



Data Sharing

Rick O. Gilmore, Melody Xu, and Karen E. Adolph

Psychologists embrace the ethical imperative of protecting research participants from harm. We argue that sharing data should also be considered an ethical imperative. Despite potential risks to participants' privacy and data confidentiality, sharing data confers benefits to participants and to the community at large by promoting scientific transparency, bolstering reproducibility, and fostering more efficient use of resources. Most of the risks to participants can be mitigated and the benefits of sharing realized through well-established but not yet widespread practices and tools. This chapter serves as a how-to manual for addressing ethical challenges in sharing human data in psychological research in ways that simultaneously protect participants and advance discovery.

THE ETHICAL IMPERATIVE TO SHARE HUMAN RESEARCH DATA

In a typical psychology study, researchers collect some data, process and analyze it, and then publish a report of their findings. Although other researchers can read the publication, typically they have no access to the original data, the context in which it was collected, or the various incarnations of the data

Work on this chapter was supported by grants from the James S. McDonnell Foundation (220020558), the Defense Advanced Research Projects Agency (HR001119S0005-MCS-FP-035), and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01-HD-094830).

<https://doi.org/10.1037/0000258-006>

Handbook of Research Ethics in Psychological Science, S. Panicker and B. Stanley (Editors)
Copyright © 2021 by the American Psychological Association. All rights reserved.

during processing and analyses (Adolph, 2020). Descriptions in the Method and Results sections of journal articles are necessarily incomplete—at best, crude road maps of data collection and data processing. No matter how detailed, graphs and statistics cannot portray the richness of the original data; they can only summarize interpretations of the data. Thus, progress in psychological science relies on a considerable amount of blind trust (Heidorn, 2008).

Moreover, the data that comprise all the steps from data collection to published report remain private. Although still potentially useful, the data go unused and the analysis scripts disappear into the lab ether. Potential resources are needlessly wasted.

Data Sharing Increases Transparency and Reproducibility

Data sharing supports transparency and reproducibility, two tenets of research that distinguish science from religion, opinion, and fake news. Indeed, data sharing is so critical for transparency and reproducibility that it is an ethical imperative for scientific progress (Gennetian et al., 2020). It is an antidote to blind trust.

However, *Psychological Science*, the flagship journal of the Association for Psychological Science (2020), and top-tier journals of the American Psychological Association (2020; e.g., *Journal of Experimental Psychology*, *Journal of Personality and Social Psychology*, *Developmental Psychology*) recommend—but do not require—data sharing. *Science*, a premiere outlet for general science, requires researchers to share only the final processed data, not the original data (American Association for the Advancement of Science, 2020). Unfortunately, without the original data, readers cannot know exactly what happened during data collection—who said and did what to whom, and how and in what context. And readers cannot know exactly what transpired over the course of data processing—how data were scored, filtered, or smoothed; which data points were removed and for what reasons; and so on. In effect, readers must trust scientific reports without sufficient means to verify them.

Moreover, without full transparency over the entire research workflow, researchers cannot identify the culprit for failures to replicate. Although outright fraud is presumably rare (Bhattacharjee, 2013; Couzin-Frankel, 2014), lack of reproducibility may result from small effects, as many doubters fear (Nosek et al., 2015). Or the problem may stem solely from lack of transparency: Truly robust results fail to be replicated because new researchers do not know exactly what the previous researchers did (Adolph, 2020; Gilmore & Adolph, 2017; Suls, 2013). Total transparency may seem a far-off, utopian ideal (Adolph et al., 2012; Nosek et al., 2012). But scientists have an ethical obligation to document procedures and findings in ways that others can readily reproduce. To maximize reproducibility, data processing, visualization, and analyses should be automated with scripts, rather than done manually with button clicks in a computer program; the scripts should be annotated and documented and shared along with the data. Preregistration is a positive step (Nosek et al., 2019; Nosek & Lindsay, 2018), but it is no panacea. Preregistrations, like published

reports, contain insufficient detail to allow exact replication of procedures and analyses, and researchers often stray from preregistered analysis plans (Claesen et al., 2019).

