

Getting Started with Data Sharing: Advice for Researchers in Education

Christine M. White*, Florida State University, Department of Psychology, Tallahassee, FL, USA; Florida Center for Reading Research, Tallahassee, FL, USA, white@psy.fsu.edu

Stephanie A. Estrera, Florida State University, Department of Psychology, Tallahassee, FL, USA; Florida Center for Reading Research, Tallahassee, FL, USA, estrera@psy.fsu.edu

Christopher Schatschneider, Florida State University, Department of Psychology, Tallahassee, FL, USA; Florida Center for Reading Research, Tallahassee, FL, USA, schatschneider@psy.fsu.edu

Sara A. Hart, Florida Center for Reading Research, Tallahassee, FL, USA; University of Waterloo, Department of Psychology, Waterloo, Ontario, Canada, sahart@fsu.edu

*Corresponding author.

ABSTRACT

Researchers in the education sciences, like those in other disciplines, are increasingly encountering requirements and incentives to make the data supporting empirical research available to others. However, the process of preparing and sharing research data can be daunting. The present article aims to support researchers who are beginning to think about integrating data sharing into their research workflows and includes a brief review of the current landscape of data sharing amongst education and special education researchers, a discussion of the benefits of

Submitted: 11 July 2024

Accepted: 02 August 2024

Published: 18 November 2024

Research in Special Education is a peer-reviewed open access journal published by the Aperio. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



OPEN ACCESS

data sharing, a curated list of existing resources designed to help researchers in education and related disciplines, a summary of the steps to data sharing throughout the research lifecycle, and responses to several common questions researchers may have when it comes to sharing research data for the first time.

Keywords: open science; data sharing; education science; education; open practices

Data sharing refers to the process of making any type of research data available to others for examination or reuse (Logan et al., 2021). This can take multiple forms, ranging from presenting processed data (such as means and other summary statistics) as a table or figure in a research publication to uploading raw, item-level datasets to an online data repository along with documentation and code to help others replicate published analyses or conduct new ones. While there is not a one-size-fits-all approach to making research data available, there are widely accepted standards for data sharing called the FAIR principles, which are intended to increase the likelihood that shared data are found and used by their intended audience, whether that is the public or a smaller group of researchers (Wilkinson et al., 2016). The FAIR principles are important because they give a framework for data sharing that allows for easy reuse of the data.

There are four FAIR principles: Findability, Accessibility, Interoperability, and Reuse. First, to be able to examine or reuse data, someone needs to be able to find them—the Findability principle. Therefore, shared data must be accompanied by metadata. Metadata are “data about data” or, in other words, data that allow others to know and search for the properties or contents of your shared data without opening the data (Logan et al., 2021). Second, once someone finds data, they need to be able to access them—the Accessibility principle. Minimally, anyone with a computer and internet access should be able to access at least descriptive information about the data. Third, shared data should be Interoperable, meaning that data and corresponding metadata are stored in a way that other computers can read them. Finally, the fourth principle, Reuse, encompasses two concepts. First, anyone who does find and access your data also needs to have all the critical information that will help them understand them and use them appropriately (i.e., metadata, documentation). The second concept is data and metadata provenance. All data and data documentation should have digital object identifiers (DOIs), with authors and other important information clearly assigned, allowing proper citation.

There are many reasons for education researchers to make data supporting empirical studies available to others. Some may commit to data sharing along with other open science practices to help create a research evidence base that is high-quality; reproducible; and freely accessible to the public and professionals, who may leverage research materials and findings for the improvement of policy and practice. Some might commit to data sharing because they have funding from granting agencies with data sharing requirements, such as the National Institutes of Health (NIH)’s Data Management and Sharing Policy, which went into effect in January 2023. Some might commit to data sharing because they wish to publish their findings in journals with data sharing policies or incentives. For example, the American Educational Research Association’s (AERA) open-access journal *AERA Open* uses the following language in their author guidelines: “In keeping with AERA policy, *AERA Open* encourages sharing of data and/or data files whenever feasible” (AERA, 2024). Some journals, such as *Exceptional Children* and the *Journal of Research on Educational Effectiveness*, have also started offering digital “badges” to highlight and encourage contributor participation in open science practices including data sharing (Center for Open Science, n.d.; Kidwell et al., 2016). Finally, some researchers may wish to share data because of its benefits to authors (e.g., increased citations; Colavizza et al., 2020) and to other researchers in one’s field (e.g., enabling others to ask novel questions with a data set they may not have resources to personally collect.)

Whatever the reason, there is a shift towards a culture of transparent and accessible education research. This culture change is already reflected in institutional policies and is unlikely to slow anytime soon. Thus, it is important for

researchers at every career stage to “get on board” with data sharing or, in other words, to understand the general process, become aware of available resources, and start to consider how they can integrate open science practices into existing procedures for generating and publishing research findings. To that end, this paper aims to describe the current landscape of data sharing in the education sciences and bring together advice and existing resources designed to help education researchers throughout the data sharing process.

DATA SHARING IN THE EDUCATION SCIENCES: A VALUE-ACTION GAP

How often are researchers in the education sciences currently sharing data? In a survey conducted by Makel and colleagues (2021), 25% of education researchers reported that their colleagues regularly engage in data sharing and 45% reported that they had personally shared data to an online public repository at least once in the past. This latter finding was replicated by Logan et al. (2024), who found that 42% of their sample of 178 education researchers reported having shared data in a repository at least once. Education researchers have also expressed positive opinions of data sharing, with 97% of Logan et al.’s respondents agreeing with the idea that data sharing is good for science and 72% of respondents expressing generally positive views on data sharing. Similar sentiments have been observed among special education researchers: a recent survey by Fleming et al. (2024) found that 57% of respondents reported having a plan to share data in the next two years. Additionally, data sharing received the highest endorsement of any open science practice, with 78% of respondents reporting a favorable opinion and only 4% reporting an unfavorable opinion.

