From Competition to Collaboration: Fostering a Culture of Data Sharing in the NIH HEAL Data Ecosystem

Heather Barnes, Brandy Farlow, Mary Hannah Marshall, Allie Gartland-Gray, Kira Bradford, Michelle Engle, and Brandi Martin

Contents

About the Authors

Heather Barnes,* PhD, is a research systems programmer/analyst at RTI International. https://orcid.org/0000-0003-3109-6482

Brandy Farlow,* MA, is a research proposal and documentation specialist at the Renaissance Computing Institute at the University of North Carolina at Chapel Hill. https://orcid.org/0009-0004-1646-9484

Mary Hannah Marshall,* MS, is a research systems programmer/analyst at RTI International. https://orcid.org/0000-0002-0427-7125

Allie Gartland-Gray, MS, is a research systems programmer/analyst at RTI International. https://orcid.org/0000-0002-0734-4071

Kira Bradford, PhD, is a senior research systems programmer/analyst at RTI International. https://orcid.org/0000-0003-2403-817X

Michelle Engle, PhD, is a research systems programmer/analyst at RTI International. https://orcid.org/0000-0002-4931-8541

Brandi Martin, PhD, is an expert witness and litigation support consultant at Raptor Pharm & Tox, Ltd. https://orcid.org/0000-0002-3929-9658

*These authors contributed equally as co-first authors.

RTI Press Associate Editor

Janelle Armstrong-Brown

Abstract

Open, accessible research data provides a foundation for scientific discovery. Despite clear data sharing benefits and increased data sharing expectations, hesitancy to share data is common in many scientific research domains. This case study highlights work underway in the National Institutes of Health (NIH)–sponsored project *Helping to End Addiction Long-term®* (HEAL) Data Ecosystem (HDE). HDE has initiated a suite of activities designed to engage HEAL-funded investigators and promote a data sharing culture. In this paper, we present results of a landscape analysis highlighting (1) common barriers to data sharing and (2) incentives identified by scholars that may help researchers overcome these barriers. We describe current strategies HDE employs to support a data sharing community.

Introduction

Our goal is to derive meaningful insights from existing research on data sharing behaviors that research discovery ecosystems may apply to program evaluation. First, we provide results from a landscape review focused on common data sharing incentives and barriers. Then we summarize outcomes from key National Institutes of Health (NIH) *Helping to End Addiction Long-term®* (HEAL) Data Ecosystem (HDE) programs to foster a data sharing community. We define data sharing herein to reflect the National Library of Medicine's definition: "Data sharing refers to the practice of making data available to other research stakeholders, including other investigators, research subjects, and the broader public" (Network of the National Library of Medicine, 2024). For this analysis, we focus on data and other outcomes expected to be shared by scientific researchers as part of the research lifecycle.

The NIH HEAL Data Ecosystem: Description and Composition

The HDE is part of the NIH HEAL Initiative, an NIH-wide effort to speed scientific solutions to stem the evolving national opioid public health crisis. HEAL-funded researchers share pain and opioid use disorder data in a wide range of formats, including imaging, animal studies, clinical data, and qualitative studies. HDE "seeks to accelerate sharing HEAL-generated data and results among the broad community of researchers, health care providers, community leaders, policy makers, and other HEAL stakeholders who can benefit from learning initiative research results" (NIH HEAL Initiative, 2021). HDE connects the HEAL community, enabling dataset search (via HEAL Data Platform and Semantic Search), analysis, and reuse for new discoveries.

HEAL funding empowers "researchers to make their HEAL-generated data FAIR (findable, accessible, interoperable, and reusable)" and promotes data sharing (NIH HEAL Initiative, 2021). The HEAL Data Platform includes a search and discovery interface powered by rich metadata and secure, cloud-based workspaces. The platform does not store data but instead interoperates with the individual HEAL-compliant repositories in which HEAL data

are deposited, providing secure access to datasets under the corresponding repository's access restrictions and approval processes. Researchers may also take advantage of HDE tools such as variable-level metadata submission tools.

The HEAL Data Stewardship Group (HEAL Stewards) includes staff members from RTI International and the Renaissance Computing Institute (RENCI) at the University of North Carolina at Chapel Hill and helps facilitate HDE. The HEAL Stewards develop researcher-facing outreach programming, including:

- organizing, leading, and maintaining HDE governance structures, including the Collective Board;
- leading outreach and engagement strategy development and implementation, including webinars, workshops, consulting, and other community member training;
- leading HEAL Semantic Search integration with the HEAL Data Platform;
- providing guidance for selecting a repository and submitting datasets and metadata; and
- developing documentation and supporting materials for interacting with HEAL Semantic Search.

The HEAL Collective Board guides HDE strategy and direction to develop methods and norms, cultivate a culture of sharing, and maximize collaboration (NIH HEAL Initiative Data Stewardship Group, n.d.).

Background and Methods

Open, accessible research data provides a foundation for scientific discovery. During the COVID-19 pandemic, data sharing across disciplines hastened vaccine development: almost half of researchers working on vaccine research (43 percent) shared data openly (Druedahl et al., 2021). Recent federal policies (in the United States and beyond) mandate data sharing plan submission to expand access to research outcomes. In 2023, the NIH implemented a revised Data Management and Sharing Policy (NOT-OD-21–013: Final NIH Policy for Data Management and Sharing; NIH, n.d.). The National Science Foundation, the National Endowment for the Humanities, and several other federal granting

agencies also now require that proposals include data management plans in grant applications.

Several researchers have examined the factors behind data sharing behaviors. For example, according to Late et al. (2024), "Supporting the scientific community, the open science agenda and fulfilling research funders' requirements motivate scholars to share their data. Impeding factors relate to the qualities of data, ownership of data, data stewardship, and research integrity" (p. 386).