Video is a better solution for comprehensive documentation of procedures. Video clips of a typical (or exemplary) data collection with real (or simulated) participants, from beginning to end, can document the procedures, displays, and surrounding context with more richness and fidelity than are possible with text-based descriptions or static images (Adolph, 2020; Adolph et al., 2017; Gilmore & Adolph, 2017; Suls, 2013). For behavioral data (on anyone doing anything), video recordings provide a detailed, dynamic record of what happened, such as what the lab room looked like, how procedures were explained, or the timing of experimental displays. For physiological, imaging, and biological data, video documentation records potentially crucial aspects of data collection, such as how researchers explained the technology, got participants to lie still in the MRI machine, put recording devices on participants, or collected blood or spit samples. We argue that video documentation of procedures should be standard in psychological science—whether or not video is a primary source of raw data (Suls, 2013). Many outlets are available for sharing procedural videos. The *Journal of Visualized Experiments* (2020) provides voiced-over video documentation of procedures, but publishing is costly. Alternatively, procedural videos can be shared alongside publications as supplemental materials in many journals for free or the cost of the article. And regardless of the existence of publications, extensive video documentation of procedures and displays can be shared for no cost at Databrary.org and on the Open Science Framework (osf.io).

In addition to documenting procedures, video excerpts provide compelling documentation of research findings in behavioral studies. Whereas a picture is worth a thousand words, a video is worth a thousand pictures (Adolph, 2020). Of course, videos can be seen differently by different eyes, so video documentation will not diminish the vigor of scientific debate, but it will shift the focus (Gilmore, 2016; Gilmore et al., 2018). Moreover, the perceived costs of video sharing to researchers (e.g., time to collect, curate, and store data) and risks to participants (the need to protect personally identifiable information from unauthorized disclosure or misuse) can be mitigated by building on previously proven technologies and established policy frameworks, which are discussed later in this chapter.

Data Reuse Leverages Resources

Psychological research is expensive. Federal agencies spend billions of taxpayer dollars annually on behavioral research. Thus, researchers have an ethical obligation to use grant-funded support to maximize benefits to research participants and to the public (Brakewood & Poldrack, 2013). Nevertheless, valuable data and the resources expended for data collection are routinely wasted. Data “shared” on researchers’ lab servers, websites, and institutional repositories may be unfindable by others. Researchers at different sites collect

similar data supported by different grants because psychology laboratories are siloed. Researchers must reinvent procedures for themselves because they do not know the details of similar procedures used in prior work.

The current situation is fraught, but data sharing can help. Data sharing speeds scientific progress because researchers can revisit and reuse existing data to ask questions outside the bounds of the original study—often questions never imagined by the original researchers. As Heidorn (2008) put it, “The true worth of the data is not determined by the cost for gathering it but in the savings incurred by someone else not needing to gather it again” (p. 290).

ETHICAL CONSIDERATIONS IN SHARING DATA

If data sharing is an ethical practice and a boon to discovery, why do so few psychological researchers share? In most areas of psychological research, data sharing is not a scientific necessity, it is a choice. It is not mandated by funders, journals, or societies, and it is not rewarded or supported by academic institutions. Moreover, researchers must overcome the inertia of the status quo to acquire the know-how and invest the time and effort to plan for sharing and to curate shared data to make them findable and reusable. And, of course, the ethical challenges in sharing human research data contribute to researchers’ reluctance to share.

Sharing by Necessity or by Choice

In some fields, researchers cannot acquire the necessary data on their own. Data collection may be infeasible, impractical, or prohibitively expensive for a single research team. Particle physics, for example, is so expensive (e.g., billions of dollars to fund the Large Hadron Collider) and physically large (the collider has a circumference of 16.6 miles) that the science requires large-scale, international data sharing (European Council for Nuclear Research, 2020). In other fields, only a shared corpus of data is sufficiently large to match the scope of the subject matter. Meteorology (weather science) and seismology (earthquake science) require international data sharing because the scope of the science eclipses the capacity of single data collection sites. So, historically, data sharing has an inverse relation with independent research. That is, the greater the difficulty of conducting research alone or in small groups, the more likely it is for that research community to share data.

What about psychological science? Sharing is necessary to collect sufficiently large and diverse samples for large-scale studies of social outcomes, health and disease, human genetics, language, neural imaging, and so on (Frank, 2020; Hood & Rowen, 2013; Inter-University Consortium for Political and Social Research, 2020; MacWhinney, 2020; Manolio et al., 2008; Poldrack & Gorgolewski, 2014). But most conventional psychological research does not require shared data. This is especially true for studies that rely on samples of convenience—undergraduates in introductory courses, people living near

the research lab, and, in the case of online data collection, people with access to the Internet and time to spare. Thus, typical research practices impose no pressure to share. For most psychology researchers, data sharing is an ethical choice, not a necessity.

