifecycle | 31 | 19% |
| Unifying multiple perspectives toward a common goal | 19 | 12% |
| **Total** | **265** | |

The term "transcending" was used in connection with the term "convergence", where convergence was defined as knowledge work that transcends boundaries, including disciplinary and organizational boundaries:

> 'Drawing on insights from several foundational publications and extending them to our field, we define convergence research as: An approach to knowledge production and action that involves diverse teams working together in novel ways—transcending disciplinary and organizational boundaries—to address vexing social, economic, environmental, and technical challenges in an effort to reduce disaster losses and promote collective well-being' [121][2].

In addition to disciplines and domains, some publications also mentioned working across geographical regions and various institutions and collaborating with non-research stakeholders, including government agencies, industry partners, and citizens. Papers discussed integration of knowledge and collaborations between academia and industry, academia and non-profit or collaborations between researchers and practitioners. Sometimes, integration referred to the researchers sharing their information with practitioners:

> 'Genetics researchers shared material with the growers and oenologists. … In all of these cases, the new knowledge arising from laboratory tests on the properties of the grape samples was integrated into the various technical and commercial perspectives of the growers, and between growers and external experts (oenologists and business managers)' [4].

Another relatively large group of definitions emphasized heterogeneity in data and approaches and their integration, sharing, and reuse (47% of papers). Similar to the previous group of definitional statements, this group used a variety of terms to describe heterogeneity and emphasized *integration* (e.g., the integration of data, information, techniques, taxonomies, theories, and so on), *combination* (e.g., combination of qualitative and quantitative data and perspectives), and *synthesis* (e.g., the synthesis of data, knowledge, or management approaches). The papers argued that it is no longer sufficient to work with one type of data and that IHCR needs methods and techniques to address data heterogeneity:

---

[2] Numbers in square brackets refer to citations in the list of sampled publications provided in the Appendix A.

"Data can originate from many independent sources, each of which may have its own semantics. It is important for interdisciplinary researchers to have a methodology by which information from a large number of sources can be associated, organized, and merged" [56].

Understanding complex systems was a relatively common goal that was identified in about 25% of the papers as part of IHCR. In some cases, the understanding of complex systems was connected to the creation and integration of large-scale heterogeneous data (e.g., [1, 22, 56, 85]). In other cases, complexity was seen as the driver of increasing interdisciplinarity and collaborations. Thus, some papers mentioned the grand challenges approach to science as a way to find new frameworks for advancing scientific research and its governance [e.g., 49, 95, 120]. Many grand challenges are multidisciplinary and their "complexity, ambiguity and uncertainty … require analysis from many different viewpoints and disciplines in ways that are best addressed by inter- or even transdisciplinary research approaches" [96].

Finally, about one fifth of the publications (19%) discussed data curation and infrastructure that supports data throughout its lifecycle as an important component of interdisciplinarity. Generating, curating, processing, and enabling access to data in a systematic way creates opportunities for forming new teams and asking new types of questions. Curated databases were seen as a source of science and scholarship that enables collaborations and interdisciplinary innovations in various domains, such as physics, chemistry, biology, astronomy, earth sciences, and others [22, 28, 99, 106, 125]. Large interdisciplinary research generates large heterogeneous data collections that need additional concerted effort of management and curation [1]. In order to stimulate new forms of interdisciplinary research, such large data collections require significant investments:

'Data collected from the fields of language acquisition and use are multi-lingual, multi-modal, multi-formatted, and derive from multiple methods of data collection (i.e., observational and experimental, cross-sectional or longitudinal). … These features result in an immensely complex set of databases often appearing in diverse formats as different labs generally practice distinct forms of data management. … Language data collections are infinitely expandable and should be used, reused and, when possible, repurposed' [21].

Overall, the publications' definitional statements in the context of data were often broad and inclusive. The underlying assumption that data is diverse and requires diverse methods and approaches allowed authors to accept a wide range of definitions of IHCR, from working side by side on a larger project and contributing individual datasets and methods to using shared data to full knowledge synthesis and data integration that enables new understanding of problems. This broad and accommodating approach supports the diversity of IHCR data and projects, but at the same time it blurs the boundaries between disciplinary and IHCR data and may weaken the argument for IHCR data curation when advocating for resources and the professionalization of the field (Klein, 2018).

## Methods, Areas, and Participants in Interdisciplinary Data Work

To understand who is involved in IHCR data work and how, papers' methods and references to participants, or actors, were coded into standardized categories. Codes reflected the traditional approach to empirical research that begins with a hypothesis or research questions and proceeds through identifying methods of data collection to testing the hypothesis or answering the questions. Even though the methods of empirical social science research have expanded over the years, the common categories that emerged from the coding implied objectivity and distance from the subjects (participants) and included participant observation, interviews, surveys, and documentary analysis (Devine & Heath, 2009). Table 3 below illustrates the frequency distribution of methods used in the papers.

**Table 3**. Methods used in sampled papers.

| Method | Number of papers | Percent of papers |
|---|---|---|
| Personal experience | 61 | 38% |
| Case study | 45 | 28% |
| Surveys and interviews | 23 | 15% |
| Other | 30 | 19% |
| **Total** | **159** | **100%** |

A large portion of the existing understanding of IHCR projects and their data came from personal experience of the authors (38% of the papers). As the name of this methods category suggests, authors of the papers were also study participants in the projects that they described (e.g., [26, 41, 56, 119]). Their backgrounds ranged from academic (students and researchers) to professional (IT, libraries, data management) to leadership (e.g., vice presidents for technology or research). The focus of such papers was often practical and described lessons learned from participating in projects or implementing programs, approaches, or technologies. In particular, the authors often discussed the challenges of data integration and heterogeneity. A typical structure of the methods section of such papers would include a description of the project and its data, challenges faced by the project, and then some details of addressing these challenges with a more detailed description provided in the results or discussion sections.

