

September 28, 2021

Introduction

The Inter-university Consortium for Political and Social Research ([ICPSR](#)) is pleased to provide this response to the National Institute of Health's [Request for Information](#) on recommended language for consent statements to participate in a study. ICPSR collects, curates, and disseminates data covering a broad spectrum of disciplines, including political science, sociology, demography, economics, education, psychology, criminology, gerontology, public health, public policy, and more. ICPSR data form the foundation for tens of thousands of research articles, reports, and books that advance science.

The collection includes over 11,000 studies comprising over 75,000 datasets and 250,000 discrete files. In addition to distributing data supported by its members, ICPSR partners with more than 20 government agencies and foundations, including the Bureau of Justice Statistics, the National Institute on Aging, the Eunice Kennedy Shriver National Institute of Child Health and Development, the Bill & Melinda Gates Foundation, and the Robert Wood Johnson Foundation, to disseminate their data, thus fulfilling mandates to make data publicly available. The special collections funded by agencies and foundations provide researchers and practitioners in these fields with key data resources as well as customized tools to support their work.

Recognizing that the increasing complexity of data requires new levels of support and guidance in their proper use, ICPSR has developed expertise in facilitating access to and secondary analysis of complex data – e.g., longitudinal data, linked data, geospatial data, video data, and data containing biomarkers. The organization distributes enhanced data and documentation to researchers in forms that facilitate reuse of complex data, and offers training and consultation.

ICPSR currently provides [advice to researchers on informed consent language](#). About ten percent of ICPSR's collection is restricted to protect the privacy of study participants, and ICPSR provides multiple, tiered modes of access to that restricted data.

ICPSR COUNCIL

Lisa Cook, Chair • Michigan State University

Dave Armstrong
Western University

Bobray Bordelon
Princeton University

Jon E. Cawthorne
Wayne State University

James Doiron
University of Alberta

Kristin R. Eschenfelder
University of Wisconsin-Madison

Mark Hansen
Columbia University

Michael Jones-Correa, Past Chair
University of Pennsylvania

Trevon Logan
Ohio State University

Lindsey Malcom-Piqueux
California Institute of Technology

Ken Smith
University of Utah

Katherine Wallman
United States Office of Management and Budget

Esther Wilder
Lehman College, City University of New York

1. Utility and useability of this resource

ICPSR applauds the work of the National Institutes of Health in promoting responsible management and sharing of data. These template consent statements will make health-related data collections more responsive to the needs and desires of study participants and promote the secure and efficient use and re-use of valuable scientific data. We will encourage our members and research community to consult and follow this guidance. This resource will assist researchers, study participants, and Institutional Review Boards in understanding how to respect study participants while generating the most scientific value from study data.

We strongly support the dissemination of NIH consent templates that include explicit language permitting responsible, confidentiality-protecting data sharing, and explain to researchers that this will help them to comply with NIH's policy on data sharing (<https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html> and https://grants.nih.gov/grants/policy/data_sharing/).

ICPSR agrees with the recommendation that this language be voluntary. We recommend that NIH articulate the value of using standardized consent statement language to decrease the temptation on the part of researchers, IRBs, and university counsel to introduce variation. Variation of consent language from recommended templates can introduce uncertainty. It also makes it more difficult to automate the control of data access in a way that is consistent with the terms of the consent statement. Standardized, modular consent statement language lends itself to automation because it can be made machine-actionable. The proposed standard components with alternate text facilitate the creation of machine actionable metadata and are to be commended -- and utilized.

ICPSR is building a Research Document Registry, as part of its Research Data Ecosystem, that will allow researchers to construct a modular consent statement for use in data collection. The consent statement will then become part of the metadata for the resulting data. Standardized and modular consent statements can become machine-actionable metadata that will embed the terms of the consent statement into automated access controls to the data. ICPSR encourages the standardization of consent statements, especially language governing secondary access to research data. Standardization will improve the reliability, security, and efficiency of data sharing and confidentiality protection.