Ethical Considerations in Sharing Human Data

Ethics boards rely on widely recognized guidelines such as the Belmont Report and the Declaration of Helsinki to ensure participants' health and well-being and to protect their privacy and the confidentiality of their data (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979; World Medical Association, 2013). However, less widely recognized considerations apply to every stage of data sharing, including the decision to collect data, data collection, data curation, data contribution, and data reuse.

Even the decision to collect new data involves ethical issues. Are new participants warranted? Although some subdisciplines in psychology rely heavily on secondary data analyses, most do not. However, like primary data collection, data reuse entails meaningful scholarship. Moreover, data reuse may be cheaper and more efficient, incur less risk, and provide more benefits to participants (e.g., by minimizing their time and effort) than primary data collection. Existing data can be culled from multiple sources and used in combination with new data to minimize participant effort or to expand a sample. Indeed, the ethics of collecting small, underpowered samples are questionable (Fleming & Reynolds, 2008), and data reuse can help address the limitations of small samples and samples of convenience. Furthermore, the bulk of psychological research focuses on only a tiny proportion of the world's population (Henrich et al., 2010), an ethical violation of a different sort. New data collection targeting underrepresented groups and widespread data sharing may help mitigate the consequences of building a human science on a narrow slice of humanity (Gilmore et al., 2020).

Perhaps the most obvious ethical challenge during data collection concerns participants' permission to share their identifiable data (Meyer, 2018). Permission to share, which can occur at any time, is distinct from consent to participate, which must occur before participation. Thus, researchers (and ethics boards) can choose whether to separate or combine these two requests. However, if researchers do not lay the foundation for sharing (e.g., obtain appropriate approval from their ethics board) and request participants' permission in real time, later sharing is often infeasible. Whereas consent to participate involves telling participants about the procedures and general aims of the study, permission to share is open ended. Shared data can be reused to address aims far beyond the scope of the original study, and new technologies (e.g., machine learning algorithms) expand the universe of potential reuses. So in seeking permission to share data, researchers should not (and likely cannot) specify how shared data will be used in the future. But researchers can and should tell participants what data will be shared, who

will have access, where data will be stored, and for how long. And researchers should inform participants about the potential benefits and risks of sharing their data.

A less obvious ethical challenge concerns the instruments and tools used to collect data. Shared data should be interoperable, in formats readable by open source or widely available tools (Wilkinson et al., 2016). Thus, use of proprietary instruments (e.g., self-report instruments copyrighted by publishers) can preclude sharing the original, unprocessed data, and proprietary data formats can require researchers to purchase the tools to enable data reuse (e.g., imaging and physiological data).

Data collection rarely incorporates curation—quality assurance, cleaning, organization, and documentation. However, researchers should curate their data during data collection and processing, rather than as an agonizing follow-up exercise after their findings are published. Ethical data sharing requires data to be findable (e.g., by tagging it with appropriate search terms and key words) and reusable by including sufficient documentation about the data and its provenance (Wilkinson et al., 2016). Thus, data sets require thoughtful curation. Sharing a spreadsheet with obscure column headers and limited documentation may meet the letter of a journal's or funder's requirements but not the spirit. Curation is less onerous if researchers collect, process, and analyze their data with an eye toward sharing.

The most critical part of data curation is to respect participants' permission to share. Participants may share personally identifiable data, including their name, genome, sexual activity, crimes, and romantic relationships. But many participants, researchers, and ethics boards prefer to share "deidentified" data stripped of personal identifiers (Joel et al., 2018). Regardless, private or sensitive information must be adequately protected (Office for Human Research Protections, 2017). At first blush, the distinction between public (e.g., birth records) and private or sensitive information (e.g., medical records) may seem obvious. However, the line is not clear cut. Publicly shared information in the United States (e.g., party enrollment of registered voters, religious affiliations) may constitute private, highly sensitive information in other countries. Similarly, the distinction between sensitive and nonsensitive information changes across generations. The current prevalence of social media platforms, for example, encourages people to publicly share private snippets of their lives in words, photographs, and videos. As evidenced by Facebook's exchange of contact list data with companies like Yahoo and Amazon, people's perception of private information and what is actually private undergo rapid and dramatic change. As an article in *The Atlantic* put it, "Facebook didn't sell your data; it gave it away" (Madrighal, 2018).

