Received: 13 May 2024 | Revised: 29 August 2024 | Accepted: 26 September 2024

DOI: 10.1002/asi.24957

RESEARCH ARTICLE

JASIST WILEY

When data sharing is an answer and when (often) it is not: Acknowledging data-driven, non-data, and data-decentered cultures

Isto Huvila[1] | **Luanne S. Sinnamon**[2]

[1]Department of ALM, Uppsala Universitet, Uppsala, Sweden

[2]School of Information, The University of British Columbia, Vancouver, British Columbia, Canada

Correspondence
Isto Huvila, Department of ALM, Uppsala Universitet, Uppsala, Sweden.
Email: isto.huvila@abm.uu.se

Funding information
Social Sciences and Humanities Research Council of Canada, Grant/Award Number: Supporting Transparent & Open Research Engagement; H2020 European Research Council, Grant/Award Number: 818210

Abstract

Contemporary research and innovation policies and advocates of data-intensive research paradigms continue to urge increased sharing of research data. Such paradigms are underpinned by a pro-data, normative data culture that has become dominant in the contemporary discourse. Earlier research on research data sharing has directed little attention to its alternatives as more than a deficit. The present study aims to provide insights into researchers' perspectives, rationales and practices of (non-)sharing of research data in relation to their research practices. We address two research questions, (RQ1) what underpinning patterns can be identified in researchers' (non-)sharing of research data, and (RQ2) how are attitudes and data-sharing linked to researchers' general practices of conducting their research. We identify and describe *data-decentered culture* and *non-data culture* as alternatives and parallels to the *data-driven culture*, and describe researchers de-inscriptions of how they resist and appropriate predominant notions of data in their data practices by problematizing the notion of data, asserting exceptions to the general case of data sharing, and resisting or opting out from data sharing.

1 | INTRODUCTION

Contemporary research and innovation policies and advocates of data-intensive research paradigms urge increased sharing of research data. Studies from a variety of scholarly and scientific domains have demonstrated the large variety of how and to what degree researchers share and reuse data. Uneven sharing and standardization of research data make data-intensive research difficult and labor-intensive. At the same time, critics have raised concerns that thinking in terms of data and data-sharing runs counter to the rationales of fields where research and knowledge production do not follow the logic of datafication. In contrast to the large number of studies of the prevalence, enablers and barriers to data-sharing, there is little cross-disciplinary research on how researchers' sharing and non-sharing of data links to their general practices of conducting and sharing research with other researchers and non-scholarly communities. Rather than being treated as integral to the broader ecology of scholarly communication, data sharing is investigated as a separate task. This may reflect a workflow in which data sharing is mediated through some kind of public infrastructure or service, which then sets this apart from other, more direct, types of informal and formal scholarly sharing.

The aim of this study is to provide insights into researchers' perspectives, rationales, and practices of (non-)sharing of research data in relation to their research practices. We draw upon the notion of "data cultures" as a sensitizing concept (Charmaz, 2003), and consider ways in which such cultures form and influence research practices, including forms of resistance. We address two research questions, (RQ1) what underpinning patterns can be identified in researchers' (non-)sharing of research data, and (RQ2) how are attitudes and data-sharing linked to researchers' general practices of conducting their research. We report findings from an interview study ($N = 16$) of senior researchers in social science and science disciplines at a major Canadian research university. This paper extends findings from an earlier paper based on these interviews, which focused on sharing of research methods information (Huvila & Sinnamon, 2022).

2 | LITERATURE REVIEW

2.1 | General practices of scholarly communication and sharing

A growing body of research investigates scholarly information sharing and research data sharing. When asked, the perceived value of openness and data sharing tends to be high (Tenopir et al., 2011; Tu & Shen, 2023) but the extent of actual sharing varies (Pampel & Dallmeier-Tiessen, 2014). A typical premise of studies investigating information and data sharing is to consider such practices beneficial to scholarship and society at large, tied to notions of transparency, accountability and research impact (cf., Kalkman et al., 2019; Kraus & Eberhard, 2022; Oliver et al., 2023b; Pilerot, 2012). As in information sharing literature in general (Pilerot, 2012), shared scholarly information and data comprises a diversity of tangibles and intangibles (Chung et al., 2016) and is facilitated by a plethora of different types of tools and infrastructures (Given & Willson, 2018). The specifics of what constitutes research data varies between studies (Gomez-Diaz & Recio, 2022), as do the definitions of data sharing that range from simply making "data" available to the full set of data curation activities, including deposition, preservation, and reuse (Kurata et al., 2022).

In contrast to the assumption that making information and data available is enough to ensure its use (Davies & Edwards, 2012), sharing involves complex negotiations and transformations of the shared information and data (Tabak & Willson, 2012; see also Huvila, 2022a, 2022b). Drawing on Latour (1999), Tabak and Willson (2012) emphasize the reciprocity and mutual shaping of information sharing and its context. Numerous studies have underlined the importance of disciplinary differences and structural factors that influence information and data sharing. Kurata et al. (2022) identified 14 categories of positions on research data and sharing among researchers, that reflect variations in stance, practices and extent of data sharing and reuse. These include views that endorse data sharing for the public good, unwillingness to share raw data, and skepticism regarding the feasibility of interpreting raw data. A major contributing factor to the complexity of views is that, besides practical and altruistic motives, journal, government, and funder policies form a major motivating factor for the interest to share data (Nugroho et al., 2015; Thelwall & Kousha, 2017). The presence of internal and external incentives and peremptory mandates makes it difficult to determine what, specifically, drives data-sharing and how the different factors influence each other.

Discussions on sharing and open research data have been criticized as reductionist, with calls made to develop data management models that are sensitive to the particularities of diverse research approaches and data types (Kraus & Eberhard, 2022). Leonelli argues that the "conception of openness as sharing is flawed" (Leonelli, 2023, p. 43) and advocates for an alternative notion of openness based on inclusion, datafication and research data sharing (Carroll et al., 2020; Oliver et al., 2023b). Researchers in humanities (e.g., Andorfer, 2015; Funari, 2014), including ethnographers (Kraus & Eberhard, 2022), and literary and cultural scholars (Drucker, 2011) but also those conducting non-empirical research in sciences (cf., Peels & Bouter, 2018) have referred to the difficulty of adapting certain research and data practices and ethical commitments to formalistic ideas of sharing data and have endeavored to develop models that are more sensitive to their needs.

2.2 | Knowledge cultures

Data sharing and scholarly communication practices do not exist in isolation, but are shaped by multiple cultural strata, which, from the perspective of an individual researcher, may or may not be aligned. These include knowledge cultures (Cetina, 2007) or paradigms (Arditi, 1994), disciplinary cultures (Becher & Trowler, 2001; Fry, 2006), and epistemic cultures (Cetina, 2007; Knorr-Cetina, 2003). Knowledge cultures exist at the societal level, establishing structures, policies, and discourses that encourage or discourage certain knowledge practices and outcomes (Arditi, 1994; Cetina, 2007). The rise of "open" movements, including open science and open data is a feature of contemporary knowledge culture. At the micro level, epistemic cultures are situated within localized knowledge settings, such as labs or research groups, and arise from: "the whole sets of arrangements, processes and principles that serve knowledge and unfold with its articulation" (Cetina, 2007, pp. 361–362). Epistemic cultures are scaffolded by societal level discourses, but are fundamentally places of situated action, and reflect the diversity and specialization of scholarship as it is practiced. Explorations of epistemic culture include relationships among human and non-human (e.g., instruments, data, repositories) actors and (mis)alignments between practices and knowledge claims (Latour & Woolgar, 1986), both of which are highly relevant in considerations of data sharing.

Disciplinary knowledge cultures are situated in the middle ground, strongly associated with epistemic cultures, and shaped by disciplinary traditions, histories, social structures (Becher & Trowler, 2001; Chung et al., 2016) and ontological and epistemological alignments (Fry, 2006; Kurata et al., 2022; Talja, 2002). Information and data sharing practices differ in term of what is shared, how and with whom, specific disciplines and across the wider academic fields of science, health, social science and humanities (Fry, 2006; Fry & Talja, 2007; Talja et al., 2022). Factors that influence this variation include differences in practices and means of sharing, a culture of sharing or not sharing, the perceived utility of and incentives to share, and access to supportive tools and infrastructures (Dallmeier-Tiessen et al., 2014; Kim & Stanton, 2015; Niu & Hedstrom, 2008).

2.3 | Data cultures and datafication

Data sharing practices are embedded within data cultures. While there is no widely accepted definition of scholarly data cultures, most reference the context, including features and attitudes, in which a range of data-related practices (e.g., production, curation, use, sharing) occur in the process of conducting research and producing knowledge (Aragona & Zindato, 2016; Oliver et al., 2023a, 2023b; Thessen & Patterson, 2011). Data cultures arise within particular domains and disciplines and reflect diverse conceptualizations of data and data sharing (Kurata et al., 2022). Similarly, as Burgess et al. (2022) note on everyday data cultures, both academic and non-academic data cultures are linked to broad knowledge cultures. Oliver et al. conducted a stratified analysis of research on data culture, identifying that researchers' practices and customs in relation to data are only one layer within a complex social, technical and cultural array (Oliver et al., 2023b).

