Developing Best Practices for Qualitative Health Data Sharing

Amy Pienta
ICPSR
USA
apienta@umich.edu

Shane M. Redman
ICPSR
USA
shanered@umich.edu

Abstract – Compared to quantitative data, qualitative data are shared with repositories and other researchers much less frequently due to additional practical and ethical considerations that must be made before sharing. As new data sharing requirements are implemented by various funders, understanding and addressing the barriers to sharing qualitative data are increasingly important. This paper describes a project that attempts to do just that. In collaboration with other institutional partners, we have developed a process that helps overcome some barriers to sharing qualitative health data, using technical and non-technical assistance in the data preparation process. In addition to describing the process we developed in this project, we summarize the lessons learned about emerging best practices in sharing qualitative health data and offer recommendations for improving qualitative data sharing. These recommendations can be implemented by individual researchers, institutions, and data repositories to ensure FAIR access to qualitative health data.

Keywords – qualitative data, deidentification, data sharing, health data, data curation

Introduction

Qualitative text data pose unique challenges for data sharing beyond more commonly shared quantitative data such as survey data. This is due largely to contextual information and direct and indirect identifying details referenced in the textual data. Qualitative data that are generated from health, biomedical, and clinical research studies are often sensitive and/or contain protected health information and thus require strict adherence to deidentification protocols and removing any personally identifiable information (PII) with attention to masking all required Protected Health Information (PHI) defined under the Health Insurance Portability and Accountability Act (HIPAA) privacy rule.

In this project, we identified a set of National Institutes of Health (NIH) grantees with completed research projects where they collected qualitative data during their research project but had not formally shared their data. The goal was to understand why qualitative data are not shared. The current NIH Data Sharing Policy, in effect since 2003, requires data sharing but applies only to projects where the direct costs of the research project exceed $500,000 in one or more project years.[1] Also, the NIH guidance and implementation around this policy have generally allowed researchers to provide a justification for not sharing their data. Thus, qualitative data, in particular, tends not to be shared during NIH-funded research projects.

As important as the lack of specific NIH policy implementation guidelines, researchers typically encounter numerous obstacles around the roles, expertise, and required resources for sharing qualitative health data. To reduce the burden placed on humans (e.g., individual researchers, data managers, and data curators) in deidentifying qualitative data, some researchers have developed software solutions to assist with detecting disclosure issues in qualitative data. The Bioethics Research Center Lab, in collaboration with the Institute for Informatics at Washington University in St. Louis (WUSTL), has developed a qualitative data sharing (QuaDS) software that supports the deidentification of qualitative (text) data. ICPSR

1 The NIH Data Sharing Policy can be found at https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html.

collaborates with WUSTL on this project and serves as the archive for pilot data run through the deidentification software. Drawing on our work as the data repository partner in this research project aimed at identifying and addressing barriers to sharing various qualitative health data, this paper summarizes the lessons learned about emerging best practices in sharing qualitative health data that have implications for improving the planning and allocation of resources required to ensure FAIR access to qualitative health data.

Overview of the Workflow from Data Provider to Repository

After first surveying a large group of qualitative researchers about their attitudes toward data sharing [1], the WUSTL team identified several researchers with qualitative text data to take part in a pilot project to test the QuaDS software. After agreeing to participate in the project, researchers first had to execute an agreement with WUSTL to share the qualitative data with them for deidentification purposes. The pilot researchers also received a detailed set of guidelines that included information about every project phase, from working with the QuaDS software to sharing the deidentified data with ICPSR for archiving. The next step was for the researcher (and any research team members) to meet jointly with a representative from WUSTL and ICPSR to discuss the project's workflow in a kick-off call. This was an opportunity for the project team to provide additional details about each phase and for the researcher to ask questions about any aspect of the project.

After the kick-off call, the researcher would work with WUSTL to run the transcripts through the QuaDS software for deidentification. It is important to note that even if researchers had done some manual deidentification to the transcripts already, they were still eligible to participate in the project as the software could catch some identifying information that may have been missed. Once a transcript was uploaded to the software, it would flag any identifying information it recognized and suggest a replacement. The researcher would review all of the suggestions made by the software and could accept or reject the suggested changes, or they could edit the replacement text. They could also flag and replace additional identifying information in a transcript the software did not catch. The researcher was asked to read through all the transcripts after the software performed its work to ensure no identifying information remained. The project's final phase involved the researcher depositing the final deidentified transcripts (i.e., the output files from the QuaDS software) and supporting documentation files to ICPSR for long-term preservation and sharing with secondary users. ICPSR would then perform its standard curation process on the qualitative data files, which included an additional disclosure risk review before making the data available to other researchers as a restricted-use dataset.