Ethical sharing requires data to be accessible (Wilkinson et al., 2016). Researchers and their ethics boards must decide when to share data, who should have access, and how access should be controlled. Embargoes until a paper goes to press, for a fixed period after data collection ends, or at the end of a grant period are justifiable. But indefinite, promissory time periods often stretch into "never" (Ascoli, 2006; Meyer, 2018). When granting access, some

repositories, journals, and researcher websites delegate control to the data contributor. That is, potential users must contact the contributor to request access. This is an ethically dubious practice if restricting access is self-serving (e.g., blocks access to scholarly rivals), and “outing” oneself to or engaging in negotiations with the data contributor may discourage certain groups of researchers (e.g., young or new investigators, underrepresented minorities, potential detractors) from requesting access. A similarly dubious practice is to allow researchers access only in exchange for coauthorship on resulting papers. Open access (without data contributors vetting reusers’ identities and purposes or mandating coauthorship) is the most ethical stance: It maximizes the benefits of research participation and ensures that access is equitable and unbiased. However, data contributors should get scholarly credit for their work. Most recognized repositories generate standard citation formats and persistent identifiers (e.g., digital object identifiers [DOIs]) for data sets, making citation easy.

The final step in the data-sharing life cycle is reuse, and data reusers also have ethical obligations and deserve ethical protections. In addition to citing the data contributor, data reusers should address data errors in ways that minimize harm to the contributor while advancing the science (Gilmore et al., 2020). If researchers add to the data set, modify it, or integrate across shared data sets, they should share the new data set, too. Researchers should be free to reuse data as they see fit, including in ways that are critical of the original data set or data contributor.

CASE STUDY: DATABRARY

Our perspectives on the ethics of data sharing come from experience creating and supporting the Databrary video data library. Databrary (databrary.org) provides a case study to illustrate how ethical challenges in sharing personally identifiable human data can be addressed and how Databrary handles situations that involve potential harm to research participants, contributors, and data reusers.

About Databrary

Databrary is the world’s only data repository specialized for storing and sharing research video. As of early 2021, the library held 72,000+ hours of shared video, and the Databrary network included 1,265 authorized researchers and 587 of their affiliates from 622 institutions around the globe. Researchers’ use of Databrary is free, including technical support and use of the Datavyu video coding tool (datavyu.org). For historical reasons, most videos depict children, but Databrary is designed to share research videos of human and nonhuman animal behavior at any age.

Video is a uniquely powerful medium for documenting procedures and findings and for training and educational purposes. Research video is also an inexpensive, widely available, and uniquely powerful source of raw material

for reuse. Video captures the richness of behavior and the subtle details of the surrounding context, and it allows multiple observers to revisit the same events with fresh eyes to ask new questions. In addition, compared with data in proprietary formats and processed data, raw research video requires relatively little information about the data provenance. Nonetheless, the virtues of video also pose tough ethical challenges. In particular, video contains personally identifiable information (e.g., people's faces, voices, names; the interiors of their homes or classrooms), and it cannot be easily anonymized without losing much of its potential for reuse.

