UC Irvine

UC Irvine Previously Published Works

Title

Epistemic Disadvantage in Open Science Practices

Permalink

https://escholarship.org/uc/item/2c38h5r8

Journal

Social Epistemology, 40(5)

ISSN

0269-1728

Authors

Alcalay, Rena

Vandekerckhove, Joachim

Publication Date

2026-09-03

DOI

10.1080/02691728.2026.2669557

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Epistemic disadvantage in open science practices

Rena Alcalay[a,†] **and Joachim Vandekerckhove**[b]

Author final version.

Alcalay, R., & Vandekerckhove, J., (2026). Epistemic disadvantage in open science practices. *Social Epistemology.*

Open science advocates aim to address epistemic injustice and global power asymmetries in research infrastructures yet underestimate structural vulnerabilities and risks embedded in data sharing practices. Drawing on an ethnographic study of the GCRF South–South Migration Inequality and Development Hub (MIDEQ), we argue that even well-intentioned systems, such as the MIDEQ hub, do not reach these vulnerabilities. To capture these overlooked harms, we employ the concept of epistemic disadvantage, a category coextensive with, but separate from epistemic injustice. Epistemic disadvantage includes ethically justified data withholding and misappropriation within consensual systems, which is not captured by existing notions of extractivism as those focus on the monetization of epistemic resources without regard for the loss incurred by the originating community. To better account for these harms, we draw on two recent epistemic attitudes: the Book of Truths (BoT), which sees scientific truths as emerging through peer review, and the Book of Conversations (BoC), which values data for the interactions and reflections it provokes. Even when epistemic values recognize knowledge as relational and situated, extractivist harms persist. We offer an ameliorated concept of extractivism to include subtler, systemic harms rooted in epistemic disadvantage.

attitudes | epistemic disadvantage | epistemic injustice | extractivism | open science

Open Science (OS) is a set of practices and ideals aimed at making scientific research more transparent, inclusive, and reproducible. However, it has been noted that the term OS functions as an umbrella term encompassing wide disputes about the future of knowledge creation and dissemination, evoking quite different understandings depending on the viewpoint of its advocates (Fecher & Friesike, 2013, p. 44). At its core, OS emphasizes the free sharing of research data, methods, software, and publications to foster wider participation and accelerate knowledge production. Indeed, a vital aspect of effectively implementing an open science is facilitating the disclosure and exchange of research components (Leonelli, 2023, pp. 24–25). However, as scholars have argued, these normative aims are often complicated by asymmetries in resources and infrastructure across the global research community and the risks associated with disseminating research data (Levin & Leonelli, 2016; Leonelli, 2022, 2023). While OS aspires to democratize scientific inquiry, its implementation frequently raises the question: Openness for whom, under what conditions, and at what cost.

Underlying this question is an unexamined tension in how researchers conceptualize the value and purpose of scientific knowledge. Recent work distinguishes between two attitudes toward scientific literature: a Book of Truths (BoT) perspective and a Book of Conversations (BoC; Davis-Stober et al., 2025). Those holding a BoT attitude see the primary purpose of the academic literature as creating a definitive, reliable repository of established knowledge, a canon that can be assessed and trusted at face value by expert and lay readers alike. By contrast, those adopting a BoC attitude view academic literature primarily as an ongoing exchange of ideas and communication among researchers, valuing provisional findings, preliminary data, and emergent questions as inherently worthwhile contributions to cumulative science, regardless of whether they achieve formal publication or definitive verification.

These divergent attitudes imply fundamentally different thresholds for what counts as acceptable epistemic risk in data sharing. Where a BoC attitude treats intermediate, unvetted data as intrinsically valuable, a BoT attitude worries that sharing raw findings risks misleading the scientific community or distorting the epistemic record. What appears to one researcher as reckless exposure to another appears as essential democratization. This attitudinal divide, we argue, underlies many tensions in contemporary OS practice. In this paper, we examine how OS initiatives are complicated not only by global asymmetries in resource and infrastructure, but also by these fundamental attitudes about what knowledge is for and how much risk is acceptable in service of scientific progress.

Against this backdrop, an ongoing concern is how to effectively 'open up' science so that it does not meaningfully contribute to inequitable data access. Toward this aim, scholars have critically examined the advantages and challenges of OS practices, particularly focusing on how these practices can inadvertently create or perpetuate epistemic inequalities between researchers from different resource levels and geographical regions. Leonelli (2022), for example, illustrates the complexities and unintended consequences of OS by examining three representative cases where the application of OS principles leads to epistemic harms—ethical and knowledge-based harms that arise from how knowledge is produced, accessed, or distributed—particularly for researchers in less-resourced environments: firstly GISAID and the Problem of "Not Open Enough", secondly Proprietary Software, and thirdly Reproducibility.

[a]Department of Science, Technology and Society, Technical University of Munich; Contact: rena.alcalay@tum.de; ORCiD: 0000-0001-9183-9917; [b]Department of Cognitive Sciences; Department of Statistics; Department of Logic & Philosophy of Science; University of California, Irvine; Contact: joachim@uci.edu; ORCiD: 0000-0003-2600-5937
[†]To whom correspondence should be addressed. E-mail: rena.alcalay@tum.de.

Acknowledgements. We gratefully acknowledge the valuable discussions and feedback from participants in the "Epistemic Diversity" reading group, organized by Rose Trappes and Nathanael Sheehan, as well as Karl Landström's insightful contributions. This manuscript has been significantly strengthened by the detailed and constructive reports from two anonymous reviewers, to whom we extend our sincere thanks. We are also deeply appreciative of the support provided by Sabina Leonelli and the whole PHIL_OS team.

Conflict of Interest. The authors have no competing interests to declare that are relevant to the content of this article.

The first case concerned the Global Initiative on Sharing All Influenza Data (GISAID), launched in 2008 to encourage the sharing of viral genome data. GISAID was rapidly expanded in 2020 to include SARS-CoV-2 data, rising to international prominence during the COVID-19 pandemic. Because sequences are tied to geographic and institutional origin, they reveal which countries have detected which variants – information well-resourced researchers could exploit. GISAID implemented restrictions on data mining and linkage to protect contributors, particularly those from lower-resourced institutions, from exploitation. These governance mechanisms were intended to ensure that data providers retained some control over how their data was used, and that proper credit and collaboration could be negotiated.

Critics, however, claimed that GISAID access restrictions posed "barriers that restrain effective data sharing," running counter to the principle of openness (Leonelli, 2022, p. 994). However, Leonelli argued that such restrictions were in fact attempts to govern equitable participation, especially by safeguarding researchers in low-resource settings from having their data appropriated by better-resourced institutions without credit or collaboration. The paradox here is striking: efforts to prevent epistemic exploitation were themselves labeled as epistemically exclusionary. The epistemic harm in this case was that by restricting data mining and linkage to govern the structure of data sharing, GISAID excluded some groups from access to knowledge. In trying to guard against exploitation, the initiative inadvertently restricted participation in the broader epistemic community. This case exemplifies a deeper tension not only in OS but in knowledge-practice more broadly: taking measures to restrict some groups from accessing knowledge can be justified – a point we will return to later.

