

To help inform the special education research community, these briefs feature information on prominent open science practices. Content comes from our series of short articles in the DR newsletter, *Focus on Research*, as well as additional content developed by DR members.

## OPEN DATA FOR RESEARCH

In the past several editions of *Focus on Research*, a number of the prominent features of Open Sciences practices have made their appearance. In this article, we focus on **open data**. Specifically, we explain what open data is and why it is beneficial for our field to share data. We also touch on some obstacles in sharing data and provide suggestions on how to get started. Due to our expertise, our primary focus is on sharing data from quantitative studies, but we also provide resources for sharing qualitative and single case design data.

### What is Open Data?

Currently, most quantitative publications provide summary data to the reader, usually in the form of means, standard deviations, and *n*'s; occasionally, authors will add information on skew and kurtosis and possibly a correlation matrix. In rare cases is this information enough to reproduce any of the analyses reported. Open data involves sharing data from a project at the participant level and the item level. Data can be shared as accompanying a specific paper, but researchers can also decide to share data from a complete project, including data that is not used in any publication.



### Primary Benefits

Sharing data with others has many benefits, both for the researchers who post their data and for the researchers using the data, and for special education research in general. For the data posters, sharing data increases the transparency of their research process (Cook et al., 2018). It enables other researchers to reproduce analyses and verify the results, which can increase trust in the outcomes of the project (Funk, 2020). Furthermore, publications that are accompanied by openly available data have been shown to get more citations (Colavizza et al., 2020; Drachen et

al., 2016). Finally, shared data sets are considered a research product, can be assigned a digital object identifier (DOI) and be added to a CV as a citable product.

Contributing to open data is a way to promote equity in research. Black, Indigenous, and People of Color (BIPOC) researchers and researchers with disabilities are less likely to get grant funding (Shavers et al., 2005; Swenor et al., 2020). By making their research data available, those who are funded can provide opportunities to these

researchers, as well as to graduate students, early career researchers, and researchers at institutions with less infrastructure for research support to obtain pilot data for further grant applications. Potentially, data sharing can also lead to new collaborations.

Having data from other projects available to other researchers can help our field advance faster (e.g., Munafò et al., 2014). Other researchers may have research questions that were not part of the original study plans. These questions might be based on just one project, but they may also be

based on a set of research projects. Current statistical advances allow researchers to combine independent data sets, both at the item (e.g., Curran & Hussong, 2009) and test score levels (e.g., van Dijk et al., 2021). This is a crucial benefit for the special education field, because by combining independent data sets, researchers can establish larger samples of populations with low incidence disabilities or behaviors. Using these larger samples, more complex research questions about these populations can be answered using advanced statistical models (Logan et al., 2021; van Dijk et al., 2020).

## How to Share Data

The gold standard in data sharing is to make your data **Findable, Accessible, Interoperable, and Reproducible** (FAIR; Wilkinson et al., 2016). There are too many specific steps to ensure FAIR data to state here (but see Logan et al., 2021 for more specific details on data sharing in education) and we will briefly touch on the broader steps.

First, Open Data involves more than just your data file. To be able to find and understand your data, other researchers need to know more details. These details are usually delineated in the metadata. Metadata can include information about the study design, the participants included, and the type of variables and information on how they were collected (Logan et al., 2021). Additionally, you should add a codebook explaining the variables, coding and recoding schemes, missing data patterns, etc. Creating the metadata and codebooks for a project take considerable time, and it is recommended to keep these updated throughout your project (Reynolds & Schatschneider, 2020).

Second, it is very important to make sure that you have permission to share data openly. This should be stated explicitly in your informed consent forms.

## RESOURCES

We believe data sharing to be a vital part of the research practices in special education, with benefits outweighing the perceived barriers. We leave you with a list of resources that might help you get started planning for sharing data.

- General overview of open science practices in special education Cook et al., 2018
- Data management for open data practices Reynolds & Schatschneider, 2020
- General review of data sharing as well as ethical considerations Meyer, 2018
- In-depth information about quantitative data sharing Logan et al., 2021
- Information about open science practices in single case research Cook et al., 2021
- Information on data sharing in qualitative research Mannheimer et al., 2019, Tsai et al., 2016
- Information on deidentification practices beyond HIPAA Edwards & Schatschneider, 2020
- Information on how to work with your IRB and prepare consent forms that allow you to share the data Shero & Hart, 2020a, b
- Making data Findable, Accessible, Interoperable, and Reusable (FAIR principles) Wilkinson 2016

If you want to share data that you have already collected and your informed consent does not state data sharing objectives explicitly, you may have to request a waiver from your IRB. In either case, we recommend working with your IRB to ensure you can share your data.