2. Gaps or additional components that should be included

Licenses for use of scientific data for commercial and other purposes should be explicitly addressed in this guidance. The guidance should include examples of different types of permissible research use of the study data. While the consent statement should not refer to particular license types, the language in the consent statement should be consistent with the specific license(s) under which data are expected to be made available. Licenses clarify how others are allowed to reuse your data, including whether the data may be redistributed, who may use the redistributed data, and whether users must provide attribution. “Open” licenses are available that maximize the freedom to access, use, modify, and share data, including for commercial use (e.g., the Creative Commons Attribution 4.0 International license; see also <https://opendatacommons.org/>). NIH guidance should provide researchers and study participants with options to exclude use by commercial entities. Licenses may be customized to protect the privacy of individuals that have provided data. Most World Bank microdata, for example, are distributed under a research license that limits use "solely for generating, and perhaps reporting, aggregated information and not for investigations into specific individuals or organizations. No reverse engineering or 'upstream research' into the sources of the data itself will be permitted" (<https://datacatalog.worldbank.org/public-licenses>).

Data Management Plans: NIH guidance should include a recommendation that consent statements be informed by the research project’s data management plan. The language regarding data sharing in the consent statement should be consistent with the data management plan. Any implied changes to the data management plan arising from the consent statement (e.g., a promise to destroy data that was not in the approved DMP when the award was made) should be submitted for approval to NIH.

Restricted access: NIH guidance should make researchers cognizant of best practices in providing access to confidential data. Understanding the options for confidentiality-protecting modes of accessing data may change the wording of (and simplify!) consent statements. For example, if the goal is to make it possible to discontinue sharing the data of an individual participant who has withdrawn consent for sharing (as may be required under GDPR), having a trusted third party maintain control of the data, and provide researchers with access but not possession, will simplify the consent statement. There is no need, in this circumstance, to qualify the participant’s ability to rescind consent. The consent statement can just assure participants that they may change their mind at a later date.

Virtual Data Enclaves hosted by a trusted third party are a useful technology for maintaining control of data. We do *not recommend* talking about VDEs in the consent statement, as that might overly constrain the mechanisms for data sharing as technology changes. For example, a VDE may be the most effective way to provide access to a dataset now, but in the future secure multiparty computing might replace VDEs for some kinds of restricted data research. However, understanding the options for sharing data securely will allow researchers to confidently promise confidentiality protection, including the option of future opt out, without unduly restricting future access and scientific use of the data.

3. Specific language proposed in the informed consent sample language

Component 1: Introduction - Description

Draft Language:

However, research could also

Proposed Revision:

*However, research could also be about unrelated diseases, conditions, or other aspects of health **and well-being**. ...Our goal is to make more research possible to learn about health, **well-being**, and disease.*

Justification:

We recommend broadening the language, so as not to preclude research on related topics that may not be unambiguously health, such as education, happiness, labor force participation, etc. or that may use health status as the independent variable rather than the dependent variable.

Draft Language:

We plan to ...

Proposed Revision:

*We plan to **preserve** your data and biospecimens for [indicate “in perpetuity” or until “used completely,” or a specific time frame, if necessary].*

Justification:

Researchers should be discouraged from promising to destroy data. If such a promise is included in the consent statement, the researcher should be expected

to explain the reasons for the planned destruction in their data management plan.

Draft Language:
Your name and ...

Proposed Revision:

*Your name **or personal identifiers** will not be on any data and biospecimens you provide. The data and biospecimens will have a code that links to your **name or personal identifiers** ... Your name **or personal identifiers** will not be on any data and biospecimens you provide. Investigators cannot link **your name or personal identifiers** to the data and biospecimens.*

Justification:

We recommend “personal identifiers” rather than “identifying information” as “identifying information” could be interpreted to mean a set of variables that would allow a researcher to re-identify data that has been “de-identified.”

Draft Language:
The code key can ...