Despite the range of topics covered by researchers in this project, the ingest, curation, and dissemination activities following deidentification were standardized by ICPSR's existing workflows. A benefit of receiving all the deidentified transcripts from the QuaDS software is that the files are all provided to the researchers and subsequently to ICPSR in the same format (.txt files). This helped streamline the curation process for each dataset as we did not need to consider how to handle or incorporate a combination of unknown file types in each deposit. At various points throughout the project, the researchers completed evaluation surveys so that we could understand how well the software performed, among other evaluations. ICPSR curators also recorded notes while processing the data to understand where the software deidentification differed from ICPSR's deidentification standard.

Types of Qualitative Health Data we Encountered

This project addressed the ethical and practical barriers to sharing transcripts generated from participant interviews or focus groups in projects funded across multiple institutes with the NIH. The qualitative text we encountered covered a wide range of topics, including receipt of treatment for psychiatric disturbances following traumatic brain injury, adolescent sexual development, fentanyl risk communication, attitudes toward medication-assisted treatment, and transcripts from mental health courts.[2] Although most of the research projects had not envisioned sharing the qualitative health data and were mostly completed

2 For a full list of the studies curated to date from the project, visit the Qualitative Data Sharing (QDS) Project Series page at ICPSR at https://www.icpsr.umich.edu/web/ICPSR/series/1780.

in one to several years prior (with at least one completed as many as ten years prior), the NIH grantees agreed to work with our project team to use the newly developed QuaDS software.

IRBs and Post Hoc Determination about Qualitative Data Sharing

In the context described above, where sharing data was not part of the original research plan, most research project teams first approached their IRBs to determine whether the qualitative data could be shared. In many cases, informed consent documents provided to study participants were silent on data sharing; they did not explicitly mention that the data would be shared, especially not outside of the research team. Indeed, some informed consent documents stated that the data would not be shared with anyone outside the research team. In most cases, the researchers/studies could not participate in this project because the transcripts could not be shared with WUSTL for deidentification using the QuaDS software. However, for all the studies where the informed consent documents were silent about data sharing or did not explicitly limit the sharing of the data, IRBs allowed a deidentified version of the data to be shared. As a first step, IRBs allowed researchers to share the data with WUSTL under a service contract for deidentifying the data through the software. Once the data were deidentified, many IRBs noted that they were no longer human subjects data and could then be shared more broadly with ICPSR and the scientific community.

Ensuring Continuity in Data Governance

An overarching goal of responsible data sharing is ensuring that when data are transferred from organization to organization, any relevant human subject protections remain in place, and the receiving organization understands the needs of the giving organization [2]. A Restricted-Use Data Deposit and Dissemination Agreement (RUDDDA) was executed between the University of Michigan, covering ICPSR's role and the organization where the research was performed. ICPSR has used this agreement for many years. Still, at the project's outset, whether the agreement would adequately cover the institutional needs of those depositing qualitative data in ICPSR was unclear.

One of the key features of the RUDDDA is that it asks the depositing institution to warrant that the data transferred to ICPSR are deidentified and any relevant Institutional Review Board conditions on data use are communicated to ICPSR. Second, the agreement communicates that ICPSR can (1) archive and make the data available to Third Parties (data users) by establishing a Restricted Data Use Agreement (RDUA) with the data user's institution and (2) process the metadata and data to improve its discoverability and use. Finally, the RUDDDA describes the obligation that ICPSR will restrict data access to its staff and provide security so the data are not disclosed to anyone outside ICPSR and any authorized users who sign an RDUA. There are also additional obligations about the enforcement of RDUAs to third-party users. Thus, any compliance and obligations from the data provider's institution are passed to the ICPSR/University of Michigan and to any subsequent institution where a data user is authorized to access the restricted data.