While data cultures are understood to be diverse and variable, the terms "data culture" or "data-driven culture" are sometimes used to denote contexts in which data are particularly valued as the basis for evidence-based decision making, transparency, and replicability (Oliver et al., 2023b). Such normative, pro-data values align well with certain academic fields, notably the hard sciences. For example, disciplines such as molecular biology (Thessen & Patterson, 2011), astronomy (Wynholds et al., 2011), biomedicine, and earth science (Pampel & Dallmeier-Tiessen, 2014) are observed to have strong intra-disciplinary data sharing cultures (Oliver et al., 2023b). However, data-centrism may sit uncomfortably in other disciplinary and epistemic cultures. Notably, Indigenous data sovereignty problematizes the notion of data and its sharing and commodification (Kraus & Eberhard, 2022; Oliver et al., 2023b; Pels et al., 2018). Nevertheless. advocacy for data-centric cultures is now a society-wide priority, playing out within organizations (Storm & Borgman, 2020), governments (Lian et al., 2023), research agencies, and institutions (Oliver et al., 2023b). Within information studies, the research data management subfield has aligned itself with these goals and is vigorously engaged in enacting data-driven cultures (Oliver et al., 2023a). Motivations for this work are explicitly linked to an increase in quality and impact of research (Borghi & Van Gulick, 2022; Donner, 2023); however, as Arditi (1994) notes, such dominant discourses may also change the nature of research and how it is carried out, and therefore should be understood in political and moral terms. The broader agenda and implications of data-centrism are also evidenced in the rise of data science, which aims to establish "integrated processes that turn data into insight" (Özsu, 2023).

2.4 | Prior research on data sharing

Within information studies, considerable research has been devoted to understanding why and how data sharing occurs, including examination of researchers' practices and infrastructures (Borgman et al., 2019; Oliver et al., 2023b). A more limited body of work focuses on the use of shared research data (Zuiderwijk et al., 2020). Surprisingly, definitions of data sharing are rare in prior work, creating a highly inclusive, but under-specified research target (Dutoit, 2017; Thoegersen & Borlund, 2022, p. 5). The principal rationales for data sharing are expressed in economic terms of public benefit and return on investment and in research terms of accelerating and improving outcomes, accountability and credibility (Oliver et al., 2023a). Despite broad consensus on these high level motivations, uptake of data sharing as a core scientific activity has been slow and uneven (e.g., Gomes et al., 2022).

Kowalczyk and Shankar (2011) categorize data sharing challenges into the practical, "how-to-do-it" issues and broader concerns relating to the nature of research and access to scientific results. Notable barriers include concerns of misinterpretation and misuse (Michener, 2015), ethical concerns of who benefits from sharing whose data (Prainsack et al., 2022), privacy and security risks, intellectual property violations and the risk of errors and bias (De Silva & Vance, 2017). Issues of effort and labor include the often uncompensated time and resources required and the perceived disproportionate benefit to the data reusers (Borgman, 2012; Tenopir et al., 2011). When researchers do not consider their data to have value for others, the willingness to invest in this work is further reduced (Mayernik, 2011; Tenopir et al., 2015; Zuiderwijk & Spiers, 2019).

Facilitators of data sharing include "personal and intrinsic motivations," expectations of improved performance and supports to reduce the effort involved (Zuiderwijk et al., 2020). To be effective, tools and infrastructures supporting sharing need to be aligned with scholarly practices (Given & Willson, 2018; Karras et al., 2021) and convenient for data depositors (Tenopir et al., 2011). For example, research communities, institutional services, and journals support and promote data sharing (Goodey et al., 2022; Lin & Strasser, 2014; Tal-Socher & Ziderman, 2020). Trust is repeatedly found as a central factor in successful sharing (e.g., Chung et al., 2016; Pilerot, 2013; Zuiderwijk et al., 2020). Data quality seems to play a moderating role among the many factors influencing sharing (Zhi et al., 2023), including effective curation. The complex reciprocality between human actors, information, and data also underline the significance of contextual (Fan et al., 2023; York, 2022) and process information (Huvila, 2022a) and sense-making (Koesten et al., 2021) as key aspects of successful sharing and (re)use. For cross-disciplinary collaborations, challenges arise from the often implicit/tacit nature of this kind of information within disciplines (Koesten et al., 2021).

The wide variation in data practices across disciplines has been cited as a barrier to data sharing (Oliver et al., 2023a). At a higher level, we could consider disciplinary variation to influence more fundamentally how data sharing is defined, practiced, and valued, or not, across disciplines (Borgman, 2012). Zuiderwijk et al. (2020) note that, across disciplines, "data is diverse in its domain, volume and type and may consequently be more or less difficult to use." They found that certain disciplinary "nuances, traditions, cultures, or 'climates'" can empower data sharing, but shared no evidence of disciplinary practices inhibiting sharing. Given that there are clearly such cases (e.g., Kraus & Eberhard, 2022; Lin, 2023; Pels et al., 2018), this result raises some questions about the nature and goals of mainstream data sharing research. Studies of data perceptions and practices within the humanities offer the most divergent perspective on data sharing.

3 | THEORETICAL CONSIDERATIONS

In this study we build on the notion of culture that is frequently used to theorize the assemblages of practices and perceptions relating to research and data practices. We draw upon the notion of "data cultures" as a sensitizing concept (Charmaz, 2003), and consider how such cultures form and influence practices of research, sharing and non-sharing of data, and resistance to normative expectations. We approach data cultures tentatively from the perspective of Knorr Cetina's notion of epistemic culture as a specific "amalgam of arrangements and mechanisms" (Knorr-Cetina, 2003) that constitute research as an undertaking operating through and with specific understandings of "data" as its key element.

Aligned with Burgess et al. (2022), we anticipate that the contemporary scholarly everyday data cultures are the sites where the global digital transformations associated with datafication are enacted in practice. Further, we presume that the data (sharing) cultures identified in this study are situated in particular sociotechnical and -cultural contexts with their specific literacies (Burgess et al., 2022). In contrast to the proposition of Burgess et al. (2022) that everyday data cultures are over-determined by contemporary techno-capitalism, within the scholarly context, we anticipate that the factors and actors involved and the extent to which they (over-)determine data sharing cultures remain unclear.

As discussed earlier in the literature review on data cultures and datafication, a common stance in research on data practices is to acknowledge and describe different approaches to data, "affirming" and "respecting" different cultures, but with the underlying aim of "animating capacity to design data sharing infrastructure and policies that are [...] acceptable to everyone" (Poirier & Costelloe-Kuehn, 2019, p. 7). In other words, the aim is often to investigate divergent disciplinary perspectives as a means to further instantiate and spread a data centric paradigm (Oliver et al., 2023b). In this study, we want to consider seriously that counter-narratives, fringe perspectives and opposition to data culture are valid options. We argue that the contemporary conceptualizations of data sharing and data culture operate through a set of assumptions, or scripts (Latour, 1992; Pelizza & Van Rossem, 2023), on the "other" in the margins or outside of data. These "scripts of alterity" (Pelizza & Van Rossem, 2023) are embedded and not immediately obvious in how data sharing is described as a desirable predominant mode of operating and how data sharing is presented as a default and desired state of affairs and the absence of data sharing as not-yet-data-sharing. To understand and earnestly respect non-data and liminal—what we call data-decentered culture—we delve into conceptual *de-inscriptions* (Pelizza & Van Rossem, 2023) in the interviews conducted with senior researchers representing different domains of science and scholarship. Rather than acting according to the scripts of the data-driven culture, researchers can engage with de-inscriptions—or de-inscribe "data"—by using it in alternative ways, both conceptually and in practice. De-inscription examines evidence of resistance and appropriation of predominant notions of data in how researchers describe their own data practices, to understand alternatives to data-centered research cultures and the implications of an ongoing shift towards data-centrism in science. De-inscriptions and their associated data cultures can be *non-data* in the sense that data is not considered a relevant empirical or conceptual category in a particular type of research, or *data-decentered* in how research material can be termed data even if it is far from being the most apposite conceptualization.

4 | MATERIAL AND METHODS

We conducted qualitative interviews ($N = 16$) of senior researchers in social science and science disciplines at a major Canadian research university. We sought experienced academics able to draw upon disciplinary knowledge and varied research activities. Our recruitment pool consisted of researchers who were named as leads in major funded projects, had shared research data through the university data portal, or were associated with a public scholarship initiative aimed at community-oriented research and dissemination. The sample was not intended to be representative or large enough for generalizability, but rather to be diverse. We collected contact information for researchers who met our criteria from publicly available resources, sent email invitations to 110 researchers and received 16 positive responses. This low response rate can be explained, in part, by the competing demands on senior faculty and the timing of the study, which took place in 2020, in the midst of the COVID-19 pandemic. We also acknowledge that the pandemic might have had an influence on researchers' attitudes on data-sharing even if the interview record did not provide any direct evidence of that. Participants were offered a $25 gift card as an honorarium.

