**TC IRB Researcher Guidance: Open Science and IRB Submissions**

**Introduction**

This guidance document consolidates essential elements of open science and consent disclosure, providing standardized consent templates and key considerations for researchers. It aims to ensure informed and voluntary participation and promote transparency, responsible data management, and adherence to ethical practices in research.

Open science policies aim to guide access to information responsibly, promoting resource sharing, idea dissemination, collaboration, and synergy in research. Ethical open science practices ensure that research data is accessible without accelerating risk, violating consent agreements, or breaching confidentiality and privacy.

**Federal Context & Restrictions**

Federal funding agencies such as the National Institute of Health (NIH) and the National Science Foundation (NSF) advocate for open data, requiring data management and dissemination plans in some grant applications. Both agencies encourage the sharing of scientific data to advance biomedical research, ensure responsible data management, and emphasize protecting participant privacy but do not explicitly state that data must be identified. Instead, researchers are empowered to choose appropriate data-sharing methods, including de-identification.

Data sharing may also be *restricted* by data use agreements (DUAs) and other agreements. These agreements may impose conditions such as limiting the use of data to specific purposes, protecting confidentiality, and restricting data redistribution. Researchers must be aware of these limitations when developing their open data plan.

### **Considerations for Making Data Open**

Making your data open depends on what participants have agreed to on the consent form. This is especially important if you want to make existing data open. The following points should be considered:

* ***Privacy and Confidentiality:*** If when initially consented, participants are told their participation will remain private (identities not disclosed) and responses will remain confidential (identities not linked to responses), openly disseminating de-identified data would not violate the agreement.
* ***No Dissemination Clause:*** If when initially consented, participants were told the data would “not be disseminated,” the researcher may violate the agreement if they openly share the data. In this case, the IRB would need to be involved.
  + Subjects may need to re-consent to allow their responses to be disseminated, and new IRB approval may be needed as the original consent agreement may change.
* *De-identify Data:* Ensure that data are de-identified to maintain subject confidentiality. This means that responses cannot be linked back to individuals' identities. This includes removing names, student IDs, subject numbers, timestamps, and anything else that could be used to uniquely identify a person.
* *Tiered Access Levels for Data Sharing:* When sharing data, consider implementing tiered access levels based on the sensitivity of the data and user credentials:
  + **Public Access:** Non-sensitive, de-identified data available to the general public.
  + **Registered Access:** Users register and agree to conditions such as no re-identification attempts.
  + **Controlled Access:** Highly sensitive data accessed by authorized researchers following a review process.

An example of this tiered access data sharing is the *All of Us* program, for which each person who will access the data must submit a separate request and acknowledge the data agreements and limitations, depending on the type of data being requested (e.g., aggregate data vs individual-level data).

Here’s sample consent language that *All of Us* has used for this purpose:

*We will create a* ***public*** *database on the All of Us website. The data in the public database will be about the group. It will not include data about individual people. It will not include your name or other data that directly identifies you. Everyone will be able to use the public database to make discoveries.*

*We will also create a scientific database. The scientific database will have individual-level data and samples. Access to this database will be* ***controlled****. Researchers will have to be approved by All of Us to use this database. These researchers may be from anywhere in the world. They may work for commercial companies, like drug companies. They may be citizen scientists. Citizen scientists are people who do science in their spare time. Their research may be on nearly any topic. You can learn more about the research being done at* [*www.joinallofus.org*](http://www.joinallofus.org)*.*

For more information, please refer to [The All of Us Consent Process](https://allofus.nih.gov/about/protocol/all-us-consent-process).

* ***Individual-Level Data Sharing:*** While sharing de-identified data is preferred, sharing of individual-level data with all ***direct identifiers*** removed may be permissible if participants provide informed consent in advance. Even when direct identifiers are removed, careful attention must be given to the presence of ***indirect identifiers*** (quasi-identifiers) that could still allow for re-identification when combined with other data sources. To ensure privacy and confidentiality, any data sharing must include safeguards such as controlled and restricted access, Data Use Agreements, and measures that minimize the risk of re-identification from both direct and indirect identifiers. These protections, along with the associated risks, should be clearly outlined in the consent document, similar to the approach taken in the All of Us program.

1. ***Direct identifiers include information that relates specifically to an individual such as the individual’s residence, including for example, name, address, Social Security Number or other identifying number or code, telephone number, e-mail address, or biometric record.***
2. ***Indirect identifiers, also known as quasi-identifiers or inferential identifiers, are characteristics that may not be unique across an entire population but can be distinctive within a particular sample. These identifiers can be combined with other data points to potentially re-identify participants in a study, posing a risk to their privacy.***

Researchers must carefully assess the presence of indirect identifiers when considering how to share their data. Even if direct identifiers are removed, indirect identifiers can still lead to re-identification when linked with external information. For this reason, researchers should take these identifiers into account when posting their data and establishing "restricted" or “controlled” access controls to mitigate the risk of re-identification.

**Table 1. Identifier Types & Considerations**



### **Recommendations for Consent Language Development**

* Balance detailed information with simplicity to ensure participants understand without being overwhelmed.
* Include provisions for potential future uses of data while ensuring ethical considerations.
* Establish processes for obtaining additional consent when new data uses alter the initial agreement or present new risks.
* Use re-consenting for significant changes and consent addendums for updates or clarifications.
* Clearly explain data de-identification processes, data sharing extents, and participant rights concerning data use.