Proposed Revision:

*The code key can only be accessed by people who have permission **and who have made substantial, enforceable commitments to protect the privacy of your data [and who will be engaged in work to enhance the data to advance scientific research].***

Justification:

This language strengthens the commitment to protect the data and helps the study participant understand how their confidential information will be protected. It also provides an option to explicitly restrict use of the code key to advance scientific research. This provides additional assurance that it will not be used inappropriately (e.g., for marketing to the study participants, for evaluating their eligibility for services or insurance, etc.).

Draft Language:
[Option #2: If the data ...

Proposed Revision:

[Option #2: If the data and biospecimens are completely delinked from identifiers and cannot be linked back to the participant, with justification provided in study's data management plan]

Justification:

Researchers should be discouraged from making data unlinkable if future linkages, e.g., to the National Death Index, would be of potential value to future scientists. If a statement is included in the consent statement that the data will not be linkable (option 2), the researcher should be expected to explain the reasons for that in their data management plan.

Component 2: Voluntary participation

Draft Language:

[Option #1: When sharing ...

Proposed Revision:

[Option #1a: When sharing of data and biospecimens will be optional (e.g., for studies that have potential benefit)]

It is your choice whether or not to let researchers share your data and biospecimens for future scientific research. If you say "yes," you can change your mind later; in that case, your data and biospecimens might still be analyzed if they have previously been shared. If you say "no," you can still fully participate in this study. Please initial next to your choice:

_____ YES, use my data and biospecimens in future scientific research

_____ NO, do NOT use my data and biospecimens in future scientific research

[Option #1b: When sharing of data and biospecimens will be optional (e.g., for studies that have potential benefit)]

It is your choice whether or not to let researchers share your data and biospecimens for future scientific research. If you say "yes," you can change your mind later and we will remove your data and biospecimens from the research data sets. If you say "no," you can still fully participate in this study. Please initial next to your choice:

_____ YES, use my data and biospecimens in future scientific research

_____ NO, do NOT use my data and biospecimens in future scientific research

Justification:

The language in the question should match the language in answers that the participant can select. We recommend using the term “future scientific research” for both. We believe that that is more likely to ensure that potential participants understand the benefits of data sharing and incorporate that into their decision-making.

Option 1b should be used when participants have the option to withdraw consent for data sharing at a later date and the research team is “sharing” data by providing access to the data, for example in a restricted data enclave, but not giving future researchers possession of the data. In that case, there is no need to qualify the research team’s ability to withdraw the data for an individual participant. It is straightforward to update the data set that researchers have access to in a restricted data enclave because the researcher does not have possession of the data.

Component 3: Discontinuation/Withdrawal.

Draft Language:

You can change your ...

Proposed Revision:

*You can change your mind about sharing your data and biospecimens at any time. If you change your mind, please contact the study team **at XXXX** to let us know.*

Justification:

Specific, concrete instructions for requesting withdrawal of consent should be included in the statement.

Component 4: Risks & Benefits

Draft Language:

[**Risks**] *When we share ...*

Proposed Revision::

[**Risks**] *When we share your data and biospecimens, there is a small risk that **your data could be associated with your identity.** We will protect your data and biospecimens as much as possible during storage and when they are shared **to prevent any unauthorized access or use that would reveal your identity.***

[**Benefits**] *You will not receive any direct benefit from sharing your data and biospecimens. However, sharing your data and biospecimens may contribute to research **and scientific advance** that helps **you and** others in the future.*

Justification:

The original risk statement suggests two distinct risks: 1) *People may get access to it who are not supposed to* and 2) *your identity could be discovered*. These are related, but they are not the same thing. Presumably what people are worried about is their identity, so we address that risk first, and indicate that preventing access or use that would permit re-identification is something we will do to try to ameliorate this risk.

We have added phrases to the benefits statement that highlight the scientific impact of data sharing and the potential benefits to all of scientific advance.

Component 5: Commercial Application

No changes recommended. See discussion above for the use of consent language consistent with specific licenses under which data may be held.