The second case contrasted the asymmetry between free and proprietary research software. OS practice tends to favor open-source tools, but these are often developed and valorized in the Global North with the means to examine larger sampled experiments or construct simulations with more computing power. Moreover, open-source software has been frequently stigmatized in scientific evaluation, especially when used by researchers in institutions lacking international preside (Vermeir et al. 2018, as cited by Leonelli, 2022, p. 994). This reflects a broader epistemic harm: epistemic tools are only effective when embedded within well-resourced infrastructures. Free tools may not carry the same scientific authority, even if technically sound, because credibility follows well-resourced distribution.

The third case addressed the reproducibility crisis, underscoring the varying methodological demands across different settings and disciplines. Standards for replicability are often drawn from highly controlled experimental settings—usually in the natural sciences—and then generalized as benchmarks for other disciplines. This becomes problematic when applied across fields with different epistemic cultures, data types, or research targets. Leonelli (2022, p. 996) warns that treating certain experimental forms as the gold standard can marginalize valid but non-conforming methodologies, such as ethnographic research or small-scale community-based studies. The epistemic harm occurs as evaluative methods are weighed with more or less credibility across domains with no strategic notation of parallel methods, goals or concepts within multiple paradigms. Methodological pluralism has been advocated as an important metric of scientific epistemology, yet OS can inadvertently enforce methodological uniformity that delegitimizes alternative forms of knowing (Cartwright, 2021).

Each of these cases exemplifies a form of epistemic harm – not necessarily the direct suppression of knowledge but rather exclusions from participation, misrecognition of valid methods, unjust credibility hierarchies, and exploitation of labor without recognition. These harms challenge the ideal that openness is always beneficial. Sometimes, more openness increases vulnerability – particularly for researching lacking the resources or institutional power to protect their contributions.

Asymmetry in infrastructure, resources, storage, and computational capacities between scientists in high-income and low-income institutions is a considerable challenge to the practicalities of 'opening up' data-intensive fields. Meaningful openness must encapsulate the equitable accessibility to all users at the same level of effectiveness (Bezuidenhout, 2025). Questions have been raised about who benefits from data sharing and its impact on equity and whether epistemic openness should be prioritized for the sake of better knowledge production even when it may come at the cost of epistemic sovereignty, credit, or justice for those producing the knowledge (Staunton et al., 2021).

An underexplored area in the democratization of scientific practice is how much risk of harm should be permitted—or is ethically justifiable—in the pursuit of epistemic or scientific goals. This reframes the question from equitable data access to epistemic harm thresholds. This question is especially pertinent in data-sharing practices where the pursuit of comparative insight requires sharing data. But sharing data exposes researchers to a serious risk that, for example, their intellectual labor may be extracted, misinterpreted, uncredited, or repurposed. Researchers in low-resourced environments may face a double bind, as the epistemic risks that enable comparative insight and cross-corridor understanding—affordable to well-funded institutions—may be inaccessible to them.

We ground this discussion in an ethnographic study of the GCRF South-South Migration Inequality and Development Hub's (MIDEQ), whose data-sharing arrangements revealed anxiety about potential epistemic extractivism where resources could be removed from their context without perceived loss of value (Landström, 2024). While we engage critically with aspects of Landström's interpretation, we argue that his observations point to a distinct form of epistemic harm: epistemically disadvantageous extractivism. Co-extensive with epistemic injustice, epistemic disadvantage better captures the structural vulnerabilities placed on researchers when the goals of cumulative science are pursued even under equitable distribution of costs and credit. It also helps clarify the epistemic harms that arise from warranted exclusions in extractivist practices that typify epistemic disadvantage (Goldstein, 2022).

In the next section, we provide a conceptual and historical overview of epistemic harm and argue for expanding the concept to better capture a wider range of harms relevant to OS. Then, we explore how two attitudes toward scientific literature, the Book of Truths and Book of Conversations, frame the value of data and the threshold for epistemic risk, when sharing data. Then we briefly summarize the evolving landscape of data sharing, setting the stage for examining the application of extractivism to these practices. Finally, we critically evaluate the extractivist framework in this context and propose an ameliorated version that better reflects the epistemic and ethical complexities of diverse OS environments.

Epistemic harm

Epistemic harm was brought to prominence in Fricker's (2007) epistemic injustice framework, as broadly ethical and knowledge-based wrongs stemming from marginalization, prejudice, ignorance, and

other forms of exclusion rooted in asymmetrical relations of power. These harms affect how agents see themselves as knowers by undermining credibility, casting doubt on legitimate evidence, hindering the development of epistemic capacities and access to further epistemic resources (Fricker, 2007, pp. 47–48). As Fricker (2007, p. 145) puts it, individuals are epistemically harmed when they are "wrongfully excluded from participation in the practice that defines the very core of the very concept of knowledge."

Since then, epistemic harms have been identified and categorized in numerous ways. For example, Kidd and Carel (2016) identified epistemic harm in medical contexts where patients are epistemically isolated, such as when diagnoses are withheld from patients. In adherence to the principle of non-maleficence, physicians sometimes withhold diagnoses of serious conditions (such as cancer) to spare patients psychological distress. Although ethically motivated, this practice deprives patients of the opportunity to make sense of their condition and participate in their care (Kidd & Carel, 2016, p. 225). Spencer (2024, p. 218) further emphasized the injustice of imposed ignorance where patients are denied resources. For instance, public misconceptions about conditions such as autism, PTSD, or OCD can prevent individuals from recognizing symptoms due to trivialization or distortions of the illness (cf. Scriver, 2019; Spencer & Carel, 2021). The resulting cognitive disadvantage impedes individuals capacity for self-understanding and agency (Spencer, 2024, p. 219).

Building on this foundation, Peña-Guzmán and Reynolds (2019) expanded the scope of epistemic harms by examining how ableism heightens the risk of medical error for people with disabilities. Ableism privileges ability and assumes that "abnormal" bodies are necessarily negative (Peña-Guzmán & Reynolds, 2019, p. 214). This affects expectations of bodies based on perceived disability status, leading to assumptions about "normal" body size, shape, motion, and function. It also influences statistical analyses of body metrics, determining "normal" blood pressure or "normal" levels of anxiety. These assumptions can result in significant epistemic harm for individuals with disabilities, as their medical needs and conditions may be misunderstood or inadequately addressed.

Attention to epistemic harms related to error was further deepened by Gardiner (2020), who examined broader patterns in how the remoteness of error possibilities can be over- and underestimated. Relevant alternatives can be socially patterned by prejudice, biases, emotions, upbringing, and social context. When patterned on non-epistemic attributes, an individual's epistemic position is harmed by rendering only certain kinds of error possibilities relevant, leading to mistreatment through lying or undue skepticism (Gardiner, 2020, p. 485). Gardiner also identifies a novel species of testimonial injustice, where assertions require evidence addressing the error possibility that the speaker is not lying or mistaken, such as in rape accusations (Gardiner, 2020, p. 506). Examples include crying wolf, conspiracy theories, and gaslighting.

The epistemic harms detailed above fall within the category of epistemic injustice. Yet, as Goldstein (2022) argued, not all epistemic harms should be evaluated as unjust. Goldstein analyzed a case in anesthesiology history involving curare, a known paralytic poison, was experimentally used to anesthetize children without other anesthetic additives. Due to an identity prejudice against children, perceived as neither fully rational or fully capable of agency, patient testimony that they were paralyzed and not numb, was ignored, resulting in a severe case of testimonial injustice. Yet the complexity of the hermeneutical gap revealed two kinds of epistemic harms: unjust and warranted. As child-patients, they were vulnerable to physicians obscuring understanding of their own pain. While physicians acknowledged the children's reports of pain, they misattributed it, failing to connect it to the absence of a numbing agent. In such cases, where exclusion is not driven by identity prejudice, the resulting harm may be merely epistemically disadvantageous, but not unjust.