After you have collected your data, it is necessary to clean and de-identify. Cleaning data should be a part of your regular data management routine (Reynolds & Schatschneider, 2020), but it never hurts to (a) check for values that are out of range (for example a reading fluency score of 300 words per minute) and (b) ensure that you are entered data consistently (such as NA or blank for all missing values). De-identifying data is a bit more complex. First, you want to make sure you deleted all the HIPAA identifiers, such as names, addresses, birth dates, phone numbers, etc. Then, you need to make sure if any of the participants might be reidentified based on variables in your data. For

example, if a school district only has one male Asian special education teacher, his data can be identified using these two variables (self-report gender and race) (See Edwards & Schatschneider, 2020 for more information).

Once your data set is ready to go, it is time to identify where you will post your dataset. Choosing a repository will depend on the type of data you want to share. A repository specific to learning differences is **LDbase** ([www.ldbase.org](http://www.ldbase.org)), specifically designed for researchers working with behavioral and achievement data. There are several repositories specialized in qualitative or video/audio data, such as the **Qualitative Data Repository** (<https://qdr.syr.edu/>) and **Databrary** ([www.databrary.org](http://www.databrary.org)). Other repositories are not discipline (e.g., ICPSR) or data specific (e.g., Open Science Framework). In addition, many grant funding agencies have their own repositories (e.g., DASH).

### Primary Obstacles and Limitations

Sharing data is the least common of all open science practices in education sciences (Makel et al., 2019). There are several reasons why special education researchers might be hesitant to share the data from their projects. One potential problem is that it is harder to deidentify data from certain special education research studies, such as studies with participants with specific, low incidence disabilities, or single case designs and qualitative studies (Mannheimer et al., 2019; Tsai et al., 2016). In these cases, sharing may still be possible if access restrictions are put in place and researchers interested in re-using the data need to apply to gain access. Additionally, for some datasets, including for single case and quantitative data, there are more stringent deidentification possibilities (see Edwards & Schatschneider, 2020, for examples), or even the possibility of posting synthetic data sets. Secondly, many researchers may be under the impression that Institutional Review Boards (IRB)

are in principle against data sharing. We believe this is erroneous. Most IRBs will be positive towards sharing data if they adhere to the Common Rule, although each institution's IRB functions differently, and we advise you work with yours to gain clarity.



**Data sharing can happen after publication of specific manuscripts, so that the main analyses are already done.**

It takes resources, both time and monetary, to get data ready to be shared. Approximately 5-10% of a research grant budget should be dedicated to preparing and sharing data (Mons, 2020).

This may deter researchers from starting. Many grant agencies, however, will allow for part of a grant budget to be assigned to a data manager. With or without grant funding, we highly recommend working with a data specialist, or at least a data science librarian at the outset of a project to get your project set up in a way that is conducive to data sharing. A final occasional reaction of researchers to sharing all their data openly is the fear of being ‘scooped’. We believe that this is a vanishingly remote possibility in our

field, although there are solutions that still allow data sharing where this is a credible fear. Data sharing can happen after publication of specific manuscripts, so that the main analyses are already done. Data can often be shared on repositories with embargo periods. Data will be stored during this period but there is no public access. With data access restrictions like these, researchers can better guide the additional research being done with their data.



Edwards, A., van Dijk, W., & Hart, S.A. (2020). 10 benefits of data sharing. Available <https://venngage.net/ps/Uz65vsS6hI/new-10-benefits-of-data-sharing>. Reuse available under a CC BY 4.0 license.