As IHCR projects could be difficult to access from the outside, summaries of personal experiences provide a valuable contribution to the discussions on interdisciplinarity. Most authors used their research training to report on the complexities of working on interdisciplinary and collaborative projects. Thus, the papers frequently mentioned participatory frameworks or action and community-oriented research even though they did not necessarily elaborate on the steps. A common way to incorporate those frameworks would be to mention it in a sentence, provide a reference, and then proceed to describe the steps of documenting the project through the more common methods of participating in meetings, conducting interviews, and engaging in discussions.

The boundaries between personal experience and other methods were sometimes blurred. Even though some studies mentioned case study as their approach, they were coded as "personal experience" because the authors were involved in the project they studied and referred explicitly to their experiences: "In this paper, we would like to discuss our experience of merging two fields or disciplines of science together… We would like here to present our experience, … to stress on certain points that we believe will help people … " [9].

Case study was used in 28% of the papers as the main method. The papers were coded as case studies when, in addition to using the words "case study", they described their methodology with more details, including selecting the case(s), engaging with participants and artifacts, and analyzing the collected data. Thus, a case study could employ a qualitative approach to examine a larger grant-funded initiative that encouraged interdisciplinarity in the humanities [62] or use mixed methods to study a doctoral program that encouraged interdisciplinarity and, more specifically, data integration [31]. Some researchers used participant observation as part of their case study (e.g., [92]), but contrary to personal experience papers, such an approach was clearly explained and supported with relevant theoretical and methodological citations.

Interviews were often used as part of the case study methodology or even personal experience papers, however, in some papers, it was the primary method. Surveys and interviews as the primary methods of data collection were used in 15% of the sample (23 papers), and these two approaches were split approximately equally. Papers with methodology coded as "other" included reviews, conceptual analysis, network analysis, and technology implementation papers [e.g., 10, 96, 116, 128]. These publications examined the existing data infrastructures and

policies in the context of collaboration and interdisciplinarity, discussed frameworks and barriers of collaboration, and, overall, promoted the role of data and data management in IHCR.

Most methodological approaches described above relied on faculty and researchers and authors themselves as the main participant (actor) groups (49% and 38% of papers respectively). Other participants included students and postdocs, stakeholders, and professionals. The latter including IT professionals (9%), librarians and archivists (6%) and data professionals (5%). The group that was coded as "stakeholders" included actors from private sector, government, and citizen communities (18% of papers). The actors that participated in research also included non-human actors, such as information and technological artifacts that served as objects of study.

In addition to methodologies described in the papers, coding included "modality" of the papers. The definition of modality is loosely based on the linguistic concept of modality that refers to how language is used to discuss the possibility and certainty of situations. Similar to how in natural languages certain parts of speech are used to communicate ability ("can"), obligation ("should" or "must"), or probability ("may / might"), papers' authors described their research ("situations") and its evidential certainty as how things *are*, how things *should be*, or how things *could be*. Based on the overall orientation of the findings and conclusions, we identified primary and secondary modalities and coded papers as **normative** (statements about what *should* be done), **prospective** (statements about what *will be* done), **evidential** (statements about how things *are* based on research and empirical evidence), **descriptive** (statements about how things *are without* empirical research evidence), and **possible** (offering a proposal or a suggestion of how things *can be*). Figure 2 below illustrates the distribution of primary paper modalities across various methodologies.
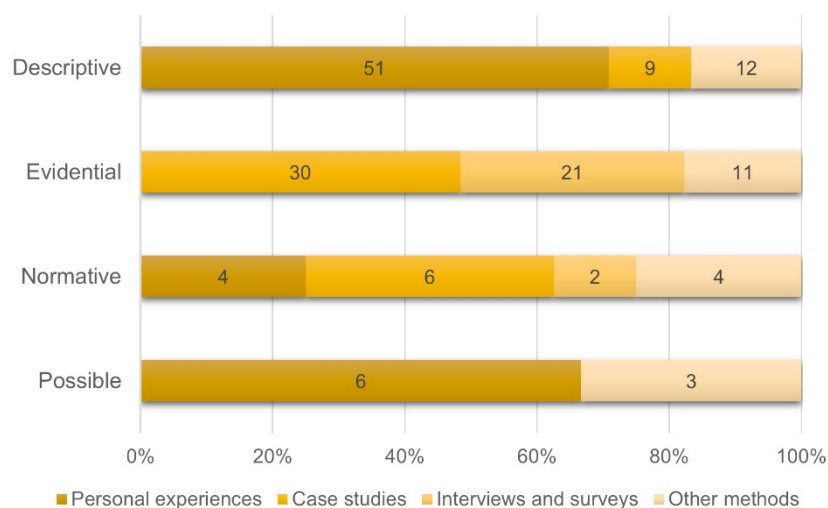


**Figure 2**. Primary modality of papers in the sample (Note: "Prospective" did not appear in the coding as a primary modality).