Based on these findings, it seems reasonable to conclude that researchers in the education sciences value data sharing and are willing to engage in it. This is promising when considered within the “theory of planned behavior” framework discussed by Fleming and colleagues, which posits that an individual’s intent to engage in an activity is a prerequisite and predictor of their actual engagement in it (Ajzen, 1985). However, an objective look at the rate of data availability in peer-reviewed research journals suggests a disconnect between intention and action when it comes to sharing research data. In a targeted review of 250 empirical articles published in special education research journals in 2020, Cook and colleagues (2023) found that only 14 articles (7%) contained any indication that supporting data had been made available, and only 3 (1.5%) had data that were actually accessible by the study team. Notably, the incidence of data sharing was much lower than that of other open science practices such as sharing study materials (21%) and publishing open access (23%). These findings are corroborated by a similar review in which Huff and Bongartz (2023) found that the rate of research data availability in papers published in educational psychology journals in 2020 was 7.16%, compared to over 60% in the general psychology journal *Cognition*. More recently, in preparation for a larger project, our team conducted an informal review of all articles published in the *Journal of Education Psychology*, *American Educational Research Journal*, and *Child Development* in 2022 and found that only 15 of 96 eligible articles contained any indication of data sharing.

Despite favorable opinions, the rate of data sharing among education and special education researchers is disproportionately low compared to engagement in other open science practices. Borrowing a term from environmental science, these findings imply the existence of a “value-action gap” in which researchers’ beliefs about the value and perhaps the social desirability of data sharing are not reflected in actual data sharing behavior (Barr, 2006). To understand this phenomenon, Ajzen’s “theory of planned behavior” describes three conditions that influence intent and subsequently action: personal attitudes, perception of norms, and perceived control over the target behavior (Ajzen, 1985). Ajzen’s theory was applied to data sharing behavior (and other two open science practices) in Fleming et al.’s (2024) survey of special education researchers, which operationalized “personal attitudes” as researchers’ ratings of the favorability and perceived benefits and drawbacks of data sharing; “perception of norms” as researchers’ estimate of how many others in their field viewed data sharing favorably; and “perceived control” as researchers’ self-reported knowledge of how to implement data sharing. In addition to personal attitudes, Fleming and colleagues found that perceived control (namely, self-reported knowledge of data sharing procedures) was a significant and

positive predictor of special education researchers' intent to share data. This suggests that one key factor underlying the value-action gap is that researchers do not feel a sense of control over engagement in data sharing because they are not confident in their knowledge of the processes and procedures involved in data sharing. Indeed, 74% of special education researchers surveyed by Fleming et al. (2024) reported having low or no knowledge of data sharing—a sentiment that has also been expressed by researchers in other disciplines. For example, El Amin and colleagues (2023) invited researchers in communication sciences and disorders to freely describe perceived barriers to data sharing, and the most frequent response was “I don't know how to share data,” mentioned by 35% of respondents. A 2018 survey of researchers across disciplines tells a similar story: Houtkoop and colleagues found that while opinions of data sharing were favorable, many respondents believed that they were not ethically allowed to share data, and half of respondents stated that they “never learned how to share data online.”

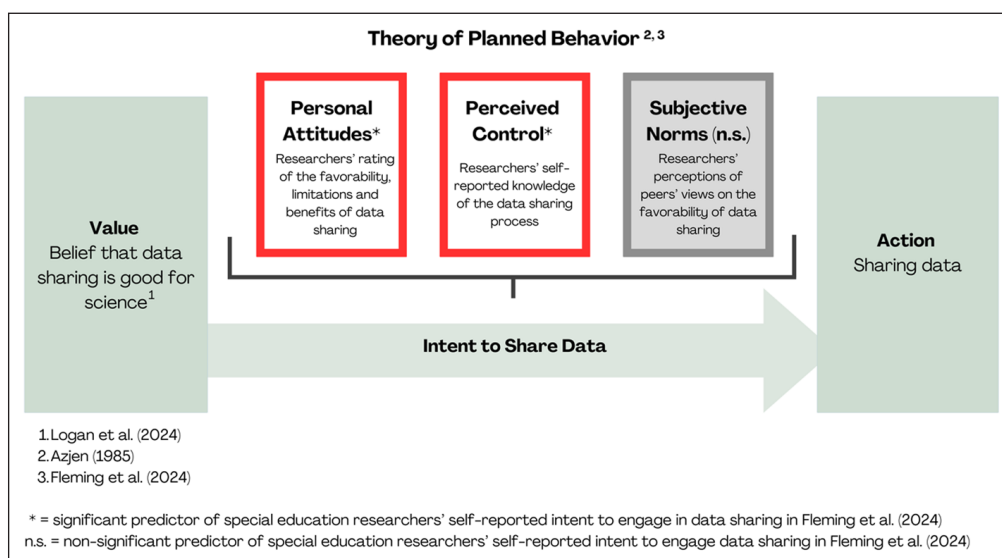


Figure 1: Applying the Value-Action Gap and Ajzen's Theory of Planned Behavior to Data Sharing.

Figure available at <https://doi.org/10.6084/m9.figshare.26252522.v1> under a CC BY 4.0 license.

The existence of the value-action gap suggests that simply telling education researchers about the importance of data sharing will not move the needle on the prevalence of open data in our field. Rather, the meta-scientific evidence on data sharing among education researchers suggests that it may be more impactful to address researchers' feelings of control over data sharing behavior by building knowledge. Thus, the next section of this manuscript aims to provide tools and tips for researchers who may be navigating the data sharing process for the first time in a succinct and approachable way. As a team with years of hands-on experience on both sides of the “data sharing equation” (sharing research data and maintaining a data repository), we hope the present guide will serve as a valuable resource for researchers who want or need to share data but are not sure where to get started.

HOW TO GET STARTED WITH DATA SHARING

As proponents of open science are aware, there are many free and reputable resources for learning about data curation, deidentification, and sharing. However, it may be difficult for education researchers to locate, navigate, and synthesize the guidelines that are specifically relevant for their discipline. Thus, our first goal is to collect and summarize a representative sample of existing scholarly publications that were developed with the intention of informing and guiding researchers in education and related fields through the process of data sharing (Table 1). While this list is non-exhaustive, it can be seen as a starting point for learning about data sharing.

