DATA MANAGEMENT & SHARING	SOP 5.1.05

1. Purpose & Policy

The purpose of this SOP is to describe the management and sharing requirements for data derived from human participants.

2. General Information

Our federal sponsors, particularly the National Science Foundation (NSF) and the National Institutes of Health (NIH) have had requirements for dissemination of research results for some time. In general, investigators are expected to share data with other researchers and are also expected to provide a Data Management Plan (DMP) or Data Management Sharing Plan (DMSP). Specific sponsor requirements can be found here:

NSF: https://www.nsf.gov/bfa/dias/policy/dmp.jsp

NIH: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html; https://sharing.nih.gov/data-management-and-sharing-policy/planning-and-budgeting-DMS/writing-a-data-management-and-sharing-plan#after

The Caltech Library has created a number of Caltech-specific DMSP resources that are available at https://library.caltech.edu/publish/data/data-management-plans. Resources include standard language, a DMSP checklist and an example DMP/DMSP.

Once an award is made and the DMP/DMSP approved by the funding NIH Institute, Center, or Office, compliance with that approved plan will be a term and condition of award. It is important that researchers ensure that their IRB Protocol and Informed Consent are consistent with the DMP/DMSP.

3. Training Requirements

There are no specific training requirements associated with data management and sharing; however, investigators should follow this guidance.

4. Data Management and Sharing Plans Including Human Participant Data and Corresponding IRB Documentation

A. Creating a Data Management and Sharing Plan

1) As described above, both the Caltech Library and the Federal Agencies (NIH and NSF) have resources and detailed descriptions of what must be included in a data management/sharing plan. NIH provides descriptions for what must be included in proposal applications both before and after January 25, 2023.

2) An NSF example Data Management Plan with customizations for Caltech can be found here: https://doi.org/10.7907/m79a-gw54. There are basically five sections to the example plan: Types of data to be produced during the project; Standards applied to data and metadata format and content; Polices for access and sharing, including protections for privacy, confidentiality, security, intellectual property, or other rights;

Policies and provisions for reuse and redistribution; and Plans for archiving data, preservation and access.

3) A NIH example Data Management and Sharing Plan with customizations for Caltech can be found here: https://doi.org/10.7907/x2g2-v221. It breaks the plan into 6 specific elements: Data Type, Related Tools, Software and/or Code, and Standards, Data Preservation, Access and Associated Timelines, Access, Distribution or Reuse Considerations, Oversight of Data Management and Sharing.

4) Considerations before Completing the DMP/DMSP with Human Participant Data

Federal sponsors encourage sharing of scientific data as soon as possible, and no later than the time of an associated publication or end of the performance period, whichever comes first. Federal sponsors also encourage researchers to make scientific data available for as long as they anticipate it being useful for the larger research community, institutions, and/or the broader public.

In deciding where and how data is made available, investigators should always consider whether access to participant scientific data should be controlled, even if the data is de-identified and lacks explicit limitations on subsequent use.

Investigators who generate scientific data derived from human participants to explain in their DMP/DMSP how the privacy, rights, and confidentiality of human research participants will be protected (i.e., through de-identification, Certificates of Confidentiality, and other protective measures).

Investigators need to remember that collection and or use of previously collected data from participants in a research study may have restrictions on them. In particular, the following categories of participants or areas of consideration should be accounted for in any DMP/DMSP:

- Participation of Indigenous Persons/Tribal Law
- Participation of Minors/Legal Concerns
- Institutional Data
- Use of Human Genomic Data
- Data Received under MTA/DTUA
- Data Received with Use Restrictions
- Data or Participants with Special Legal Considerations
- Data Received Pursuant to Caltech Subawards/contracts
- Data Received Pursuant to a Caltech Prime Award/Collaboration

Data carrying explicit limitations on subsequent use require access controls be implemented and monitored. All such limitations must be included in the DMP/DMSP. Please see the section regarding IRB application, below, for assistance with data from participants in these categories.