Data sharing hesitancy has repercussions for open science. Pearson (2003) argued that limited result sharing can lead to a decline in the open exchange of ideas, hindering scientific progress. Delayed data sharing impedes progress in health care research, potentially resulting in increased costs (Vickers, 2006). Research on data sharing behavior among scientists indicates that hesitancy may be rooted in several factors, including a lack of certainty about how, when, and where to share data and a competitive research culture in which high-impact publications drive career advancement. In addition, many researchers worry their discoveries will be "scooped," possibly resulting in loss of credit. Conversely, factors that incentivize data sharing include receiving full credit for their findings, adequate training in open science practices, and fostering a collaborative research culture.

Ensuring HEAL-funded researchers share research outcomes is a critical objective for the HEAL Stewards. Conducting a landscape analysis serves the HEAL Stewards' efforts to connect researcher-facing programs and activities with evidence-based practice. Our approach to the literature review involved first identifying existing research studies that address data sharing factors in two general categories: (1) barriers/disincentives and (2) benefits/incentives. In compiling a list of publications related to data sharing, we first gathered articles in a range of research disciplines to explore how scholars currently define key data sharing barriers and incentives. The initial database search for existing literature on data sharing benefits and barriers included a broad spectrum of research disciplines and data types; however, the final set of references samples more intensively from health, biomedical, and social sciences research. Biomedical

data sharing likely differs from that in nonmedical fields; however, the scope of this review did not include an intentional differentiation of factors between fields. The literature search was designed to identify how scholars have described the most common barriers and incentives across a range of disciplines.

While not generalizable to all research fields, the factors identified in the landscape analysis cluster around common incentives and barriers that we anticipate will be helpful to HDE staff planning new or evaluating existing programming. The analysis and recommendations described herein aim to support an informed evaluation of how well HDE activities align with current best practice and where there may be room for improvement and expansion. Most HDE activities were launched before the literature review was conducted; therefore, the evidence-based factors we identify here are primarily intended to be informative. That is, the analysis will support developing key metrics to assess HEAL Stewards' outreach efforts.

Defining the Landscape: Data Sharing Incentives and Barriers

Barriers to Data Sharing

Despite clear benefits and recent technological advances that help streamline the process, data sharing remains stubbornly low in the sciences (Houtkoop et al., 2018; Pearson, 2003; Vines et al., 2013). A survey by Hipsley and Sherratt (2019) found that only 14 percent of investigators shared biological imaging data. Low data sharing rates occur even in federally funded research, suggesting that barriers may exist beyond a lack of awareness of how and why to share data. Researchers may encounter any number of barriers, including legal and ethical restrictions, time constraints, a lack of incentives, and the fear that sharing their data may result in scooping or exploitation (Hipsley & Sherratt, 2019; Houtkoop et al., 2018; Pearson, 2003; Tenopir et al., 2011). Legal and ethical restrictions may limit sharing certain types of data that could identify participants or endanger rare species if released (Duke & Porter, 2013; Pearson, 2003). Technical issues, such as limited storage options for large amounts of data, may pose

logistics challenges, although improved infrastructure and software have begun to mitigate this challenge (Farley et al., 2018; Stephens et al., 2015).

Fear of Being Scooped, Career Advancement, and Citations

The fear of scooping—the idea that other researchers may exploit findings if results are shared prematurely—appears frequently in literature on barriers (Hipsley & Sherratt, 2019; Houtkoop et al., 2018; Pearson, 2003).

Concerns about loss of publication opportunities—which are critical for building academic reputation, applying for tenure, and securing grants—serve as disincentives to sharing (Callaway, 2019; Walsh & Hong, 2003). Publications represent costs in terms of time and funding; researchers who hesitate to share may view data as a proprietary resource (Barczak et al., 2022). Having research ideas scooped may threaten a researcher's ownership over their work (Callaway, 2019) or may damage an early career academic's reputation (Teixeira da Silva & Dobránszki, 2015); lack of attribution also causes some researchers concern (Devriendt et al., 2021).

In one survey of cell biologists, over 75 percent reported fear of getting scooped. Anxieties are heightened in rapidly moving fields like cell and molecular biology, where experiments can be designed, executed, and published within weeks (Pearson, 2003). Additionally, online data repositories, preprint servers, and electronic journal submissions may enable competitors to generate early manuscript versions and publish results ahead of the original researcher (Teixeira da Silva & Dobránszki, 2015). In a "winner-takes-all" culture, where reputation and careers hinge on high-profile, first-author publications, this sense of competitiveness (Barczak et al., 2022) exacerbates data sharing hesitancy. Some researchers respond by limiting prepublication communications altogether (Adams et al., 2018; Walsh & Hong, 2003) or by delaying data sharing to secure the first opportunity to present their findings (Hulsen, 2020; Mozersky et al., 2021).

In addition to fears of exploitation or scooping, researchers express concern about the need to prioritize their career advancement, which in

scientific fields depends heavily on publishing. For early career researchers, who may struggle to receive credit for their research contributions (Hardy, 2021; Hutchings et al., 2020), a perceived lack of credit may foster data sharing hesitancy. Soeharjono and Roche (2021) noted that researchers "report [more] benefits (47.9%) and neutral outcomes (43.6%) than costs (21.4%) from openly sharing data…[but] early career researchers were more likely to report costs" (p. 750). Career advancement opportunities tended to be less abundant for early career researchers (Hutchings et al., 2020). Hutchings et al. (2020) propose "a shift away from the traditional criteria of academic promotion, which includes research outputs, to one which is inclusive of a researcher's data sharing history and the availability of their research dataset for secondary analysis" (p. 26).