PAPER	RESEARCH AREA	FOCUS	DETAILS
Lewis, C. (2024a). <i>Data management in large-scale educational research</i> . Chapman & Hall. Free online version accessible at: https://datamgmtinedresearch.com/	Education	Data management	Extensive guide for data management throughout educational research lifecycle; specific guidance for data sharing in Chapter 16
Logan, J. A., Hart, S. A., & Schatschneider, C. (2021). Data sharing in education science. <i>AERA Open</i> , 7, 23328584211006475. https://doi.org/10.1177/23328584211006475	Education	Data sharing	Step-by-step guide and responses to common concerns
Neild, R. C., Robinson, D., & Agufa, J. (2022). Sharing study data: A guide for education researchers. Toolkit. NCEE 2022–004. <i>National Center for Education Evaluation and Regional Assistance</i> . https://ies.ed.gov/ncee/pubs/2022004/pdf/2022004.pdf	Education	Data sharing	Toolkit from the National Center on Educational Effectiveness
Van Dijk, W., Schatschneider, C., & Hart, S. A. (2021). Open science in education sciences. <i>Journal of Learning Disabilities</i> , 54(2), 139–152. https://doi.org/10.1177/0022219420945267	Education	Open science practices	Overview and how-to guide for open science practices including data sharing; see Figure 1 (p. 144) for infographic representation of data sharing process
Cook, B. G., Fleming, J. I., Hart, S. A., Lane, K. L., Therrien, W. J., van Dijk, W., & Wilson, S. E. (2022). A how-to guide for open-science practices in special education research. <i>Remedial and Special Education</i> , 43(4), 270–280. https://doi.org/10.1177/07419325211019100	Special education	Open science practices	How-to guide for open science practices including data sharing (p. 274–276)
Kalandadze, T. & Hart, S. A. (2022). Open developmental science: An overview and annotated reading list. <i>Infant and Child Development</i> , 33(1), e2334. https://doi.org/10.1002/icd.2334	Developmental science	Open science practices	A reading list for learning about open science practices including data sharing (p. 8–9)
Edwards, A. A., van Dijk, W., Tripodi, S. J., & Hart, S. A. (2023). Data sharing for randomized controlled trials in social work. <i>Research on Social Work Practice</i> . Online first publication. https://doi.org/10.1177/10497315231186799	Social work	Data sharing	How-to guide for sharing data from randomized controlled trials
Meyer, M. (2018). Practical tips for ethical data sharing. <i>Advances in Methods and Practices in Psychological Science</i> , 1(1), 131–144. https://doi.org/10.1177/2515245917747656	Psychology	Data sharing	Practical tips for data sharing
Gilmore, R. O., Kennedy, J. L., & Adolph, K. E. (2018). Practical solutions for sharing data and materials from psychological research. <i>Advances in Methods and Practices in Psychological Science</i> , 1(1), 121–130. https://doi.org/10.1177/2515245917746500	Psychology	Data and materials sharing	Practical tips for data sharing

Table 1: A Curated List of Existing Guides to Sharing Research Data.

While these guides contain nuanced information for their respective target audiences, the general steps to preparing and sharing research data are highly similar across them and can be abstracted into five general steps: Permission Check, Data Preparation, Data Deidentification, Metadata Creation, and Data Upload. Thus, our second goal is to present an abstracted version of these five steps and provide resources for each. In Figure 2, we summarize these steps, while Table 2 provides a non-exhaustive list of online resources for each step. Note that while the steps are presented in sequence, different aspects of preparation are relevant at different points throughout the research lifecycle (i.e., before, during, or after data collection), as visualized in Figure 2. In other words, some steps are specifically relevant to preparing existing data to be shared (e.g., data deidentification), while others can (and should) be incorporated into standard research procedures before and during data collection (e.g., using consent form language that permits data sharing, adhering to good data management practices, and creating and maintaining detailed documentation).

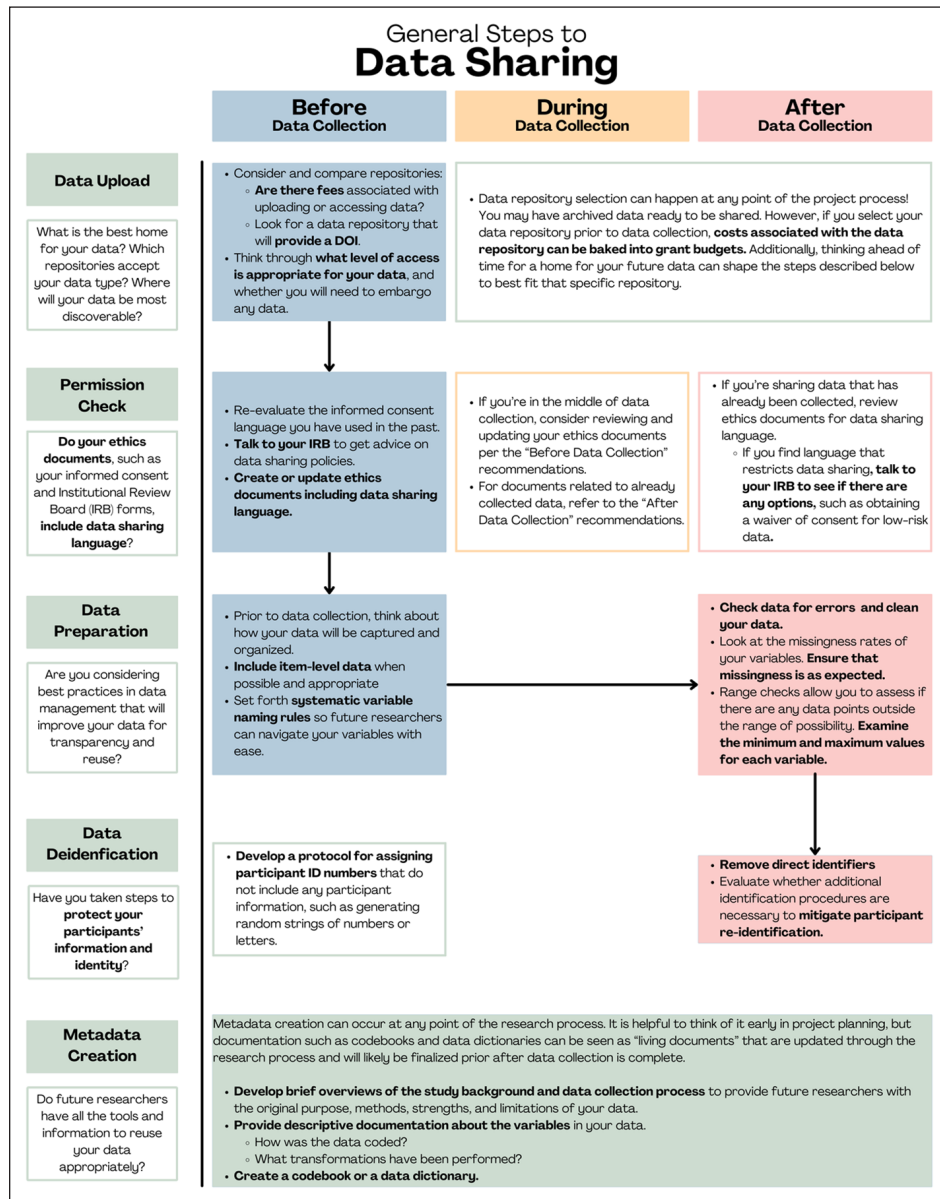


Figure 2: General Steps to Data Sharing.