B. Relating the Data Management Plan to an IRB Protocol Application and Informed Consent

The IRB Protocol and Informed Consent are specific to a particular project. Generally, when the IRB discusses sharing within the context of that project, it is focused on sharing of data amongst collaborators and not the type of broad data sharing expected of the federal sponsors. However, there have been cases where broad sharing has been requested or required by sponsors as part of the IRB protocol application and the protocol and informed consent have been designed and approved to accommodate this broad sharing.

As can be seen in the draft and example plans, investigators proposing to generate and broadly share scientific data derived from human participants must include a number of factors in the management plan that are also essential components of an IRB protocol application and Informed Consent. The IRB is charged with reviewing the IRB protocol, including project and future data sharing plans to ensure participants are informed of risks and protected to the extent possible.

This SOP provides the elements of the NIH sharing plan and aligns them with the NSF elements and the IRB protocol application and informed consent elements:

ALIGNMENT DATA SHARING PLAN AND IRB APPLICATION/INFORMED CONSENT

Section	NIH	NSF	IRB Application
Data Type	Data type: Summarize the types and estimated volume of scientific data expected to be generated during the project. Identify which data will be preserved and shared, including the rationale for this decision. Identify any metadata or other documentation that will be made available to facilitate the interpretation of the scientific data.	Types of data to be produced during the project	**Protocol Section: Samples & Data and Caltech and Partner Orgs** It is essential that the investigator include the data description in their IRB protocol and identify all human participant or protocol data that will ultimately be broadly shared.
Related Tools	Related tools, software, code: Describe any tools and software that might be needed to access and manipulate data.		This element is not specific to the IRB protocol; however, if particular tools, software or code are necessary to access data, this may be a factor limiting access and potentially protecting participant data.
Standards	Standards: State the common data standards applicable to the data/metadata to be shared. If no consensus standard exist, the researcher must state this in their plan.	Standards applied to data and metadata format and content	This element is not specific to the IRB protocol.
Data Preservation, Access, Timelines	Data preservation, access, timelines: Provide the name of a repository where scientific data/metadata will be archived, as applicable. Describe how the data will be findable and identifiable. Describe when the data will be made available and for how long it will be available.	Plans for archiving data, preservation and access	**Protocol Section: Samples & Data and Caltech and Partner Orgs** **Informed Consent: Future Use/Broad Consent** The repository name/site, the specific participant data that will be deposited, and the time and length of accessibility must be included in the IRB protocol and the Informed Consent.
Access, Distribution, Reuse Considerations	Access, distribution, reuse considerations: The NIH expects investigators to maximize appropriate sharing of data. In this section of their plan, investigators must include any data use limitations affecting such broad sharing.	Polices for access and sharing, including protections for privacy, confidentiality, security, intellectual property, or other rights Policies and provisions for reuse and redistribution	**Protocol Section: Samples & Data and Caltech and Partner Orgs** **Informed Consent: Future Use/Broad Consent** This section is particularly relevant to the IRB protocol and Informed Consent. Investigators should consider whether the participant data were received with any restrictions (Materials Transfer Agreement- MTA or Data Transfer Use Agreement- DTUA, GDPR, PIPL) or whether a pre-existing Informed Consent restricts broad sharing of participant data. If there are restrictions, or if restrictions must be imposed to protect the privacy, rights and/or confidentiality of the human participants, these must be described in the IRB protocol and the Informed Consent. **Relevant Contacts: IRB, OTTCP, OSR, Procurement**

Oversight	Oversight of data management: This is the compliance aspect of the management plan. The investigator will be responsible for this.		The investigator will be responsible for this. **It is essential that an approved IRB protocol and Informed Consent be in place before participant data is collected or shared. The researcher must SHARE and/or protect the data consistent with the DMP/DMSP, the IRB Protocol, and the Informed Consent.**

C. Completing the IRB Protocol Application and Informed Consent Document

1) Considerations before Completing the IRB Application

Federal sponsors encourage sharing of scientific data as soon as possible, and no later than the time of an associated publication or end of the performance period, whichever comes first. Federal sponsors also encourage researchers to make scientific data available for as long as they anticipate it being useful for the larger research community, institutions, and/or the broader public.