Individual interviews were conducted by the authors via video conference using the Zoom platform; audio recordings were saved and fully transcribed and shared with informants for member checking. Interviews lasted between 45 and 60 min. Informants are listed in Table 1 with their career stage, presumed gender and field of research indicated.

The interviews were semi-structured and designed to elicit the impacts of a changing socio-technical landscape on scholarly communication practices. Questions on data making and sharing were posed alongside questions regarding scholarly information practices involving, for example, social media, video, podcasts, and visual information. The interviewees were asked to reflect upon the strengths and limitations of the different practices as means of disseminating and sharing out their research and sharing in others' research as a support to their work. The semi-structured nature of the interviews meant that the discussions on data making and sharing emerged in different parts of the interviews and the exact wording of the questions varied. By treating data sharing as only one example of different approaches to share research, the interview protocol did not establish a particular set of expectations or norms regarding data culture, as may be the case with studies that examine research data practices in isolation (e.g., Borghi & Van Gulick, 2021; Tenopir et al., 2020). We present an analysis of responses from informants when asked to describe their practices related to research data, including its strengths and limitations as a means of sharing and/or re-using research. Additionally, we draw upon respondents' general descriptions of their research processes and materials, whether they used data-oriented terminology or alternate framings.

Interviews were preliminarily coded in QDA Miner Lite to facilitate a content analysis based on constant comparative method (Glaser & Strauss, 1967) and close reading (DuBois, 2003) of the data. We identified and grouped the data thematically, employing the notion of data culture as the central sensitizing concept.

5 | FINDINGS

5.1 | Shift towards data-centrism

The analysis of the interview record showed a breadth of views from how data-making and data-sharing are integral to research efforts to how they lie far outside of certain epistemes. A common backdrop reflected in all interviews is a shift towards data-centrism and normalization of sharing research materials and data. Even in fields that are not oriented towards data-sharing (e.g., R15-SCANTHRO), many researchers adopt such

TABLE 1 Informants ($N = 16$) interviewed.

PID	Career stage	Gender (presumed)	Field
R1-IS	Associate professor	M	Information Science
R2-EDU	Professor	F	Education
R3-AGRI	Professor	M	Agricultural Science
R4-AGRI	Professor	F	Agricultural Science
R5-BIO	Professor	M	Biology
R6-ANTHLING	Associate professor	M	Anthropology and Linguistics
R7-LING	Associate professor	F	Linguistics
R8-POL	Professor	F	Political Science
R9-LANGEDU	Professor	F	Language Education and Applied Linguistics
R10-EDU	Associate professor	F	Education
R11-OCC	Associate professor	M	Occupational Science and Therapy
R12-HGEOG	Professor	M	Human Geography
R13-EDU	Associate professor	F	Education
R14-HGEOG	Professor	F	Human Geography
R15-SCANTHRO	Associate professor	F	Sociocultural Anthropology
R16-CS	Professor	F	Computer Science

practices, for instance, by publishing their research material that does not end up in publications. Across the disciplines, increasing awareness that data might be consulted by someone else in the future incentivises data curation and documentation, to add an "extra level of clarity." This helps others and also the researcher himself when he returns to his data (R5-BIO).

The shift is not, however, straightforward. Data-sharing does not necessarily correlate with high re-use. Interviewee R5-BIO had "done some re-use of data" but mostly shared out her research results and the software used to generate them rather than utilizing what others had shared. Several others had not shared or re-used data (e.g., R6-ANTHLING, R9-LANGEDU, R10-EDU, R11-OCC, R12-HGEOG) and those who had primarily reused purposefully produced data, for instance, government datasets (R12-HGEOG, R13-EDU). Shared research data is often used for teaching rather than new research (R5-BIO, R7-LING, R11-OCC). In such cases, the specific data is less important than having something to exemplify, for example, a method (R5-BIO). Also while the interviewees referred to many established and emerging approaches, no clear consensus emerged on the meaning of data and data sharing in practice. "Data sharing" seems to be primarily viewed as something that takes place online and involves datasets rather than data products or research archives, with some exceptions. For example, interviewee R3-AGRI recalled non-digital pre-internet age data sharing using "data rich books" with data tables.

The multiplicity of outputs and residues of research work means that partial sharing of research data is not necessarily a prohibitive problem for reuse. In cases where incomplete data is available, it may be possible to (re)create parts of it (R7-LING). Similarly, when research data is already published in a specific repository, it is enough to share tools and analysis procedures (R13-EDU).

However, the interviewees also identified common problems. Ethical and legal barriers (R11-OCC), and unclear rules and guidelines (R8-POL, R13-EDU) were a typical conundrum. A crucial complication is that many ethical problems might not be solvable at all even if ethical pre-clearance functions as a "a big switch" (R11-OCC) with the capacity to resolve many of them. Another typical issue relates to metadata and the adequacy of contextual information. Available data is sometimes at an aggregate level, which is not as useful in statistical analyses as "raw data" (R11-OCC). Sometimes, data lacks adequate documentation for particular reuse scenarios (R5-BIO). The lack of adequate metadata and paradata is problematic especially for the evidential potential of data and its usefulness in replication studies. As interviewee R3-AGRI remarks, the replication issue is more complex than can be solved by sharing data. It requires not only sharing information on methods used but for the receiver to master them. A related challenge for replication and the evidential use of data is not knowing if everything has been kept and shared. Sharing a transcript of an audio or video file may not be enough if the original study had access to recordings (R7-LING). In linguistic research, replication with shared stimuli requires that the population understands the stimuli in same manner—and it can be difficult to find a comparable population and/or comparable stimuli (R7-LING). In addition, there is likely to be tacit and undocumented knowledge that guided the researchers who did the original analysis.

However, the interviewees also described some examples of means to improve the richness of data sharing. A contemporary approach to data sharing used by interviewee R5-BIO is to share statistical data using R Shiny presentations to share datasets together with analysis procedures. Data can also be shared in aggregate forms through presentations. Interviewee R3-AGRI explained how he often experiments with different ways of presenting data and after finding a satisfactory design, hands it over to a person skilled at creating the final visualization.

All in all, even if there is a move towards data-centrism, it is like William Gibson suggested of the future: already here, but unevenly distributed. While an option to sharing data in journals is becoming more common, it does not apply to all journals (R14-HGEOG). Interviewee R5-BIO noted that his journal and journals he works with mandate data sharing while R1-IS4 noted that his journal does not prioritize publishing data but rather survey instruments and adding a dataset link. His impression was that, at least in his own field, the push towards data-sharing comes from (general) research funders rather than from journals (R14-HGEOG) or intra-disciplinary authorities. The same applies to social anthropology journals (R15-SCANTHRO). A key prerequisite is a shared episteme including a common understanding of what is adequate quality data and what are appropriate procedures to produce it.

5.2 | Three forms of research data cultures

The interviewees indicated diverse reasons for sharing and non-sharing of data, which we roughly categorized as direct benefits, norms, and culture, and external and internal drivers. These are presented in the following sections, in relation to three major constellations of thinking and acting in relation to data and data-sharing, which we approach as data cultures. These are analytical aggregates compiled from the entire corpus of interview material rather than representing a classification of individual interviewees. None of the interviewed individuals fall completely in one profile and most have multiple affinities. While some of the interviewees appeared to be comfortable with their stance in relation to data and data-sharing, many showed signs of struggling with the notion of data, and the extent to which their research materials could be framed as data in the epistemic context of their research. The "data" the interviewees mentioned working with ranged from numeric information (e.g., R3-AGRI) and text (R7-LING) to audio and video (R7-LING), interview transcripts (R11-OCC, R14-HGEOG), photographs (R15-SCANTHRO), databases (R11-OCC), research instruments (R5-BIO, R7-LING) and beyond. Even if sharing was not always considered straightforward, the basic attitude towards data sharing was positive. No one was explicitly against sharing data but there were considerable differences in how the interviewees perceived what is data, what is shareable, and how.

The following sections draw upon the interview data to outline three forms of research data culture: data-driven, non-data, and data-decentered. These exist in the context of a general pull and shift towards greater data-centrism, and thus represent varying degrees of convergence/alignment and resistance.

5.2.1 | Data-driven culture

We use the term data-driven culture to refer to an arrangement of perspectives and practices associated with research that is explicitly based on working with data, which is considered, a priori, to be a valuable, shareable asset. In data-driven culture, the approaches to data sharing range from "dumping" data online (R1-IS) to using major disciplinary, topical (R3-AGRI) and multi-disciplinary (R5-BIO) online data repositories. Interviewee R14-HGEOG envisioned a "one stop shop" (R14-HGEOG) where all data and everything else needed to replicate a study could be found. In the analyzed interview record, data-driven culture is typically linked to large-scale quantitative research (e.g., R3-AGRI, R5-BIO), but its influence is visible also in the responses of qualitative researchers. Interviewee R3-AGRI, an agricultural scientist, is an example of a researcher whose entire career revolved around collecting, systematizing and making available data in his field of study for his own and others' research. The work spanned publishing data-rich monographs to developing first local and later online databases making available data from a large network of collaborators.