What we mean by epistemic disadvantage draws on Goldstein (2022), who initially characterized it through three conditions: (1) harms are not deliberate, (2) speakers lack precision of concepts to effectively communicate their experiences, and (3) the exclusion is warranted in the epistemic practice (Goldstein, 2022, pp. 1861–1878). Spencer (2024, p. 224) has helpfully amended the first condition, arguing that harms need not be deliberate for it to be an instance of epistemic injustice. By distinguishing epistemic disadvantage from injustice, Goldstein engaged a broader category of structurally rooted harms within asymmetrical epistemic relations. Indeed, various frameworks for understanding structural epistemic harms have been proposed but much remains to be explored and clarified.[1]

For example, Kristie Dotson (2011; 2014) highlighted the limitation of Fricker's framework in capturing structural harms, except in the case of hermeneutical gaps.[2] To remedy this, Dotson (2014, p. 115) defined epistemic oppression as persistent epistemic exclusions preventing reasoners from contributing to knowledge production. Influenced by Bartunek and Moch (1987) schemata for organizational change, Dotson identified three levels of systemic exclusions: the first-order level, involving inconsistencies in system utilization, but the schemata (system) are not itself problematic. Changing first-order exclusions may not require changing entire beliefs or values about credibility assessments, but perhaps by educating about inefficient credibility assessments (Dotson, 2014, p. 12). Second-order exclusions result from insufficient shared resources, such as pre-existing gaps in collective understanding that obscure the interests of particular knowers and may necessitate change to the prevailing conceptual resources (Dotson, 2014, p. 14). Third-order exclusions compromise epistemic agency from the inadequacy of the entire system, where change would require 'transcending the schemata' (Bartunek & Moch, 1994, p. 24, as cited by Dotson, 2014, p. 5).

Building on Dotson's work, Pohlhaus (2020, p. 237) examined epistemically oppressive systems that include, rather than exclude, people. She highlighted how systems can be exploitative by coercively and non-reciprocally extracting epistemic labor in ways that benefit some while harming others. When systems are deployed inconsistently, contain targeted gaps, or are altogether inapt for up-taking the experiences of marginalized groups, even partial or full inclusion can produce substantial epistemic harm. Some inclusions can be epistemically exploitative, extracting epistemic labor solely for the benefit of others (Pohlhaus, 2020, p. 241).

Indeed, Alcoff's (2022) concept of epistemic extractivism captures epistemic harm arising from both exclusions and exploitative forms of inclusion. Drawing on the Marxist distinction between

[1] See Leonelli in this issue. Additionally, it has been debated whether testimonial injustice can take a structural form. Wanderer (2017) argued against recognizing it as a distinct category, while Anderson (2012) defended its legitimacy.

[2] Other limitations have been discussed by, for example, Rebecca Mason (2011). Indeed, Medina (2017) offers a nuanced framework for understanding hermeneutical injustice, identifying four types—two of which are especially relevant to structural dynamics. The first classifies semantically produced source where harms result from the absence of labels (e.g., sexual harassment). The second designates structural or institutional dynamics that prevent use of hermeneutical resources and expressive styles (e.g., categories on questionnaires). Third and fourth are the dimensions of breadth—how widely the injustice is distributed—and depth—how severely it undermines the affected group's capacity to make and share meaning (pgs. 45–47).

use-value or exchange-value (rather than normative-value commitments), she defines epistemic extractivism as the practice of extracting monetized value from epistemic resources without regard for the loss incurred by the originating community (Alcoff, 2022, p. 4). This practice is deeply rooted in colonial history and perpetuated by ongoing global inequalities as exemplified by bioprospecting. In such cases, corporations leverage legal mechanisms, like patent rights, to claim superior knowledge in developing resources such as food or health additives, effectively disregarding the traditional knowledge of indigenous communities (p. 9). A notable case of bioprospecting is the Neem tree, long valued in South Asia for its pharmacological properties. When its use as a pest control agent drew corporate interest, particularly from Western and Japanese firms, it triggered a patent dispute. These corporations stripped the Neem of its cultural context to meet international patent standards, presenting it as "original" and "innovative" (p. 11). Alcoff showed how such extraction not only alters the ontology of knowledge resources but also displaces indigenous communities. Transnational pharmaceutical companies often seize land for agro-development, forcing displacement and turning local populations into labor for the very industries exploiting their knowledge (p. 7).

While the epistemic harm may appear self-evident, it is important to clarify that extractivism operates on the assumption that epistemic resources can be isolated, commodified, and assigned market value, reducing knowledge to raw data stripped of interpretation and context. Alcoff (2022) argues that this reduction constitutes epistemic injustice by commodifying data—and those who produce it—rather recognizing knowledge as relational and situated (cf. Leonelli, 2016). The injustice lies in unreflective, non-collaborative methods that ignore the epistemic agency of those from whom knowledge is taken, enabling the exploitative appropriation of epistemic labor (cf. Tanesini, 2022).

While Alcoff's concept of epistemic extractivism offers a powerful lens for identifying structural harms in knowledge practices, not all instances of harms in data sharing practices are captured by this framework. We will argue that the mechanics of extraction—taking something from a source, potentially altering its context or value to the originator—can also occur outside market-driven contexts and are not always unwarranted, leading to what we term 'epistemically disadvantageous extractivism'. This distinction becomes clearer in contexts like MIDEQ, where internal data-sharing arrangements reflect a Book of Conversations attitude, treating data as relational and intrinsically valuable, rather than reducible to use-value. Such arrangements resist commodification but still raises questions about epistemic harms and exclusions.

While frameworks of epistemic injustice and disadvantage help us identify where and how harms occur, they do not fully explain *why* well-intentioned systems (that is, systems designed with awareness of power asymmetries and equipped with equitable protocols) still generate exclusion. A complementary insight emerges from examining researchers' deeper, often implicit assumptions about what scientific knowledge is for and what level of uncertainty and risk is acceptable in its circulation. As we will argue, different epistemic attitudes toward the purposes of scientific literature imply different evaluative thresholds: they shape how researchers interpret the same data-sharing practice, whether they see it as democratizing or extractive. To understand how OS initiatives harm, we must examine these underlying attitudes and how they shape tolerance for epistemic risk that data sharing demands.

Attitudes toward scientific publication

Having established that epistemic harm can arise from how knowledge is produced and distributed, we now turn to the epistemic attitudes that shape how such harms are tolerated or mitigated in scientific publication. In this section, we elaborate on two recently proposed attitudes toward scientific literature: the Book of Truths attitude (BoT) and Book of Conversations attitude (BoC). We argue that the BoC attitude values data for the interactions and reflections provoked, rather than for its role in producing finalized outputs. In contrast, the BoT attitude prioritizes scientific truths, which are more likely to emerge through peer review processes that vet and revise the inferences and conclusions drawn from data. As we will show in the section on the harm of extractivism, this distinction complicates labeling some epistemic harms in data sharing practices unjust.