In deciding where and how data is made available, investigators should always consider whether access to participant scientific data should be controlled, even if the data is de-identified and lacks explicit limitations on subsequent use.

Investigators who generate scientific data derived from human participants to explain in their DMP/DMSP how the privacy, rights, and confidentiality of human research participants will be protected (i.e., through de-identification, Certificates of Confidentiality, and other protective measures).

Investigators need to remember that federal, Tribal, or state laws, regulations, or policies, or existing or anticipated agreements may impose restrictions upon distribution of data.

- Indigenous Persons/Tribal Law: If participant research involves American Indian, Alaska Native, or Indigenous populations, investigators must secure appropriate agreements with tribal authorities before using and sharing that data. Please contact irb@caltech.edu for assistance.

- Minors/Legal Concerns: If participant research involves minor populations, researchers must secure appropriate consent before using and sharing that data. Please contact irb@caltech.edu for assistance.

- Institutional Data: If participant research includes data collected at Caltech from a student, staff, or Faculty member related to their employment or status as a student, the data may be restricted from disclosure. See SOP: Institutional Data.

- Human Genomic Data: If participant research data includes human genomic data: Investigators are expected to submit the data to a repository acceptable under the Genomic Data Sharing Policy. See Where to Submit Genomic Data. Human genomic data is expected to be shared according to NIH's Data Submission and Release Expectations, but no later than the end of the performance period, whichever comes first. Please contact irb@caltech.edu for assistance.

- GDPR/PIPL: Persons from the European Union (EU), European Economic Area (EEA), the United Kingdom (UK) or the People's Republic of China have rights to access, correct, or delete their personal data, however these rights may be limited in certain circumstances. The individuals may restrict the types of activities the research team can do with their data or object to the use of their

data. Importantly, such persons can withdraw their consent to use their personal data, however, if the data are anonymized (i.e., see IRB SOP: Anonymous, De-identified and Coded Data for definition of anonymized), such data can be broadly shared and withdraw of consent will not be an option.

- MTA/DTUA: If participant research data are received under a Materials Transfer Agreement or a Data Transfer Use Agreement, the conditions of the MTA or DTUA must be reviewed and a copy of the MTA/DTUA provided with the IRB application. Please also provide a copy of any Informed Consent used to collect the data in question. Please contact MTA@caltech.edu and/or irb@caltech.edu and/or security@caltech.edu for assistance.

- Use Restrictions: If participant research data require any other type of use agreement (e.g. agreements to secure or store data in a particular way), please provide a copy of the agreement and a copy of any Informed Consent used to collect the data in question. Contact irb@caltech.edu and/or security@caltech.edu for assistance.

- Legal Considerations: If collection or use of participant research data are regulated pursuant to a federal, state or local law, please note this in your IRB application and contact irb@caltech.edu for assistance.

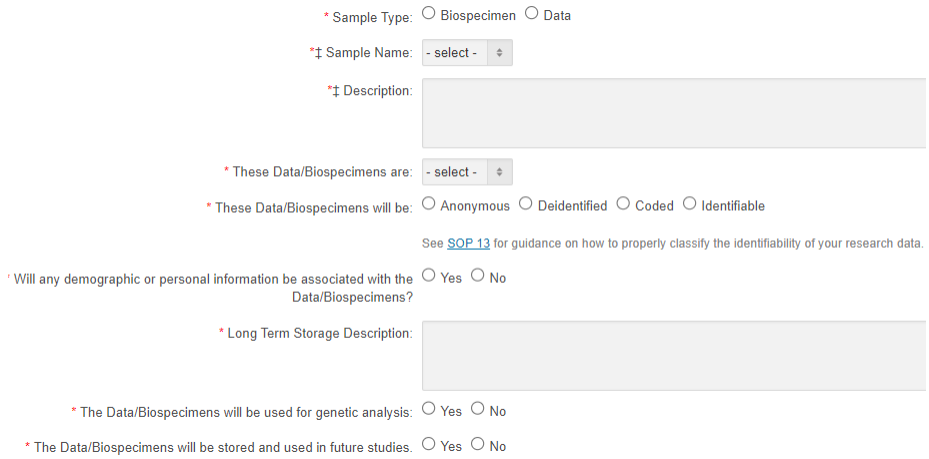
- Caltech Subawards/contracts: If participant research data are provided to Caltech pursuant to subaward or subcontract, please provide the subaward number in your IRB application and a copy of any Informed Consent used to collect the data in question. You may contact irb@caltech.edu for assistance.