Characteristic to data-driven culture is assenting to the *ideals of sharing data and its perceived beneficiality*. Some interviewees had participated in projects that shared data (R11-OCC) or in establishing small (R6-ANTHLING, R15-SCANTHRO) and major data-sharing initiatives (R3-AGRI, R5-BIO). The most explicit underlying reason to share research data among the interviewees was expressed in terms of altruism and the belief that sharing and transparency are good per se. Beyond these, personal fascination and the perceived interestingness and value of data and the information it contains can motivate to share (R8-POL, R12-HGEOG).

Data-driven research acknowledges also the specific practical benefits of sharing and reusing data. A core tenet of the data-driven mindset is that one researcher or research group might not be able to exhaust the analytical potential of a particular dataset (R8-POL). Using existing data can be more cost-effective than collecting data from scratch (R3-AGRI). Having data available provides opportunities to "work with data" (R3-AGRI), that is, reanalyze it (R1-IS), conduct meta-analyses (R3-AGRI, R4-AGRI), reproduce earlier studies (R3-AGRI), longitudinal analysis (R8-POL), approach data with new research questions (R8-POL) and analysis methods (R11-OCC), and "mobilizing" historical datasets (R6-ANTHLING). Interviewee R3-AGRI noted that in his discipline, to "reach high" a researcher needs a lot of data to infer new things. On a narrow base, "the thing will fall down" (R3-AGRI). Even if much of the purported benefits focus on sharing with others, documenting data properly and sharing in a repository makes it available also for the researcher himself (R5-BIO). Interviewees R1-IS commented that it is more challenging to gather and clean data and get it into an analyzable form than to run hypothesis testing.

Besides expected benefits and an altruistic ideal, data-driven culture is also underpinned by a sense of *external imperatives of data availability and sharing*. For some, sharing was a personal priority (e.g., R5-BIO), but the basic shared assumption is that the availability of data is a standard practice (e.g., R3-AGRI) rather than a matter of individual choice. A part of the imperative is implicit. Public funding was discussed as a reason to share research with society, which instigates data-sharing (R8-POL). Interviewee R8-POL described sharing as her "responsibility." Another interviewee (R13-EDU) noted that she keeps her data for 5 years as evidence for examination purposes. Explicit mandates also influence data publishing and sharing (R5-BIO). Interviewees referred to directives and encouragement from funders (R10-EDU), journals (R5-BIO, R10-EDU) and colleagues in the context of particular research projects (R14-HGEOG) as a reason to share or consider sharing data.

Another distinctive feature of the data-driven culture is how *it builds on the exemplar of data-intensive research powered by standardized datasets*. Engaging in data-oriented lines of research unfolds as a key practical reason for sharing research data (R4-AGRI). Trying to replicate earlier studies helps to understand data and identify gaps in methods sections (R7-LING). The "one-stop shop" envisioned by one of the interviewees exemplifies an ideal of what standards and data-empowered *modus operandi* might imply for research, although as the interviewee notes, linking of resources and infrastructures could also become problematic if it ends up being someone's monopoly (R14-HGEOG). Sharing is facilitated if protocols are in place to show researchers how it is done in practice (R7-LING). Openness requires "making [data] available without people having to ask for it" (R7-LING).

In the data-driven culture, the *major barriers to data-sharing tend to be framed as practical rather than epistemic*. Uneven access to data from different sources makes aggregating data and, for example, conducting meta-analyses difficult (R4-AGRI). Technical and conceptual obsolescence, such as unreadability of old file formats, is a major hindrance for sharing and reusing data (R5-BIO). Data size can be prohibitive to data sharing, for example with sound and video files (R7-LING). Maintaining the confidentiality and security of study participants makes data sharing difficult, especially with vulnerable study participants (R10-EDU). Data use licenses can hinder data sharing. For example, in some governmental and private datasets, the use license might not allow resharing of the data (R11-OCC). Lack of permission to (re)share data because of the stipulations of the original owner or provider forms a comparable obstacle (R11-OCC) that might apply to any third-party data. While economic benefits can motivate data reuse and sharing data, it can also become a barrier to sharing if the data collector is worried about "free-riding" (R1-IS), where others benefit without doing the work. Similarly, variations in data quality can both facilitate and hinder sharing and reuse (R1-IS). Ensuring that shared data is of sufficient quality (termed, e.g., as "good enough replication" R7-LING) demands time and other resources. Differences in the organization and structuring of data is another practical challenge in reuse of datasets, as interviewee R1-IS explained with reference to comparing statistical data collected at different levels of aggregation across countries. Finally, low data literacy can hinder data sharing and reuse within disciplinary communities and externally (R1-IS). Teaching students how to use shared data is an important feature of data-driven culture. For instance in linguistics, being data literate enough to (re)use previously developed stimuli means understanding that it demands a thorough mastery of the earlier work (R7-LING).

Overall, even if the data-driven culture unfolds as an analytically distinct mindset, the analysis shows that there are different ways of being data-driven. The interviewees expressed conflicting attitudes towards data sharing and management. Some of the interviewees criticized the level of documentation expected. It is, as interviewee R5-BIO put it, "ridiculous" as most of the information is not needed and authors will not comply. Rather, aiming at perfect would lead to no documentation and consequently undermine the realistic possibility to obtain good quality data (R5-BIO).

5.2.2 | Non-data culture

Even if notions of data and data-sharing are applicable to many disciplines, we identify non-data culture as research and knowledge sharing practices and perspectives that exist apart from data. Several interviewees noted that their research, as interviewee R2-EDU put it plainly, "does not involve data." This applies to most theoretical research and studies based on other's published results (R2-EDU) but is also applicable to fields where primary research material is not self-evidently characterizable as "data." While few explicitly rejected the relevance of data, there was a palpable contrast between data-driven work (especially, e.g., R3-AGRI, R5-BIO) and the struggle of others in discussing research materials as "data" (e.g., R1-IS, R6-ANTHLING, R9-LANGEDU, R15-SCANTHRO). References to data and data sharing within the non-data culture stood out as speculative and concerning, rather than integral to the research endeavor.

In the analyzed interview transcripts, ethnographic anthropology provides an illustrative example of a discipline with a complex relation to conceptualizations of research materials as "data" and how the interviewees produced their own *de-inscriptions* of data (cf., Pelizza & Van Rossem, 2023). Albeit referring to "data," one of the interviewees expressed that he is "sceptical/allergic to the idea of sharing primary data" (R6-ANTHLING), and was clearly discussing epistemically very different kinds of materials from the straightforward understanding of shareable and (re)usable resources several others were describing. Another interviewee noted that even asking to share this "data" would be "really bad form" (R15-SCANTHRO). Partly, "field notes, historically, are kind of the dirty laundry" (R6-ANTHLING) of the discipline, material that is not tidy enough for sharing, but also something that can be prohibitively difficult to share because of the importance of having been personally engaged in the ethnographic situation. "[I]t is a dataset, but it can't, in a sense, be dissociated from the experience of creating it, so it's not something that can be circulated freely" (R15-SCANTHRO). The de-inscription used by the first interviewee (R6-ANTHLING) was to provide an alternative conceptualization of "data." The second (R15-SCANTHRO) did it by placing her research practices and materials outside of data; and even if accepting that it could be called data—that it is not sharable in a useful or ethically appropriate manner.

Comparable perspectives to the importance of personal engagement and having been there were expressed also by interviewees from other fields of social research however typically without resorting to such radical de-inscriptions. The contextuality of research makes sharing material sometimes difficult in linguistics (R7-LING). Also interviewee R1-IS reasoned that not all of his research material "lends itself to large datasets being dumped online that would enable other people to make different sense of the data than what I made." Sometimes, the interviewees remarked that a part of the material they use in their research could be shared. However, this would apply primarily to particularly shareable, for instance, quantitative observations, or material left out of publications with potential secondary cultural relevance (R1-IS, R15-SCANTHRO). With the comment—"[b]ut it's not framed as like a dataset, so to speak" (R15-SCANTHRO) this interviewee acknowledges and also resists data-oriented perspectives, which do not reflect the nature of their own research practices and materials.

5.2.3 | Data-decentered culture

Data-decentered culture—or as discussed later, the space between data-driven and non-data cultures—refers to a constellation of views and practices with imaginable, but often subtle and contradictory affinities to thinking about research materials as data that is sharable. While some data-decentered research explicitly refers to sharing and working with "data," not everyone would, a priori, describe their research material as such. Also, when referring to data in data-decentered culture it does not necessarily have the same implications as in the data-driven culture. It lacks a comparable imperative to preserve and share data. Interviewee R14-HGEOG remarked that "I could anonymize my data and put it up" (R14-HGEOG) indicating willingness but also lack of disciplinary routine to do so. Moreover, when data is kept, researchers use project-specific or generic solutions.