**Sample Consent Language**

Here is sample language to be considered and added under the **HOW WILL THE RESULTS BE USED** section of the consent document:

*Data with all identifiers removed may be used for future projects that focus on any topic and may be unrelated to this study. This new data may be made available to the public via the Internet and an open database. This information will not have your name or other personally identifiable information included (i.e. it will be de-identified). Therefore, the data we share with the public will be free of information that would link your responses to it.*

***(If you know in advance how the data will be made publicly available, you can link to that website/information about the repository that will be used.)***

*The deidentified data will be uploaded to the [REPOSITORY NAME] data repository, and as such many questions about the repository/open data can be answered here [FAQ/ABOUT PAGE URL]*

*If you would like to participate in the study but are concerned about the public nature of the data, you can opt out of having your data made publicly available while still participating in the study. In this instance, your data will be viewed only by the researchers, and will not be included in the dataset that will be made available to the public. Please initial if you do or do not consent to have your de-identified data made publicly available:*

* *\_\_\_\_\_\_I give my consent to having my de-identified data made publicly available.*
* *\_\_\_\_\_\_I* ***do not*** *consent to having my de-identified data made publicly available.*

**Sample IRB Application Language**

Here is sample language to be considered and added under **Section IV: Confidentiality Procedures & Participant Privacy** of the IRB application:

*De-identified individual-level data, with all direct and indirect identifiers removed, may be used for future research projects, potentially focusing on topics unrelated to this study. This de-identified data may be shared publicly via the Internet and uploaded to an open-access database. No personally identifiable information will be included, ensuring that the shared data cannot be linked back to participants.*

*The de-identified data will be uploaded to the [REPOSITORY NAME] data repository. Additional information regarding the repository and the data-sharing process is available at [FAQ/ABOUT PAGE URL].*

*Participants will have the option to opt out of having their de-identified data made publicly available while still participating in the study. In such cases, their data will only be accessible to the research team and will not be included in the publicly accessible dataset.*

**Navigating IRB Processes**

1. **Engage with the IRB Early:** If you plan to share data that were not initially intended to be open, communicate with the IRB as early as possible to address any necessary modifications or approvals. This is especially important for data that may involve sensitive information or indirect identifiers.
2. **Ensure Consistency Across Documentation:** Make sure that your IRB application, consent forms, and any relevant data-sharing agreements clearly reflect the possibility of sharing data openly. This includes outlining your de-identification procedures and addressing potential risks associated with indirect identifiers in the consent form, if applicable.
3. **Repository Compliance:** If you plan to use a specific data repository, ensure it complies with the institution's data-sharing policies and IRB requirements. Provide a link to the repository’s privacy and security protocols to demonstrate compliance with data protection standards.
4. **Data Use Agreements:** Consider the use of Data Use Agreements (DUAs) to set clear terms regarding who can access the data, how the data will be used, and any restrictions on further sharing. If applicable, outline the terms of the DUA in your IRB submission.
5. **Coordinate with Other Departments:** Recognize that your data-sharing proposal may also require review and approval from other institutional departments, such as Teachers College Information Technology (TCIT), the Office of General Counsel (OGC), or other compliance offices. Engage these stakeholders early to prevent delays in your approval process.
6. **De-identification of Data:** Prior to sharing, ensure that all direct and indirect identifiers that could link data back to individual participants are removed. Include a detailed description of your de-identification process in your protocol to demonstrate compliance with privacy standards.

**Other Considerations**

While open science promotes transparency and data accessibility, it is important to acknowledge the financial and logistical challenges it may present, particularly for researchers with limited funding. Researchers concerned about the costs of data-sharing repositories should consider cost-effective or free options such as the Inter-university Consortium for Political and Social Research (ICPSR), Open Science Framework (OSF), or Qualitative Data Repository (QDR). These platforms provide no-cost or low-cost data-sharing services that comply with open science principles, making them accessible even for unfunded projects.

Additionally, before depositing data into any external repositories, researchers should consult with TCIT and the OGC to ensure compliance with institutional policies on data security and privacy. This step helps to avoid any potential conflicts with institutional data management requirements.

**Conclusion**

Open data is a cornerstone of the open science movement, aiming for greater transparency in scientific research. Researchers should collaborate with TC IRB to navigate the complexities of data sharing while upholding the privacy and confidentiality of research participants. By following these guidelines, researchers can contribute to the advancement of science ethically and responsibly.

**Resources**

* [TC IRB Data Security Plan](https://www.tc.columbia.edu/media/administration/institutional-review-board-/guide-amp-resources---documents/Data-Security-Plan.pdf)
* [TC IRB Informed Consent Form Template](https://www.tc.columbia.edu/media/administration/institutional-review-board-/irb-submission-templates---documents/Informed-Consent-Template_2020_TC-IRB.docx)
* [TC IRB Data Sharing Form Template](https://www.tc.columbia.edu/media/administration/institutional-review-board-/irb-submission-templates---documents/Data-Sharing-Form-Template.docx)
* [TC IRB Data Release Form Template](https://www.tc.columbia.edu/media/administration/institutional-review-board-/irb-submission-templates---documents/Data-Release-Consent-Form-Template.docx)
* [Final NIH Policy for Data Management and Sharing](https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html)
* [NSF's Preparing Your Data Management and Sharing Plan](https://new.nsf.gov/funding/data-management-plan#nsfs-data-sharing-policy-1c8)
* [ICPSR (Inter-university Consortium for Political and Social Research)](https://www.icpsr.umich.edu/web/pages/index.html)
* [OSF (Open Science Framework)](https://osf.io/)
* [QDR (Qualitative Data Repository)](https://qdr.syr.edu/)