Collaborating on publications supports younger academics' advancement; however, efforts to circumnavigate hesitancy by co-authoring face challenges. Melbourne researcher Josh Hardy (2021) recounts, from an effort to co-author research publications with overlapping studies, "Rather than being redundant, our experiments had validated each other's finding in different viruses and strengthened the result of both experiments. However, coordinating publications is not always straightforward. Many journals do not have clear mechanisms for co-submission and do not sufficiently support the model" (p. 2). Hardy argues that if early career researchers are to engage in collaborative research, "more scientific journals need to support and have guidelines for reviewing and accepting joint submissions" (Hardy, 2021, p. 3). In addition to transforming publication models, academic culture should reward researchers for engaging in research collaborations (Hutchings et al., 2020).

Strategies for measuring research impact also drive data sharing behaviors. Citation metrics, for example, tend to define a researcher's scientific stature. A 2019 representative sample of United States and

Canadian institutions found that 40 percent of the research-intensive institutions had impact factor language in retention, promotion, and tenure package documentation (McKiernan et al., 2019).

Data Equity and Access

Data sharing can introduce data equity challenges, further exacerbating hesitancy. Common data equity concerns relate to sensitive data handling and information access. Finding a balance between open and accessible data sharing and privacy/sensitivity concerns remains a challenge (Sardanelli et al., 2018; Vickers, 2006). Addressing researcher, patient, and community concerns is critical to data sharing, particularly as patients and/or research participants are increasingly recognized as the rightful owners of their data (Hulsen, 2020; Vickers, 2006). Regulations governing health data privacy, including the Health Insurance Portability and Accountability Act, constrain open data sharing. Survey findings suggest overcoming sensitive data barriers may require articulating explicit norms, incentives, Institutional Review Board processes, and levels of trust around open data (Hipsley & Sherratt, 2019; Houtkoop et al., 2018).

Other data considerations include creating equitable policies governing appropriate data sharing, particularly with respect to low-resource communities. Clear agreements and effective sensitive data policies help ensure responsible and ethical data sharing (Hulsen, 2020; Vickers, 2006). Pratt and Bull (2021) highlighted five data sharing barriers in low-resource communities, including (1) lack of infrastructure and technology necessary to use and analyze data; (2) lack of research credit for data reuse; (3) inaccessible research outcomes (publications, presentations, and data); (4) population-specific stigmas; and (5) other adverse consequences to communities.

Incentives for Data Sharing

Research on data sharing highlights the many benefits to fostering an open data sharing culture. Sharing research data creates opportunities for collaboration and knowledge-building (Adams et al., 2018; Barczak et al., 2022) and supports reproducibility and reuse (Berman et al., 2015; Houtkoop et al., 2018; Wilkinson et al., 2016). Secondary data analysis often leads to

cross-disciplinary discoveries (Reichstein et al., 2019; Stephens et al., 2015). Data sharing accelerates public health research on topics such as disease outbreaks and climate change (Sarabipour et al., 2019; Tse et al., 2020). Moderating the often-competitive research culture, supporting researchers' career advancement, providing credit for research contributions, and optimizing publication/citation opportunities are some of the most common themes that the literature on incentives for data sharing addresses.

Open Research Culture, Career Advancement, and Citations

One of the primary drivers of sharing behavior is funding agency mandates. Federal policy now requires data sharing for publicly funded research. Many international publishers have also adopted policies requiring funded researchers to provide data access as soon as possible (Barczak et al., 2022; Chawinga & Zinn, 2019). Given appropriate support, researchers tend to share data more willingly.

In addition to policy-driven sharing, researchers choose to share data for various reasons. Barczak et al. (2022) observed that researchers recognize community benefits from sharing (mutual support and collaboration). In fact, collaboration is often a silver lining to sharing data, despite researcher misgivings. Sharing may thus be perceived as a counterweight to the fear of scooping. The sharing process and a shared commitment to open science practices within a research community or discipline help limit disincentives. Laine (2017) reported on a project in which a culture of open data encouraged traditional competitors to collaborate and "focus their projects on different research themes to avoid direct competition" (p. 6). Melero and Navarro-Molina (2020) highlighted the promise of increased citations as one positive outcome, but beyond these direct benefits, the concept of openness as a moral/ethical good is also a cultural factor that supports data sharing (Lounsbury et al., 2021).

Data sharing incentives include training (Houtkoop et al., 2018); funding that covers repository fees/costs; credit in the form of citations (Melero & Navarro-Molina, 2020); and a clear process for sharing data (Hipsley & Sherratt, 2019). The promise of increased

citations convinces some researchers to make datasets available (Curty et al., 2016; Gomes et al., 2022); however, researchers must understand where and how to share data. Devriendt and colleagues (2021) identify incentives that need to be present for researchers to feel comfortable sharing data, including *credit/recognition*, *transparency*, *reciprocity*, and *trust*. Hipsley and Sherratt (2019) explored key drivers and reported that financial rewards in any form increase data sharing behavior. Soeharjono and Roche (2021) examined both barriers and incentives, reporting that researchers interviewed tended to experience a sense of personal reward after sharing, although this is less a tangible incentive than a general benefit. Soeharjono and Roche also found that career benefits (advancement and stature) may serve as incentives.

Research Access, Efficiency, and Impacts

Zuiderwijk and colleagues (2020) examine key factors incentivizing data sharing, finding in a broad literature review that incentives depend on researcher background (discipline), but formal data access requirements/policies, such as data sharing mandates, serve as a key driver. In addition, automatic dataset publication (research efficiency) and institutional financial support help improve data sharing rates. A wide range of personal incentives also drive sharing, including researcher commitments to (1) reproducibility, (2) a culture of sharing, (3) advancing research in their field (research impact), and (4) validating results. Zuiderwijk and colleagues found that favorable conditions for

sharing also include access to the following resources: appropriate research data repositories; shorter embargo periods; minimal risk to participant privacy; rewards and recognition for publication and data sharing; increased citations; social influence; more research collaborations; experience/skills in sharing; and using data types that support sharing (are easy to convert to open formats).

