DRUM Human Participant Consent Policy Proposal

Authors: Alicia Hofelich Mohr, Valerie Collins, Shanda Hunt

September 28, 2021

Background

On September 6, 2021, FORCE11 and COPE published formalized recommendations for ethical challenges related to the sharing and publication of datasets. Their documents regarding risk address sensitive and protected data. The joint recommendations include careful screening of datasets before publication, removal of problematic datasets, and clear policies around these matters: “Data repositories must establish clear terms of use that address ethical violations and how the data publisher will respond if it comes to light that the public availability of a dataset presents a risk to human subjects, endangered species or sites, specific communities or society.”

Sharing human participant data raises ethical considerations that don’t exist with other types of research. Sharing human participant data - when shared at the individual-level rather than in aggregate - increases the risk of re-identification, or exposing a participant’s identity. Additionally, if the data are deposited into a public repository, there are no restrictions on how the data are used or who accesses it - this may be in direct conflict with what the participants agreed to in the consent form, or they may be completely unaware that this is how their data is being used.

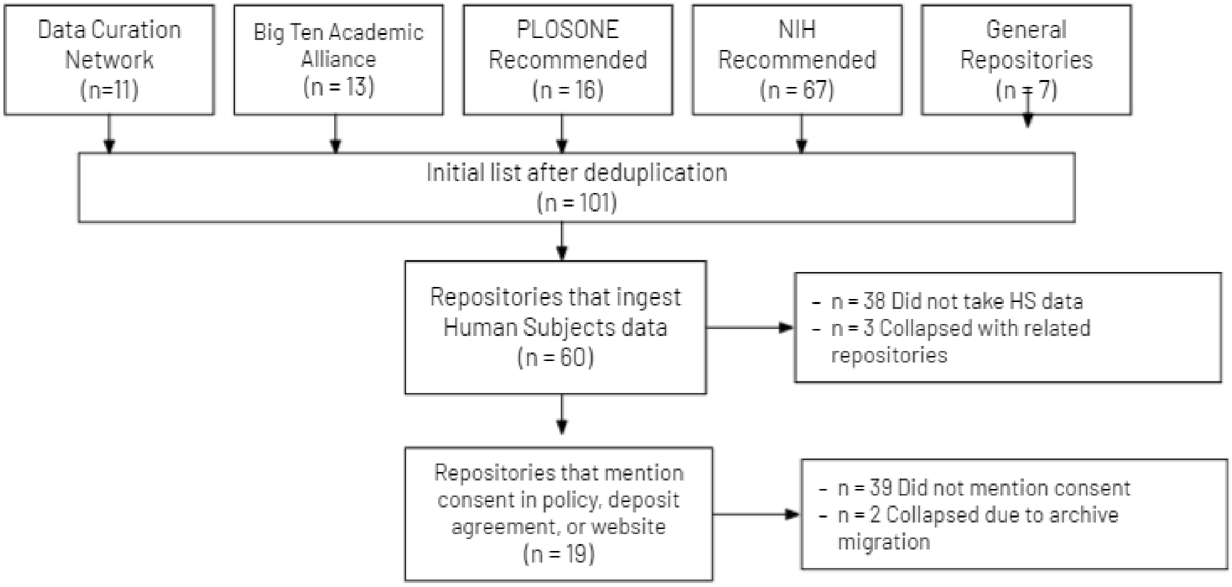
Beyond all the data files we see as curators (i.e., audio interviews, transcripts, psychophysiological data, spreadsheets of survey responses, etc.) are the humans behind them - the people who willingly gave their personal information for the sake of research and knowledge. Both Shanda and Alicia (the two DRUM human participant curators) come from research backgrounds - they've interviewed participants, asking questions about sexual behavior and health and coaching them through reliving and evaluating painful personal traumas. These experiences drive us to be good stewards of the informational gifts participants give us through their data, and the emotional labor and time participants have put into our studies. To us this means both ensuring their information is protected and their efforts have an impact beyond the study. Informed consent offers participants autonomy in the research process, and for marginalized people and populations this is especially crucial. At a minimum, they should be made aware of how and where their data will be shared.