Davis-Stober et al. (2025) discussed two cognitive attitudes to describe replication and the normative role of peer-reviewed literature. Each attitude characterizes the extremes of a spectrum: on one end, an attitude toward scientific publication as a Book of Truths (BoT) and on the other an attitude of publishing as a Book of Conversations (BoC). Scientists holding the BoT believe that the purpose of the academic peer-reviewed literature is to create a definitive collection of established truths. For them, the primary purpose of the literature is to serve as a reliable repository of knowledge, which may be accessed, trusted at face value, and put into practice by experts and lay readers (e.g., policymakers and journalists) alike. Examples of the BoT may be found in the many publications on the reproducibility of published research (e.g., Munafò et al. 2017) and indeed in the very use of "replicability" as a property of a paper.

At the other extreme lies the BoC attitude. Those who hold this attitude see academic literature primarily as a means of communication and an ongoing exchange of ideas among scientists. Here, the purpose of the literature is to facilitate dynamic dialogue. Each claim within such a literature is to be considered carefully as a contribution to this conversation, but not necessarily believed to be definitively true or replicable. Emphasis here is less on the robustness of individual statements and more on the literature serving as an accurate and detailed record of what research was conducted and what ideas were proposed. While this might accelerate discovery, it could also result in inefficiencies due to inconsistent results or make the literature difficult for non-experts to navigate without field-specific expertise to assess the robustness of claims. A notable example of the BoC is due to Dang and Bright (2021), who argue that it would hamper the progress of science if scientific claims were held to the same stringent standards as assertions of facts.

The postulation of this spectrum by Davis-Stober et al. (2025) is intended to be descriptive, not normative. However, to the extent that consumers and producers of the academic literature differ in how they view its purpose, guidance on the proper and contextual application of these attitudes would be valuable for scientific readers and writers. We think a BoC attitude reasonably claims that data in repositories hold intrinsic value due to the significance from the interactions and reflections they provoke. In contrast, a BoT attitude prioritizes tangible outputs, such as publications that contain fully formed scientific claims, over the interactive potential with data.

These two attitudes imply different evaluative responses to the risks of sharing data.[3] Since the BoC attitude intrinsically values

[3] Risk attitudes have been empirically studied by economists since the 1950s (cf. Arrow 1951).

Table 1. Concept Taxonomy Table

Concept	Definition	Key Feature	Example in OS Context
Epistemic Injustice	Wrongful exclusion from knowledge practices due to identity prejudice, credibility deficits, or power asymmetries	Unjust by definition; rooted in prejudice	OS initiatives require data to be in specific formats, languages, or accessible through platforms that exclude researchers without institutional bandwidth to prepare their work
Epistemic Disadvantage	Harm arising from exclusions that are warranted within epistemic practice, not driven by prejudice	Not unjust, but ethically problematic	A research team withholds data to protect against exploitation, preventing collaborative insights
Extractivism	Appropriation of knowledge resources without regard to loss incurred by originators; typically involves commodification	Market-driven; knowledge is reduced to exchangeable value	Corporations patent indigenous knowledge without benefit-sharing
Epistemically Disadvantageous Extractivism	Extraction of knowledge that occurs within ethically-structured systems	Occurs even with protocols designed to overturn asymmetrical power relations; no commodification required	A research team shares qualitative data with consent and in good faith to a repository. Another team downloads the data, conducts secondary analysis, and, despite formal agreements about attribution, publishes the findings that misrepresent the original context. (*See "The harm of extractivism" for detailed analysis.)*

intermediately published data, such as datasets shared before formal publication, preliminary findings presented at conferences, or even insights exchanged on social media platforms, it tolerates uncertainty in favor of broader epistemic participation. From this perspective, data should be published even if it does not meet conventional thresholds of completeness, significance, or methodological rigor. The rationale is that imposing arbitrary publication standards risks censoring potentially valuable datasets that could, over time, contribute to the emergence of new research questions or even entire fields of study. Rather than filtering data through rigid criteria at the point of dissemination, this attitude encourages treating each dataset as a provisional contribution – a piece of a larger epistemic puzzle. Researchers and readers are then invited to interpret and contextualize the data post hoc, integrating it into broader conversations as relevance and connections emerge.

From the BoT perspective, the value of data is tightly coupled to the final analysis presented in a published paper. On this view, a paper's legitimacy hinges on the logical coherence between the data and the claims it supports. Publishing a paper where the evidential basis is weak risks misleading the scientific community and distorting the epistemic record. The BoT attitude thus upholds a high threshold for epistemic risk, insisting that only well-substantiated findings enter the literature, and indeed the publication process serves as a critical filter.

Compared to BoT, the BoC attitude has a lower threshold of risk to ensure there is greater access to intermediate stages of science whereas the BoT attitude maintains a higher threshold for what can be shared. To illustrate, imagine a Little Free Library where neighbors are encouraged to take and leave books freely. Sharing data is like placing a book in the Little Free Library. With a BoC attitude, the goal is to maximize circulation, even if that means accepting risks like misuse or misinterpretation. By contrast, a BoT attitude sees premature sharing as a threat to knowledge integrity, risking introducing incomplete or misleading information into the epistemic record.[4]

These are different risk tolerances. From the BoT perspective, data gains epistemic value primarily through its integration into peer-reviewed publications. Sharing raw data prematurely risks not only introducing incomplete beliefs, but also having one's work scooped, which can have serious professional consequences.

Indeed, there is a sense in which a BoC attitude runs orthogonal to the prevailing credit systems in academia, which reward individual authorship and high-impact publications. This creates a tension: researchers may want to share data for the sake of epistemic progress, but are disincentivized by the risk of being scooped, misinterpreted, or uncredited. If someone takes a book (or data), repurposes it, and receives credit or publication without acknowledgment, they have effectively extracted the value of someone else's intellectual labor. Yet, if one adopts the view that scientific research is, at times, a broad and ongoing conversation, then value should be placed not solely on finalized publications, but also on the extent of a researcher's contributions – including data shared in repositories.

As discussed in the introduction, the normative aims of OS are complicated by global disparities in resources and infrastructures. Attitudes toward data sharing can deepen epistemic inequalities, making efforts to promote inclusivity through risk-taking ethically and practically fraught, especially when they risk reproducing marginality.

The contemporary scientific landscape has seen a significant push towards transparent and accessible data practices, in part shaped by concerns relating to the replication crisis and the resulting OS movement. At the forefront of this have been changes in data sharing practices (Munafò et al., 2017). This shift in part reflects a growing recognition that making research data openly

Research interests evolved toward risk aversion, with particular interest in the differences between risk preferences of men and women (e.g., Wilson and Daly 1985; Charness & Gneezy, 2012; Eckel & Grossman, 2002; 2008; Croson and Gneezy 2009; Nelson, 2012, 2018; Lemaster and Strough 2014), race (e.g., Finucane, Slovic, Mertz, Flynn, & Satterfield, 2000; Kahan, Braman, Gastil, Slovic, & Mertz, 2013), and the impact of stereotype threat on risk attitudes (e.g., Carr and Steele 2010; Adamus, Grežo, and Dudeková 2021). More generally, cognitive attitudes have been argued to play a major role in determining how values are weighted when making scientific decisions (e.g., Elliott and Willmes 2013). Following Elliott and Willmes (2013) and Davis-Stober et al. (2025), we think of scientific attitudes as evaluative responses directed toward a variety of content, including propositions, models, theories, or hypotheses.

[4] An alternative perspective might be that the BoT attitude, per se, says nothing about how one would view unpublished data—much like it says nothing about how one should view flowers. But the distinction emerges because the BoC attitude does give special value to unpublished research artifacts.

available is crucial for methodological verification, reproducibility of findings, and the acceleration of scientific discovery through data reuse.