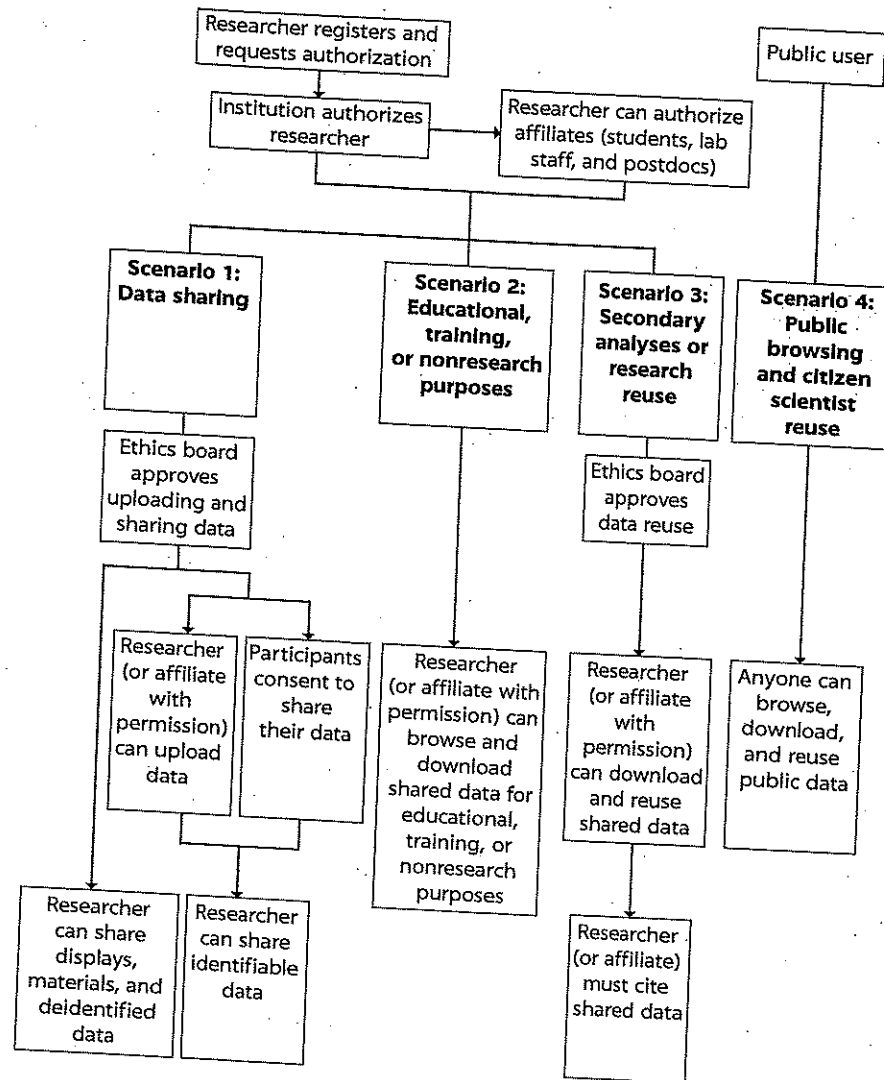
Databrary's ethical framework protects participants' privacy and the confidentiality of their shared data. It also protects the rights of data contributors and data reusers while ensuring the integrity of the data. The policy framework rests on the Databrary access agreement (Databrary, 2020a), which is a legal contract between researchers, their institutions, and Databrary's host institution, New York University (NYU). To become "authorized investigators" with access to Databrary, researchers must promise to respect participants' wishes about sharing, to treat other researchers' data with the same high standards as they treat their own, and to be responsible for the treatment of data by people they supervise ("affiliates" such as students, staff, and collaborators). Institutions are responsible for their researchers' conduct on Databrary, so most institutions limit authorization to faculty or full-time independent researchers. In contrast to most data use agreements, the Databrary access agreement gives researchers open access to all shared data in the library, and it allows them to contribute data themselves with the appropriate permissions. Fundamentally, the Databrary policy framework for sharing and protecting data rests on two pillars: enabling open access among a restricted community of researchers and obtaining explicit permission to share from participants.

Open Access to Shared Data Among a Restricted Community of Researchers

Databrary limits access to all nonpublic data to its community of authorized researchers. Among this community, access is open to all shared data, so would-be users need not contact contributors to request access to their data or specify to contributors how they will use the data. Reciprocally, contributors do not know who accessed their data or for what purposes, at least until a publication citing the data set appears.

As shown in the top left of Figure 6.1, when researchers apply for authorization, Databrary staff work with the institution to approve the access agreement. Some institutions require an ethics board to review the agreement; most handle authorization like a formal contract or data use agreement. After institutions authorize them, researchers can, in turn, give their students, staff, and trainees access to their own data or to all shared data on Databrary. Requiring a formal contract that holds institutions responsible for their authorized investigators, and authorized investigators responsible for their trainees

FIGURE 6.1. Scenarios for Sharing and Reusing Data and Materials Stored on Databrary



Note. The figure summarizes the main steps involved in four different data sharing and use scenarios involving Databrary.

and staff, limits accessibility to a known community of researchers, bolsters accountability, and provides protections for shared data.

Data Security

NYU (2020) has classified data stored on Databrary as moderate risk and evaluates and maintains security measures accordingly. To ensure data safety, storage and backup are geographically distributed. Moreover, Databrary is a member of the Data Preservation Alliance for the Social Sciences

(<https://www.data-pass.org/>), a group of social science data repositories that share data and metadata across groups to reduce the likelihood of data loss. NYU committed to storage and preservation in perpetuity.

What about unauthorized access? Each account is password protected, and Databrary requires researchers to follow the security practices recommended by their institutions, such as downloading data only to protected computers on campus or to encrypted hard drives on laptops. Data breaches or unauthorized access in the labs of researchers who use shared data are the responsibility of those researchers and their institutions. Like other authorized investigators, Databrary staff have no access to unshared private data uploaded by researchers.

Change in Researchers' Status

What happens if authorized investigators change institutions, retire, or die? Or if a student leaves the lab or gets a faculty position? Researchers and institutions are responsible for maintaining their own status and that of their affiliates. They cannot access other people's shared data without appropriate institutional authorization. Institutions decide whether to allow former faculty to retain control over data they contributed or whether to appoint a data steward. Databrary will continue to share contributed data unless directed to do otherwise by the original institution to ensure that changes in the status of data contributors do not disrupt its availability to others.