Within DRUM, there are 62 datasets with human participant data (as of 2020). Because of the sensitive nature of human participant data, and the potential issues with publishing such data in a publicly accessible repository, DRUM assesses all human participant data before acceptance into the repository. As part of this process, we review data for de-identification concerns, and we also ask researchers to share with us the consent form that participants signed as they joined the study. We've encountered problematic language regarding the sharing of data in some consent forms. They mention records rather than data, and promise that responses will only be seen by the researcher. Even when the data is mentioned, the language can still be problematic, as in: "Data will only be shared in aggregate," "Data will only be seen by the research team," and "Data will be destroyed."

Institutional review boards have historically provided sample language that does not consider the possibility of sharing the data to broad audiences. Since researchers look to their IRBs for guidance on consent forms, we typically see this default language in the consent forms we receive. Additionally, DRUM has been inconsistent in our data acceptance practices due to changes in submission requirements over time (i.e., we now require that the consent form is submitted) and lack of a clear policy regarding participant consent for data sharing. Thus, from 2019 to 2020, we performed a scan of consent policies found on other data repository websites to inform development of a new DRUM policy.

Repository Scan Results & Recommendations

The methodology focused on an environmental scan of data repository websites with the aim of identifying repositories that ingest human participant data. The initial list was created from the Data Curation Network, Big Ten Academic Alliance, PLOS ONE and NIH recommended repositories, and well-known general repositories. After duplicates and out-of-scope repositories were removed, we ended with 19 repositories for the analysis.



We scanned for consent language in deposit agreements and on all repository webpages. We located 15 deposit agreements. Twelve stated that data sharing should be consistent with informed consent. Of those 12, nine went above and beyond to state that an ethics review board must also verify that data sharing is allowable or that limitations established in the consent form must be shared with the repository. We also scanned entire repository websites, and found many provide additional guidance on informed consent resources and requirements outside the deposit agreements.

Based on these results, in a 2020 Medical Library Association presentation, we made the following recommendations for development of human participant consent policies:

- Human participant consent policy language should be in the deposit agreement and match any language about consent on the website.
- Any consent form language guidance on the repository website should advise explicit data sharing language in the consent form.
- Require that depositors submit the participant consent form with their data submission, and establish how to evaluate those consent forms.
- If unsure how to reconcile the consent form with the type of data sharing that your repository does, defer to the IRB. Keep in mind, however, that IRBs in general are less knowledgeable about data sharing and can downplay the significance of informed consent. **Note: our experiences of referring researchers back to the IRB for data sharing determinations have resulted in more permissive data sharing than recommended by us and other groups such as FORCE11 and COPE, with a primary focus on de-identification of the data, rather than explicit consent language.*
- Have an established practice in place for how to handle participant agreements (these sometimes exist when a consent form isn't required).

Policy Proposal

Original language in black text; proposed additions in pink text.

Before you submit (About the Data Repository page)

Check that your data meet criteria for acceptance into DRUM, including:

- Data must be authored by at least one University of Minnesota researcher with an active U of M Internet ID.
- Data must be non-restricted data that DO NOT contain any private, confidential, or other legally protected information (e.g., personal identifiable information).
- If you are uploading human participant data, you will be asked to provide the consent form, information sheet, or other participant agreement, and a Digital Conservancy curator will evaluate the submission according to the criteria specified in our Human Participant Data Policy and Guidelines.
- Data must be deposited for open access. Authors will have the option of restricting access for a maximum of two years (see our End-user access policy).
- Data are digital, and each file must not exceed 150 GB. Larger data files are considered on a case-by-case basis. Please contact us to discuss larger data sets.
- Data must include adequate documentation describing the nature of the data at an appropriate level for purposes of reuse and discovery. All data receive curatorial review and data that are incomplete or not ready for reuse may not be accepted into the repository.