Although less comprehensive than Zuiderwijk and colleagues' literature review, Laine's (2017) broad exploration of data sharing incentives confirms increased citations and publications benefits. Similarly, Woods and Pinfield (2021), in a literature review, categorized key data sharing incentives thematically, including:

> the need to build on existing cultures and practices, meeting people where they are and tailoring interventions to support them; the importance of publicizing and explaining the policy/service widely; the need to have disciplinary data champions to model good practice and drive cultural change; the requirement to resource interventions properly; and the imperative to provide robust technical infrastructure and protocols, such as labeling of data sets, use of DOIs [digital object identifiers], data standards and use of data repositories. (p. 1)

Literature Summary

Table 1 summarizes the frequently mentioned data sharing barriers and the incentives that may help mitigate these barriers.

Table 1. Common data sharing barriers and incentives

Barriers	Incentives
Fear of scooping Many researchers fear being scooped, losing career advancement opportunities and publication rights on their findings, if they openly share data before publication. This may cause numerous problems, including withholding ideas prepublication, which hinders scientific progress.	**Fostering a culture of open science** Sharing data helps move the needle toward open science practices, which improves access to data, publications, and other research products generated through publicly funded studies (Zuiderwijk et al., 2020). **Promoting research in the field** One of the goals of federal data sharing policies is to ensure that publicly funded research drives new knowledge and discoveries. Sharing your data helps build awareness of the key issues in your research area (Zuiderwijk et al., 2020). **Reproducibility and validation** Sharing data supports the likelihood that others can reproduce and confirm your research results, leading to increased credibility and data reuse (Soeharjono & Roche, 2021).

Barriers	Incentives
Credit for early-career researchers; career progression While mid- to late-career researchers report numerous benefits from collaborative research, early-career researchers are more inclined to report costs. Data sharing credit becomes vital in these instances because collaborative research is a staple of early career research. In conjunction with this, considerations for journal article co-submissions and data sharing rewards would help early career researchers gain appropriate credit for their research contributions, increasing their willingness to share.	**Up to a 25% increase in citations** Research indicates that making datasets available alongside publications can boost citation counts by up to 25%, enhancing the impact of your study on the field (Colavizza et al., 2020). **Productivity, reputation, and career advancement** Sharing data provides expanded opportunities to have your data sets reused and cited in other publications, supporting your case for promotion and increasing the impact of your research (Soeharjono & Roche, 2021). **Research efficiency** Preserving data in open repositories ensures long-term sustainability of your research products. Data archives provide persistent identifiers, supporting access to your datasets long after the grant cycle is complete (Soeharjono & Roche, 2021).
Publication barriers Citation metrics and publications have traditionally impacted tenure and promotability, but dataset citations have yet to be tracked as closely. Because of this emphasis, researchers may worry about their lack of promotability incentive or about not gaining proper attribution for their data.	**New publishing opportunities** Sharing data associated with your publications encourages other researchers to explore your findings, increasing citations and leading to possible future publications (Bock et al., 2005). **Co-authorship and collaboration** Serendipitous discovery of other researchers' data can enhance collaboration and spark new collaboration and co-authorship opportunities (Soeharjono & Roche, 2021).
Data access in low-resource communities Higher-income countries (as well as better funded universities, and higher-income communities) can more easily access adequate infrastructure and resources to analyze and use data. The investment to properly analyze data may be out of reach for lower-income countries, institutions, and communities.	**Reciprocal access to open data** A culture of data sharing supports reuse, ensuring that all researchers can find and use data that may be of benefit (Park & Gabbard, 2018). **Positive impacts on public health** Broad access to publicly funded research helps improve health outcomes, expanding knowledge about public health problems and research-based solutions (Hutchings et al., 2020).
Patient concerns Each patient owns their data, and though this data can help advance medical research, it comes with risks to the data owners (re-identification, privacy breaches, data misuse and stigmatization), which may limit both patient and researcher willingness to share data.	**Sensitive data-handling guidance** Some datasets must remain private, but de-identification protocols can make study-level metadata available through research data platforms that enable "research at scale" (Hulsen, 2020, p.6).

Fostering a Culture of Data Sharing in the HEAL Data Ecosystem Through Engagement and Outreach

Overview of Data Sharing in HDE

HDE's design is informed by a distributed data system model. Distributed systems vary widely in their implementation but tend to include differentiated governance and geographically dispersed infrastructure components. HDE serves as a centralized metadata catalog, providing users tools to discover and easily access HEAL-funded data. As a distributed data ecosystem, the HDE operation relies on HEAL-compliant digital repositories for study dataset storage. Researchers preparing to deposit data may select from an abbreviated list of prevetted repositories. Researchers generally have some flexibility to choose the repository that best suits their data. The HEAL Data Platform aggregates metadata from HEAL studies and serves as a central discovery portal. **Figure 1** illustrates the HDE's primary components, which include HEAL-supported researchers, data repositories, and community stakeholders.

HDE promotes data sharing, setting research-data producer expectations to share data with the ecosystem. Each HEAL-funded study is expected to submit study-level and variable-level metadata. Other HEAL-specific implementation steps include:

- Register on the HEAL Data Platform and submit necessary metadata.
- Use HEAL common data elements.