A foundational framework guiding these efforts is based on the FAIR Data Principles, which stipulate that data should be Findable, Accessible, Interoperable, and Reusable (Wilkinson et al., 2016). The aim of the FAIR principles is to enhance the value of research data for both human users and automated systems, and to facilitate across-study integration and secondary analyses (Research Data Alliance FAIR Data Maturity Model Working Group 2020). The practical implementation of FAIR principles is an ongoing effort, with a focus on moving from abstract principles to concrete, actionable practices within existing research communities and data infrastructures (e.g., Mons et al., 2017; European Commission, n.d.; Billah et al., 2025), including through policies at journals, funding agencies, and professional societies such as the AGU (Stall, McEwen, Wyborn, Hoebelheinrich, & Bruno, 2020).

While there has been wide support for these ideas in the scientific community, the transition to widespread and effective data sharing is not without its challenges. Researchers often face social or cultural barriers, and myriad concerns have been raised about publishing datasets prior to publication, including issues of patient privacy in the medical context, misinterpretation of data, the risk of being scooped, loss of future publication opportunities, lack of credit or remuneration for one's labor, exposure to error, or potential misuse of data (Tenopir et al., 2020). Furthermore, the practicalities of preparing, curating, and archiving data can be resource-intensive, requiring time, specific skills, and funding that are not always readily available (Tenopir et al., 2011). These challenges are further compounded by concerns about extractivist data-sharing arrangements, particularly in global research contexts where data flows from under-resourced to well-resourced institutions without equitable benefit-sharing – as will be examined in greater detail in the following section (Kasy & Abebe, 2021; Lehuede, 2021).

Within the context of large-scale research collaborations, formalized data sharing practices are increasingly common, although there is no widely adopted set of norms or habits on which researchers could fall back. Foundational to the goal of formal data sharing is the establishment of consortium-wide agreements or data sharing agreements (e.g., O'Hara, 2020). These typically stipulate data governance, access rights for members and external parties, ethical compliance (especially for sensitive data), intellectual property, and publication policies, aiming to establish clear expectations and facilitate collaborative work (Budin-Ljøsne et al., 2013).

While the benefits of such collaborative data sharing—including accelerated discovery, enhanced statistical power, and the ability to address complex questions beyond the scope of single institutions—are widely acknowledged (Poldrack & Gorgolewski, 2014), challenges persist. Such challenges include ensuring robust data security and harmonizing data from disparate sources, but especially also navigating complex ethical and legal landscapes (the latter particularly in international projects with varying regulations), and fostering a culture of trust and timely data release (Budin-Ljøsne et al. 2013). Research highlights a trend towards using common data platforms and establishing clear data governance structures, like Data Access Committees, to manage these intricate sharing arrangements, balancing the collaborative needs of the consortium with broader open science goals (Shabani, Dyke, Joly, & Borry, 2015).

The harm of extractivism

In what follows, we unpack the specific epistemic harms at play, arguing that some of MIDEQ's data-sharing arrangements reflect a BoC attitude, one that is disposed to data as relational, situated, and intrinsically valuable. This perspective reveals a distinct category of epistemic harms rooted in epistemic disadvantage. Recognizing this calls for an expanded scope of extractivism, one that encompasses the broader range of epistemic harms.

Funded by the UKRI, the GCRF South-South Migration Inequality and Development Hub (MIDEQ) project studied migration and inequality in the global South, aiming to shift the focus of knowledge production from the global North to the regions where South-South migration occurs (Landström, 2024, p. 394; cf. Crawley, Garba, & Nyamnjoh, 2022). Collaborating with institutions in 12 countries across 5 continents, MIDEQ was organized into 12 research teams: six focused on specific migration corridors linking origin and destination countries (e.g., Haiti–Brazil, Jordan–Egypt, Nepal–Malaysia, Ghana–China) and six thematic work-packages addressing issues such as gender inequalities, migrant perceptions, knowledge and decision-making, and political mobilization. Additional work-packages implemented development interventions based on MIDEQ's findings, conducted cross-corridor surveys, and explored the relationship between migration and inequality through arts-based, practice-centered research (MIDEQ, n.d., as cited by Landström, 2024, pp. 394–395).

Internal data sharing within the MIDEQ Hub was structured to enable comparative research across various migration corridors, promoting mutual learning among research teams (Landström, 2024, p. 396). This exchange was central to MIDEQ's knowledge production processes: for researchers at either ends of a migration corridor to understand the dynamics of migration between their respective countries, reciprocal data sharing was essential (Landström, 2024, p. 396). Thematic work packages, tasked with conducting cross-corridor analyses, relied on close collaboration with the country teams and required access to their data, and vice versa. Thus, the success of both corridor-specific and thematic research hinged on a continuous, cooperative flow of information.

Notably, the MIDEQ Hub's data management protocol was not formalized until about 2 years into the project timeline.[5] The internal data-sharing protocols were the result of negotiations among principal investigators (PIs) and co-investigators (Co-PIs), some of whom expressed concerns about data sharing but ultimately accepted the agreed-upon terms. However, other teams remained unwilling to share data even after the protocol was established.

Karl Landström ethnographic research (2020–2023) offers a detailed account of the epistemic and organizational tensions within the large-scale, international research collaboration. Landström's (2024, p. 399) interviews revealed anxieties among some teams around data control, usage, and proper attribution for epistemic labor. One striking example of these revealed anxieties came from an interview with one of the Co-PIs, who spoke about a previous experience on a different project in which qualitative interview data had been contributed to a shared repository, only to find that an external, international collaborator had extracted and published selective portions, stripped of their local context and without crediting the original contributors.

Landström observed that internal data-sharing structures of

[5] The project officially started on December 1, 2018, with a basic data management plan in place. The full data management plan was then implemented in December 2020 following input and feedback from the co-directors of the Hub, the Hub's data management working group, the Co-PIs, and then finally a Hub-wide seminar for final input and alignment across all the members of the Hub (from conversation and email correspondence with Karl Landström April–May 2025).

MIDEQ recreate and exemplify epistemically oppressive structures. While the project clearly aimed to foster equitable collaboration, its organizational design reproduced systemic exclusions. Teams led by senior researchers with funding from institutions in the Global North operated alongside teams based in the Global South, often the very countries under study. This asymmetry replicated global power imbalances into the project's epistemic infrastructure. Even when MIDEQ's management introduced mechanisms to mitigate first-order exclusions—such as negotiating data-sharing protocols and allowing teams to opt out of data sharing—deeper, second- and third-order exclusions persisted. These stem from pre-existing disparities in research capacity, institutional recognition, and access to epistemic resources. Such conditions can perpetuate, what Coady (2010) calls, distributive epistemic injustice: a failure to equitably allocate the epistemic goods of knowledge, both in terms of what knowledge is generated and the material conditions of who generates it.[6] As Pohlhaus (2020) emphasized, even the inclusion of diverse research teams can be exploitative when embedded in oppressive systems, especially when those systems apply equal standards where equitable ones are required.

Furthermore, researchers at MIDEQ expressed concern about the possibility that their data would be repurposed without regard for the context in which it was produced or acknowledgment of those who had generated it (Landström, 2024, p. 400). Stripping data from its interpretative context reflects reductive attitudes of knowledge as modular and transferable, rather than relational and situated. When epistemic labor is appropriated unreflectively and non-collaboratively, a decontextualized value system is imposed onto the data itself. Extracting and redeploying knowledge in this way is to enact a form of epistemic violence – one that privileges utility over understanding, and instrumentalization over reciprocity. It is this logic that lies at the heart of epistemic extractivism.