A large proportion of papers in the sample were descriptive (72 papers, 45% of the sample). In other words, they did not conduct systematic research to collect evidence and support their argumentation. While not necessarily anecdotal, these papers mostly relied on the immediate experience to illustrate their findings and recommendations. Understandably, the majority of descriptive papers used "personal experience" as their methodology (51 out of 72 papers that were coded as descriptive). The second largest category was evidence-based (62 papers, 38% of the sample), and such evidence came primarily from case studies, surveys, and interviews. Eleven percent of the papers used the normative language, making recommendations or suggesting certain courses of action, e.g., "In order to set up this community of practice, integration process needs to be done in a one–two addressing both "institutional" and "scientific" OH dynamics" [16]. Nine

papers that were coded as "possible" modality discussed current and future trends in data and IHCR and how certain technical implementations could improve collaborations:

"The system is still under development, but we believe it represents solid progress towards the goal of making curated database technology available to those who need it the most: namely, scientific database curators and consumers of scientific data, where provenance, annotation, citation, and versioning are key requirements that currently need to be revisited for each new database." [28]

Eight papers had more than one prominent modality, for example, several descriptive papers also made normative or prospective statements. Thus, one paper that used case study as its methodology and reported findings on data curation in interdisciplinary projects, also suggested a change in how institutions approach data and metadata:

"What this article suggests is that to develop and sustain data and metadata curation as self-perpetuating activities, institutional goals and relationships need to change in ways that recognize the multiple facets of data curation as an institutional issue." [6]

To get a sense of the most common areas of interdisciplinary efforts, papers were categorized using the broadest classification of the branches of science (see the Methodology section above). If the publication mentioned areas belonging to a physical science only (e.g., it discussed interdisciplinary work in astronomy or physics or combined oceanography and chemistry), the paper would be coded as physical science. If the paper mentioned areas that belong to physical science and life sciences (e.g., geology and biology), it would be coded as "physical science + life sciences". The order of terms indicates the primacy or prominence of the domain in the publication, i.e., in the earlier example the physical science would appear to be more prominent in the discussions of methods, team composition, or paper results, than the life sciences. Table 4 below shows the frequency distribution of areas of interdisciplinarity:

**Table 4**. Areas of interdisciplinary work.

| Standardized area | Additional area characteristics | Number of papers | Percent of papers |
|---|---|---|---|
| Physical science+ | polar research, physics, climate science, geoscience, genomics, STEM, astrobiology, ecology, material science, health | 52 | 33% |
| Applied sciences | agriculture, manufacturing, big data, informatics, biomedical research, public health, computer science, engineering, libraries | 27 | 17% |
| Data work | management and curation, open data, sharing, visualization, GIS | 25 | 16% |
| Life sciences or life sciences+ | biology, biodiversity, environmental research, ecology, neuroscience, sustainability science, hydrology | 22 | 14% |
| Social sciences or social sciences+ | archaeology, linguistics, anthropology, sociology, economics, communication | 15 | 9% |
| Humanities | digital humanities | 12 | 8% |
| Other | mathematics, public sector, disaster and risk | 6 | 4% |

| | | |
|---|---|---|
| management | | |
| **Total** | **159** | **100%** |

As can be seen from the table above, a large portion of the papers (33%) had physical science as the most prominent component of the interdisciplinary efforts, with more than half of the papers in that group (28) focusing on physical science exclusively and another almost half (24) combining physical science with life sciences, applied sciences, social science, and data work.

Two other relatively large groups (17% and 16% respectively) were coming from the areas of applied sciences and data work. The latter is of particular interest as it demonstrates the importance of attention to data within IHCR research. While all papers in our sample addressed data curation in some form, these papers focused on data as a cross-cutting topic and considered data as an interdisciplinary entity that brings together various disciplines and enables them to collaborate. Most of these papers framed the challenges of data management and curation in IHCR and data openness and sharing as topics that require interdisciplinarity and collaboration. Some papers also discussed new types of data, such as social media data, remote sensing data, or big data as opportunities for interdisciplinary research, particularly, as an opportunity to merge computational sciences with the social, life, or physical sciences (e.g., [115], [140]).

## Approaches to IHCR Data

Data curation was a crucial component of all IHCR efforts described in the sample. Even though not many papers used the word "curation," almost all of them addressed some aspects of curation and management as they discussed metadata, data quality and integrity, stewardship, preservation, and long-term sustainability of interdisciplinary research. The analysis of findings and conclusions of the papers have demonstrated that data curation in IHCR encompasses many aspects that need to be addressed for it to be efficient and successful. These aspects are discussed below in two larger themes: **socio-cultural** and **data-technological**.

The **data-technological theme** in findings and conclusions addressed taking care of interdisciplinary data and building infrastructure to support it. It included challenges of digitization, metadata and documentation, sharing and preservation, quality, and ownership. Thus, regarding metadata, IHCR projects, many of which were still at the earlier stages of maturity, were found to rely on incomplete data descriptions, especially in the context of data re-use. Team members could not use accepted disciplinary metadata standards, so instead of using standardized structures to describe data, they engaged in ad-hoc negotiations and explanations of what was in the data and used informal communication channels to sporadically share metadata with others. Moreover, the implementation of metadata tools and standards did not fit the needs of IHCR:

> "These CENS [the Center for Embedded Networked Sensing] cases illustrate how taking metadata as a formalized representation of data glosses over many nuances of interaction and communication around data and metadata. Formal metadata records that conform to established standards are almost nonexistent in the day-to-day work of CENS researchers, and the different priorities of interdisciplinary collaborators work against the implementation of single-disciplinary standards, such as EML, in communal data systems. … The data management tools intended to facilitate EML implementation proved unusable due to incompatibility with existing local practices and infrastructures." [53]

Proposed solutions to the metadata problems varied across publications. In some publications the recommendations emphasized education and training of researchers on the appropriate metadata standards and their use and facilitated discussions on how to improve the

standards to fit with the IHCR needs [e.g., 1, 83, 114, 151]. Others acknowledged that the approach to metadata currently fits more with libraries and data management rather than the practices of researchers and other stakeholders. As IHCR involves collaborations across institutional and professional boundaries in addition to disciplinary boundaries, understanding that practices are different for government agencies, industry, and public sector was also important.