Figure 1. Overview of the HEAL Data Ecosystem's components
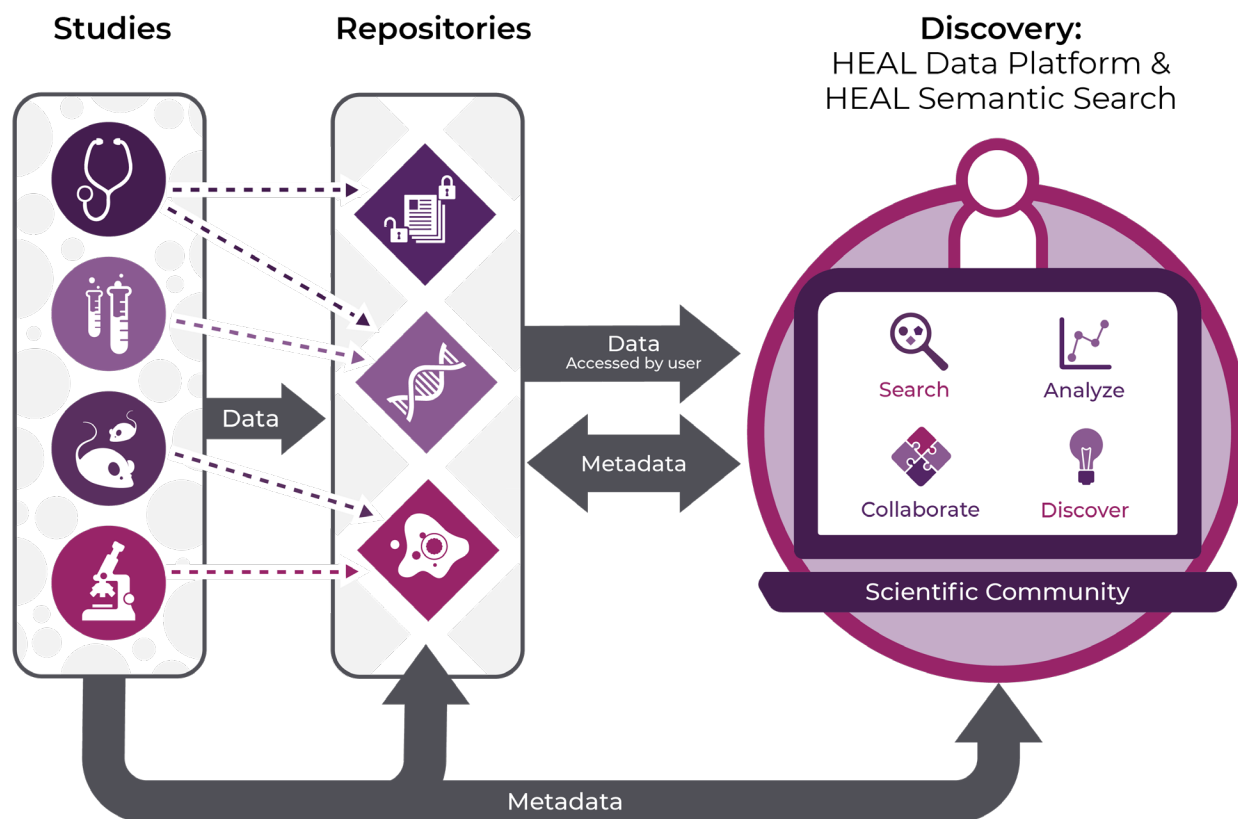
- Use broad consent language.
- Indicate the planned HEAL-compliant repository.

HDE supports activities that foster a sense of community around data sharing. These activities are discussed in more detail in subsequent sections.

HDE Outreach and Engagement Activities

As of fall 2024, approximately 20 percent of HEAL-funded studies have selected a repository. As HDE has evolved, various stakeholders have identified factors that tend to promote or inhibit participation in data sharing activities. In addition to the factors identified by scholars (described in the previous sections), HEAL Stewards have identified common researcher questions that, when addressed, help foster HDE-wide data sharing participation:

Questions related to *why* to share data:

- Does my study type fit within the NIH HEAL Initiative's sharing requirements?

- If I am working on a study that does not require data sharing, are there ways to participate in the HEAL processes to increase transparency in my work?

Questions related to *how* to share data:

- How do I comply with the 2023 NIH Data Management and Sharing Policy?

- What are the FAIR principles? And how do they affect my data sharing protocols?

- How do I use required common data elements and other metadata standards?

- How do I create research documentation, such as README.txt files?

Questions related to *where* to share data:

- Would a generalist or specialized repository be more suitable for my data?

- Which repositories specialize in my study's data type?

Questions related to *when* to share data:

- When in the research lifecycle should I select a repository?
- Should I share data before my study/grant has ended?

To help researchers navigate these concerns and move toward successfully sharing data, the HDE has developed a suite of services and tools to connect researchers with just-in-time resources for overcoming barriers and enhancing incentives. HDE services include a wide range of in-person and asynchronous support, from direct assistance with selecting a HEAL-compliant repository to webinars on navigating sensitive data. HEAL Stewards encourage study teams to implement FAIR principles (devised by Wilkinson et al., 2016) in their data management strategies. HEAL Stewards' webinars and consulting activities aim to address researcher concerns and questions, explain how to participate effectively in HDE, connect researchers to the optimum repositories for their study, and provide efficient data sharing guidance. The HEAL Collective Board, comprising more than 20 active HEAL-funded investigators, meets regularly to advise HEAL Stewards and help promote a culture of data sharing throughout the HEAL research community.

In **Table 2**, we list HEAL Stewards' outreach and support services to date, identify the barriers or incentives to which they most directly correspond, and provide an assessment of the programs' effectiveness in addressing the barriers.