A major distinction between the data-driven and data-decentered cultures is that in the latter, even if research material would be called data, *there is much more friction and variation in how different types of materials are considered to be collectable, usable, and shareable*. Data-decentered culture is also deeply concerned with attempts to standardize data. From the perspective of an episteme lacking consensus on how much material or data is enough for research, the most apparent problem is that all conceivable data is never collected or made available (R7-LING). In parallel, even when talking about research material as "data," qualitative data, in particular, can feel too situated and personal to consider sharing and circulating it freely (R13-EDU, R15-SCANTHRO). Qualitative data is "premised upon one's own engagement in an ethnographic situation, so it's very personal work, it's not a dataset, per se, that can be shared in a sort of—I mean, it is a data set, but it can't, in a sense, be dissociated from the experience of creating it, so it's not something that can be circulated freely" (R15-SCANTHRO). In contrast, for example, a corpus of (Twitter, now X) tweets can be far less controversial to share (R1-IS). For this reason, sharing data can feel possible only within a project rather than outside (R13-EDU).

Some interviewees also made distinctions between open data-sharing "proper" and sharing photographs online and depositing research material in an archive (e.g., R15-SCANTHRO, R6-ANTHLING). As R15-SCANTHRO notes, such resources are not "framed as like a dataset." However, at the same time photographs could also qualify as shareable material comparable to, for example, survey data, but unlike field notes that interviewee R15-SCANTHRO preferred to keep for herself. This allowed her to be sensitive to ethical and epistemic concerns and also share a "component" of her research "that seems publicly shareable" with the community. A parallel epistemic concern is that "the kind of data" a researcher has does not "get[…] shared that way." For example, some data does not "lend[…] itself to large datasets being dumped online that would enable other people to make different sense of the data than what [the researcher] made" (R1-IS). However, in some cases, the interviewees agreed that their "data" could be shared if properly anonymized (R14-HGEOG). Moreover, the interviewees also called for clarity about what can and what cannot be shared. Interviewee R10-EDU emphasized that also non-sharing of data needs to be acceptable.

In data-decentered culture the shareability of research material, whether called data or not, is underpinned by a variety of *often unsolved practical, moral and epistemic issues*. Quantitative data was considered easier to share than qualitative "data," which raised various legal, ethical, and epistemological difficulties even if the researchers were positively inclined towards sharing (e.g., R6-ANTHLING, R15-SCANTHRO). In contrast to data-driven culture where ethical and legal concerns exist but are typically turned into solvable problems and managed away, in data-decentered culture they are more likely to be viewed as "vicious" problems without clear-cut solutions.

Ethical and legal concerns result in some forms of data being more likely to be shared than others. For instance, social anthropological data "allows for all kinds of ethical misdeeds" (R15-SCANTHRO). Privacy concerns and legislation can make sharing audio recordings difficult, and therefore, they might be more often deleted than, for instance, survey data (R8-POL). Similarly, interview transcripts can be difficult to share (R11-OCC, R10-EDU); whereas, survey data might be less problematic (R10-EDU). As R7-LING noted, all speech and language data have political associations that lead to legitimate concerns about revealing participant identities, such that only simple "push button" data is easy to share. Because especially research with humans is not a technical endeavor, researchers' expressions of discomfort (e.g., R14-HGEOG) about sharing certain kinds of data should be taken seriously. Sharing a field researcher's personal diary, "that's what field notes are," is "kind of a problem" (R15-SCANTHRO).

Copyright can limit the possibility to share some types of data further (R12-HGEOG). Such legal issues can be difficult to solve (e.g., finding copyright holders or ensuring that no one is alive) (R12-HGEOG). Sharing data internationally can be hindered by different parameters and stipulations of how and what can be shared (R13-EDU). Ethical and legal concerns can lead to only a part of the data being shared and/or potentially identifying information left out (R8-POL) which has obvious implications to the (re)usability of the data.

A feature related to the multiplicity and diversity of experienced challenges in dealing with data is that *data sharing can be one-sided and planned rather than an ongoing practice*. Interviewee R5-BIO had shared a lot of data and had championed open data policies in his field and journals he had edited but was less sure whether or not he had "done some re-use of data." Others, such as interviewees R6-ANTHLING and R7-LING, had not accessed or produced open data but might have thought how to publish individual or multiple datasets collected in the past. Many had not shared data so far at the time of the interviews (e.g., R6-ANTHLING, R9-LANGEDU, R10-EDU, R11-OCC, R12-HGEOG). Interviewee R6-ANTHLING, an anthropologist, had for a long time planned to make the data from his PhD research available online for others to interrogate, but had not yet, although he had done a lot of work "harnessing and mobilizing historical fieldwork datasets" (R6-ANTHLING). Some agreed that they would be in principle "open to doing that" (R1-IS, also R7-LING) or in general terms considering that "we need to be more transparent about data" (R4-AGRI, also R6-ANTHLING), and sometimes with intentions to share research data (R6-ANTHLING, R8-POL, R12-HGEOG). Interviewee R7-LING mentioned that her lab is open to data-sharing and they share when specifically asked, if the ethical permissions required for the project are in place. Also, interviewee R13-EDU reported that she was asked for data once and would have happily shared it, but in the end, the query did not lead to anything.

In data-decentered culture *data sharing often takes place through project- and dataset-specific websites (R15-SCANTHRO) and via generic library-provided solutions (R15-SCANTHRO) rather than centralized and standardized disciplinary repositories* common in the data-driven culture. Some of the interviewees referred to sharing when asked by personal handover (R7-LING, R13-EDU). Finally, a typical trait in this data-decentered discourse is to *make comparisons to fields that are presumed to have longer and more extensive engagement with datafication*. In some fields, like language education (R9-LANGEDU), data-sharing is still new and not established as a norm. In contrast, R6-ANTHLING and R7-LING suggested that data-sharing is much more common in, for instance, linguistics and even if it is not mandated, the option to share data is rapidly becoming a norm (R7-LING).

6 | DISCUSSION

The present study has delved into researchers' perspectives, rationales and practices of (non-)sharing of research data in relation to their research practices. We have identified (RQ1) three data cultures (data-driven culture, non-data culture, and data-decentered culture) that underpin researchers' (non-)sharing of research data and are linked (RQ2) to their epistemic and practical understandings of their research work. Our study reinforces much of the prior research in terms of drivers, enablers and obstacles to data sharing (cf., e.g., Borgman, 2012; Michener, 2015; Prainsack et al., 2022; De Silva & Vance, 2017; Tenopir et al., 2011). Similarly, the disciplinary patterns of more active participation in data sharing among researchers in the sciences than in other areas, and a relatively low percentage of active data sharing and use among interviewees maps onto prior work (e.g., Gomes et al., 2022).

In parallel to confirming earlier observations, our study has two major novel contributions. First, while the analysis recognizes the diversity of practices and perspectives on data, it extends earlier work on data-driven culture (Oliver et al., 2023a, 2023b) by showing how it has become normative and dominant. It is shaping discourse even in fields where research has not been conceptualized in terms of generating, collecting, analyzing, and managing data, such as ethnographic anthropology (cf., Kraus & Eberhard, 2022). We also show that different forms of resistance exist, which may be aimed at preserving or foregrounding important values and goals of epistemic and disciplinary cultures. Acknowledging the normative dominance of data-driven culture and diverse forms of resistance allows us to recognize that trying to fit all research into a big tent of open research data may have unintended consequences, especially for non-data or data-decentered cultures. This emphasizes the importance of more work that actually tries to establish definitions of data (cf., Kraus & Eberhard, 2022; e.g., as Chao et al., 2015) and data sharing (cf., Dutoit, 2017; Thoegersen & Borlund, 2022)—and to consider the implications of datafication. The present dominance of data-driven culture is operating on the level of what Cameron (2021) describes as the fifth mode of standardization. The variations in research practices are systematized and normalized by framing research material as "data," which is a priori extractable to distinct entities, acontextual enough to be shared and reused by others, manageable in repositories, and has independent value outside of the practice where it originates. At the same time it upholds an artificial division between a (good) data-driven and (bad) non-data culture that refuses to acknowledge the diversity of how research works.

Another contribution and difference in relation to earlier research is that our data culture perspective does not try to map specific disciplines onto these cultures, as prior research on the classification of research fields or disciplines has typically tried to do (e.g., Fry & Talja, 2007; Gregory et al., 2023; Tenopir et al., 2020). Drawing hard parallels between disciplines and their data cultures risks leading to stereotyping that can be counter-productive to understanding and supporting the diversity and specificity of data practices. The same applies to comparable stereotyping in terms of classifying researchers categorically as theoreticians and empiricists (Birnholtz & Bietz, 2003; Borgman, 2012; Dutoit, 2017). The approach to data cultures adopted in this study recognizes the stratified and sometimes competing nature of cultural influences within research, which means there is substantial variation in data sharing practices in terms of a multitude of local data (sub-)cultures within the broad constellations sketched out in this study.