## REFERENCES

- Colavizza, G., Hrynaskiewicz, 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>
- Cook, B. G., Lloyd, J. W., Mellor, D., Nosek, B. A., & Therrien, W. J. (2018). Promoting open science to increase the trustworthiness of evidence in special education. *Exceptional Children*, 85(1), 104–118. <https://doi.org/10.1177/0014402918793138>
- Cook, B. G., Fleming, J. I., Hart, S. A., Lane, K. L., Therrien, W. J., van Dijk, W., & Wilson, S. E. (2021). A how-to guide for open-science practices in special education research. *EdArXiv*. <https://edarxiv.org/zmeba/>
- Cook, B. G., Johnson, A. H., Maggin, D., Therrien, W., Barton, E., Lloyd, J. W., Reichow, B., Talbott, E., & Travers, J. C. (2021). Open Science and Single-Case Design Research. *EdArXiv*. <https://doi.org/10.1177/0741932521996452>
- Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, 14(2), 81–100. <https://doi.org/10.1037/a0015914>
- Drachen, T., Ellegaard, O., Larsen, A., & Dorch, S. (2016). Sharing data increases citations. *Liber Quarterly*, 26(2).
- Edwards, A., & Schatschneider, C. (2020). De-Identification Guide (Version 1). *figshare*. <https://doi.org/10.6084/m9.figshare.13228664.v1>
- Funk, C. (2020, February 12). Key findings about Americans' confidence in science and their views on scientists' role in society. *Pew Research Center*. <https://www.pewresearch.org/fact-tank/2020/02/12/key-findings-about-americans-confidence-in-science-and-their-views-on-scientists-role-in-society/>
- Logan, J., Hart, S., & Schatschneider, C. (2021). Data Sharing in Education Science. *EdArXiv*. <https://doi.org/10.35542/osf.io/2x3cu>
- Makel, M. C., Hodges, J., Cook, B. G., & Plucker, J. (2019). Questionable and open research practices in education research.
- 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>
- Munafò, M., Noble, S., Browne, W. J., Brunner, D., Button, K., Ferreira, J., Holmans, P., Langbehn, D., Lewis, G., Lindquist, M., Tilling, K., Wagenmakers, E.-J., & Blumenstein, R. (2014). Scientific rigor and the art of motorcycle maintenance. *Nature Biotechnology*, 32(9), 871–873. <https://doi.org/10.1038/nbt.3004>
- Reynolds, T., & Schatschneider, C. (2020). The Basics of Data Management (Version 1). *figshare*. <https://doi.org/10.6084/m9.figshare.13215350.v1> ([1])
- Shavers, V. L., Fagan, P., Lawrence, D., McCaskill-Stevens, W., McDonald, P., Browne, D., McLinden, D., Christian, M., & Trimble, E. (2005). Barriers to racial/ethnic minority application and competition for NIH research funding. *Journal of the National Medical Association*, 97(8), 1063–1077.
- Shero, J., & Hart, S. A. (2020a). Informed consent template (Version 1). *figshare*. <https://doi.org/10.6084/m9.figshare.13218773.v1>
- Shero, J., & Hart, S. A. (2020b). Working with your IRB: Obtaining consent for open data sharing through consent forms and data use agreements (Version 1). *figshare*. <https://doi.org/10.6084/m9.figshare.13215305.v1>
- Swenor, B. K., Munoz, B., & Meeks, L. M. (2020). A decade of decline: Grant funding for researchers with disabilities 2008 to 2018. *PLOS ONE*, 15(3), e0228686. <https://doi.org/10.1371/journal.pone.0228686>
- Tsai, A. C., Kohrt, B. A., Matthews, L. T., Betancourt, T. S., Lee, J. K., Papachristos, A. V., Weiser, S. D., & Dworkin, S. L. (2016). Promises and pitfalls of data sharing in qualitative research. *Social Science & Medicine*, 169, 191–198. <https://doi.org/10.1016/j.socscimed.2016.08.004>
- van Dijk, W., Schatschneider, C., Al Otaiba, S., & Hart, S. (2021). Assessing measurement invariance across multiple groups: When is fit good enough? *EdArXiv*. <https://doi.org/10.35542/osf.io/kxcyt>
- van Dijk, W., Schatschneider, C., & Hart, S. A. (2020). 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.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>



**A special thank you to the following DR members for their contribution to this OPEN SCIENCE BRIEF, ISSUE 3:**

- Corey Peltier, Andrew Heuer, & Bree Jimenez (Eds).
- Bryan G. Cook, Lydia A. Beahm, & William J. Therrien