Explicit Permission to Share

The second pillar of the Databrary policy framework involves securing explicit permission to share data. As illustrated in Scenario 1, shown in Figure 6.1, data contributors must secure permission from their ethics board and research participants. The ethics board must agree that data can be stored on Databrary and shared according to participants' wishes, and participants must specify their preferred level of sharing. If participants say no to sharing beyond the original research team, the data default to "private." Researchers must manually change the permissions setting if participants agree to share all (or portions) of their data with other Databrary researchers ("authorized users"), or in addition to allow video excerpts to be shown ("learning audiences"), or to share their data with anyone in any context ("public"). Researchers can decide at any point to restrict sharing (e.g., if a participant inadvertently revealed sensitive information or displayed a potentially embarrassing behavior), but they cannot share at broader levels than participants specified. For minors (< 18 years of age), parents must give permission for their children's data to be shared, and children capable of assent must do so in addition. Databrary offers template permission forms for participants and video tutorials and scripts to illustrate how researchers can request permission from participants (Databrary, 2020b). In most cases, permission to share can be separated from consent to participate in the study. Thus, researchers can request permission to share after the

study is completed and participants have a clearer understanding of the protocol than before they began the study.

All files—including private ones—are stored together to ensure the integrity of the data set. Moreover, the entire data set remains accessible only to the original research team until they are ready to share it. Thus, researchers can treat Databrary as their lab data server and remote storage for both private and shared data. To ensure that contributors are cited and credited for their work, the system generates a standard citation for shared data sets that includes a persistent identifier (DOI).

Participants change their mind about sharing, or about their specified level of sharing. Under the European Union General Data Protection Regulation (2016), research participants may ask to have their data removed. However, only the original data contributor has information that links participants' identities to their data on Databrary. Accordingly, research participants who wish to retract permission to share data must contact the original researcher. Even if the original researcher can determine which specific files to delete or make private, Databrary makes no promise that data previously shared can be retrieved from other researchers who accessed it prior to the participant's change of heart. These limits on the "right to be forgotten" must be conveyed to participants at the outset.

Minors come of age. In the United States, only adults can consent to participate in research or give permission for a minor to participate. The same principle extends to permission to share data. Institutions and ethics boards must decide whether to require that consent or permission be obtained again for minors who come of age in order to share their data on Databrary. U.S. federal agencies do not require recontact. Regardless, as with adult participants, information linking minors' identity to their data on Databrary is held (or not) by the original data contributor, and Databrary cannot promise to retrieve previously shared data.

Using Databrary

By restricting access to authorized researchers and sharing data with explicit permission, Databrary demonstrates that personally identifiable video data can be openly shared in ways that protect participants. As illustrated in Figure 6.1, authorized investigators and their affiliates can browse and download shared data and reuse it for various purposes. Only files marked "authorized users," "learning audiences," or "public" are accessible outside of the original research team. The Databrary access agreement provides assurance that files will be downloaded to secure platforms (e.g., computers inside a locked laboratory) and viewed only by people to whom authorized researchers grant access.

Many use cases for video do not involve research; see Scenario 2 in Figure 6.1. Researchers may want to view procedure videos to train their staff, access coding manuals and spreadsheets to verify a coding scheme, download videos to decide whether a data set is appropriate for their reuse case, or simply browse shared data to learn what their colleagues are doing or to look

for inspiration. Alternatively, researchers may want to show an existing video clip or to create their own excerpt from another researcher's raw data to illustrate a point in a classroom lecture or presentation (i.e., they wish to show the clip to broader audiences for educational purposes). Videos shared on Databrary for use in teaching and presentations are designated "learning audiences," meaning that participants gave permission for these materials to be shown by authorized investigators to audiences outside of a research laboratory. Such nonresearch use cases typically do not require approval from an institutional review board.

However, as illustrated in Scenario 3 of Figure 6.1, researchers may want to reuse shared data to address a specific research question (secondary data analysis) or to demonstrate feasibility for a grant proposal. In this case, they must obtain ethics board approval. The range of potential questions is limited only by the researcher's imagination. Ideally, researchers who create new findings from shared data will share those data back with the research community. The only requirement is to cite the Databrary resources used for a project using the recommended citation format.

Finally, as illustrated in Scenario 4 of Figure 6.1, like many websites, Databrary provides access to selected data and materials to any member of the public with a web browser. Publicly accessible data are reviewed by the data contributor to ensure that materials are not sensitive, do not violate copyright, and pose minimal risk to participants. Public (nonauthorized) users cannot upload data.