Fluidity and lack of consistency and accepted norms and practices were common in the discussions of IHCR data practices. Words such as "new", "evolving", "uncertain", and "dynamic" were quite common in the descriptions of IHCR data work. This acknowledgement of both the uncertainty and the diversity of practices and stakeholders in IHCR data led many authors to include in their recommendations a call for studying and consistently mapping the existing practices in interdisciplinary research. Some practices were similar to discipline-oriented research, for example, inconsistent or relatively poor metadata practices, lack of data sharing due to concerns of proper credit, confidentiality, or lack of time, and highly individualized approaches to data storage and management. The practices that were rather unique to IHCR included incorporating heterogeneous data into the analysis, considering the future uses of data as part of the data management planning, and actively negotiating publication venues and authorship norms.

Large datasets and new forms of data, such as social media data, have been seen in the papers as a nexus of interdisciplinary and collaborative activities. These types of data, that can be collected, stored, and maintained separately from any specific group of researchers, support multiple uses and perspectives. The shift toward data-driven exploration and research further separates data collection and curation activities from analysis and points to a need to have trained data managers and curators as a group of professionals who become collaborators and stakeholders in research. Such professionals could improve the interdisciplinary use of big data by not only creating metadata and addressing data quality and preservation, but also by addressing the concerns of privacy, security, and responsible use of data and shifting the practices towards openness and synergy in theories and methodological approaches.

Infrastructure and technologies that support data activities were another very common theme in the sample. At least one fourth of the papers (40) mentioned cyberinfrastructure and technological support in connection to data in interdisciplinary research. The goal of such support was to "transform shared data as a core modality for research and discovery" [119]. In 159 papers analyzed for this study there were 83 distinct projects that ranged from small group projects to large international collaborations, and each of these projects discussed their own technologies and tools in one way or another. Most of them relied on their own custom-built tools for data management and preservation, justifying the development and customization with the lack of necessary functionality. The need to create customized software for each project may be justified in the initial funding applications, especially, when the funding is cyberinfrastructure-oriented, however, it raises questions of sustainability, wider adoption, and interoperability.

Some of the reasons for custom infrastructure were tied to the professional preferences of developers and data managers. For example, some papers argued for the use of semantic technologies, while others insisted on relational databases or other approaches as a better choice [e.g., 87, 125, 155]. In other cases, the choice depended on how the projects were organized and what expertise was included. Thus, with less expertise in software development the projects had to re-use some of the existing software and rely on customized connections between components rather than full development [e.g., 50, 155]. Overall, the use of existing technical solutions was associated with the use of institutional resources and lack of external funding. This was more common to the projects driven by libraries and other non-research stakeholders.

Another data-technological challenge commonly discussed was interoperability, which can be defined as linking and exchange across datasets, databases, or repositories that are used in IHCR projects. Many publications acknowledged the existence of multiple diverse data collection and management efforts and the diversity of their underlying infrastructure. They also acknowledged that integration and interoperability are the next steps in IHCR infrastructure

development (e.g., [69[). At the same, as one paper that focused on global interoperability within and across disciplines stated, interoperability efforts must involve large coordination, international, and even with that goodwill and top-down mandates are not enough:

> "... certainly that scientific data sharing endeavour must be science□driven. This is recognized by all, and it is very important to keep it in mind because it means that technology should obviously be exploited, but to serve the scientific aims and not as a driver. Similarly, responding to top□down injunctions to share data or to use specific methods or support infrastructures is not enough to ensure that the data will be useful and usable. … this requires at some stage international agreement and can involve different projects and organizations which may have their own aims. Such 'standardisation' discussions are also a social process where fostering broad support often requires some coordination by trusted community members." [64]

Another suggested model of interoperable data infrastructure redistributed the effort between researchers, libraries, and data repositories and archives. Such a model engages more professional groups in the efforts of creating, sustaining, and coordinating data infrastructures:

> "In the making of a data infrastructure, the division of tasks between different information service providers needs to be re-negotiated. We present a federal data infrastructure with a layered architecture including a FrontOffice–BackOffice model. This model allows to articulate different roles in the interaction with research communities, the acquisition of expert knowledge, and the provision of data management services." [51]

The **socio-cultural theme** focused primarily on people and relationships. IHCR projects pose an additional challenge in coordination and project management, and many publications reported the initial lack of collective shared understanding and vocabulary that comes from common professional training and socialization that helps with establishing common data practices. In addition to coordinating their methodological or theoretical heterogeneity, the teams, especially those that collaborate across geographic boundaries, had to coordinate their schedules, communication preferences, and working styles. Sometimes the decision was to maintain the division of labor and allow each researcher to do their work separately, i.e., collect, process, and analyze data, to be later coordinated by the top manager or principal investigator. Others developed strategies to overcome the barriers of coordination:

> "A lack of mutual knowledge hindered the distributed team's ability to coordinate the work process. To compensate, they devised strategies to increase awareness. First, they relied on the participation of the full team in documenting fixes in the SPR database. Second, they developed informal rules which created a searchable knowledge base specifically for this project. Finally, they modified their e-mailing behavior, using broadcast messages to increase their team members' knowledge of their work status." [97]

IHCR projects discussed in the publications varied significantly in how they viewed and assigned data-related roles and responsibilities, but many publications indicated that strong relationships promoted strong data curation and improved the quality of research data. Thus, researchers could be considered data producers or users depending on whether they used their own or others' data. They also could be considered curators when they contributed to creating or improving community- or government-generated data. Similarly, other team members, including data managers and computational scientists, contributed to data production, curation, and dissemination. Many of them had more than one role throughout the project, but the

findings from the papers also indicated that varying roles created conflicting expectations regarding data curation. In such cases it was particularly important to establish a professional culture of shared ownership of the research so that everyone contributes to defining the roles and expectations:

> "Scientists with different curation roles, given common curation tasks, lack a consensus for selecting data quality criteria for genome data curation. Scientists' data quality expectations change as their work roles are pluralistic and evolving, and the curators must strive to keep up with newer or emerging skills. Identification of these differences can help develop data management architectures to support role-based community curation." [78]

Several other strategies that have been described as successful in overcoming barriers to coordination and interdisciplinary collaboration included: 1) establishing a shared ethos of research and openly discussing team members values and commitments with emphasis on respect and empathy; 2) using visual communication and other information organization techniques to capture the nuances of heterogeneous data work, 3) discussing data work and activities and each individual's roles and contributions to it, 4) providing incentives for data curation and sharing within the team, and 5) creating opportunities for sharing tacit knowledge either by physically co-locating team members or by creating virtual spaces for spontaneous interactions.

The theme of people and relationships also included such complexities of doing research in interdisciplinary and collaborative teams as lack of sufficient funding and other resources, the need to interact with a broader range of stakeholders, and the need for new or innovative approaches to training and workforce. Many publications concluded that IHCR and its data requires specialized funding frameworks. Thus, long-term data curation is often not budgeted in IHCR projects, even though the data that is collected from such projects can be highly valuable in future studies. The overhead of working with IHCR data is often underestimated due to its complexity, especially when knowledge practices shift towards co-production, or collaboration between researchers and other stakeholders:

> "Because coproduction of knowledge takes time and resources to do well and is a process that is not well understood there are currently a limited numbers of scientists who undertake it …, contributing to a gap between the number of people producing usable climate science and the demand from users for that information…" [112]

Interactions with a broader range of stakeholders in data curation help to improve processes, share resources, and engage in intellectual exchanges. This seemed more common to collaborations that involved libraries as their efforts tend to have less external funding and more consortium-based and collective bargaining efforts. As the libraries recognize the benefit of joint use of resources and expertise, they strived to engage in discussions across institutions about their role and value in data curation:

> "Like the pieces of a puzzle, one department may have the data, and another have the resources to curate it; put them together and a project impossible for one becomes achievable with the collaborative partnership of both. … The responsibility for data curation can fall to individuals within a department or institution, but librarians are a more logical choice. … Valuable content is often lost simply because no one is designated the responsibility of perpetuating it, or those who are designated the responsibility lack key resources that would allow them to be successful stewards of their data. University libraries can bring those resources to the table, especially their expertise as caretakers of information, allowing them to serve as beneficial partners and facilitators in collaborative efforts." [101]

IHCR projects have also demonstrated a diversity of ways librarians and data managers engage in the data curation practices. Even though often the engagement did not happen until the data needed a home as part of the publication process or complying with open data requirements, publications that described researcher-library collaborations called for an engagement with researchers as early as possible.

# Discussion

This study confirms and reinforces the previous findings that IHCR data work takes many forms and requires significant additional effort. Such effort involves building both infrastructure and relationships and finding ways to coordinate not only across disciplines, but also across institutional, geographic, and other boundaries. The diversity of IHCR projects is reflected in the diversity of data work, which is heterogeneous in many aspects, including definitions, methods, participants, and curation approaches.

Not surprisingly, the discussions of how interdisciplinary science helps to address societal challenges often focus on disciplines and sharing data across them rather than on broader contexts of working across institutions, professions, and stakeholders. Disciplines continue to offer stable career identities; they do not require scholars and researchers to quickly re-define themselves in light of newer problems, data, and tools. The stability of disciplinary perspectives and frameworks and their permeating socio-cultural structures remains a significant barrier to IHCR (Abbott, 2002; Klein, 2018):

> "A long historical process has thus resulted in a more or less steady, institutionalized structure in American academia: a social structure of flexibly stable disciplines, to which is attached an extremely complex and loose cultural structure of disciplines, the whole permeated by a perpetual hazy buzz of interdisciplinarity." (p. 215)

Personal experience and case studies were the primary methods of gathering evidence to understand and advance IHCR data curation. Predominance of these methods in the sample and a large number of descriptive studies indicates that this area is still at an early stage of accumulating evidence and information. While the number of IHCR projects and initiatives continues to proliferate, their overall nature and the nature of their data is so diverse that it evades comparisons and identification of stable patterns. Moreover, as data is part of complex assemblages with multiple components, charting those components and their interactions at varying levels is a challenging task that may require different, e.g., critical or sense-making, paradigms (Kitchin & Lauriault, 2014; Poirier & Costelloe-Kuehn, 2019).