However, the MIDEQ internal data sharing arrangements do not neatly fit this characterization of extractivism, where data is appropriated unreflectively and non-collaboratively in a decontextualized value system. Although internal data-sharing arrangements can create both ethical and epistemic risks, the interview with the Co-PI is not an example of extractivism but rather the anxiety of potential extractivism. Alone, the anxiety of the Co-PI is not demonstrative of extractivism. However, as Landström pointed to in conversation, perhaps a slightly different epistemic harm arises, one that echoes "testimonial smothering" (à la Dotson, 2011) in that the Co-PI researcher truncated possible courses of actions, like sharing data, to avoid potential epistemic violence. Landström (p. 406) writes:

> One MIDEQ Co-Investigator based in the global South described the MIDEQ Hub's data-sharing requirements as extractive and as an exercise of colonial power. They argued that the overarching design of the MIDEQ Hub, in which some research teams are assigned data-generating roles and other teams assigned roles in which they were to use the data for comparative analysis, encouraged extractivist research practices. . . The co-investigator described being skeptical of the data-sharing requirements of the MIDEQ Hub and that they, together with the other researchers working as part of their team, decided to only share summaries of their data rather than full datasets.

The Co-PI immediately protected themselves from extractivist practices but also curtailed their participation in the wider collaborative efforts of the Hub. To be precise, then, the harm would be more accurately said to arise from a form of 'aleatoric smothering,' curtailing actions to avoid the likelihood of violence.

Additionally, Landström (2024, p. 407) notes that the data-sharing arrangements ensured that the less resource-rich teams had greater access and control over the epistemic resources that other teams wanted to access. Indeed, there were several "resistance" strategies that corridor teams employed:

> This type of resistance was mentioned in interviews with several other MIDEQ researchers and became evident across the institutional ethnography as well. Some research teams found ways to subvert and resist taking part in data-sharing practices they objected to. As noted above, some research teams curtailed the data-sharing expectations by only sharing summaries of their data. Others decided to only share data in the original language that it was produced, thus raising barriers for use of the data by researchers who did not speak the local language (OBS). These examples illustrate that researchers can exercise their control over the data they produced to resist potentially epistemically extractivist research practices.

Hence, the MIDEQ internal data arrangements gave rise to the inverse of extractivist practices. Teams that resisted data-sharing showed that they were in control over the generation of, and access to, the prerequisite epistemic goods needed for the cross-cutting/comparative thematic analysis. While Landström's (2024) analysis helpfully identifies the structural and procedural data-sharing arrangements that raised concerns about epistemic injustice within MIDEQ, the specifics of the case open the question of how structural and attitudinal shifts in epistemic practice alters the landscape of harm. The MIDEQ internal data-sharing arrangements also shows researchers who are sensitive to the reality of epistemic injustice in OS, and who took thoughtful procedural steps to mitigate and counteract it. There is a sense in which MIDEQ arrangements did not treat knowledge as modular and transferable, but instead as relational, situated, and intrinsically valuable.

We suggest that the MIDEQ demonstrated a BoC attitude toward data collection. Recall that the BoC attitude values data for the interactions and reflections provoked, aligning with a view of knowledge as relational and situated. By foregrounding the interdependence between knowers and the conditions of knowing, this attitude encourages research practices that treat data as embedded in social, cultural, and ethical contexts. The internal data-sharing arrangements of MIDEQ exemplifies this attitude in the way in which corridor teams were given full control to share what data and to whom.

However, even with better epistemic attitudes or protocols, some extractivist harms that we see as epistemically disadvantageous (à la Goldstein, 2022) might persist, harms like withholding data, a form of extraction, from the collective. There is justifiable extraction that arises from ethical attitudes and procedures leading to "epistemic disadvantage" for the collective and unjustifiable extraction constituting epistemic injustice.

[6] Irzik and Kurtulmus (2024) distinguish between primary and secondary forms of distributive epistemic injustice. A primary form occurs when scientific institutions fail to provide knowledge that people have an objective interest in, or do so unfairly prioritizing some groups over others despite having the necessary resources. A secondary form concerns the unequal distribution of the conditions necessary for knowledge production itself, such as access to healthcare, education, or legal protections (p. 5). This concept is distinct from Fricker's (2017) notion of discriminatory epistemic injustice, which centers on prejudice—either direct or indirect—a distinctive class of wrongs in which someone is "ingenuously downgraded and/or disadvantaged in respect of their status as an epistemic subject" (p. 53).

When data is withheld in a project like MIDEQ, where comparative analysis across migration corridors depends on reciprocal data sharing, the refusal to share data can undermine the epistemic goals of the collaboration. Researchers—and by extension, the broader scientific community—may be denied access to epistemic goods they have legitimate reason to seek. This explicitly happened in MIDEQ and was the response from several research teams worried about extractivist research practices.

Denying access to epistemic resources can be harmful, but in epistemically oppressed environments, such harms may be justifiable. In migration studies, a corridor team may be warranted in excluding other teams from participation by withholding data on the basis that there are concerns about misuse or misrepresentation. When the exclusion is grounded in ethical attitudes rather than social-political-economic advantage, the resulting harm is better understood as an epistemic disadvantage rather than distributive or discriminatory epistemic injustice. That is, asymmetrical relations can warrant exclusions. In this case, the warranted exclusion pertains to withholding data amidst reciprocal sharing arrangements situated in existing unjust hierarchies and so necessitating an epistemic harm to the scientific community.

A second direction in which epistemic harm may be warranted, and thus better described as an epistemic disadvantage, is misappropriation within a consensual and ethically motivated system. Even when data is shared with consent and in good faith, extractivist harms can still occur. In collaborative research hubs like MIDEQ, consent does not eliminate the risk of misrepresentation, loss of credit, or unintended consequences. This is not a failure of ethics, but a structural vulnerability: even in systems that treat data as relational and intrinsically valuable, contributions may be taken up in ways that diverge from the contributor's expectations.

Solutions proposed by Landström would not have fully alleviated these epistemic harms. For instance, he suggests that extractivist practices could have been mitigated through early establishment of clear, mutually agreed-upon data-sharing protocols. He further argues that researchers with the deepest insight into the data generation process, and the most understanding of the context, should play a central role in analysis and dissemination (Landström, 2024, p. 401). As he notices, this would have required a fundamental shift in how credibility and authority are distributed within research collaborations: academic seniority and institutional prestige, and toward a more context-sensitive model of epistemic authority (Landström, 2024, pp. 401–402). Without such a shift, even well-intentioned protocols risk reinforcing the very hierarchies they aim to dismantle.

Such calls for implementing participative governance in data infrastructures offer one avenue for reducing extractivist risks. Sheehan and Leonelli (2024, p. 9) argued that participative governance involves multilingual tools, engaging communities in the design processes, and developing flexible policies that are accessible and understandable to all stakeholders. Such approaches can increase transparency and stakeholder involvement in how data is accessed and used, thereby reducing the risk of misappropriation.

However, even robust participative governance cannot eliminate certain epistemic risks inherent to data sharing. A research team's data may still be misunderstood, scooped, or repurposed in ways that disadvantage the originators. These risks are not resolved by governance structures alone. Rather, they point toward what we term *epistemically disadvantageous extractivist harms*: harms rooted in structural misalignments of epistemic attitudes and risk.