Data Errors Discovered During Reuse

Some researchers may worry that sharing makes it more likely that a critic or rival will find errors in the work. Indeed, open sharing *should* accelerate the detection of mistakes along with the accumulation of robust findings. This is one way that openness improves science. Open data sharing in Databrary allows contributors to fix a mistake as soon as it is discovered, and the corrected version becomes immediately available to the research community. The possibility of error detection and correction should give the community greater confidence in the validity of findings based on shared data.

CONCLUSION

Sharing research data involves ethical mandates to protect research participants from harm. But the ethics of data sharing extends beyond participant well-being. Researchers also have ethical obligations to the scientific community and to the public at large. The ethical principles of beneficence and justice compel researchers both to protect participants and to maximize benefits to them and to others (Brakewood & Poldrack, 2013; National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). Thus, the most defensible ethical stance for researchers is to share data

as broadly as participants allow, but to protect the data with appropriate restrictions on who can gain access. Open sharing of sensitive and identifiable psychological data that balances these ethical imperatives can be achieved. Databrary and similar restricted-access data repositories show how. Our aim should be to accelerate discovery by shedding light on the dark data of psychological science and practice that typically go unshared (Heidorn, 2008).

REFERENCES

- Adolph, K. E. (2020). Oh, behave! Presidential address, XXth International Conference on Infant Studies, New Orleans, LA, May 2016. *Infancy*, 25(4), 374–392. <https://doi.org/10.1111/infa.12336>
- Adolph, K. E., Gilmore, R. O., Freeman, C., Sanderson, P., & Millman, D. (2012). Toward open behavioral science. *Psychological Inquiry*, 23(3), 244–247. <https://doi.org/10.1080/1047840X.2012.705133>
- Adolph, K. E., Gilmore, R. O., & Kennedy, J. L. (2017, October). Science brief: Video data and documentation will improve psychological science. *Psychological Science Agenda*. <https://www.apa.org/science/about/psa/2017/10/video-data>
- American Association for the Advancement of Science. (2020). Science journals: Editorial policies. <https://www.sciencemag.org/authors/science-journals-editorial-policies#unpublished-data-and-personal-communications>
- American Psychological Association. (2020). Journal of Personality and Social Psychology: General submission guidelines: Open science. <https://www.apa.org/pubs/journals/psp?tab=4>
- Ascoli, G. A. (2006). The ups and downs of neuroscience shares. *Neuroinformatics*, 4(3), 213–216. <https://doi.org/10.1385/NI:4:3:213>
- Association for Psychological Science. (2020). 2020 submission guidelines: Open practices statement. https://www.psychologicalscience.org/publications/psychological_science/ps-submissions#OPS
- Bhattacharjee, Y. (2013, April 26). The mind of a con man. *The New York Times*. <https://www.nytimes.com/2013/04/28/magazine/diederik-stapels-audacious-academic-fraud.html>
- Brakewood, B., & Poldrack, R. A. (2013). The ethics of secondary data analysis: Considering the application of Belmont principles to the sharing of neuroimaging data. *NeuroImage*, 82, 671–676. <https://doi.org/10.1016/j.neuroimage.2013.02.040>
- Claesen, A., Gomes, S., Tuerlinckx, F., Vanpaemel, W., & Leuven, K. U. (2019, May 9). Preregistration: Comparing dream to reality. <https://psyarxiv.com/d8wex/>
- Couzin-Frankel, J. (2014, May 30). Harvard misconduct investigation of psychologist released. *Science News*. <https://www.sciencemag.org/news/2014/05/harvard-misconduct-investigation-psychologist-released>
- Databrary. (2020a). *About Databrary: Authorized access*. <https://databrary.org/about/agreement.html>
- Databrary. (2020b). *Databrary support: Ethics-related support*. <https://databrary.org/support/irb.html>
- European Council for Nuclear Research (CERN). (2020). *The Large Hadron Collider*. <https://home.cern/science/accelerators/large-hadron-collider>
- European Union General Data Protection Regulation. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016*.
- Fleming, D. A., & Reynolds, D. (2008). Ethical human-research protections: Not universal and not uniform. *The American Journal of Bioethics*, 8(11), 21–22. <https://doi.org/10.1080/15265160802516864>
- Frank, M. C. (2020). *Frequently asked questions: Why should I contribute to Wordbank?* <http://wordbank.stanford.edu/faq#why-should-i-contribute-to-wordbank>