4. Hurdles or barriers to wider use of this resource by the community

We recommend that NIH provide **guidance that specifically addresses key legislation** or regulations that might raise concerns about sharing data. Researchers will need guidance about specific content to include in consent statements so that, if the data for their research project are covered by a particular statute or regulation, they will still meet the letter or spirit of NIH data sharing and consent policies. We present brief examples of such guidance for FERPA and HIPAA. Examples of other legislation that may require guidance are listed below.

FERPA: Under the Family Educational Rights and Privacy Act (FERPA), disclosures of data must take place within a data sharing agreement with the educational institution, so the researcher will need to execute such an agreement separately from the consent statement. In addition:

1. FERPA assumes that data will be destroyed when no longer needed for the specified purpose. Retaining the data beyond that point requires explicit consent to do so.

2. Consents must include explicit permission to disclose to any third parties, including researchers conducting secondary analysis.
3. FERPA does not allow uses of data beyond the specified purpose, so the stated purpose in the consent should be written to encompass all intended use of the data, including secondary analysis of the data.

HIPAA: If individual consent is required, it will have to include the elements of a HIPAA authorization (summarized below) in addition to the usual elements of an informed consent. This can be included in a single document. The purpose should be described broadly enough to permit meaningful reproducibility and secondary analyses of data.

1. Give a specific description of the personal health information (PHI) involved and the purpose of each requested use or disclosure.
2. List the people who may execute the use or disclosure of the PHI (this may be individual names or categories of people).
3. List the people to whom they may disclose the PHI, and the purpose. (For example, a study sponsor.)
4. If the PHI may be re-disclosed to people or organizations which aren't governed by HIPAA, explain who and why.
5. Include an expiration date for the authorization (though this can be indefinite, or can be an event such as "the end of the research study.").
6. Include an explanation of the consequences of failing to sign the authorization (for example, that the individual will not be able to receive treatment as part of the research project).
7. Include the individual's right to revoke their authorization and how to do it (this may be separate from revocation of consent).
8. Signature of the individual and date.

NIH should provide guidance that outlines the circumstances under which a researcher may obtain data covered by a regulation *without* going through the consent process. For example, under FERPA, use of educational records for specific categories of research does not require consent. And under HIPAA, studies which only require access to existing records (or review of records preparatory to research) may be approved by an IRB or Privacy Board without requiring individual consent or a HIPAA authorization from the individuals involved.

Additional examples of legislation that may affect the content of consent statements and the sharing of data:

- Family Violence Prevention and Services Act
- Individuals with Disabilities Education Act (IDEA)
- McKinney-Vento Homeless Assistance Act
- Medicaid Act (Title XIX of the Social Security Act)

- The Public Health Service Act
- The Privacy Act of 1974
- Runaway and Homeless Youth Act
- Title IV-A of the Social Security Act, Temporary Assistance for Needy Families (TANF)
- Title IV-D, Federal Parent Locator Service (Social Security Act, Section 453)
- Violence Against Women Reauthorization Act of 2013
- CIPSEA

We recommend that NIH consider the development of **an online tool that works with research data collection and management software** (e.g., RedCap, Blaise, Illume, Qualtrics, Survey Monkey, and MyDataHelps), making it easy for people to export the consent statement into other software to be used for subsequent data collection.

We recommend that NIH link its guidance on consent statements to **ClinicalTrials.gov** so that people are aware of the guidance as soon as they design their trial.

We recommend that NIH support **outreach and training materials for IRBs**, so that they default to this consent language as best practice for responsible scientific research.

We recommend that NIH encourage standardization of the consent statement to allow consents to be mapped to **machine-actionable metadata**. That would allow data collection software to appropriately tag individual observations so that they would be protected and managed in a way that is consistent with the consent selected by the study and the study participant. If someone checks no, the data collection software could automatically firewall that participant's data. These benefits require three additional steps beyond the recommendations themselves 1. Widespread use of standardized language, 2. Standardization of metadata that captures this language and its inclusion as part of research datasets, and 3. Development of software tools that make this metadata useful for research data management. Creation of standardized metadata and relevant software tools would encourage widespread use of the standardized language recommended here.