This is precisely why the concept of extractivism must be expanded. Previous accounts focus on commodification and exploitation, but extractivist harm can also arise even with shared values that reject commodification and despite institutional safeguards against epistemic oppression. Yet even within this framework, extractivist harms can persist – the key difference is that this harm is no longer rooted in injustice, but in disadvantage. Recognizing this distinction allows us to see that epistemic harm is not eliminated by relational ethics; it is transformed. If we fail to acknowledge this transformation, we risk overlooking the subtle ways in which epistemic labor remains vulnerable even in our most well-intentioned systems.

This discussion is particularly relevant for illuminating novel social-political-epistemic dynamics in emerging OS environments, where infrastructure and norms are still in formation. As ethical principles evolve alongside these environments, we must recognize that not all epistemic harms rise to the level of injustice – yet they still demand ethical attention. While concepts of epistemic injustice are crucial for identifying and addressing structural inequities in OS, they do not fully capture the nuanced challenges of enacting OS principles in practice, especially when those principles are guided by thoughtful, relational attitudes toward scientific knowledge. Given the complexity of these harms, there is a pressing need to develop a taxonomy that distinguishes between different scales and types of epistemic harm.[7] Harms such as exclusion, isolation, and extractivism often point to injustice, but they can also occur in ethically complex contexts—such as data-sharing repositories—where the harm may be disadvantageous rather than unjust. In these settings, sharing data may be necessary to understand complex phenomena like migration, even when it risks epistemic harm. No amount of ethical guidance can fully eliminate these harms; instead, we must learn to navigate them with sensitivity to context, power, and the evolving norms of collaborative knowledge production.

In summary, previous accounts of extractivism, such as articulated by Alcoff, typically frame it as the commodification of knowledge – where dominant groups extract epistemic resources from marginalized communities for economic gain, leaving the latter disadvantaged and disempowered. This model presumes a clear power asymmetry and a capitalist logic of exploitation. In contrast, the case under review—MIDEQ's internal data-sharing practices—operates within a collaborative framework driven not by profit, but by the epistemic necessity of understanding complex migration dynamics. While extractivist harms still occur, they are not rooted in economic exploitation but in the structural vulnerabilities of shared epistemic labor. This distinction matters. To accurately assess the ethical stakes of data sharing in such environments, we must expand the concept of extractivism to account for contexts where harm arises from relational, rather than commodified, knowledge practices. Only then can we develop a more nuanced ethical vocabulary for navigating the complexities of open science and collaborative research.

References

Alcoff, L. M. (2022, November). Extractivist epistemologies. *Tapuya: Latin American Science, Technology and Society*, *5*(1). Retrieved from http://dx.doi.org/10.1080/

[7] Scholars are increasingly recognizing the need for a more systematic taxonomy of epistemic harms—one that distinguishes between degrees of harm and clarifies when such harms rise to the level of injustice. This includes efforts to define a qualitative notion of harm (Beckers et al. 2022, 2024) and differentiate epistemic wrongs from epistemic harms in context of epistemic injustice (Dunne and Kotsonis 2024).

25729861.2022.2127231 doi: 10.1080/25729861.2022.2127231

Bartunek, J. M., & Moch, M. K. (1987, December). First-order, second-order, and third-order change and organization development interventions: A cognitive approach. *The Journal of Applied Behavioral Science*, *23*(4), 483–500. Retrieved from http://dx.doi.org/10.1177/002188638702300404 doi: 10.1177/002188638702300404

Bartunek, J. M., & Moch, M. K. (1994, February). Third-order organizational change and the western mystical tradition. *Journal of Organizational Change Management*, *7*(1), 24–41. Retrieved from http://dx.doi.org/10.1108/09534819410050795 doi: 10.1108/09534819410050795

Bezuidenhout, L. (2025, March). Critically unpacking the concept of equity for open science. *Social Epistemology*, 1–18. Retrieved from http://dx.doi.org/10.1080/02691728.2025.2475460 doi: 10.1080/02691728.2025.2475460

Billah, T., Cho, K. I. K., Borders, O., Chung, Y., Ennis, M., Jacobs, G. R., ... Bouix, S. (2025, April). Enabling fair data stewardship in complex international multi-site studies: Data operations for the accelerating medicines partnership® schizophrenia program. *Schizophrenia*, *11*(1). Retrieved from http://dx.doi.org/10.1038/s41537-025-00560-x doi: 10.1038/s41537-025-00560-x

Budin-Ljøsne, I., Isaeva, J., Knoppers, B. M., Tassé, A. M., Shen, H.-y., McCarthy, M. I., & Harris, J. R. (2013, June). Data sharing in large research consortia: experiences and recommendations from engage. *European Journal of Human Genetics*, *22*(3), 317–321. Retrieved from http://dx.doi.org/10.1038/ejhg.2013.131 doi: 10.1038/ejhg.2013.131

Cartwright, N. (2021, November). Rigour versus the need for evidential diversity. *Synthese*, *199*(5-6), 13095–13119. Retrieved from http://dx.doi.org/10.1007/s11229-021-03368-1 doi: 10.1007/s11229-021-03368-1

Coady, D. (2010, June). Two concepts of epistemic injustice. *Episteme*, *7*(2), 101–113. Retrieved from http://dx.doi.org/10.3366/epi.2010.0001 doi: 10.3366/epi.2010.0001

Crawley, H., Garba, F., & Nyamnjoh, F. (2022, June). Editorial introduction — migration and (in)equality in the global south: Intersections, contestations, and possibilities. *Zanj: The Journal of Critical Global South Studies*, *5*(1/2). Retrieved from http://dx.doi.org/10.13169/zanjglobsoutstud.5.1.0001 doi: 10.13169/zanjglobsoutstud.5.1.0001

Dang, H., & Bright, L. K. (2021, April). Scientific conclusions need not be accurate, justified, or believed by their authors. *Synthese*, *199*(3-4), 8187–8203. Retrieved from http://dx.doi.org/10.1007/s11229-021-03158-9 doi: 10.1007/s11229-021-03158-9

Davis-Stober, C. P., Sarafoglou, A., Aczel, B., Chandramouli, S. H., Errington, T. M., Field, S. M., ... Gundersen, O. E. (2025, January). How can we make sound replication decisions? *Proceedings of the National Academy of Sciences*, *122*(5). Retrieved from http://dx.doi.org/10.1073/pnas.2401236121 doi: 10.1073/pnas.2401236121

Dotson, K. (2011). Tracking epistemic violence, tracking practices of silencing. *Hypatia*, *26*(2), 236–257. Retrieved from http://dx.doi.org/10.1111/j.1527-2001.2011.01177.x doi: 10.1111/j.1527-2001.2011.01177.x

Dotson, K. (2014, January). Conceptualizing epistemic oppression. *Social Epistemology*, *28*(2), 115–138. Retrieved from http://dx.doi.org/10.1080/02691728.2013.782585 doi: 10.1080/02691728.2013.782585

European Commission. (n.d.). *European open science cloud | open science – research & innovation.* Retrieved from http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud (Accessed May 30, 2025.)

Fecher, B., & Friesike, S. (2013, December). Open science: One term, five schools of thought. In *Opening science* (p. 17–47). Springer International Publishing. Retrieved from http://dx.doi.org/10.1007/978-3-319-00026-8_2 doi: 10.1007/978-3-319-00026-8_2

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford: Oxford University Press.