Figure available at <https://doi.org/10.6084/m9.figshare.26252402.v1> under a CC BY 4.0 license.

STEP	WHAT IS INVOLVED?	RESOURCES
Permission Check	Adding or checking for data sharing language in ethics documents (e.g., informed consent or other Institutional Review Board (IRB) forms)	<p>Templates:</p> <ul style="list-style-type: none"> • Open Science Framework. (2016). <i>IRB and consent form examples</i>. https://osf.io/g4jfv/wiki/home/ • Shero & Hart (2020a). Informed consent template. https://doi.org/10.6084/m9.figshare.13218773.v1 • Shero & Hart (2020b). IRB protocol template. https://doi.org/10.6084/m9.figshare.13218797.v1 <p>Guides:</p> <ul style="list-style-type: none"> • Gilmore et al. (2018). Practical solutions for sharing data and materials from psychological research. https://doi.org/10.1177/2515245917746500 <ul style="list-style-type: none"> ◦ Table 2: “Resources for seeking permission to share identifiable data and recordings” (p. 124) • Shero & Hart (2022). Working with your IRB: Obtaining consent for open data sharing through consent forms and data use agreements. https://doi.org/10.6084/m9.figshare.13215305.v2
Data Preparation	Adhering to good data management practices throughout project; performing simple summary statistics such as range checks and missingness calculations when creating final datasets for sharing	<p>Services:</p> <ul style="list-style-type: none"> • Inter-university Consortium for Political and Social Research. (2020). <i>ICPSR curation levels</i>. https://www.icpsr.umich.edu/files/datamanagement/icpsr-curation-levels.pdf • Digital Curation Centre. (2024). <i>Services</i>. https://www.dcc.ac.uk/services <p>Guides:</p> <ul style="list-style-type: none"> • Inter-university Consortium for Political and Social Research. (2024). <i>Guide to social science data preparation and archiving</i>. https://www.icpsr.umich.edu/web/pages/deposit/guide/ • Lewis (2024a). <i>Data management for large-scale education research</i>. https://datamgmtinedresearch.com/ • Reynolds et al. (2022). The basics of data management. https://doi.org/10.6084/m9.figshare.13215350.v2 <p>Trainings:</p> <ul style="list-style-type: none"> • Center for Open Science. (n.d.). <i>Events & webinars</i>. https://www.cos.io/events • Logan & Hart (2023). Data management for data sharing: Workshop for the Purdue University Emerging Perspectives in Early STEM Learning. https://doi.org/10.6084/m9.figshare.24460651.v1 <ul style="list-style-type: none"> ◦ Authors provide bespoke data management for data sharing workshops on request. • Center for Open Science (n.d.). <i>Openness and reproducibility research practices training</i>. https://www.cos.io/services/training
Data Deidentification	Removing direct identifiers and evaluating whether additional de-identification procedures are needed to mitigate the possibility of participant re-identification	<p>Guides:</p> <ul style="list-style-type: none"> • LDbase (n.d.). <i>Data deidentification</i>. https://ldbase.org/data-sharing-resources/best-practices/data-de-identification • Logan et al. (2021). Data sharing in education science. https://doi.org/10.1177/23328584211006475 <ul style="list-style-type: none"> ◦ “After data have been collected: De-identify your data” (p. 5) • Meyer (2018). Practical tips for ethical data sharing. https://doi.org/10.1177/2515245917747656 <ul style="list-style-type: none"> ◦ “Do be thoughtful when considering risks of re-identification” (p. 133) • Schatschneider et al. (2021). Deidentifying data. https://doi.org/10.6084/m9.figshare.13228664.v2 <p>Tools:</p> <ul style="list-style-type: none"> • Johns Hopkins University Data Services. (2016). <i>Applications to assist in de-identification of human subjects research data</i>. https://dataservices.library.jhu.edu/resources/applications-to-assist-in-de-identification-of-human-subjects-research-data/

STEP	WHAT IS INVOLVED?	RESOURCES
Metadata Creation	Providing descriptive documentation about the variables in your data, how they are coded, and what transformations have been performed (i.e., creating a codebook or “data dictionary”)	<p>Guides:</p> <ul style="list-style-type: none"> • Lewis. (2024b). <i>Data management</i>. https://cghlewis.com/categories/data-management/ • Logan & Hart (2023). Data management for data sharing: Workshop for the Purdue University Emerging Perspectives in Early STEM Learning. https://figshare.com/articles/presentation/Data_Management_for_Data_Sharing_Workshop_for_the_Purdue_University_Emerging_Perspectives_in_Early_STEM_Learning_/24460651 • Logan et al. (2021). Data sharing in education science. https://doi.org/10.1177/23328584211006475 <ul style="list-style-type: none"> ◦ “Preparing data for sharing: Document your study” (p. 6)
Data Upload	Selecting an appropriate repository/level of access for your data	<p>Guides:</p> <ul style="list-style-type: none"> • Gilmore et al. (2018). Practical solutions for sharing data and materials from psychological research. https://doi.org/10.1109/NYSDS.2016.7747817 <ul style="list-style-type: none"> ◦ Table 1: “Some resources for sharing data and materials” (p. 123) • Meyer (2018). Practical tips for ethical data sharing. https://doi.org/10.1177/2515245917747656 <ul style="list-style-type: none"> ◦ Table 1: “Governance attributes of some social-science and general data repositories” (p. 136) • Neild et al. (2022). Sharing study data: A guide for education researchers. https://ies.ed.gov/ncee/pubs/2022004/pdf/2022004.pdf <ul style="list-style-type: none"> ◦ Table 2: “At-a-glance summary of data repositories” (p. 27) • National Institutes of Health. (n.d.). <i>Selecting a data repository</i>. https://sharing.nih.gov/data-management-and-sharing-policy/sharing-scientific-data/selecting-a-data-repository <p>Policies:</p> <ul style="list-style-type: none"> • Databrary. (2020). <i>Databrary access agreement</i>. https://databrary.org/about/agreement/agreement.html • Inter-university Consortium for Political and Social Research. (2024). <i>Access and dissemination</i>. https://www.icpsr.umich.edu/web/pages/datamanagement/lifecycle/access.html • LDbase. (n.d.) <i>Frequently asked questions</i>. https://ldbase.org/data-sharing-resources/faq/general • Open Science Framework. (2023). <i>Control your privacy settings</i>. https://help.osf.io/article/285-control-your-privacy-settings