Inquiring into the practices of resistance and scripts of alterity (Pelizza & Van Rossem, 2023) in the analysis directs attention to three major patterns in the interview transcripts. First, the interviewees engage in *problematization of the notion of "data" and its relevance in their research*. It was not always clear if they outright rejected datafication in relation to their work, or if they are just uncertain of the meaning of data sharing and what kinds of materials are included. The big tent approach to research data management tries to frame almost everything as data to include all research. Some researchers try to go along with this, others resist the idea. Second, several interviewees were *asserting exceptions to the general* *case for data sharing*. This form of resistance insists that not all data should be shared. It highlights issues of ownership, relationality, and privacy that would make data sharing unacceptable. Interestingly, the research data management discourse itself has begun to acknowledge the need of certain exceptions and adopted principles that resist the tenets of the data-driven culture. The European Union Horizon 2020 programme guidelines on FAIR (as in FAIR principles of making research data Findable, Accessible, Interoperable, and Reusable, see Wilkinson et al., 2016) data posit that data should be "as open as possible and as closed as necessary" (European Commission, 2016, p. 8) opening up space for a general exception. The parallel CARE principles for Indigenous data governance (Carroll et al., 2020; Gupta et al., 2023), which reframes data sharing from an instrumental practice to an ethical and relational practice, represents a more systematic and comprehensive script of alterity. However, as scripts that frame data sharing in fundamentally very different terms compared to the doctrine of the data-driven culture, they are hardly reconcilable and it is questionable whether data or data sharing ever can be both FAIRful and CAREful. This irreconcilability emphasizes that all data cultures follow their own scripts posing exceptions to each others' adages. This does not need to mean, however, that the differences cannot be navigated (cf., Pels et al., 2018; cf., Kraus & Eberhard, 2022) or that it is impossible to find practical workarounds to share "data" to some extent, even across data cultures. The crux is that when it is done, it is important to acknowledge that such workarounds are mere workarounds.

Third and finally, some of the interviewees expressed their *reluctance to engage in data sharing or chose to opt out* entirely. This script is interesting because it does not necessarily oppose data sharing, but for whatever reason, situates this as a practice outside the realm of their scholarly work.

Evidence of these forms of resistance could raise the need for new conceptions of data sharing. Comparable to Leonelli, who argues that the "conception of openness as sharing is flawed" (Leonelli, 2023, p. 43) and advocates for an alternative notion of openness based on inclusion, the conception of sharing "data" as molding it into datasets and setting them free is erroneous for many researchers. It makes sense in data-driven culture but is hardly relevant in others. In the analysis, the preference for domain-specific repositories in data-driven culture versus the lack of repositories and use of generic solutions in non-data culture and data-decentered culture points to this direction. Acknowledging that generic mechanisms can sometime be adequate, the lack of purpose-built infrastructures signals the risk that the real sharing might happen somewhere else, or that there is no meaningful data to share at all and that "sharing" is facework of adhering to the normative expectations stemming from the data-driven culture. To counter such tendencies and risks, it is crucial to acknowledge that "data sharing" and "sharing" are culturally determined and to refrain from universalist assumptions that lead to, as one of the interviewees put it, "ridiculous" (R5-BIO) requirements stemming from an instrumentalist data management assumption that all fields have research data and it is just a matter of defining what that is. Rather than expecting everyone to share data from their projects and enforcing mandatory data-sharing policies across all research, efforts should focus on addressing the calls for better data and rigorous data-making (e.g., Jarrahi et al., 2023) by incentivizing the production of high-quality purpose-built datasets in data-driven contexts and fields where it makes sense and where gathering and cleaning data is a separate major effort from hypothesis testing (cf., R1-IS). In the Canadian context, where this study was conducted, the national "Tri-Council" research funding agencies actively encourage data sharing principles and practices in line with the data-driven culture, but have not yet imposed clear mandates across funding programs (Government of Canada, 2021).

Likewise, the lack of adherence to a normative data culture should not be taken a priori as a sign of a lack of skills or bad behavior. As the analysis shows, not all research deals with data at all, "data" can be a more or less fitting concept to describe research material, and the eventual sharing of "data" can be possible or impossible, relevant or irrelevant. Research materials take many forms for good reasons (cf., Kraus & Eberhard, 2022). All "data" should perhaps not be archived or shared at all for ethical (cf., Pels et al., 2018) and also for diverse practical reasons. We should assume that research relies on epistemological and ontological perspectives in considering data in relation to research practices and outcomes. This may include assumptions about properties of data that may or may not align with their research practices.

Even in data-centric fields some of the expectations of the data-driven culture are incompatible, because researchers do not necessarily see the main goal of their research to be the production and sharing of data. The question is, how can the science itself take precedence? Similar to Frohmann's (1999) argument that focusing on publications shifts emphasis away from the doing of science, focusing on the data can be a problem. Other problems arise from diverging assumptions about ethical rules and dilemmas. As the interviews (R6-ANTHLING, R8-POL, R11-OCC, R15-SCANTHRO) suggest, eventual risks and problems with privacy, disclosure, and unethical secondary data use can be difficult to anticipate if data is shared openly. The training of large-language models provides a recent example of the unanticipated reuse of open data (Choksi & Goedicke, 2023). Much of this problem can be traced back to the cultural-historical alignment between datafication and an increase in access through digital platforms. There is a mismatch between how big data in business has a strong association with value and competition; whereas, the shift to data-centric research in academia has been aligned with the open movement, with the goal of transparency and broader societal benefits from public research investment (Kitchin, 2014). The linking of datafication and openness makes the shift more complex for some fields. The conundrum has become increasingly difficult to manage in relation to the evolving data privacy legislation (Corte, 2018) and commitments to Indigenous data sovereignty. Especially in fields working with little data that lacks potential for linking, the absence of apparent benefits of the data perspective underlines the potential risks.

As a whole, a reasonable question to ask is whether data-decentered culture is a "real" data culture or rather a liminal space between data-driven culture and non-data culture where non-data culture is engaging in boundary work (Houf, 2021) with data-driven culture trying to explore to what extent data-driven ideals and discourse can be accommodated to frame non-data. Data-decentered culture is not a state of deprivation but rather a field of political struggle. Currently money and prestige are associated with the dominant data-driven paradigm meaning that the researchers outside the paradigm have to make a choice.

7 | CONCLUSIONS

In this study we have contrasted the data-driven culture with non-data and data-decentered data cultures. The findings emphasize the critical importance of taking seriously the urges to acknowledge that many of the contemporary universalizing tenets of research data management and sharing are seriously flawed in many contexts of research. While there are barriers and apparent reluctance to appropriate management of research materials for sharing, reuse and statutory archiving, it would be important to acknowledge the occasional absence of data and the plurality of what data and data sharing mean, and what are appropriate practices of caring for different types of "data." This should also be recognized in data-sharing policies to avoid pushing researchers to avoid risks and share data in order to comply with a mandate even in such research that embraces a data-decentered or no-data perspective. A challenge is that data is associated with prestige in the sense that "real research" has data and data has economic value. Therefore, it may be important to establish other categories of "research stuff" that can be recognized as having value.

In parallel there is a dire need to incentivize data sharing in data cultures where it contributes to advancing the aims of the scientific enterprise. However, rather than aiming at forcing everyone to share data, the emphasis should be on incentivizing the production of high-quality purpose-built datasets and encourage others merely to keep their research materials for a reasonable period of time, in some cases for some years and in others forever deposited in an archive, to meet statutory obligations but also for eventual future use, even if the goal would never be to share it with anyone.

Our study has evident limitations. The group of interviewees is not a representative sample but represents a single institution and senior scholars, who are actively engaged in collaborative and/or public-facing research. What is notable is, however, that even within this group, we found considerable variation consistent with prior research and provided new substantial understanding of the sharing of research data.

ORCID

Isto Huvila https://orcid.org/0000-0001-9196-2106

REFERENCES

Andorfer, P. (2015). *Forschen Und Forschungsdaten in Den Geisteswissenschaften* (Technical Report 10). DARIAH-DE.

Aragona, B., & Zindato, D. (2016). Counting people in the data revolution era: Challenges and opportunities for population censuses. *International Review of Sociology*, *26*(3), 367–385. https://doi.org/10.1080/03906701.2016.1244926

Arditi, J. (1994). Geertz, Kuhn and the idea of a cultural paradigm. *British Journal of Sociology*, *45*(4), 597. https://doi.org/10.2307/591885

Becher, T., & Trowler, P. R. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines* (2nd ed.). Open University Press.