Results and Recommendations

The landscape review will help inform HDE's ongoing efforts to address data sharing challenges. Programs implemented before the review have generated both positive outcomes and areas for improved alignment with researcher needs. Understanding where researchers face challenges will contribute to HDE refinements and expansion. We recommend the following strategies to fine-tune HDE's alignment with the key factors that support a data sharing culture:

1. HDE should continue to build on early successes. For example, Fresh FAIR webinars, one-to-one consultations, and ongoing outreach programming have resulted in demonstrated increases to HDE participation. Evaluating efforts considering common researcher barriers and incentives helps program administrators understand the nuances of data sharing behavior. Much work remains to foster a sense of community with respect to data sharing. Additional planned efforts focused

Table 2. HEAL Data Ecosystem activities addressing common barriers to data sharing

HDE Outreach Activities	Relevant Data Sharing Factors	Program Outcomes and Assessment
Fresh FAIR webinar series and HEAL tutorials Webinars include: *Demystifying Data Sharing; Common Data Elements (CDEs); Protecting Privacy in HEAL Research: A Deep Dive Into Data De-identification*	• Patient concerns and ethical responsibilities • Fear of scooping • Where/when/how/why to share data; positive impacts on public health • Reproducibility and validation	There have been over 4,000 registrants for HEAL webinars and over 300 registrants for the three HEAL tutorials. This initiative has been one of the most successful, enabling the team to reach thousands of participants. One strength of this outreach strategy is its focus on a wide range of potential barriers to HEAL investigators. The format enables participants to ask questions and receive live guidance from Subject Matter Experts and guest speakers. Future plans include expanding the tutorial programming to provide additional data sharing guidance to HEAL-funded investigators.
PI checklist "tracker" The tracking tool lets researchers look up their study in the HEAL Data Platform and determine which steps they have completed in the data sharing process.	• Research efficiency • Where/when/how/why to share data	The checklist tracker fosters increased researcher engagement with HDE registration and repository selection. While the checklist tracker is a relatively new initiative, the site has logged over 450 tracker interactions to date. In addition, the tracker was used 114 times at the 2024 Annual NIH HEAL Initiative Scientific Meeting, where it was piloted. The HEAL Stewards plan additional tool refinements as they receive investigator feedback.

(continued)

HDE Outreach Activities	Relevant Data Sharing Factors	Program Outcomes and Assessment
Researcher 1–1 consultations Consultations may include general data management and HDE participation concerns; and using community resources, such as DMPTool.	• Where/when/how/why to share data • Considerations around sharing sensitive data • Fostering open science culture	HEAL Stewards have completed over 185 individualized consultation meetings since June 2022. Consultations increased threefold since the 2024 Annual PI Meeting as a result of the HEAL Stewards' "Get the Data" targeted initiative. The HEAL Stewards continue to refine engagement best practices with researchers. Ensuring that PIs share data in a repository and participate in HDE remain core goals. Current challenges include difficulty identifying publications (and subsequently datasets) that belong to HEAL-funded investigators.
Annual NIH HEAL Initiative Scientific Meeting programming Activities include a PI raffle; support with platform registration; and repository selection assistance.	• Fostering open science culture • Where/when/how/why to share data • Funding or resources gaps	PI meeting programming sparked an increase in overall PI engagement with HDE, including an uptick in HEAL Data Platform registrations and repository selection assistance requests. In calendar year 2024, there were 100 visits to the PI booth, including 27 repository selection consultations.
HEAL Collective Board The Collective Board, which is composed of HEAL researchers, guides the strategy and direction of the HDE to develop methods and norms, cultivate a culture of sharing, and maximize collaboration to enable translational discoveries directly benefiting patients in the opioid use disorder and pain communities. The Collective Board meets monthly to discuss ongoing data sharing challenges and provide insights on researcher concerns and needs.	• Fostering open science culture • Where/when/how/why to share data • Patient concerns and ethical responsibilities	The HEAL Collective Board has met over two dozen times to provide insights on HDE, metadata, repository selection, and other topics important to HEAL researchers. Areas for expansion and refinement include developing meeting topics most relevant to HEAL Collective Board members and HEAL researchers generally.
HealDataFAIR.org The HDE website includes extensive guidance for researchers on implementing FAIR data standards and provides a PI checklist, metadata standards, and repository selection guidance.	• Fostering open science culture • Where/when/how/why to share data • Co-authorship and collaboration • Promoting research in the field	The HealDataFAIR.org resources page has had over 9,000 views from September 2021 to September 2024, with a 57.8% engagement rate. The most viewed resources include the Checklist for HEAL-Compliant Data and the HEAL Data Repository Selection Guide, both tools to help individual researchers make decisions on data sharing. HEAL Stewards continue to refine and add materials in response to identified researcher needs.
Sensitive data decision tree A decision tree helps researchers make decisions about sharing sensitive human subjects data.	• Patient privacy and responsible data sharing	This resource has improved researcher access to individualized consulting and asynchronous resources on data privacy and de-identification. In addition to the online resource, the HEAL Stewards have refined standard operating procedures to address concerns about patient data via customized consulting calls.
Data Asset Inventory The Data Asset Inventory is an annual survey sent to HEAL studies. Results provide HEAL Stewards with data on HEAL researcher research practices and outcomes.	• FAIR data practices • Reproducibility and validation	Responses from 200+ study teams facilitate targeted outreach planning activities. Although we received a substantive number of responses, the current total study population is well over 1,000. HEAL Stewards would like to connect with additional study teams to learn about the types of data they generate and the challenges that may be unique to pain and opioid use disorder research. In addition, some qualitative data collection may be useful to get a clearer understanding of HEAL researcher needs and unique barriers.

Notes: DMPTool = Data Management Plan Tool; FAIR = findable, accessible, interoperable, and reusable; HDE = HEAL Data Ecosystem; HEAL = *Helping to End Addiction Long-term*®; NIH = National Institutes of Health; PI = principal investigator.

on (1) identifying potential dataset-associated publications and (2) targeting PI engagement to support HEAL-funded study teams and address data sharing challenges.

2. The HEAL Stewards and NIH HEAL Initiative leadership should continue to refine HDE programs and services in response to Collective Board and specific researcher input about their unique barriers. In particular, the team recognizes the challenge of tracking research outcomes over time and the lack of tracking mechanisms linking publications with data management plans.