Table 2: General Steps to Data Sharing and Curated Resources.

Our final goal is to expand upon three practical decisions that researchers in education need to make when committing to make research data available: where to share data, when to share it, and who should be responsible for the process of preparing and sharing data.

WHERE TO SHARE YOUR DATA

There are five categories of places an investigator can share their data. Three categories involve sharing in an online data repository, including a domain-general repository, a domain-specific repository specializing in a content area, and a domain-specific repository specializing in a data format. The remaining two categories, sharing as supplementary materials of a journal article and sharing as “data are available upon request,” are less desirable and are not recommended.

Domain-general or generalist repositories accept data from any scientific discipline or data type (e.g., the Open Science Framework [OSF; <https://osf.io>]; the Inter-university Consortium for Political and Social Research [ICPSR; <https://www.icpsr.umich.edu/>]; or an institutional repository maintained by a university library or other entity, such as the

University of Illinois DataBank [<https://databank.illinois.edu>] or any of the 131 repositories in the Dataverse community [<https://dataverse.org>]). They are an appropriate and useful option for investigators who belong to a discipline, or have a data type, that is not accepted into a domain-specific repository. Domain-specific repositories store data from a specific discipline and/or in a specific format. For example, LDbase (<https://ldbase.org>) accepts quantitative data relating to education, learning, and development, importantly including learning disabilities data (Hart et al., 2024; Hart et al., 2020), while Databrary (<https://nyu.databrary.org>) is specialized for identifiable audio and video data (Gilmore et al., 2016). There are two key benefits to sharing your data in a domain-specific repository. First, because domain-specific repositories are developed by and for a particular community based on knowledge of the types of project structures and data types commonly used in this field, they are designed to support easy and intuitive sharing. More importantly, because domain-specific repositories are narrower in scope and often feature a pre-defined set of metadata terms of specific relevance to a particular field, they may enhance the ability of other researchers and practitioners in your field to find and reuse your data. If uploading to a domain-specific repository is not feasible, note that some repositories allow researchers to link to data shared elsewhere. For example, although LDbase does not accept video data, a researcher with both quantitative behavioral data and video data could store their quantitative data on LDbase and their video data on Databrary that is linked to using LDbase, enhancing findability for researchers and professionals in education. Databrary also encourages this approach for data types they do not accept (e.g., neuroimaging data; Gilmore et al., 2018). You may also utilize this approach if your institution encourages or requires investigators to share data in an institutional repository that are typically domain-general (see Carlson, 2020, for a brief list of university repositories in the U.S.). To comply with this expectation while still optimizing the findability and accessibility of your data, it may be possible to share data in a domain-specific repository and link this back to your institution's repository.

As mentioned previously, there are two data sharing options we do not recommend. The first is sharing data as supplemental materials to a published article. Although certainly this is sharing your data, it does not meet the FAIR data sharing principles as it restricts the findability of your data to only researchers in your field who might read your article. Given the prevalence of paywalls in scientific research journals, data in the supplemental materials can be blocked to those without access. Finally, some authors “share” their data by indicating within a published article that “data are available upon request” to the authors. This approach is not recommended as it severely limits the findability of your data and can lead to complications with accessibility when there are changes to personnel or institutional affiliations down the road. Further, accumulating evidence shows that response rates to requests for data available on request are notoriously low and may be biased (Acciai et al., 2023; Tedersoo et al., 2021). Thus, this option is not considered a useable form of data sharing.

In conclusion, we recommend sharing data in an online repository whenever feasible. This is not only the most FAIR method of data sharing but aligns best with federal funder recommendations. For example, the 2023 NIH Data Management and Sharing policy “promotes the use of established data repositories” for depositing data (NIH, 2023), while the newest Institute of Education Sciences (IES) Public Access Plan states that Data Sharing and Management Plans (DSMPs) for new awards “should include a plan to share data in public repositories” (NCER, 2024). Please refer to the “Data Upload” section of Table 2 to find resources listing and comparing existing repositories on key features and how well they satisfy federal funding agency requirements.

Note that the timing of deciding on strategy and location for data deposit may vary depending on context, as visualized in Figure 2. For example, if you are applying for federal funding for a project, you may need to decide on a data repository before the project has started (e.g., to include a Data Sharing and Management Plan with your application; NCER, 2024). This can be useful for making sure your data management procedures result in data and documentation that align with the standards of the repository where you will ultimately share them. Alternatively, you may have a data set from a former project that you now wish to share, in which case your decision of where to share may be guided by features of the existing data set (e.g., file size and structure, data types.)

WHEN TO SHARE YOUR DATA

There is no one answer to the question “when should I share my data?” For many investigators, it is the case that you can decide to share your data when you want to. Open science practices encourage sharing data to accompany a published manuscript to support the credibility of the research, and doing so can also benefit you as the investigator with increased citations of that work (Christensen et al., 2019). Given that many projects in education are larger than a single manuscript, an investigator may choose to share data after the primary aims of a project have been met, often after the researchers have published the main findings of their project.

For federally funded investigators in the United States, it is becoming the norm that there is not a choice in when you share your data. Federal funding agencies are updating their data management and sharing policies to reflect guidance from the White House Office of Science and Technology Policy’s 2022 memorandum that requires researchers to make their data freely available by the time of the publication (Nelson, 2022). For example, the Institute of Education Sciences has announced a new Public Access Plan starting in Fiscal Year 2025 that will require investigators to share data immediately upon publication (NCER, 2024). Researchers pursuing and awarded federally funded grants should refer to the grant agencies’ most recent data management and sharing policies for specific timeline requirements for sharing data.