Historically, it was commonplace that it could take many months to execute a RUDDDA before data could be transferred to ICPSR. In some instances, it could take more than a year. The RUDDDAs established through this project have been relatively quick to establish. Fourteen RUDDDAs have been fully executed through this project, with the average time from RUDDDA submission to the University of Michigan's Office of Research and Sponsored Projects to full execution being about 4.3 weeks (30 calendar days). However, there is quite a bit of variation in this step of the process, with two RUDDDAs being executed in less than one week and two other RUDDDAs taking more than two months to execute. Many factors affect the processing time of these agreements, including the types and extent of modifications requested by the data-providing organization and the queue of agreements to be reviewed at both the data-providing organization and the University of Michigan.

Preparing to Deposit Qualitative Health Data to a Data Repository

A common barrier to sharing data often expressed by researchers is the time it takes to deposit data to a repository [3]. To facilitate a smooth deposit process and reduce the burden on researchers depositing their data, we provided two main sources of deposit assistance throughout this

project. First, in collaboration with our project partners, we developed instructions for each researcher depositing data at ICPSR. This document walked researchers through the online deposit process at ICPSR, with screenshots of each screen the researcher would encounter. Second, we held a kick-off call with every researcher participating in this project. In the kick-off call, we provided researchers with an overview of the deposit process, provided additional information about the process of executing a RUDDDA, and allowed the researcher to ask us any questions before they started the deposit process. Researchers were also encouraged to contact ICPSR with any questions as they arose throughout the deposit process. We found that very few researchers contacted us with issues related to depositing their data, which we attribute to the detailed instruction document and kick-off call before the deposit process began.

Curating Qualitative Data for Reuse

One of the main concerns with the reuse of data, especially qualitative data, is protecting the identities of study participants [4]. Compared to quantitative data, which is much more structured and easier for a computer algorithm to find and replace problematic variables systematically, qualitative data requires more human effort to find parts of the data that pose a disclosure concern. Thus, a large part of the qualitative data curation process at ICPSR typically involves identifying and mitigating disclosure risks. As part of this project, however, ICPSR anticipated that the original research team and the QuaDS software would de-identify the qualitative data. Thus, we expected to curate the incoming data more quickly than typical qualitative data we might receive outside this project.

In addition to ethical protection of human subject participant data concerns, qualitative data curation involves retaining context so the data can be reused through quality documentation [5]. ICPSR's curation of the qualitative data involved (1) creating a study-level metadata catalog entry from information supplied by the depositing organization, (2) ensuring all files deposited are openable, readable, and match what is expected, (3) reviewing the files for the presence of direct identifiers, (4) creating documentation and converting to PDF/A and additional formats for ease of use, (5) assigning a persistent identifier (ICPSR uses DOI), (6) associating any related citations to the data, and (7) ensuring the long-term preservation of the Archival Information Package (AIP) following the Open Archival Information System (OAIS) and releasing a Dissemination Information Package. In addition to these curation steps, ICPSR performs two quality checks (QC) of the curator's work on the data and documentation: one by a peer, which is lengthier, followed by a second, briefer review by a senior curator. For qualitative data, the QC process can be especially lengthy, given that verification of deidentification involves the peer curator reviewing the full text. Thus, while both QCs were performed for the first several qualitative studies archived at ICPSR through this project, the high level of accuracy of the pre-curation deidentification phase using the QuaDS software led us to eliminate the lengthier peer QC and allow the senior curator to only spot check the files.

Table 1 below shows that the 14 qualitative data collections deposited to date contain an average of 20 data files and 254 pages of transcripts per study. Interestingly, on average, the data collections that received the reduced curation process (which eliminated the peer QC stage) were larger than those that received full curation. The average data collection that received reduced curation had about 23 transcripts with 330 pages. In comparison, the average data collection that received full curation had about 16 transcripts with 178 pages. The qualitative data studies have required an average of 66 hours to perform the curation steps listed above, including one or both QCs. The later set of larger studies, where we eliminated the lengthier QC, took only 33 hours to curate on average, less than half of the time required to curate the studies with both QC stages (which took 84 hours on average).