    - Learn how to document your data using a readme.txt template.
- The data should be in a final publishable state. For active or changing data, use a U of M storage solution listed on the IT@UMN website.

Data collection policy (DRUM Policies page)

Check that your data meet criteria for acceptance into DRUM, including:

- Data must be authored by at least one University of Minnesota researcher with an active U of M Internet ID.
- Data must be non-restricted data that DO NOT contain any private, confidential, or other legally protected information (e.g., personal identifiable information). See Human Participant Data Policy if you are uploading data collected from or about humans.
- Data must be deposited for open access. Authors will have the option of restricting access for a maximum of two years (see our End-user access policy).
- Data are digital, and each file must not exceed 150 GB. Larger data files are considered on a case-by-case basis. Please contact us to discuss larger data sets.
- Data must include adequate documentation describing the nature of the data at an appropriate level for purposes of reuse and discovery. All data receive curatorial review and data that are incomplete or not ready for reuse may not be accepted into the repository.
    - Learn how to document your data using a readme.txt template.
- The data should be in a final publishable state. For active or changing data, use a U of M storage solution listed on the IT@UMN website.

Human Participant Data Policy (to be after Data Collection Policy)

Data involving human participants will be subject to additional review to ensure that it meets our requirements for public sharing. Data that we determine is insufficiently de-identified or which has not properly consented participants to the sharing of individual-level data in a public repository may be rejected from DRUM, or may be subject to removal after acceptance.

See our Human Participant Data Guidelines for additional details.

Human Participant Data Guidelines (to be after Preservation Policy)

- If you are uploading human participant data:
    - You will be asked to provide the consent form, information sheet, or other participant agreement.
    - A Digital Conservancy curator will evaluate the submission for the following criteria:
        - The dataset has been de-identified by removing direct and indirect identifiers. A Digital Conservancy curator may make further de-identification recommendations.
        - The consent form, information sheet, or other participant agreement should not contain any language that restricts sharing data. Examples of restrictive language include:
            - "Data will only be shared in aggregate."

      - "Data will only be seen by the research team."
      - “Data will be destroyed.”
    - The consent form, information sheet, or other participant agreement should explicitly state that data will be shared at the individual level and that data will be shared publicly (anyone on the Internet can access the data). Please note that “records” are not equivalent to data.
  - If the dataset does not meet the above criteria, a Digital Conservancy curator will evaluate the language in the consent form versus potential harm caused by sharing without participants’ knowledge. See evaluation rubric below:
    - If the data is not at all sensitive in nature and clearly de-identified, the curator will likely accept the dataset into DRUM and share resources for best practices for the future.
      - Consent form language consultation (e.g., clear data sharing language, altering UMN templates, etc.)
      - Data sharing information sheet for full transparency with research participants
    - If the data is sensitive or unable to be de-identified, the curator will recommend participant re-consent or redirect to another repository that allows for restricted access to the data.
    - If a researcher or research team continues to submit datasets that have not acquired appropriate consent after two educational opportunities that led to acceptance of the data into DRUM, the Digital Conservancy reserves the right to reject the data. We will make recommendations for alternative repositories that allow for restricted access to the data.
    - If a Certificate of Confidentiality is mentioned in the consent form, the language about data sharing in DRUM would need to be very explicit or the dataset will be rejected. We will make recommendations for alternative repositories that allow for restricted access to the data.
  - Contact us before you start your submission to ensure that your consent form, information sheet, or other participant agreement aligns with open data sharing practices.

Preservation policy (DRUM Policies page)
The Data Repository accepts data files in all digital formats, however, digital preservation support will be provided at different levels for specific formats as specified in the UDC preservation policy. In order to meet various compliance requirements, and to support the use of data for scholarship, the data housed in the Repository will be preserved at the appropriate level for at least 10 years following acceptance. Criteria for longer preservation periods will be established in accordance with University policy on research data management. Published data that is found to violate UDC or DRUM policy can be removed by the repository administrators at any time in accordance with the UDC withdrawal policy.