In our experience, some researchers have noted concerns about “scooping” (i.e., when another person uses data or findings from a project before the original researcher has a chance to publish them) given recently changed timelines for data sharing. Despite scooping being a commonly cited concern, there is very little empirical research on this topic, and thus it is difficult to say how often it really occurs (in the context of data sharing or otherwise). Antidotally, we have found that when secondary data users approach a shared dataset, they bring their own theoretical leanings and methodological approaches to their research questions, making true scooping likely very rare. To assuage concerns over “scooping,” however, it should be noted that unless required by a funding agency, data do not have to be made freely available to any user immediately after data collection is complete. Many data repositories have an embargo feature, which allows researchers to temporarily restrict public access to shared data. Embargoing data means that you have uploaded your dataset to a repository and there is an electronic record of its existence, but only you (and specific individuals you approve) have access to the data. Note that setting an embargo differs from uploading a restricted-use dataset or making data available upon request because it is a temporary measure and implies that your data will ultimately be made accessible to the public. Typically, researchers embargo data for a certain time frame (e.g., one year), but an embargo may also be tied to a particular goal (e.g., when all main aims of a grant have been published). Thus, an embargo may be set to lift automatically after a specified period or be lifted intentionally by the PI. Once the embargo is lifted, the data then become available to the public. This form of data sharing does satisfy some data sharing requirements and is preferable to not making data available in an online repository at all, because at the very least the metadata of the data will be available and useable to the community immediately. For more information on data repositories’ policies and procedures surrounding data embargo, see the Data Upload section of Table 2.

WHO PREPARES AND SHARES YOUR DATA

Preparing research data to be shared can be a daunting task, particularly if your data are extensive, structurally complicated, or at high risk of re-identification. Although the principal investigator (PI) is the person ultimately responsible for the data being shared, it is not recommended to go through this process alone (or to delegate this task to one other person such as a graduate student or data manager). If possible, the PI should consider whether it is feasible to assemble a team to help with and check each step of data cleaning and deidentification.

Additionally, consider taking advantage of available resources, either online or at your institution. Regarding the latter, one option may be to arrange a meeting with staff in your university library system, who often have first-hand experience with data sharing and repositories. As discussed in an article by Mannheimer et al. (2019) focused on qualitative data sharing, institutional librarians can be invaluable allies while you are developing and implementing a plan to make research data publicly accessible and can provide support in endeavors such as including working with your institution's IRB and helping ensure that shared data are properly de-identified and have adequate documentation and metadata.

Many online resources are also available. LDbase provides free checklists and code to help check that your data is fully de-identified and hosts workshops on topics such as data management and deidentification. For example, the Center for Open Science (COS) offers an "Openness and Reproducibility Research Practices Training" curriculum that includes training in data management/accessibility and complying with the FAIR principles, while ICPSR offers data curation services (see Table 2). However, note that the convenience of resources often comes at a cost of time and/or money: the COS training costs between \$800 and \$3000 depending on the intensity of curriculum selected, while ICPSR's for-a-fee data curation services take between 6 to 18 weeks from data submission to data availability depending on data curation level selected. If you have written a data sharing line item into your grant budget and are not in a hurry to share your data, this may be a good option.

The timing of when to decide who will be involved in managing, cleaning, and sharing your data will depend again on the context in which you are sharing data (e.g., before or after data collection), the properties of your data (e.g., how complex or sensitive they are and thus how intensive they will be to de-identify) and the financial and administrative support available to you. However, it is ideal to start having these conversations early, especially if you plan to budget for data management, curation support, or repository fees in a grant budget.

CONCLUSION

Requirements to make data supporting educational research projects and publications publicly available are becoming increasingly prevalent as scientific culture shifts towards prioritizing an evidence base that is more transparent and accessible. While some researchers may view these requirements and incentives as simply new boxes to check in order to receive federal funding or publish in certain journals, the reality is that data sharing benefits both individuals and the scientific community. First, as part of the open science movement, data sharing encourages the production of high-quality and replicable research, which ultimately makes scientific progress more trustworthy, efficient, and cost-effective. Openly available data enable the development of a cumulative knowledge base within a scientific discipline. Openly available data can aid in the conduct of meta-analyses, as it is often the case that a published article does not contain all the necessary information needed for a meta-analysis. It is also possible to rigorously combine multiple original data sets using integrative data analysis (Curran & Hussong, 2009) to increase statistical power and model the presence and interaction of individual differences. Importantly for researchers in special education, combining data in this way is advantageous for studying low base-rate conditions, increasing the absolute number of affected individuals in the sample and thus improving the stability of statistical models and reducing the influence of outliers compared to any single study (Curran & Hussong, 2009). Additionally, the integration of longitudinal data can afford researchers the ability to study development over longer time periods than may be feasible when data are collected independently (Bainter & Curran, 2015).

Openly accessible data also support pedagogy and democratize participation in science by enabling students and early career researchers to pose research questions and perform secondary analyses on data they may not have time or resources

to personally collect. Researchers may locate these datasets by conducting keyword searches directly in data repositories or by using existing resources that are designed to help researchers discover data for secondary analysis, such as the Partnership for Expanding Education Research in STEM (PEERS) Data Hub (<https://www.icpsr.umich.edu/web/pages/peersdatahub/index.html>), which maintains both curated lists of public- and restricted-access data and has offered free webinars on this topic (PEERS Data Hub, 2024). Finally, because datasets shared via an online repository receive a DOI and are thus citable, data sharing can increase recognition for individual researchers (Colavizza et al., 2020).

Education researchers face both universal and unique barriers to sharing data, as discussed by Logan et al. (2024), and there is no doubt that venturing into data sharing for the first time is a daunting enterprise. However, given these benefits and the field's movement towards open and accessible science, it is in education researchers' best interest to at the very least be informed about data sharing and, ideally, to begin incorporating data management, cleaning, and sharing into existing workflows. While meta-scientific research in this area is still in its infancy, recent findings suggest that although there are many resources available specifically intended to help education researchers learn about best practices for sharing data, researchers' reported knowledge of data sharing and verifiable participation in data sharing is low. This paper is an attempt to help bridge the value-action gap by connecting education researchers with these resources; however, further meta-scientific research on the barriers and predictors of data sharing is also in order.