- Gennetian, L. A., Tamis-LeMonda, C. S., & Frank, M. C. (2020). Advancing transparency and openness in child development research: Opportunities. *Child Development Perspectives*, 14(1), 3–8. <https://doi.org/10.1111/cdep.12356>
- Gilmore, R. O. (2016). From big data to deep insight in developmental science. *Wiley Interdisciplinary Reviews: Cognitive Science*, 7(2), 112–126. <https://doi.org/10.1002/wcs.1379>
- Gilmore, R. O., & Adolph, K. E. (2017). Video can make behavioural science more reproducible. *Nature Human Behaviour*, 1(7), Article 0128. <https://doi.org/10.1038/s41562-017-0128>
- Gilmore, R. O., Cole, P. M., Verma, S., van Aken, M. A. G., & Worthman, C. M. (2020). Advancing scientific integrity, transparency, and openness in child development research: Challenges and possible solutions. *Child Development Perspectives*, 14(1), 9–14. <https://doi.org/10.1111/cdep.12360>
- 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>
- Heidorn, P. B. (2008). Shedding light on the dark data in the long tail of science. *Library Trends*, 57(2), 280–299. <https://doi.org/10.1353/lib.0.0036>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Hood, L., & Rowen, L. (2013). The Human Genome Project: Big science transforms biology and medicine. *Genome Medicine*, 5(9), Article 79. <https://doi.org/10.1186/gm483>
- Inter-University Consortium for Political and Social Research. (2020). *Find data*. <https://www.icpsr.umich.edu/web/pages/ICPSR/index.html>
- Joel, S., Eastwick, P. W., & Finkel, E. J. (2018). Open sharing of data on close relationships and other sensitive social psychological topics: Challenges, tools, and future directions. *Advances in Methods and Practices in Psychological Science*, 1(1), 86–94. <https://doi.org/10.1177/2515245917744281>
- Journal of Visualized Experiments. (2020). *JoVE's objectives and criteria for publication*. <https://www.jove.com/publish/editorial-policies/#4>
- MacWhinney, B. (2020). *TalkBank: Principles of sharing*. <https://talkbank.org/share/principles.html>
- Madrugal, A. C. (2018, December 19). Facebook didn't sell your data; it gave it away. *The Atlantic*. <https://www.theatlantic.com/technology/archive/2018/12/facebook-failures-and-also-its-problems-leaking-data/578599/>
- Manolio, T. A., Brooks, L. D., & Collins, F. S. (2008). A HapMap harvest of insights into the genetics of common disease. *The Journal of Clinical Investigation*, 118(5), 1590–1605. <https://doi.org/10.1172/JCI34772>
- Meyer, M. N. (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 Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html>
- New York University. (2020). *Electronic data and system risk classification policy*. <https://www.nyu.edu/about/policies-guidelines-compliance/policies-and-guidelines/electronic-data-and-system-risk-classification.html>
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., . . . Yarkoni, T. (2015). Scientific standards: Promoting an open research culture. *Science*, 348(6242), 1422–1425. <https://doi.org/10.1126/science.aab2374>

- Nosek, B. A., Beck, E. D., Campbell, L., Flake, J. K., Hardwicke, T. E., Mellor, D. T., van 't Veer, A. E., & Vazire, S. (2019). Preregistration is hard, and worthwhile. *Trends in Cognitive Sciences*, 23(10), 815–818. <https://doi.org/10.1016/j.tics.2019.07.009>
- Nosek, B. A., & Lindsay, D. S. (2018, February 28). Preregistration becoming the norm in psychological science. *APS Observer*. <https://www.psychologicalscience.org/observer/preregistration-becoming-the-norm-in-psychological-science>
- Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, 7(6), 615–631. <https://doi.org/10.1177/1745691612459058>
- Office for Human Research Protections. (2017). *Federal policy for the protection of human subjects ("Common Rule")*. <https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html>
- Poldrack, R. A., & Gorgolewski, K. J. (2014). Making big data open: Data sharing in neuroimaging. *Nature Neuroscience*, 17(11), 1510–1517. <https://doi.org/10.1038/nn.3818>
- Suls, J. (2013). Using "cinéma vérité" (truthful cinema) to facilitate replication and accountability in psychological research. *Frontiers in Psychology*, 4, Article 872. <https://doi.org/10.3389/fpsyg.2013.00872>
- 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, L., 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), Article 160018. <https://doi.org/10.1038/sdata.2016.18>
- World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, 310(20), 2191–2194. <https://doi.org/10.1001/jama.2013.281053>