Gardiner, G. (2020, July). Relevance and risk: How the relevant alternatives framework models the epistemology of risk. *Synthese*, *199*(1-2), 481–511. Retrieved from http://dx.doi.org/10.1007/s11229-020-02668-2 doi: 10.1007/s11229-020-02668-2

Goldstein, R. B. (2022, January). Epistemic disadvantage. *Philosophia*, *50*(4), 1861–1878. Retrieved from http://dx.doi.org/10.1007/s11406-021-00465-w doi: 10.1007/s11406-021-00465-w

Kasy, M., & Abebe, R. (2021, March). Fairness, equality, and power in algorithmic decision-making. In *Proceedings of the 2021 acm conference on fairness, accountability, and transparency* (pp. 576–586). ACM. Retrieved from http://dx.doi.org/10.1145/3442188.3445919 doi: 10.1145/3442188.3445919

Kidd, I. J., & Carel, H. (2016, February). Epistemic injustice and illness. *Journal of Applied Philosophy*, *34*(2), 172–190. Retrieved from http://dx.doi.org/10.1111/japp.12172 doi: 10.1111/japp.12172

Landström, K. (2024, May). On epistemic extractivism and the ethics of data-sharing. *Philosophy of the Social Sciences*, *54*(5), 387–411. Retrieved from http://dx.doi.org/10.1177/00483931241255253 doi: 10.1177/00483931241255253

Leonelli, S. (2016). *Data-centric biology: A philosophical study*. University of Chicago Press. Retrieved from http://dx.doi.org/10.7208/chicago/9780226416502.001.0001 doi: 10.7208/chicago/9780226416502.001.0001

Leonelli, S. (2022, May). Open science and epistemic diversity: Friends or foes? *Philosophy of Science*, *89*(5), 991–1001. Retrieved from http://dx.doi.org/10.1017/psa.2022.45 doi: 10.1017/psa.2022.45

Leonelli, S. (2023). *Philosophy of open science*. Cambridge University Press. Retrieved from http://dx.doi.org/10.1017/9781009416368 doi: 10.1017/9781009416368

Levin, N., & Leonelli, S. (2016, October). How does one "open" science? questions of value in biological research. *Science, Technology, & Human Values*, *42*(2), 280–305. Retrieved from http://dx.doi.org/10.1177/0162243916672071 doi: 10.1177/0162243916672071

MIDEQ. (n.d.). *Migration for development and equality.* Retrieved from https://www.mideq.org/en/ (Accessed May 30, 2025.)

Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B., & Wilkinson, M. D. (2017, February). Cloudy, increasingly fair; revisiting the fair data guiding principles for the european open science cloud. *Information Services and*

Use, *37*(1), 49–56. Retrieved from http://dx.doi.org/10.3233/ISU-170824 doi: 10.3233/isu-170824

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., ... Ioannidis, J. P. A. (2017, January). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1). Retrieved from http://dx.doi.org/10.1038/s41562-016-0021 doi: 10.1038/s41562-016-0021

O'Hara, A. (2020). *Model data use agreements: A practical guide.* In *Handbook on Using Administrative Data for Research and Evidence-Based Policy*, edited by Cole, Dhaliwal, Sautmann, and Vilhuber. Retrieved from https://admindatahandbook.mit.edu/book/v1.0/dua.html (Accessed May 17, 2025. No DOI in CrossRef for handbook chapter; metadata from manuscript reference list.)

Peña-Guzmán, D. M., & Reynolds, J. M. (2019, September). The harm of ableism: Medical error and epistemic injustice. *Kennedy Institute of Ethics Journal*, *29*(3), 205–242. Retrieved from http://dx.doi.org/10.1353/ken.2019.0023 doi: 10.1353/ken.2019.0023

Pohlhaus, G. (2020, May). Epistemic agency under oppression. *Philosophical Papers*, *49*(2), 233–251. Retrieved from http://dx.doi.org/10.1080/05568641.2020.1780149 doi: 10.1080/05568641.2020.1780149

Poldrack, R. A., & Gorgolewski, K. J. (2014, October). Making big data open: data sharing in neuroimaging. *Nature Neuroscience*, *17*(11), 1510–1517. Retrieved from http://dx.doi.org/10.1038/nn.3818 doi: 10.1038/nn.3818

Research Data Alliance FAIR Data Maturity Model Working Group. (2020). *Fair data maturity model: Specification and guidelines.* Research Data Alliance. Retrieved from https://doi.org/10.15497/RDA00050 doi: 10.15497/RDA00050

Scriver, A. (2019). *I have post-traumatic stress and didn't know it — and you might, too.* Healthline. Retrieved from https://www.healthline.com/health/mental-health/didnt-know-about-cptsd (Online article.)

Shabani, M., Dyke, S. O. M., Joly, Y., & Borry, P. (2015, December). Controlled access under review: Improving the governance of genomic data access. *PLOS Biology*, *13*(12), e1002339. Retrieved from http://dx.doi.org/10.1371/journal.pbio.1002339 doi: 10.1371/journal.pbio.1002339

Sheehan, N., & Leonelli, S. (2024, December). Reconciling data actionability and accountability in global health research: The case of sars-cov-2. *Global Public Health*, *20*(1). Retrieved from http://dx.doi.org/10.1080/17441692.2024.2436422 doi: 10.1080/17441692.2024.2436422

Spencer, L. J. (2024, September). Institutional epistemic isolation in psychiatric healthcare. *Social Epistemology*, *39*(2), 215–228. Retrieved from http://dx.doi.org/10.1080/02691728.2024.2403620 doi: 10.1080/02691728.2024.2403620

Spencer, L. J., & Carel, H. (2021, April). 'isn't everyone a little ocd?': The epistemic harms of wrongful depathologization. *Philosophy of Medicine*, *2*(1). Retrieved from http://dx.doi.org/10.5195/pom.2021.19 doi: 10.5195/pom.2021.19

Stall, S., McEwen, L., Wyborn, L., Hoebelheinrich, N., & Bruno, I. (2020, January). Growing the fair community at the intersection of the geosciences and pure and applied chemistry. *Data Intelligence*, *2*(1-2), 139–150. Retrieved from http://dx.doi.org/10.1162/dint_a_00036 doi: 10.1162/dint_a_00036

Staunton, C., Barragán, C. A., Canali, S., Ho, C., Leonelli, S., Mayernik, M., ... Wonkham, A. (2021, November). Open science, data sharing and solidarity: who benefits? *History and Philosophy of the Life Sciences*, *43*(4). Retrieved from http://dx.doi.org/10.1007/s40656-021-00468-6 doi: 10.1007/s40656-021-00468-6

Tanesini, A. (2022, March). Intellectual vices in conditions of oppression: The turn to the political in virtue epistemology. In *The political turn in analytic philosophy* (pp. 77–104). De Gruyter. Retrieved from http://dx.doi.org/10.1515/9783110612318-006 doi: 10.1515/9783110612318-006

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

Tenopir, C., Rice, N. M., Allard, S., Baird, L., Borycz, J., Christian, L., ... Sandusky, R. J. (2020, March). Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLOS ONE*, *15*(3), e0229003. Retrieved from http://dx.doi.org/10.1371/journal.pone.0229003 doi: 10.1371/journal.pone.0229003

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016, March). The fair guiding principles for scientific data management and stewardship. *Scientific Data*, *3*(1). Retrieved from http://dx.doi.org/10.1038/sdata.2016.18 doi: 10.1038/sdata.2016.18