ACKNOWLEDGMENTS

This work was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B200020 to the Florida Center for Reading Research at Florida State University. The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences or the U.S. Department of Education. Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under award numbers R01HD095193 and P50HD052120. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This research was undertaken, in part, thanks to funding from the Canada Excellence Research Chairs Program.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Christine M. White  orcid.org/0000-0002-8913-2821

Florida State University, Department of Psychology, Tallahassee, FL, USA; Florida Center for Reading Research, Tallahassee, FL, USA

Stephanie A. Estrera  orcid.org/0000-0003-3340-1006

Florida State University, Department of Psychology, Tallahassee, FL, USA; Florida Center for Reading Research, Tallahassee, FL, USA

Christopher Schatschneider  orcid.org/0000-0002-1700-7685

Florida State University, Department of Psychology, Tallahassee, FL, USA; Florida Center for Reading Research, Tallahassee, FL, USA

Sara A. Hart  orcid.org/0000-0001-9793-0420

Florida Center for Reading Research, Tallahassee, FL, USA; University of Waterloo, Department of Psychology, Waterloo, Ontario, Canada

AUTHOR CONTRIBUTION

Christine White: Conceptualization, Visualization, Writing – Original Draft, Writing – Review & Editing. **Stephanie Estrera:** Conceptualization, Visualization, Writing – Original Draft, Writing – Review & Editing. **Sara Hart:** Conceptualization, Writing – Review & Editing, Funding Acquisition. **Chris Schatschneider:** Conceptualization, Writing – Review & Editing, Funding Acquisition.

REFERENCES

- Acciai, C., Schneider, J. W., & Nielsen, M. W. (2023). Estimating social bias in data sharing behaviours: An open science experiment. *Scientific Data*, *10*(1), 233. <https://doi.org/10.1038/s41597-023-02129-8>
- AERA Open (2024). *Submission guidelines*. Sage Publications. <https://us.sagepub.com/en-us/nam/aera-open/journal202293#submission-guidelines>
- Ajzen, I. (1985). From intention to actions: A theory of planned behaviour. In J. Kuhl & J. Beckmann (Eds.), *Action-control: From cognition to behavior* (pp. 11–39). Springer.
- Bainter, S. A. & Curran, P. J. (2015). Advantages of integrative data analysis for developmental research. *Journal of Cognition and Development*, *16*(1), 1–10. <https://doi.org/10.1080/15248372.2013.871721>
- Barr, S. (2006). Environmental action in the home: Investigating the ‘value-action’ gap. *Geography*, *91*(1), 43–54. <https://doi.org/10.1080/00167487.2006.12094149>
- Carlson, J. (2020). *Institutional data repositories: An important option for complying with data sharing requirements*. <https://dx.doi.org/10.3998/2027.42/163716>
- Center for Open Science (n.d.). *Events & webinars*. <https://www.cos.io/events>
- Center for Open Science (n.d.). *Open science badges*. <https://www.cos.io/initiatives/badges>
- Center for Open Science (n.d.). *Openness and reproducibility research practices training*. <https://www.cos.io/services/training>
- Christensen, G., Dafoe, A., Miguel, E., Moore, D. A., & Rose, A. K. (2019). A study of the impact of data sharing on article citations using journal policies as a natural experiment. *PloS one*, *14*(12). <https://doi.org/10.1371/journal.pone.0225883>
- Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. *PloS one*, *15*(4). <https://doi.org/10.1371/journal.pone.0230416>
- Cook, B. G., Fleming, J. I., Hart, S. A., Lane, K. L., Therrien, W. J., van Dijk, W., & Wilson, S. E. (2022). A how-to guide for open-science practices in special education research. *Remedial and Special Education*, *43*(4), 270–280. <https://doi.org/10.1177/07419325211019100>
- Cook, B. G., Van Dijk, W., Vargas, I., Aigotti, S. M., Fleming, J. I., McDonald, S. D., Richmond, C. L., Griendling, L. M., McLucas, A. S., & Johnson, R. M. (2023). A targeted review of open practices in special education publications. *Exceptional Children*, *89*(3), 238–255. <https://doi.org/10.1177/00144029221145195>
- Curran, P. J. & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, *14*(2), 81. <https://doi.org/10.1037/a0015914>
- Databrary. (2020). *Databrary access agreement*. <https://databrary.org/about/agreement/agreement.html>
- Digital Curation Centre (2024). *Services*. <https://www.dcc.ac.uk/services>
- Edwards, A. A., van Dijk, W., Tripodi, S. J., & Hart, S. A. (2023). Data sharing for randomized controlled trials in social work. *Research on Social Work Practice*. Online first publication. <https://doi.org/10.1177/10497315231186799>
- El Amin, M., Borders, J. C., Long, H. L., Keller, M. A., & Kearney, E. (2023). Open Science practices in communication sciences and disorders: A survey. *Journal of Speech, Language, and Hearing Research*, *66*(6), 1928–1947. https://doi.org/10.1044/2022_JSLHR-22-00062