The reliance on customized infrastructure reported in the papers can be seen as another significant barrier in IHCR data curation. If every IHCR project creates its own environment that stores, processes, and provides access to data, the data remains isolated. The incentives to build rather than re-use and adapt technologies and open interoperable standards contributes to the increasing heterogeneity rather than convergence of data and tools. Moreover, it promotes inequity in data practices as under-resourced researchers and data managers will miss out on data analytics and management or use the so-called shadow IT, i.e., disparate independent technologies that suit their needs (Newell et al., 2007). With some exceptions such as the European Open Science Cloud[3] or the Dataverse Project[4], a range of standardized open source or easily available platforms and tools is still needed to support IHCR data curation.

Focusing on a wide range of disciplines with a noticeable preference for physical sciences, IHCR data-related papers discussed an expanded range of professions and spheres of expertise that go beyond researchers or scientists. At the same time the professions of data managers, librarians, and IT remained less visible in the literature. As the majority of papers focused on

---

[3] https://eosc-portal.eu/
[4] https://dataverse.org/

faculty and researchers, the role of other professionals and stakeholders and their contributions to IHCR was studied and acknowledged less. Large interdisciplinary efforts and new organizations, such as synthesis centers, were among those who tried to formalize the roles of experts with different backgrounds and acknowledge data managers as active participants in research:

> "Synthesis-center data-management specialists help with working groups before, during, and after meetings to acquire and organize data, compile databases and models, and offer the opportunity to make the most out of the data with which they work. Synthesis-center staff members also assist in the publication of the synthesized data, thereby continuing the cycle" [13].

Due to the early stages of research on IHCR data curation, approaches to data work and recommendations from the analyzed papers were rather high-level. Thus, IHCR data curation needs improved and more consistent approaches to metadata; it needs to pay attention to the emerging forms of data, such as big data or social media data. IHCR data work also creates a larger overhead in management, coordination, and interaction. It requires a more nuanced understanding of the dynamic roles and responsibilities of data producers, managers, and users and diverse stable systems of recognition and reward for the data work. At the same time the specific of how these recommendations can be implemented were rather scarce in the literature.

The recommendations for IHCR data curation in the sample went beyond curation in a narrower sense, i.e., beyond the issues of metadata, data quality, or preservation. They included requests for more funding and more technical expertise, suggestions to spend time to build trust among team members, and emphasis on data work as the foundation of research. They also discussed the need for more or better standards and training that could improve interdisciplinary data work. Most importantly, some papers called for an increased involvement of librarians and data professionals in IHCR as a path toward addressing many challenges of interdisciplinary data.

There are still many unknowns in IHCR data practices and an attempt to map them is only the first step in understanding and improving IHCR data curation. A larger research agenda could expand the documentation of various projects and include studies that compare and evaluate the differences in disciplinary and IHCR data practices and clarify and test the assumptions and hypotheses that are being put forward as the result of case studies and project reports. The topics and questions outlined below can form such a larger research agenda on IHCR data practices:

- **Drivers of interdisciplinarity.** Surveys of IHCR initiatives demonstrate an increasing push for interdisciplinarity both from the top, i.e., funding agencies and university administrations, and from the bottom, i.e., from faculty (Jacobs & Frickel, 2009). Most of this push relies on a set of commonly accepted assumptions that IHCR and its data are better than disciplinary research. While such assumptions can be a good motivator, it is important to further examine what motivates individual researchers to engage in more resource-intensive interdisciplinary initiatives and how the outcomes of those initiatives, including data quality and availability, compare to those from disciplinary research.

- **Failures of IHCR data.** The literature that describes IHCR projects prioritizes reporting successes rather than failures. While some papers reported lessons learned, challenges, and suggestions for improvement, overall, there is a positive publication bias in the literature. For IHCR data to become the foundation of new knowledge and ways to address societal problems, more information is needed about negative outcomes and failures of IHCR data. Acquiring, cleaning, and organizing data are typically the most labor-intensive aspects of IHCR projects, they need to be reported more objectively and with their own measurement and evaluation frameworks.

- **Interdisciplinarity and collaboration structures.** IHCR is supported through a variety of academic and administrative structures, including the synthesis centers mentioned above, research centers and labs that exist within universities, and externally funded projects and programs. Which structures are the most effective in supporting IHCR and generating new forms data and outputs of high quality? What forms of IHCR organization are the most persistent and beneficial to a broader range of stakeholders, including researchers, data and IT professionals, students, and others? What forms ensure equity in the division of labor as well as data access and long-term availability?

- **Sharing of data and knowledge.** The papers in the sample argued that shared and open data are one of the cornerstones of IHCR. How is sharing of IHCR data different from sharing and openness that are being promoted in many individual disciplines? Is data repository an effective form of sharing for IHCR? How can openness be promoted and supported across disciplinary and institutional boundaries? The IHCR literature was not clear or specific in addressing these questions.

- **Professional training and career paths.** This last item is somewhat connected to all other items discussed above. The expanding range of experts that contribute to research and knowledge production, especially, to its interdisciplinary and collaborative forms, has been already well documented. More details are needed on how such an expansion contributes to IHCR data and knowledge outcomes, what data-related skills various experts need, and how data curation labor can be divided fairly and effectively and what fairness and effectiveness in data curation look like (Shankar et al., 2020).