- Caltech Prime Award/Collaboration: If participant research data are provided to Caltech pursuant to a specific sponsored award or collaboration agreement, please provide the award number in your IRB application and a copy of any Informed Consent used to collect the data in question. You may contact irb@caltech.edu for assistance.

Data carrying explicit limitations on subsequent use (federal, Tribal or state laws, MTAs, DTUAs, restrictive informed consent) require access controls be implemented and monitored. All such limitations must be included in the IRB protocol and in the Informed Consent, as necessary.

2) Information to include in IRB protocol, and Informed Consent document, and where to find the information in the IRB Protocol application:

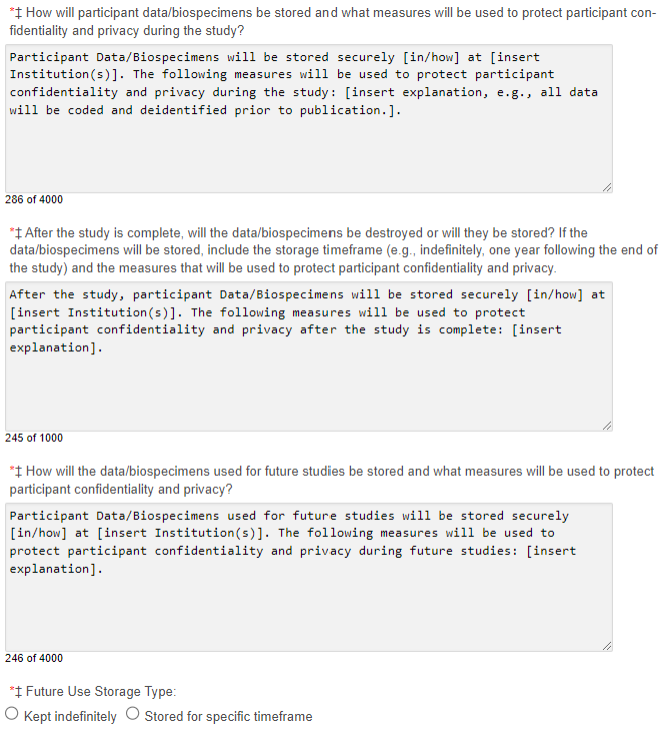
   a. Data Type
   Protocol section: Samples & Data and Protocol Information



b. Data Preservation, Access, Timelines & Access, Distribution, Reuse Considerations

Protocol section: Samples & Data and Protocol Information

Summarize how the data will be stored and the measures used to ensure participant confidentiality. Will the data be anonymous, de-identified, or coded?
Protocol section: Samples & Data

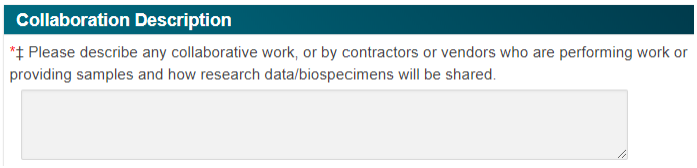


See IRB SOP: Anonymous, De-identified and Coded Data for additional guidelines on classifying the identifiability of data.

1. Summarize how the participant's data is expected to be used.
   Protocol section: This information can be included in multiple areas of the protocol. Protocol Information, Samples & Data