3. Future programming should build on lessons learned through engagement activities and landscape analyses (including data asset inventories), which point to a continued need for outreach, online resources/guidance, and instructional programming. Researchers at all levels, particularly new researchers, benefit from data management support. In addition to consulting services, fostering community-wide connections and ensuring researchers are aware of existing resources at institutions will help study teams cross the data sharing finish line and cultivate a vibrant and collaborative research community.

Conclusion

One of HDE's core objectives is to implement evidence-driven strategies for building a culture of research data sharing and collaborative discovery. Existing literature helps provide a foundation

for HDE system growth and refinement to meet researchers' needs and address their challenges; however, additional data drawn from HEAL researcher feedback would help the HEAL Stewards fine-tune programming to meet specific needs and address challenges. Although substantial challenges to improving data sharing participation rates remain, the HEAL Stewards' outreach and engagement activities have been foundational in addressing some of the common barriers to sharing and fostering a collaborative research culture throughout the HEAL community. Much work remains to be done to align HDE programs fully with gaps in researchers' capacity to meet data sharing expectations. Exploring research evidence around data sharing behaviors—including the most common barriers to and incentives for sharing data—supports outreach programming and helps address researcher concerns. Practical guidance and services that address known barriers and provide targeted participation incentives are essential to foster researchers' ability and willingness to share data and digital assets.

Data Availability Statement

In this publication, we do not report on, analyze, or generate any data.

Generative AI Use

We confirm that we did not use generative AI tools/ services to author this submission.

References

Adams, M. J., Harris, R. N., Grant, E. H. C., Gray, M. J., Camille Hopkins, M., Iverson, S. A., Likens, R., Mandica, M., Olson, D. H., Shepack, A., & Waddle, H. (2018). Prepublication communication of research results. *EcoHealth*, *15*(3), 478–481. https://doi.org/10.1007/s10393-018-1352-3

Barczak, G., Hopp, C., Kaminski, J., Piller, F., & Pruschak, G. (2022). How open is innovation research? An empirical analysis of data sharing among innovation scholars. *Industry and Innovation*, *29*(2), 186–218. https://doi.org/10.1080/13662716.2021.1967727

Berman, H. M., Gabanyi, M. J., Groom, C. R., Johnson, J. E., Murshudov, G. N., Nicholls, R. A., Reddy, V., Schwede, T., Zimmerman, M. D., Westbrook, J., & Minor, W. (2015). Data to knowledge: How to get meaning from your result. *IUCrJ*, *2*(1), 45–58. https://doi.org/10.1107/S2052252514023306

Bock, G.-W., Zmud, R. W., Kim, Y.-G., & Lee, J.-N. (2005). Behavioral intention formation in knowledge sharing: Examining the roles of extrinsic motivators, social-psychological forces, and organizational climate. *Management Information Systems Quarterly*, *29*(1), 87–111. https://doi.org/10.2307/25148669

Callaway, E. (2019). Scooped in science? Relax, credit will come your way. *Nature*, *575*(7784), 576–577. https://doi .org/10.1038/d41586-019-03648-4

Chawinga, W. D., & Zinn, S. (2019). Global perspectives of research data sharing: A systematic literature review. *Library & Information Science Research*, *41*(2), 109–122. https://doi.org/10.1016/j.lisr.2019.04.004

Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. *PLoS One*, *15*(4), e0230416. https://doi.org/10.1371/journal.pone.0230416

Curty, R., Yoon, A., Jeng, W., & Qin, J. (2016). Untangling data sharing and reuse in social sciences. *Proceedings of the Association for Information Science and Technology*, *53*(1), 1–5. https://doi.org/10.1002/pra2.2016 .14505301025

Devriendt, T., Borry, P., & Shabani, M. (2021). Factors that influence data sharing through data sharing platforms: A qualitative study on the views and experiences of cohort holders and platform developers. *PLoS One*, *16*(7), e0254202. https://doi.org/10.1371/journal.pone .0254202

Druedahl, L. C., Minssen, T., & Price, W. N. (2021). Collaboration in times of crisis: A study on COVID-19 vaccine R&D partnerships. *Vaccine*, *39*(42), 6291–6295. https://doi.org/10.1016/j.vaccine.2021.08.101

Duke, C. S., & Porter, J. H. (2013). The ethics of data sharing and reuse in biology. *Bioscience*, *63*(6), 483–489. https://doi.org/10.1525/bio.2013.63.6.10

Farley, S. S., Dawson, A., Goring, S. J., & Williams, J. W. (2018). Situating ecology as a big-data science: Current advances, challenges, and solutions. *Bioscience*, *68*(8), 563–576. https://doi.org/10.1093/biosci/biy068

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

Hardy, J. (2021, July 8). *Fear of being scooped is fuelling the replication crisis in research*. Times Higher Education (THE). https://www.timeshighereducation.com/ opinion/fear-being-scooped-fuelling-replication-crisis -research

Hipsley, C. A., & Sherratt, E. (2019). Psychology, not technology, is our biggest challenge to open digital morphology data. *Scientific Data*, *6*(1), 41. https://doi .org/10.1038/s41597-019-0047-0

Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V. M., Nichols, T. E., & Wagenmakers, E. J. (2018). Data sharing in psychology: A survey on barriers and preconditions. *Advances in Methods and Practices in Psychological Science*, *1*(1), 70–85. https://doi.org/10 .1177/2515245917751886

Hulsen, T. (2020). Sharing is caring—Data sharing initiatives in healthcare. *International Journal of Environmental Research and Public Health, 17*(9), 3046. https://doi.org/10.3390/ijerph17093046

Hutchings, E., Loomes, M., Butow, P., & Boyle, F. M. (2020). A systematic literature review of researchers' and healthcare professionals' attitudes towards the secondary use and sharing of health administrative and clinical trial data. *Systematic Reviews*, *9*(1), 240. https:// doi.org/10.1186/s13643-020-01485-5