TABLE 1.
Attributes of the Qualitative Data Deposited and Average Length of Curation (n=14 studies received; n=11 studies curated)

		Avg	**Min**	**Max**
Number of transcript/data files	Total (=14)	20	3	40
	Full Curation (=7)	16	9	31
	Reduced Curation (=7)	23	3	40
Number of transcript pages	Total (=14)	254	39	610

	Full Curation (=7)	178	39	399
	Reduced Curation (=7)	330	66	610
Curator work hours logged	Total (=11)	66	15	159
	Full Curation (=7)	84	41	159
	Reduced Curation (=4)	33	15	44

Review of the Results of the Software Deidentification Process by Data Curators

The QuaDS deidentification software is programmed to flag the 18 HIPAA Safe Harbor identifiers as well as other non-HIPAA information that may be used to identify participants, including geographic areas larger than a state, numbers, organization/institution names, race, ethnicity, nationality, and indigenous status, and sexual orientation, among others.[3] After going through the QuaDS software at WUSTL, the deidentified data are deposited to ICPSR by the researcher. Data curators at ICPSR then review the data manually to ensure the deidentification performed by the software meets ICPSR's disclosure risk review standards before sharing the data with other researchers. Overall, we found that the deidentification software performed quite well [6], but some issues are being improved in future versions.

One issue found is that shortened names within the text, regarding individuals and organizations, are not always identified by the software. For example, the software may flag "Susan," but if the interviewee referred to this person as "Sue" later in the transcript, it would not be flagged. Similarly, "Hopkins" may refer to Johns Hopkins University but would not be flagged if the full name is not used.

Another issue found by the data curators is that the software does not always flag neighborhoods, street names, or other context clues (e.g., descriptions of hospitals or unique job positions) that may be identifiable. For example, the "house on Orange" may be missed as a street identifier. Finally, the software was sometimes overly conservative at flagging possible identifiers, especially related to numbers in the text, since numbers are frequently an identification risk. For example, inches of snow and interview numbers already randomly assigned by the PI were masked as identifiable by the software. In these cases, ICPSR deferred to the researcher's choice of keeping the numbers masked since we did not have the raw data to revert the masking process performed in the software. In other words, ICPSR could mask additional information that the software and/or researcher missed. Still, ICPSR could not revert any unnecessary masking already done without the original data files.

There were some key benefits to the overall process of using both software operated by the research team producing the data and review by data curators of the repository. First, the software evolved throughout the project using the feedback provided by the curators about the small errors found, which helped improve the software. Second, the data running through the deidentification software saved the curators quite a bit of time in masking the data compared to qualitative data outside of this project that have not been deidentified by the software. The information masked by the software was easily identifiable during the review and was efficiently applied throughout the remainder of the transcript. That said, the review by curators makes clear the need for some human element in the deidentification process for text data.

Challenges Related to the CoreTrustSeal Repository Role

Repositories wishing to accept and manage qualitative health data will be guided by improving usability by standardizing the data and metadata, preserving the data, and providing access to the user community. As described above, qualitative health data presents unique curation challenges in that contextual information contained within the rich text may be used indirectly to disclose the data, and additional restrictions on use must be managed. For CoreTrustSeal-certified repositories, including ICPSR, some robust processes and policies are applied to incoming data deposits to ensure qualitative health data are accessible and can be reused outside the original research team. These data repository functions have costs, and repositories will offer different solutions and cost models around qualitative health data.

3 For the 18 HIPAA-defined categories, visit https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#standard.

In this project, we learned several friction points exist across the distributed roles across the three organizations that provided public access to qualitative health data from the NIH research projects -- the organization collecting data, the organization where the data were deidentified, and the data repository. The data provider's organization and the organization offering the software to help deidentify the qualitative health data worked together to produce de-identified health data following the HIPAA guidelines for removing direct identifiers. Pre-deposit work does not remove the obligation or work of the repository in examining the data to ensure a minimal deidentification standard was met. For qualitative health data, the disclosure review work can represent many hours of curation work to ensure that all direct identifiers have been removed.

With the goal of the incoming qualitative health data being available to authorized third-party users, ICPSR adhered to its usual disclosure risk review (DRR) process to minimize the risk of any inadvertent disclosure of data. ICPSR's disclosure risk review process found minor inconsistencies around the redaction of direct identifiers. The ICPSR Data Stewardship Policy Committee, which has the authority to exempt data from the ICPSR DRR process according to its policy, decided verification of deidentification was necessary across all the qualitative health data ingested throughout the project. Nonetheless, the high accuracy of the deidentification work before deposit [6] minimized this labor-intensive process and reduced the cost incurred for the repository staff work.[4] Further, because ICPSR agreed to manage access to the qualitative health data using restricted-use data processes, this allowed ICPSR to also rely on prohibiting third-party data users from disclosing any data via the execution of a restricted data use agreement.