- Fleming, J. I., Wilson, S. E., Espinas, D., van Dijk, W., & Cook, B. G. (2024). Special education researchers' knowledge, attitudes, and reported use of open science practices. *Remedial and Special Education*. Online first publication. <https://doi.org/10.1177/07419325241237268>
- Gilmore, R. O., Adolph, K. E., & Millman, D. S. (2016, August). Curating identifiable data for sharing: The databrary project. In *2016 New York Scientific Data Summit (NYSDS)* (pp. 1–6). IEEE. <https://doi.org/10.1109/NYSDS.2016.7747817>
- Gilmore, R. O., Kennedy, J. L., & Adolph, K. E. (2018). Practical solutions for sharing data and materials from psychological research. *Advances in Methods and Practices in Psychological Science*, *1*(1), 121–130. <https://doi.org/10.1177/2515245917746500>
- Hart, S.A., Schatschneider, C., Reynolds, T.R., Calvo, F.E., Brown, B.J., Arsenault, B., Hall, M.R.K., van Dijk, W., Edwards, A.A., Shero, J.A., Smart, R. & Phillips, J.S. (2020). LDbase. <http://doi.org/10.33009/ldbbase>
- Hart, S. A., Schatschneider, C., Reynolds, T., & Calvo, F. (2024). A community data sharing resource: The LDbase data repository. *Journal of Learning Disabilities*. Online first publication. <https://doi.org/10.1177/00222194241254091>
- Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V., 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>
- Huff, M. & Bongartz, E. C. (2023). Low research-data availability in educational-psychology journals: No indication of effective research-data policies. *Advances in Methods and Practices in Psychological Science*, *6*(1). <https://doi.org/10.1177/25152459231156419>
- Inter-university Consortium for Political and Social Research (2020). *ICPSR curation levels*. <https://www.icpsr.umich.edu/files/datamanagement/icpsr-curation-levels.pdf>
- Inter-university Consortium for Political and Social Research (2024). *Access and dissemination*. <https://www.icpsr.umich.edu/web/pages/datamanagement/lifecycle/access.html>
- Inter-university Consortium for Political and Social Research (2024). *Guide to social science data preparation and archiving*. <https://www.icpsr.umich.edu/web/pages/deposit/guide/>
- Johns Hopkins University Data Services (2016). *Applications to assist in de-identification of human subjects research data*. <https://dataservices.library.jhu.edu/resources/applications-to-assist-in-de-identification-of-human-subjects-research-data/>
- Kalandadze, T. & Hart, S. A. (2022). Open developmental science: An overview and annotated reading list. *Infant and Child Development*, *33*(1), e2334. <https://doi.org/10.1002/icd.2334>
- Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L. S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, S., Errington, T. M. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLOS Biology*, *14*(5), Article e1002456. <https://doi.org/10.1371/journal.pbio.1002456>
- LDbase (n.d.). *Frequently asked questions*. <https://ldbbase.org/data-sharing-resources/faq/general>
- Lewis, C. (2024a). *Data management in large-scale education research*. Chapman & Hall.
- Lewis, C. (2024b). *Data management*. <https://cghlewis.com/categories/data-management/>
- Logan, J. A. R., Hart, S. A., & Schatschneider, C. (2021). Data sharing in education science. *AERA Open*, *7*. <https://doi.org/10.1177/23328584211006475>
- Logan, J. A. R. & Hart, S. A. (2023). Data management for data sharing workshop for the Purdue University Emerging Perspectives in Early STEM Learning. figshare. <https://doi.org/10.6084/m9.figshare.24460651.v1>
- Logan, J. A. R., Hanson, A., Swanz, A., & Ceviren, A. B. (2024). *Education researchers' barriers and attitudes toward data sharing*. EdArxiv. <https://doi.org/10.35542/osf.io/8y6fd>
- Makel, M. C., Hodges, J., Cook, B. G., & Plucker, J. A. (2021). Both questionable and open research practices are prevalent in education research. *Educational Researcher*, *50*(8), 493–504. <https://doi.org/10.3102/0013189X211001356>

- Mannheimer, S., Pienta, A., Kirilova, D., Elman, C., & Wutich, A. (2019). Qualitative data sharing: Data repositories and academic libraries as key partners in addressing challenges. *American Behavioral Scientist*, 63(5), 643–664. <https://doi.org/10.1177/0002764218784991>
- Meyer, M. (2018). Practical tips for ethical data sharing. *Advances in Methods and Practices in Psychological Science*, 1(1), 131–144. <https://doi.org/10.1177/2515245917747656>
- National Center for Educational Research (2024, June 6). IES releases a new public access plan for publications and data sharing: What you need to know. *Inside IES Research*. <https://ies.ed.gov/blogs/research/post/ies-releases-a-new-public-access-plan-for-publications-and-data-sharing-what-you-need-to-know>
- National Institutes of Health (2023). *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 (n.d.). *Selecting a data repository*. <https://sharing.nih.gov/data-management-and-sharing-policy/sharing-scientific-data/selecting-a-data-repository>
- Neild, R. C., Robinson, D., & Agufa, J. (2022). Sharing study data: A guide for education researchers. Toolkit. NCEE 2022–004. *National Center for Education Evaluation and Regional Assistance*. <https://ies.ed.gov/ncee/pubs/2022004/pdf/2022004.pdf>
- Nelson, A. (2022). *Memorandum for the heads of executive departments and agencies: Ensuring free, immediate, and equitable access to federally funded research*. Office of Science and Technology Policy.
- Open Science Framework (2023). *Control your privacy settings*. <https://help.osf.io/article/285-control-your-privacy-settings>
- Open Science Framework (2016). *IRB and consent form examples*. <https://osf.io/g4jfv/wiki/home/>
- Partnership for Expanding Education Research in STEM (PEERS) Data Hub (2024). <https://www.icpsr.umich.edu/web/pages/peersdatahub/index.html>
- Reynolds, T., Schatschneider, C., & Logan, J. A. R. (2022). The basics of data management. figshare. <https://doi.org/10.6084/m9.figshare.13215350.v2>
- Schatschneider, C., Edwards, A., & Shero, J. A. (2021). Deidentifying data (Version 2). figshare. <https://doi.org/10.6084/m9.figshare.13228664.v2>
- Shero, J. A. & Hart, S. A. (2020a). Informed consent template (Version 1). figshare. <https://doi.org/10.6084/m9.figshare.13218773.v1>
- Shero, J. A. & Hart, S. A. (2020b). IRB protocol template (Version 1). figshare. <https://doi.org/10.6084/m9.figshare.13218797.v1>
- Shero, J. A. & Hart, S. A. (2022). Working with your IRB: Obtaining consent for open data sharing through consent forms and data use agreements (Version 2). figshare. <https://doi.org/10.6084/m9.figshare.13215305.v1>
- Tedersoo, L., Küngas, R., Oras, E., Köster, K., Eenmaa, H., Leijen, Ä., Pedaste, M., Raju, M., Astapova, A., Lukner, H., Kogermann, K., & Sepp, T. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific Data*, 8(1), 192. <https://doi.org/10.1038/s41597-021-00981-0>
- Van Dijk, W., Schatschneider, C., & Hart, S. A. (2021). Open science in education sciences. *Journal of Learning Disabilities*, 54(2), 139–152. <https://doi.org/10.1177/0022219420945267>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 1–9. <https://doi.org/10.1038/sdata.2016.18>