CIRM GUIDANCE FOR COMPLETING A DATA SHARING AND MANAGEMENT PLAN (DSMP)

A template for creating a DSMP is provided as part of an Application in the CIRM Grants Management System

This guidance outlines the elements to be addressed in a Data Sharing and Management Plan (DSMP) within four (4) pages or less. A DSMP should reflect the proposed approach to data management and sharing at the time it is prepared and be updated during the course of the award/support period to reflect any changes in the management and sharing of scientific data (e.g., new scientific direction, new repository option, timeline revision). CIRM will review plans with a Data Advisory with the goal of arriving to the best DSMP which can be appropriately incorporated into project milestones during pre-funding administrative review (PFAR). For some programs and data types, CIRM has developed specific data sharing expectations (e.g., scientific data to share, relevant standards, repository selection, timelines) that apply and should be reflected in a DSMP. When no additional CIRM data sharing expectations apply, researchers should propose their own approaches to data sharing and management in a DSMP. CIRM requires data management and sharing practices to be consistent with the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles and reflective of practices within specific research communities.

A. DATA TYPE

a) A general summary of the types and estimated amount of scientific data to be generated and/or used in the research.

Describe data in general terms that address the type and amount/size of scientific data expected to be collected and used in the project (e.g., 256-channel EEG data and fMRI images from ~50 research participants). Descriptions may indicate the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing that has occurred (i.e., how raw or processed the data will be).

***Guidance*:**

CIRM's regulations with CIRM-Funded Technology define "Data" as: Scientific, clinical, or technical recorded information derived during the Project Period of an Award, regardless of form or the media on which it may be recorded, but not any of the following: financial, administrative, management data, other information incidental to contract administration, preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. "Data" excludes physical objects (e.g., laboratory samples).

Even those scientific data not used to support a publication are considered scientific data and within the DSMP guidelines scope. We understand that a lack of publication does not necessarily mean that the findings are null or negative; however, indicating that scientific data are defined independent of publication is sufficient to cover data underlying null or negative findings.

Additional Guidance:

Research projects vary widely in the types of data produced. In this section, you will describe the categories, amounts, and degree of processing of your data.

***Example Answer*:**

This project will produce _____[Data type, e.g., imaging, sequencing, experimental measurements] data generated/obtained from _____[e.g., instrument, method, survey, experiment, data repository]. Data will be collected from ____[number] research participants/specimens/experiments, generating ____[number] datasets totaling approximately ____[amount of data] in size. The following data files will be used or produced in the course of the project: _[list input data files, intermediate files, and final, post-processed files]. Raw data will be transformed by _____[analysis, method] and the subsequent processed dataset used for statistical analysis. To protect research participant identities, _____[e.g., individual, aggregated, summarized] data will be made available for sharing.

b) A description of which scientific data from the project will be preserved and shared.

***Guidance*:**

CIRM does not anticipate that researchers will preserve and share all scientific data generated in a study. Researchers should decide which scientific data to preserve and share based on ethical, legal, and technical factors that may affect the extent to which scientific data are preserved and shared. Provide the rationale for these decisions.

***Example Answer*:**

Based on _____[ethical, legal, technical] considerations, the following data produced in the course of the project will be preserved and shared: _____[list] **OR** All data produced in the course of the project will be preserved and shared.

c) A brief listing of the metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) that will be made accessible to facilitate interpretation of the scientific data.

***Example Answer*:**

To facilitate interpretation of the data, _____[e.g., metadata*, documentation, study protocols, data collection instruments] will be shared and associated with the relevant datasets.

*data that provide additional information intended to make scientific data interpretable and reusable.

B. RELATED TOOLS, SOFTWARE AND/OR CODE

An indication of whether specialized tools are needed to access or manipulate shared scientific data to support replication or reuse, and name(s) of the needed tool(s) and software.

***Guidance*:**

The file formats in which data are saved in a digital format can be divided into two general categories:

- Proprietary - The specification of the data encoding format is not released or restricted in some way. Proprietary formats can only be easily opened and manipulated by particular software tools.
- Open - The specification of the data encoding format which can be used and implemented by anyone. Open formats can often be easily opened and manipulated by a large number of software tools.

***Example Answer*:**

- ***If no specialized tools are needed to access or manipulate the data:***

  _____[Data type - Imaging data, survey data, etc] data will be made available in _____[csv, txt, dicom, etc] format and will not require the use of specialized tools to be accessed or manipulated.

- ***If specialized tools are needed to access or manipulate the data:***

  _____[Data type] data will be made available in _____format, which requires the use of specialized tools, such as _____[include list of tools] to be accessed and manipulated.