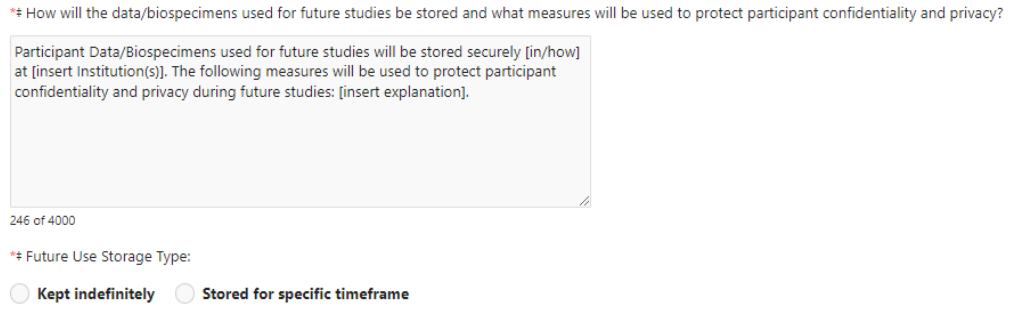
2. Summarize how the participant's data is expected to be shared. Use the Collaboration Section to identify any repositories or broad sharing in addition to collaboration for this project (if any). Describe any data sharing (either for this project and for future long term sharing pursuant to a DMP or DMSP) including who will be doing it, and how the research data will be shared. Include the names of collaborators (if any) and the repositories or data sharing websites used.

   Protocol section: Caltech and Partner Orgs



3. Will the participant's data be used in future studies? Limitations on subsequent use of data should be included. Factors that may affect subsequent use of data should be properly communicated to the participant.

   Protocol section: Samples & Data



<u>Informed Consent Language for Future Use</u>

This study will require the storage and maintenance of your anonymized data or biospecimen for secondary research that may be different from this study or for non-research purposes. Your data or biospecimens may be shared with other investigators, research, academic or medical institutions, and/or placed into repositories. Your anonymized data and/or biospecimens may be stored for [TIME FRAME OR INDEFINITELY]. You will not be informed about the details of any specific research studies that might be conducted. This includes not knowing the purpose of the research. You will also be consenting to a research study that you may not otherwise have chosen to consent to. No clinically relevant research results will be disclosed to you. If you have any questions regarding your rights regarding storage and use of your data and/or specimen in the future, please contact [PI].

Although the researchers will take every reasonable precaution to keep any identifiable data we collect from you confidential, they cannot guarantee this. If you

have concerns about the confidential use of the data the researchers plan to collect from you, you should not participate in this study.

The future use of your data and/or biospecimen in research may result in new products, tests, or discoveries which may have potential commercial value. You do not retain any property rights to these materials. As such, you would not share in any financial benefits from these products, tests, or discoveries.

____ I consent to my data and/or biospecimens being saved and used for future research.

By signing this form, you allow the study researchers to make your data or information available to the Caltech Institutional Review Board (IRB), Institutional Official, Caltech Administration, any Collaborator IRB, and regulatory, law enforcement or funding agencies as required by law.

5. Sharing the Data

The implementation of the DMP/DMSP must be consistent with the plan, the IRB protocol and the Informed Consent. The IRB is available to assist, in coordination with the Library, IMSS, and the Office or Research Administration to ensure that the requirements are clear and can be implemented and maintained by the researcher.

Steps for Sharing the Data

1. DMP/DMSP
2. IRB Protocol and Informed Consent; Including MTA, DTUA or any other agreement controlling distribution of the data
3. De-identify data pursuant to #1 and #2.
4. Remove any metadata from the data set that is not directly relevant to the study or that may divulge information not intended (e.g., photographs that may contain GPS coordinates, type of device taking the picture with details about its settings, size of image, data/time picture taken, author or collaborator identities, dates of document creation/modification, and sound file metadata that might include creator's name or other information about the file).
5. Share data pursuant to #1 and #2

6. Contacts

For questions regarding…	**Contact**
➢ IRB Protocols and Informed Consent	IRB irb@caltech.edu or 626-395-8448
➢ Creating a DMP or DMSP ➢ Data management guidance ➢ Choosing a data repository ➢ Publishing and preserving data	Caltech Library Services data@library.caltech.edu or 626-395-3405
➢ Proposal submission	Office of Research Administration

➢ Budgeting and allowable costs	https://researchadministration.caltech.edu or 626-395-2882
➢ Data security ➢ Best practices	IMSS security@caltech.edu or 626-395-4631
➢ GDPR / PIPL	Director of Compliance tye.welch@caltech.edu or 626-395-8633