# Conclusion

This paper examined approaches to data curation in publications that focus on interdisciplinary and highly collaborative research. While curation was not the most prominent term in data-related publications, discussions about how to take care of data was a prominent component in interdisciplinary research and curating and integrating heterogeneous data were included into the definitions of interdisciplinarity. The research on IHCR data curation is at its early stages, it accumulates evidence via personal experiences and descriptive studies. Consequently, the recommendations on how to curate IHCR data are sometimes not specific enough to be adopted into practice. Shifting from individual project reporting to larger data collection efforts and incorporating comparative, critical, and other methodologies into this research domain can help broaden the research agenda and generate deeper, more generalizable knowledge.

# Acknowledgements

# References

Abbott, A. (2002). The disciplines and the future. In S. G. Brint (Ed.), *The Future of the City of Intellect: The Changing American University*. Stanford University Press.

Akers, K. G., Sferdean, F. C., Nicholls, N. H., & Green, J. A. (2014). Building support for research data management: Biographies of eight research universities. *International Journal of Digital Curation*, *9*(2), 171–191. https://doi.org/10.2218/ijdc.v9i2.327

Baron, J. S., Specht, A., Garnier, E., Bishop, P., Campbell, C. A., Davis, F. W., Fady, B., Field, D., Gross, L. J., Guru, S. M., Halpern, B. S., Hampton, S. E., Leavitt, P. R., Meagher, T. R., Ometto, J., Parker, J. N., Price, R., Rawson, C. H., Rodrigo, A., Sheble, L., A., & Winter, M. (2017). Synthesis centers as critical research infrastructure. *BioScience*, *67*(8), 750–759. https://doi.org/10.1093/biosci/bix053

Barry, A., Born, G., & Weszkalnys, G. (2008). Logics of interdisciplinarity. *Economy and Society*, *37*(1), 20–49.

Becher, T. (1981). Towards a definition of disciplinary cultures. *Studies in Higher Education*, *6*(2), 109–122. https://doi.org /10.1080/03075078112331379362

Bennett, L.M., and Gadlin, H. (2012). Collaboration and team science: From theory to practice. Journal of Investigative Medicine, 60(5):768–775.

Borgman, C. L., Darch, P. T., Sands, A. E., Pasquetto, I. V., Golshan, M. S., Wallis, J. C., & Traweek, S. (2015). Knowledge infrastructures in science: Data, diversity, and digital libraries. *International Journal on Digital Libraries*, *16*(3), 207–227. https://doi.org/10.1007/s00799-015-0157-z

Calhoun, C. (2017). Interdisciplinarity in the social sciences: Area studies, quantitative methods, and problem-oriented research. In R. Frodeman, J. T. Klein, & R. Pacheco (Eds.), The Oxford handbook of interdisciplinarity (pp. 117- 130). Oxford: Oxford University Press.

Cheruvelil, K. S., & Soranno, P. A. (2018). Data-Intensive Ecological Research Is Catalyzed by Open Science and Team Science. *BioScience*, *68*(10), 813–822. https://doi.org/10.1093/biosci/biy097

Cox, A. M., & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*, *46*(4), 299–316. https://doi.org/10.1177/0961000613492542

Devine, F., & Heath, S. (2009). *Doing Social Science: Evidence and Methods in Empirical Research*. Macmillan International Higher Education.

Dogan, M. (1996). The hybridization of social science knowledge. *Library Trends*, *45*(2), 296–314.

Esteva, M., Xu, W., Simone, N., Gupta, A., & Jah, M. (2020). Modeling data curation to scientific inquiry: A case study for multimodal data integration. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*, 235–242. https://doi.org/10.1145/3383583.3398539

Fry, C. V., Cai, X., Zhang, Y., & Wagner, C. S. (2020). Consolidation in a crisis: Patterns of international collaboration in early COVID-19 research. *PLOS ONE*, *15*(7), e0236307. https://doi.org/10.1371/journal.pone.0236307

GEO (2021) *'About us.'* https://www.earthobservations.org/geo_community.php

Gibbons, M. (1999). Science's new social contract with society. *Nature*, *402*(6761), 11–18. https://doi.org/10.1038/35011576

Huutoniemi, K., Klein, J. T., Bruun, H., & Hukkinen, J. (2010). Analyzing interdisciplinarity: Typology and indicators. *Research Policy*, *39*(1), 79–88. https://doi.org/10.1016/j.respol.2009.09.011

Jacobs, J. A. (2013). In defense of disciplines. Chicago: University of Chicago Press.

Jacobs, J. A., & Frickel, S. (2009). Interdisciplinarity: A Critical Assessment. *Annual Review of Sociology*, *35*, 43–65.

Hirsch, P. D., & Brosius, J. P. (2013). Navigating complex trade-offs in conservation and development: An integrative framework. *Issues in Interdisciplinary Studies, 31*, 99-122.

Karasti, H., Baker, K. S., & Halkola, E. (2006). Enriching the Notion of Data Curation in E-Science: Data Managing and Information Infrastructuring in the Long Term Ecological Research (LTER) Network. *Computer Supported Cooperative Work*, *15*(4), 321–358. https://doi.org/10.1007/s10606-006-9023-2

Kitchin, R., & Lauriault, T. (2014). *Towards critical data studies: Charting and unpacking data assemblages and their work*.