Birnholtz, J. P., & Bietz, M. J. (2003). Data at work: Supporting sharing in science and engineering. In *Proceedings of the 2003 international ACM SIGGROUP conference on supporting group work* (pp. 339–348). ACM. https://doi.org/10.1145/958160.958215

Borghi, J. A., & Van Gulick, A. E. (2021). Data management and sharing: Practices and perceptions of psychology researchers. *PLoS ONE*, *16*(5), e0252047. https://doi.org/10.1371/journal.pone.0252047

Borghi, J., & Van Gulick, A. (2022). Promoting open science through research data management. *Harvard Data Science Review*, *4*(3), 2644–2653. https://doi.org/10.1162/99608f92.9497f68e

Borgman, C. L. (2012). The conundrum of sharing research data. *JASIST*, *63*(6), 1059–1078. https://doi.org/10.1002/asi.22634

Borgman, C. L., Scharnhorst, A., & Golshan, M. S. (2019). Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. *JASIST*, *70*(8), 888–904. https://doi.org/10.1002/asi.24172

Burgess, J., Albury, K., McCosker, A., & Wilken, R. (2022). *Everyday data cultures*. Polity Press.

Cameron, F. R. (2021). *The future of digital data, heritage and curation*. Routledge. https://doi.org/10.4324/9781003149606

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE principles for indigenous data governance. *Data Science Journal*, *19*, 43.

Cetina, K. K. (2007). Culture in global knowledge societies: Knowledge cultures and epistemic cultures. *Interdisciplinary Science Reviews*, *32*(4), 361–375. https://doi.org/10.1179/030801807X163571

Chao, T. C., Cragin, M. H., & Palmer, C. L. (2015). Data practices and curation vocabulary (DPCVocab): An empirically derived framework of scientific data practices and curatorial processes. *Journal of the Association for Information Science and Technology*, *66*(3), 616–633.

Charmaz, K. (2003). Qualitative interviewing and grounded theory analysis. In J. Holstein & J. F. Gubrium (Eds.), *Inside interviewing: New lenses, new concerns* (pp. 311–330). SAGE.

Choksi, M. Z., & Goedicke, D. (2023). Whose text is it anyway? Exploring BigCode, intellectual property, and ethics. https://doi.org/10.48550/arXiv.2304.02839

Chung, E. K., Kwon, N., & Lee, J. (2016). Understanding scientific collaboration in the research life cycle: Bio- and nanoscientists motivations, information-sharing and communication practices, and barriers to collaboration. *JASIST*, *67*(8), 1836–1848. https://doi.org/10.1002/asi.23520

Corte, D. L. (2018). The European right to data protection in relation to open data. In B. van Loenen, G. Vancauwenberghe, & J. Crompvoets (Eds.), *Open data exposed* (pp. 127–148). T.M.C. Asser Press. https://doi.org/10.1007/978-94-6265-261-3_7

Dallmeier-Tiessen, S., Darby, R., Gitmans, K., Lambert, S., Matthews, B., Mele, S., Suhonen, J., & Wilson, M. (2014). Enabling sharing and reuse of scientific data. *New Review of Information Networking*, *19*(1), 16–43. https://doi.org/10.1080/13614576.2014.883936

Davies, T., & Edwards, D. (2012). Emerging implications of open and linked data for knowledge sharing in development. *IDS Bulletin*, *43*(5), 117–127. https://doi.org/10.1111/j.1759-5436.2012.00372.x

De Silva, P. U. K., & Vance, C. K. (2017). *Scientific scholarly communication*. Springer. https://doi.org/10.1007/978-3-319-50627-2

Donner, E. K. (2023). Research data management systems and the organization of universities and research institutes: A systematic literature review. *Journal of Librarianship and Information Science*, *55*(2), 261–281. https://doi.org/10.1177/09610006211070282

Drucker, J. (2011). Humanities approaches to interface theory. *Culture Machine*, *12*, 1–20.

DuBois, A. (2003). Close reading: An introduction. In F. Lentricchia & A. DuBois (Eds.), *Close reading: A reader* (pp. 1–40). Duke University Press.

Dutoit, M. (2017). Research data in scholarly practices: Observations of an interdisciplinary Horizon2020 project. In J. Kamps (Ed.), *Research and advanced technology for digital libraries* (pp. 623–627). Springer.

European Commission. (2016). *H2020 programme guidelines on FAIR data management in horizon 2020* (Technical Report). Directorate-General for Research & Innovation.

Fan, L., Lafia, S., Li, L., Yang, F., & Hemphill, L. (2023). DataChat: Prototyping a conversational agent for dataset search and visualization. *Proceedings of the Association for Information Science and Technology*, *60*(1), 586–591. https://doi.org/10.1002/pra2.820

Frohmann, B. (1999). The role of the scientific paper in science information systems. In M. E. Bowden, T. B. Hahn, & R. V. Williams (Eds.), *History of information science: Proceedings of the 1998 conference on the history and heritage of science information systems*. Information Today.

Fry, J. (2006). Scholarly research and information practices: A domain analytic approach. *Information Processing & Management*, *42*(1), 299–316. https://doi.org/10.1016/j.ipm.2004.09.004

Fry, J., & Talja, S. (2007). The intellectual and social organization of academic fields and the shaping of digital resources. *Journal of Information Science*, *33*(2), 115–133. https://doi.org/10.1177/0165551506068153

Funari, M. (2014). Research data and humanities: An European overview. *JLIS*, *5*(1), 209–239. https://doi.org/10.4403/jlis.it-8927

Given, L. M., & Willson, R. (2018). Information technology and the humanities scholar: Documenting digital research practices. *JASIST*, *69*(6), 807–819. https://doi.org/10.1002/asi.24008

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine.

Gomes, D. G. E., Pottier, P., Crystal-Ornelas, R., Hudgins, E. J., Foroughirad, V., Sánchez-Reyes, L. L., Turba, R., Martinez, P. A., Moreau, D., Bertram, M. G., Smout, C. A., & Gaynor, K. M. (2022). Why don't we share data and code? Perceived barriers and benefits to public archiving practices. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1987), 20221113. https://doi.org/10.1098/rspb.2022.1113

Gomez-Diaz, T., & Recio, T. (2022). Research software vs. research data I: Towards a research data definition in the open science context. *F1000Research*, *11*, 118. https://doi.org/10.12688/f1000research.78195.2

Goodey, G., Hahnel, M., Zhou, Y., Jiang, L., & Chandramouliswaran, I. (2022). *The state of open data 2022* (Report). Digital Science, Springer Nature, Figshare.

Government of Canada, Innovation, Science and Economic Development. (2021). *Tri-Agency research data management policy* (Report). Author. Retrieved from https://science.gc.ca/site/science/en/interagency-research-funding/policies-and-guidelines/research-data-management

Gregory, K., Ninkov, A., Ripp, C., Roblin, E., Peters, I., & Haustein, S. (2023). Tracing data: A survey investigating disciplinary differences in data citation. *Quantitative Science Studies*, *4*, 1–51. https://doi.org/10.1162/qss_a_00264

Gupta, N., Martindale, A., Supernant, K., & Elvidge, M. (2023). The CARE principles and the reuse, sharing, and curation of indigenous data in Canadian archaeology. *Advances in Archaeological Practice*, *11*(1), 76–89. https://doi.org/10.1017/aap.2022.33

Houf, J. R. (2021). Boundary work and boundary objects: Synthesizing two concepts for moments of controversy. *Journal of Technical Writing and Communication*, *51*(3), 293–312. https://doi.org/10.1177/0047281620947355

Huvila, I. (2022a). Improving the usefulness of research data with better Paradata. *Open Information Science*, *6*(1), 28–48. https://doi.org/10.1515/opis-2022-0129

Huvila, I. (2022b). Making and taking information. *JASIST*, *73*(4), 528–541. https://doi.org/10.1002/asi.24599

Huvila, I., & Sinnamon, L. (2022). Sharing research design, methods and process information in and out of academia. *Proceedings of the Association for Information Science and Technology*, *59*(1), 132–144. https://doi.org/10.1002/pra2.611

Jarrahi, M. H., Memariani, A., & Guha, S. (2023). The principles of data-centric AI. *Communications of the ACM*, *66*(8), 84–92. https://doi.org/10.1145/3571724

Kalkman, S., Mostert, M., Gerlinger, C., van Delden, J. J. M., & van Thiel, G. J. M. W. (2019). Responsible data sharing in international health research: A systematic review of principles and norms. *BMC Medical Ethics*, *20*(1), 21. https://doi.org/10.1186/s12910-019-0359-9

Karras, O., Groen, E. C., Khan, J. A., & Auer, S. (2021). Researcher or crowd member? Why not both! The open research knowledge graph for applying and communicating CrowdRE research. In *2021 IEEE 29th international requirements engineering conference workshops (REW)* (pp. 320–327). IEEE. https://doi.org/10.1109/REW53955.2021.00056

Kim, Y., & Stanton, J. M. (2015). Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis. *JASIST*, *67*(4), 776–799. https://doi.org/10.1002/asi.23424

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures & their consequences*. SAGE.

Knorr-Cetina, K. (2003). *Epistemic cultures: How the sciences make knowledge*. Harvard University Press.