Recommendations for Improving Qualitative Health Data Sharing

We highlight some of the lessons learned from the workflow of this project to outline some key points for improving FAIR access to qualitative health data. Our interactions with the research teams and through the process of becoming data stewards of the qualitative data from their NIH projects suggest some areas where the research community and repositories can work together more effectively.

A. Early planning and preparation during data collection

Data sharing is best started before data collection begins [7]. Starting early ensures that potential challenges to data sharing are addressed proactively before data is collected. Researchers should plan for data sharing and address IRB and informed consent issues, as well as resources and roles related to data sharing, at the outset of collecting data. Information about how qualitative data will be managed and shared is recorded in the data management plan, where the researcher establishes how they will maintain trustworthiness around the data being collected. Although data management plans, originating in design to capture management of quantitative data, have numerous problems in their application to qualitative data [8], capturing information about the scope of the data collection and how openly the researcher plans to share the data is valuable.

An important aspect of this planning phase is developing a human subjects protection plan that both protects the confidentiality of the human subjects and provides information about plans for data sharing. The present study generally did not include studies where informed consent stated that no data would be shared outside the research team. Unfortunately, this is quite a common occurrence in research, and Institutional Review Boards and universities vary in their information to improve this when data are being collected. An all too often pattern is for researchers to be directed to past approved informed consent statements, limiting language being passed on to the next generation of studies. There are better models for informed consent that address this need, especially found in data repositories. As funding agency requirements and norms, and expectations around data management plans and data sharing strengthen, it is more likely that data repositories will be consulted early in the research process about informed consent. Data repositories, too, will want to take a proactive role in educating the research community about planning and implementing data

4 In an evaluation study of the Natural Language Processing pipeline used in the QuaDS software Reference [6] finds a consistent F1-score of ~0.90. After adjusting the pipeline following an error analysis, the F1-score increased to 0.96 in a second iteration.

sharing at these earliest parts of the research life cycle.

B. Development of Depositor Guidelines and Resources

Reference [9] points out that most data repositories do not offer guidelines or support for the preparation of qualitative data. The intervention offered throughout this project suggests this is a fruitful area for further development. We learned that researchers we engaged in the deposit process understood and responded to the data governance needs of the repository. We helped this along by providing technical assistance early in the workflow with the projects and sharing the full process for depositing qualitative data at the start. Consequently, studies moved through the RUDDDA phase quicker than ICPSR has observed in the past and were able to deposit the data with few challenges.

We can offer a few observations about the scope of the qualitative data that should be deposited. The materials that should be deposited to a repository that will later be shared with secondary users are similar to any other data deposit where the context around the study matters and should be captured. Researchers should include all the raw but deidentified data files in their deposit and any documentation, allowing other researchers to understand and analyze the data collection independently. When data are collected using Computer Assisted Qualitative Data Analysis Software (CAQDAS) such as NVIVO, ATLAS, or Dedoose, researchers should be encouraged to include the final coding tree and any useful memo or protocol files in their deposit. This will aid in replicability and verification as well as reuse.

C. Balancing Approach to Curation Resources

Access to high quality data and metadata requires a substantial investment of resources. This is acutely true for qualitative data, which relies on conveying the context of the study and preserving the narrative as much as possible. Considerable resources are often needed to ensure that data can be accessed after the data are collected. This project is among the first to capture the number of human hours for curating the data when working in an environment with tools and processes to work efficiently. Data curation costs were lower than those of similar work happening outside the project due to the implementation of the QuaDS software as a helpful first pass at the deidentification of the qualitative data. Also, the repository's commitment to establishing DUAs with data users further added protection against unintentional data disclosure through the reuse process. Although we have not directly compared the required hours of work across the studies archived under this project with studies archived outside the project as yet, the 11 studies curated to date under this project indicate that the pre-deposit assistance and the quality of the deidentification from the QuaDS software have the potential to create significant efficiencies when it comes to ingesting and curating qualitative data.