Klein, J. T. (1985). The Evolution of a Body of Knowledge: Interdisciplinary Problem-Focused Research. *Science Communication*, *7*(2), 117–142. https://doi.org/10.1177/016402595007002002

Klein, J. T. (1990). *Interdisciplinarity: History, Theory, and Practice*. Wayne State University Press.

Klein, J. T. (1996). *Crossing boundaries: Knowledge, disciplinarities, and interdisciplinarities*. University Press of Virginia.

Klein, J. T. (2008). *Evaluation of Interdisciplinary and Transdisciplinary Research. A Literature Review*. *35*(2). https://doi.org/10.1016/j.amepre.2008.05.010

Klein, J. T. (2018). "Advancing" Interdisciplinary Studies: The Boundary Work of Integrating, Complexifying, and Professionalizing. *Issues in Interdisciplinary Studies*, *36*(2), 45–67.

Knorr Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Harvard University Press.

Kowalczyk, S., & Shankar, K. (2011). Data sharing in the sciences. *Annual Review of Information Science and Technology*, *45*(1), 247–294.

Larivière, V., Pontille, D., & Sugimoto, C. R. (2021). Investigating the division of scientific labor using the Contributor Roles Taxonomy (CRediT). *Quantitative Science Studies*, *2*(1), 111–128. https://doi.org/10.1162/qss_a_00097

Laughton, P., & du Plessis, T. (2013). Data curation in the World Data System: Proposed framework. *Data Science Journal*, *12*(8), 56–70.

Maienschein, K., Parker, J., Laubichler, M., & Hackett, E. (2018). Data management and data sharing in science and technology studies. *Science, Technology, & Human Values*, *44*, 016224391879890. https://doi.org/10.1177/0162243918798906

Mokrane, M., & Parsons, M. A. (2014). Learning from the International Polar Year to Build the Future of Polar Data Management. *Data Science Journal*, *13*. https://doi.org/10.2481/dsj.IFPDA-15

Moran, J. (2010). *Interdisciplinarity*. Routledge. https://1lib.us/book/734194/41a8ef

National Research Council. (2015). *Enhancing the Effectiveness of Team Science* (N. J. Cooke & M. L. Hilton, Eds.). NAP. https://doi.org/10.17226/19007

National Science Foundation (NSF). (2016). *What is Convergence? Convergence Research at NSF*. https://www.nsf.gov/od/oia/convergence/index.jsp

Nature (2021) 'Research collaborations bring big rewards: The world needs more' (Editorial). (*Nature*, *594*(7863), 301–302. https://doi.org/10.1038/d41586-021-01581-z

Newell, S., Wagner, E. L., & David, G. (2007). Clumsy information systems: A critical review of enterprise systems. In K. C. Desouza (Ed.), *Agile Information Systems: Conceptualization, Construction, and Management*. Routledge.

Norris, T. B., & Mader, C. C. (2019). Data Curation for Big Interdisciplinary Science: The Pulley Ridge Experience. *Journal of eScience Librarianship*, *8*(2), Article 2. https://doi.org/10.7191/jeslib.2019.1172

O'Rourke, M., Crowley, S., & Gonnerman, C. (2016). On the nature of cross-disciplinary integration: A philosophical framework. *Studies in History and Philosophy of Biological and Biomedical Science, 56*, 62-70.

Palmer, C. L. (2001). *Work at the boundaries of science: Information and the interdisciplinary research process*. Springer Science & Business Media.

Parsons, M. A., Godøy, Ø., LeDrew, E., de Bruin, T. F., Danis, B., Tomlinson, S., & Carlson, D. (2011). A conceptual framework for managing very diverse data for complex, interdisciplinary science. *Journal of Information Science*, *37*(6), 555–569. https://doi.org/10.1177/0165551511412705

Poirier, L., & Costelloe-Kuehn, B. (2019). Data sharing at scale: A heuristic for affirming data cultures. *Data Science Journal*, *18*(1), 48. https://doi.org/10.5334/dsj-2019-048

Porter, A. L., Roessner, J. D., Cohen, A. S., & Perreault, M. (2006). Interdisciplinary research: Meaning, metrics and nurture. *Research Evaluation*, *15*(3), 187–195. https://doi.org/10.3152/147154406781775841

Powell, K. (2021). The broken promise that undermines human genome research. *Nature*, *590*(7845), 198–201. https://doi.org/10.1038/d41586-021-00331-5

Shankar, K., Jeng, W., Thomer, A., Weber, N., & Yoon, A. (2020). Data curation as collective action during COVID-19. *Journal of the Association for Information Science and Technology*, *72*(3), 280–284. https://doi.org/10.1002/asi.24406

Stichweh, R. (1992). The Sociology of Scientific Disciplines: On the Genesis and Stability of the Disciplinary Structure of Modern Science. *Science in Context*, *5*(1), 3–15. https://doi.org/10.1017/S0269889700001071

Yarmey, L., & Khalsa, S. J. (2014). Building on the International Polar Year: Discovering Interdisciplinary Data Through Federated Search. *Data Science Journal*, *13*. https://doi.org/10.2481/dsj.IFODA-13

Weingart, P. (2010). A short history of knowledge formations. In R. Frodeman, J. T. Klein, & C. Mitcham (Eds.), The Oxford handbook of interdisciplinarity (pp. 3-14). Oxford: Oxford University Press.