Additional Guidance:

If applicable, specify how needed tools can be accessed, (e.g., open source and freely available, generally available for a fee in the marketplace, available only from the research team) and, if known, whether such tools are likely to remain available for as long as the scientific data remain available.

***Example Answer*:**

- The _____tool, which can be used to _____is available free of charge through _____[source name]
- The _____tool, which can be used to _____is available for a fee of _____through _____[source name].
- Custom tools to _____will be/have been developed by the research team.
- Requests for these tools should be directed to _____[include details of members of the research team].

- These tools will be shared openly via ____.

C. STANDARDS

An indication of what standards will be applied to the scientific data and associated metadata (i.e., data formats, data dictionaries, data identifiers, definitions, unique identifiers, and other data documentation).

***Guidance*:**

While many scientific fields have developed and adopted common data standards, others have not. In such cases, the DSMP may indicate that no consensus data standards exist for the scientific data and metadata to be generated, preserved, and shared.

Additional Guidance:

A *standard* specifies how exactly data and related materials should be stored, organized, and described. In the context of research data, the term typically refers to the use of specific and well-defined formats, schemas, vocabularies, and ontologies in the description and organization of data. However, for researchers within a community where more formal standards have not been well established, it can also be interpreted more broadly to refer to the adoption of the same (or similar) data management-related activities or strategies by different researchers and across different projects.

It is possible that your work will employ multiple formal standards or a mix of formal standards and other data management strategies. You should be as specific as possible when describing the standards used for each type of data included in your proposal.

***Example Answer*:**

To facilitate their efficient use, all of our data and materials will be structured and described using the following standards:

- ***If there are formal data standards for some/all of the data:***

  Whenever possible, we will use _____[common data elements, standardized survey instruments, etc] to structure and organize our data.

  Our _____data will be structured and described using the _____standard, which has been widely adopted in the _____community. [Add additional information about this standard, if applicable - e.g. implementation in data repositories, utility in combining/reusing datasets]

- ***If there are not formal standards:***

  Formal standards for _____data have not yet been widely adopted. However, our data and other materials will be structured and described according to best practices.

  Data will be stored in common and open formats, such as _____for our _____data. Information needed to make use of this data [e.g. the meaning of variable names, codes, information about missing data, other metadata etc] will be recorded in _____[data dictionaries/codebooks] that will be accessible to the research team and will subsequently be shared alongside final datasets.

  Information about our research process, including the details of our analysis pipeline will be maintained contemporaneously, using _____[lab notebooks, protocols, etc]. This information will be accessible to all members of the research team and will be shared alongside our data.

D. DATA PRESERVATION, ACCESS, AND ASSOCIATED TIMELINES

This section of the DSMP provides an outline of the plans and timelines for data preservation and access, including:

a) **The name of the repository(ies) where scientific data and metadata arising from the project will be archived.**

***Guidance*:**

CIRM has provided additional information to assist in selecting suitable repositories for scientific data resulting from CIRM funded research: find link to resources here.

Selecting a Data Repository

1. In some cases, CIRM will identify particular data repositories (or sets of repositories) to be used to preserve and share specific types of data. These requirements will be described in the applicable Program Announcement (PA) or Request for Applications (RFA). For data subject to such requirements, researchers must use the designated data repository(ies).
2. For data generated from research for which no data repository is specified by CIRM (as described above), researchers are encouraged to select a data repository that is appropriate for the data generated from the research project and is in accordance with the desired characteristics, taking into consideration the following guidance:
    i. Primary consideration should be given to data repositories that are discipline or data-type specific to support effective data discovery and reuse (see CIRM resource).
    ii. If no appropriate discipline or data-type specific repository is available, researchers should consider a variety of other potentially suitable data sharing options:
        - Small datasets (up to 2 GB in size) may be included as supplementary material to accompany articles submitted to PubMed Central (see PMC policies).
        - Data repositories, including generalist repositories (see CIRM resource) or institutional repositories, that make data available to the larger research community, institutions, or the broader public.
        - Large datasets may benefit from cloud-based data repositories for data access, preservation, and sharing.

***Example Answer*:**

All dataset(s) that can be shared will be deposited in _____[Add appropriate data repositories] OR _____ [Add appropriate subject or disease repositories]

b) How the scientific data will be findable and identifiable

Guidance:

Via a persistent unique identifier or other standard indexing tools.