Conclusions

By setting up a workflow with resources and tools to support qualitative data sharing, we expanded access to several qualitative health data collections that otherwise would not have been shared. This work also allowed us to understand where additional resources and attention are needed to realize open science goals for qualitative data. Many papers have pointed to early and careful planning to conserve limited resources and avoid future data sharing barriers. We add the observation that previously constructed IRB plans and informed consent processes that did not consider data sharing limit the potential for sharing qualitative data. However, we also found willingness to meet the rapidly changing expectations around data access, and many projects were able to transfer and share data using a repository by paying careful attention to data deidentification and when the basic data governance structure could be maintained.

Repositories are advised to meet researcher needs by filling in some of the missing information about planning and resourcing data sharing so that qualitative data do not remain adrift outside the repository sphere. We also recognize that repositories are feeling constrained by limited resources and need tools and processes that allow them to meet the demand for accessible data. This project demonstrates that applying a full range of solutions to qualitative data is helpful for repository costs. As part of this, we provided enough standardization for data to be used well but did not focus on creating a perfectly curated data package. Whether this strategy impacts use negatively remains to be seen.

As the NIH implements its new data sharing policy on January 25, 2023 [10], this project points out what is entailed in ensuring FAIR access to qualitative data. The policy calls for data to be shared using repositories to ensure long-term access, but what is missing is the data type-specific practices and required resources that will ensure qualitative data are a meaningful part of the NIH data ecosystem. Our project underscores that the QuaDS software, the RUDDDA, deposit technical assistance, and the modifiable curation approach of the repository worked together to create an environment for sharing qualitative health data and can serve as a model for future data sharing

REFERENCES

[1] J. Mozersky, T. McIntosh, H.A. Walsh, M.V. Parsons, M. Goodman, and J.M. DuBois, "Barriers and facilitators to qualitative data sharing in the United States: A survey of qualitative researchers," *PLoS One*, vol. 16, no. 12, e0261719, 2021. doi.org/10.1371/journal.pone.0261719

[2] S. Mannheimer, A. Pienta, D. Kirilova, C. Elman, and A. Wutich, "Qualitative data sharing: Data repositories and academic libraries as key partners in addressing challenges," *American Behavioral Scientist*, vol. 63, no. 5, pp. 643-664, 2019. doi.org/10.1177/0002764218784991

[3] B. Saunders, J. Kitzinger, and C. Kitzinger, "Anonymising interview data: Challenges and compromise in practice," *Qualitative Research*, vol. 15, no. 5, pp. 616-632, 2015. doi.org/10.1177/1468794114550439

[4] S. Karcher, D. Kirilova, C. Pagé, and N. Weber, "How data curation enables epistemically responsible reuse of qualitative data," *The Qualitative Report*, vol. 26, no. 6, pp. 1996-2010, June 2021. doi.org/10.46743/2160-3715/2021.5012

[5] S. Mannheimer, "Data curation implications of qualitative data reuse and big social research," *Journal of eScience Librarianship*, vol. 10, no. 4, e1218 pp. 1-25, 2021. doi.org/10.7191/jeslib.2021.1218

[6] A. Gupta, A. Lai, J. Mozersky, X. Ma, H. Walsh, and J.M. DuBois, "Enabling qualitative research data sharing using a natural language processing pipeline for deidentification: moving beyond HIPAA Safe Harbor identifiers," *JAMIA Open*, vol. 4, no. 3, pp. 1-10, July 2021. https://doi.org/10.1093/jamiaopen/ooab069

[7] L. Corti, V. Van den Eynden, L. Bishop, and M. Woollard, *Managing and Sharing Research Data: a Guide to Good Practice*, Sage Publishing, 2014. https://doi.org/10.25607/OBP-1540

[8] M.G. Antonio, K. Schick-Makaroff, J.M. Doiron, L. Sheilds, L. White, and A. Molzahn, "Qualitative data management and analysis within a data repository," *Western Journal of Nursing Research*, vol. 42, no. 8, pp. 640–648, 2020. doi.org/10.1177/0193945919881706

[9] A.L. Antes, H.A. Walsh, M. Strait, C.R. Hudson-Vitale, and J.M. DuBois, "Examining data repository guidelines for qualitative data sharing," *Journal of Empirical Research on Human Research Ethics*, vol. 13, no. 1, pp. 61-73, 2018. doi.org/10.1177/1556264617744121

[10] Final NIH Policy for Data Management and Sharing, NIH, 2020. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html