Laine, H. (2017). Afraid of scooping—Case study on researcher strategies against fear of scooping in the context of open science. *Data Science Journal*, *16*(0), 29. Advance online publication. https://doi.org/10.5334/dsj -2017-029

Late, E., Skov, M., & Kumpulainen, S. (2024). To share or not to share? Image data sharing in the social sciences and humanities. *Information Research: An International Electronic Journal*, *29*(2), 386–400. https://doi.org/10 .47989/ir292834

Lounsbury, O., Roberts, L., Goodman, J. R., Batey, P., Naar, L., Flott, K. M., Lawrence-Jones, A., Ghafur, S., Darzi, A., & Neves, A. L. (2021). Opening a "can of worms" to explore the public's hopes and fears about health care data sharing: Qualitative study. *Journal of Medical Internet Research*, *23*(2), e22744. https://doi.org/10 .2196/22744

McKiernan, E. C., Schimanski, L. A., Muñoz Nieves, C., Matthias, L., Niles, M. T., & Alperin, J. P. (2019). Use of the Journal Impact Factor in academic review, promotion, and tenure evaluations. *eLife*, *8*, e47338. https://doi.org/10.7554/eLife.47338

Melero, R., & Navarro-Molina, C. (2020). Researchers' attitudes and perceptions towards data sharing and data reuse in the field of food science and technology. *Learned Publishing*, *33*(2), 163–179. https://doi.org/10 .1002/leap.1287

Mozersky, J., McIntosh, T., Walsh, H. A., Parsons, M. V., Goodman, M., & DuBois, J. M. (2021). Barriers and facilitators to qualitative data sharing in the United States: A survey of qualitative researchers. *PLoS One*, *16*(12), e0261719. https://doi.org/10.1371/journal.pone.0261719

National Institutes of Health. (n.d.). *NOT-OD-21–013: Final NIH policy for data management and sharing.* https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html

National Institutes of Health *Helping to End Addiction Long-term®* Initiative. (2021). *About the HEAL Data Ecosystem | NIH HEAL Initiative*. https://heal.nih.gov/data/heal-data-ecosystem

National Institutes of Health *Helping to End Addiction Long-term®* Initiative Data Stewardship Group. (n.d.). *HealDataFAIR.org.* https://www.healdatafair.org

Network of the National Library of Medicine. (2024). *Network of the National Library of Medicine glossary.* https://www.nnlm.gov/guides/data-glossary/data-sharing

Park, J., & Gabbard, J. L. (2018). Factors that affect scientists' knowledge sharing behavior in health and life sciences research communities: Differences between explicit and implicit knowledge. *Computers in Human Behavior*, *78*, 326–335. https://doi.org/10.1016/j.chb.2017.09.017

Pearson, H. (2003). Competition in biology: It's a scoop! *Nature*, *426*(6964), 222–223. https://doi.org/10.1038/426222a

Pratt, B., & Bull, S. (2021). Equitable data sharing in epidemics and pandemics. *BMC Medical Ethics*, *22*(1), 136. https://doi.org/10.1186/s12910-021-00701-8

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Sarabipour, S., Debat, H. J., Emmott, E., Burgess, S. J., Schwessinger, B., & Hensel, Z. (2019). On the value of preprints: An early career researcher perspective. *PLoS Biology*, *17*(2), e3000151. https://doi.org/10.1371/journal.pbio.3000151

Sardanelli, F., Alì, M., Hunink, M. G., Houssami, N., Sconfienza, L. M., & Di Leo, G. (2018). To share or not to share? Expected pros and cons of data sharing in radiological research. *European Radiology*, *28*(6), 2328–2335. https://doi.org/10.1007/s00330-017-5165-5

Soeharjono, S., & Roche, D. G. (2021). Reported individual costs and benefits of sharing open data among Canadian academic faculty in ecology and evolution. *Bioscience*, *71*(7), 750–756. https://doi.org/10.1093/biosci/biab024

Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., Iyer, R., Schatz, M. C., Sinha, S., & Robinson, G. E. (2015). Big data: Astronomical or genomical? *PLoS Biology*, *13*(7), e1002195. https://doi.org/10.1371/journal.pbio.1002195

Teixeira da Silva, J. A., & Dobránszki, J. (2015). Potential dangers with open access data files in the expanding open data movement. *Publishing Research Quarterly*, *31*(4), 298–305. https://doi.org/10.1007/s12109-015-9420-9

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

Tse, E. G., Klug, D. M., & Todd, M. H. (2020). Open science approaches to COVID-19. *F1000Research*, *9*, 1043. https://doi.org/10.12688/f1000research.26084.1

Vickers, A. J. (2006). Whose data set is it anyway? Sharing raw data from randomized trials. *Trials*, *7*(1), 15. https://doi.org/10.1186/1745-6215-7-15

Vines, T. H., Andrew, R. L., Bock, D. G., Franklin, M. T., Gilbert, K. J., Kane, N. C., Moore, J.-S., Moyers, B. T., Renaut, S., Rennison, D. J., Veen, T., & Yeaman, S. (2013). Mandated data archiving greatly improves access to research data. *The FASEB Journal*, *27*(4), 1304–1308. https://doi.org/10.1096/fj.12-218164

Walsh, J. P., & Hong, W. (2003). Secrecy is increasing in step with competition. *Nature*, *422*(6934), 801–802. https://doi.org/10.1038/422801c

RTI Press Publication No. OP-0095-2501. Research Triangle Park, NC: RTI Press.

https://doi.org/10.3768/rtipress.2025.op.0095.2501

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., Bonino da Silva Santos, L., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Woods, H. B., & Pinfield, S. (2021). Incentivising research data sharing: A scoping review [version 1; peer review: 3 approved]. *Wellcome Open Research*, *6*, 355. https://doi.org/10.12688/wellcomeopenres.17286.2

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