Koesten, L., Gregory, K., Groth, P., & Simperl, E. (2021). Talking datasets—Understanding data sensemaking behaviours. *International Journal of Human-Computer Studies*, *146*, 102562. https://doi.org/10.1016/j.ijhcs.2020.102562

Kowalczyk, S., & Shankar, K. (2011). Data sharing in the sciences. *ARIST*, *45*(1), 247–294. https://doi.org/10.1002/aris.2011.1440450113

Kraus, W., & Eberhard, I. (2022). Managing data, managing contradictions: Archiving and sharing ethnographic data. In M. Burkhardt, D. van Geenen, C. Gerlitz, S. Hind, T. Kaerlein, D. Lämmerhirt, & A. Volmar (Eds.), *Interrogating datafication: Towards a praxeology of data* (pp. 185–206). transcript Verlag. https://doi.org/10.14361/9783839455616

Kurata, K., Matsubayashi, M., & Mine, S. (2022). Identifying the complex position of research data and data sharing among researchers in natural science. *SAGE Open*, *7*(3), 2158–2440. https://doi.org/10.1177/2158244017717301

Latour, B. (1992). Where are the missing masses? The sociology of a few mundane artifacts. In W. E. Bijker & J. Law (Eds.), *Shaping technology/building society: Studies in sociotechnical change* (pp. 225–258). MIT Press.

Latour, B. (1999). *Pandora's hope: Essays on the reality of science studies*. Harvard University Press.

Latour, B., & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts*. Princeton University Press.

Leonelli, S. (2023). *Philosophy of open science*. Cambridge University Press.

Lian, Z., Oliver, G., & Chen, Y. (2023). Genres of online COVID-19 information and government information culture: A comparative case study. *Library & Information Science Research*, *45*(4), 101263. https://doi.org/10.1016/j.lisr.2023.101263

Lin, C.-S. (2023). Exploring humanities researchers' perceptions of “data”: A phenomenological approach. *Proceedings of the Association for Information Science and Technology*, *60*(1), 1052–1054. https://doi.org/10.1002/pra2.942

Lin, J., & Strasser, C. (2014). Recommendations for the role of publishers in access to data. *PLoS Biology*, *12*(10), e1001975. https://doi.org/10.1371/journal.pbio.1001975

Mayernik, M. S. (2011). Metadata tensions: A case study of library principles vs. everyday scientific data practices. *Proceedings of ASIS&T-AM*, *47*(1), 1–2. https://doi.org/10.1002/meet.14504701337

Michener, W. K. (2015). Ecological data sharing. *Ecological Informatics*, *29*, 33–44. https://doi.org/10.1016/j.ecoinf.2015.06.010

Niu, J., & Hedstrom, M. (2008). Documentation evaluation model for social science data. *Proceedings of the American Society for Information Science and Technology*, *45*(1), 1–11. https://doi.org/10.1002/meet.2008.1450450223

Nugroho, R. P., Zuiderwijk, A., Janssen, M., & de Jong, M. (2015). A comparison of national open data policies: Lessons learned. *Transforming Government: People, Process and Policy*, *9*(3), 286–308.

Oliver, G., Cranefield, J., Lilley, S., & Lewellen, M. (2023a). Data cultures: A scoping literature review. *Information Research*, *28*(1), 3–29.

Oliver, G., Cranefield, J., Lilley, S., & Lewellen, M. J. (2023b). Understanding data culture/s: Influences, activities, and initiatives. *Annual Review of Information Science and Technology*, *75*(3), 201–214. https://doi.org/10.1002/asi.24737

Özsu, M. T. (2023). Data science—A systematic treatment. *Communications of the ACM*, *66*(7), 106–116. https://doi.org/10.1145/3582491

Pampel, H., & Dallmeier-Tiessen, S. (2014). Open research data: From vision to practice. In S. Bartling & S. Friesike (Eds.), *Opening science* (pp. 213–224). Springer. https://doi.org/10.1007/978-3-319-00026-8_14

Peels, R., & Bouter, L. (2018). The possibility and desirability of replication in the humanities. *Palgrave Communications*, *4*(1), 1–4. https://doi.org/10.1057/s41599-018-0149-x

Pelizza, A., & Van Rossem, W. R. (2023). Scripts of alterity: Mapping assumptions and limitations of the border security apparatus through classification schemas. *Science, Technology, & Human Values*, *49*(4), 794–826. https://doi.org/10.1177/01622439231195955

Pels, P., Boog, I., Henrike Florusbosch, J., Kripe, Z., Minter, T., Postma, M., Sleeboom-Faulkner, M., Simpson, B., Dilger, H., Schönhuth, M., von Poser, A., Castillo, R. C. A., Lederman, R., & Richards-Rissetto, H. (2018). Data management in anthropology: The next phase in ethics governance? *Social Anthropology*, *26*(3), 391–413. https://doi.org/10.1111/1469-8676.12526

Pilerot, O. (2012). LIS research on information sharing activities—People, places, or information. *Journal of Documentation*, *68*(4), 559–581. https://doi.org/10.1108/00220411211239110

Pilerot, O. (2013). A practice theoretical exploration of information sharing and trust in a dispersed community of design scholars. *Information Research*, *19*(4).

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

Prainsack, B., El-Sayed, S., Forgó, N., Szoszkiewicz, L., & Baumer, P. (2022). Data solidarity. In *White paper*. Governing Health Futures.

Storm, M., & Borgman, H. (2020). Understanding challenges and success factors in creating a data-driven culture. In T. X. Bui (Ed.), *Proceedings of the 53rd Hawaii international conference on system sciences* (pp. 5399–5408). University of Hawai'i at Mānoa.

Tabak, E., & Willson, M. (2012). A non-linear model of information sharing practices in academic communities. *Library & Information Science Research*, *34*(2), 110–116. https://doi.org/10.1016/j.lisr.2011.11.002

Talja, S. (2002). Information sharing in academic communities: Types and levels of collaboration in information seeking and use. *New Review of Information Behaviour*, *3*, 143–159.

Talja, S., Tenopir, C., & Late, E. (2022). Desired affordances of scholarly E-articles: Views from scholars based on open-ended answers. *Libri*, *72*(1), 67–81. https://doi.org/10.1515/libri-2021-0012

Tal-Socher, M., & Ziderman, A. (2020). Data sharing policies in scholarly publications: Interdisciplinary comparisons. *Prometheus*, *36*(2), 116–134. https://doi.org/10.13169/prometheus.36.2.0116

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE*, *6*(6), e21101. https://doi.org/10.1371/journal.pone.0021101

Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., & Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS ONE*, *10*(8), e0134826. https://doi.org/10.1371/journal.pone.0134826

Tenopir, C., Rice, N. M., Allard, S., Baird, L., Borycz, J., Christian, L., Grant, B., Olendorf, R., & Sandusky, R. J. (2020). Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLoS ONE*, *15*(3), e0229003. https://doi.org/10.1371/journal.pone.0229003

Thelwall, M., & Kousha, K. (2017). Do journal data sharing mandates work? Life sciences evidence from Dryad. *Aslib Journal of Information Management*, *69*(1), 36–45.

Thessen, A., & Patterson, D. (2011). Data issues in the life sciences. *ZooKeys*, *150*, 15–51. https://doi.org/10.3897/zookeys.150.1766

Thoegersen, J. L., & Borlund, P. (2022). Researcher attitudes toward data sharing in public data repositories: A meta-evaluation of studies on researcher data sharing. *Journal of Documentation*, *78*(7), 1–17. https://doi.org/10.1108/JD-01-2021-0015

Tu, Z., & Shen, J. (2023). Value of open research data: A systematic evaluation framework based on multi-stakeholder survey. *Library & Information Science Research*, *45*(4), 101269. https://doi.org/10.1016/j.lisr.2023.101269

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

Wynholds, L., Fearon, D. S., Jr., Borgman, C. L., & Traweek, S. (2011). When use cases are not useful: Data practices, astronomy, and digital libraries. In *Proceedings of the JCDL*. ACM.

York, J. (2022). *Seeking equilibrium in data reuse: A study of knowledge satisficing* (Thesis). https://doi.org/10.7302/6170

Zhi, F., Zhang, M., Zhang, S., Cheng, C., & Shen, T. (2023). Can social capital and planned behaviour favour an increased willingness to share scientific data? Evidence from data originators. *The Electronic Library*, *41*(4), 456–473. https://doi.org/10.1108/EL-01-2023-0005

Zuiderwijk, A., Shinde, R., & Jeng, W. (2020). What drives and inhibits researchers to share and use open research data? A systematic literature review to analyze factors influencing open research data adoption. *PLoS ONE*, *15*(9), e0239283. https://doi.org/10.1371/journal.pone.0239283

Zuiderwijk, A., & Spiers, H. (2019). Sharing and re-using open data: A case study of motivations in astrophysics. *International Journal of Information Management*, *49*, 228–241. https://doi.org/10.1016/j.ijinfomgt.2019.05.024

How to cite this article: Huvila, I., & Sinnamon, L. S. (2024). When data sharing is an answer and when (often) it is not: Acknowledging data-driven, non-data, and data-decentered cultures. *Journal of the Association for Information Science and Technology*, 1–16. https://doi.org/10.1002/asi.24957