***Example Answer*:**

The _____[Insert repository name] provides metadata, persistent identifiers (i.e., insert whether DOI, handles, other), and long-term access. This repository is supported by _____[Insert funder/organization] and dataset(s) are available under a _____[Insert license information] **OR** through a request process _____[Insert information about request process].

c) **When the scientific data will be made available to other users** (i.e., the larger research community, institutions, and/or the broader public) **and for how long.**

***Guidance*:**

CIRM requires scientific data be shared as soon as possible, and no later than time of an associated publication or end of the project period, whichever comes first. Researchers are required to consider relevant requirements and expectations (e.g., data repository or platform policies, award record retention requirements, journal policies) as guidance for the minimum time frame scientific data should be made available. CIRM requires 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. Identify any differences in timelines for different subsets of scientific data to be shared.

***Example Answer*:**

Data will be made available as soon as possible or at the time of associated publication, always before the end of the project period.

Data will be made available in accordance with data repository _____[insert data repository name] policies.

E. ACCESS, DISTRIBUTION, OR REUSE CONSIDERATIONS

a) **Describe any applicable factors affecting subsequent access, distribution, or reuse of scientific data related to:**

- ***Informed consent*** (e.g., disease-specific limitations, particular communities' concerns).
- ***Privacy and confidentiality protections*** (i.e., de-identification, Certificates of Confidentiality, and other protective measures) consistent with applicable federal, Tribal, state, and local laws, regulations, and policies.

***Guidance*:**

CIRM expects that in drafting DSMPs, researchers maximize the appropriate sharing of scientific data generated from CIRM-funded research, consistent with privacy, security, informed consent, and proprietary issues.

Additional Guidance:

Certain kinds of data, especially human subjects' data, require extra preparation before they can be shared to ensure participant privacy. In this section, you will describe your approach to preparing human subjects data for sharing and note any additional restrictions or policies that will impact access to your data.

Whether access to scientific data derived from humans will be controlled (i.e., made available by a data repository only after approval).

- Any restrictions imposed by federal, Tribal, or state laws, regulations, or policies, or existing or anticipated agreements (e.g., with third party funders, with partners, with Health Insurance Portability and Accountability Act (HIPAA) covered entities that provide Protected Health Information under a data use agreement, through licensing limitations attached to materials needed to conduct the research).
- Any other considerations that may limit the extent of data sharing.

b) ***(If applicable)*** **Describe your approach to preparing human subjects data for sharing and note any additional restrictions or policies that will impact access to your data.**

Guidance:

Certain kinds of data, especially human subjects data, require extra preparation before they can be shared to ensure participant privacy. In this section, you will describe your approach to preparing human subjects data for sharing and note any additional restrictions or policies that will impact access to your data. If you are working with human subjects you should also describe how you will address data management and sharing in your informed consent process. You will also need to describe your methods for ensuring privacy and confidentiality, including how you will de-identify your data. If you have decided that a controlled access repository (where researchers must apply to access data) is a better fit for your data than an open repository, you should describe the repository's access procedures. Finally, if there are any other laws, policies, or existing agreements that impact your ability to share your data they should be described here.

***Example Answer*:**

- ***For researchers working with human subjects data***

  In order to ensure participant consent for data sharing, IRB paperwork and informed consent documents will include language describing plans for data management and sharing data, describing the motivation for sharing, and explaining that personal identifying information will be removed.

  To protect participant privacy and confidentiality, shared data will be de-identified using the _____method. [Describe de-identification method, noting any other applicable laws or policies such as HIPAA].

- ***For researchers selecting controlled access repositories***

  Given the sensitive nature of the dataset, de-identified human subjects data will be made available in _____data repository, which restricts access to the data to qualified investigators with an appropriate research question who sign a data use agreement. [Describe data repository access methods and security measures].

F. OVERSIGHT OF DATA MANAGEMENT AND SHARING

Indicate how compliance with the DSMP will be monitored and managed, frequency of oversight, and by whom (e.g., titles, roles, qualifications).

Guidance:

This section should address titles and roles overseeing data management and sharing, within the investigator team or as key personnel.

List the roles responsible for data capture, metadata production, data quality, storage and backup, data archiving, and data sharing. Include the name (if available), title, affiliation, and ORCIDs where possible.

If this is a collaborative project across institutions, explain how data management tasks will be addressed across partners.

Identify which individual (or role) will be responsible for implementing, updating, and revising the DSMP.

Explain how the necessary resources (for example personnel time) to prepare the data for sharing/preservation have been budgeted. Consider and justify any resources needed to adhere to the DSMP. These may include curating data and developing documentation, infrastructure necessary to provide local management and preservation, and data deposit fees.

***Example Answer*:**

The following individuals [or just the position titles if unknown] will be responsible for data collection, management, storage, retention, and dissemination of project data, including updating and revising the DSMP when necessary.

- Name, Position Title, Host Institution, ORCID, email, role in project data sharing/management

G. DSMP BUDGET JUSTIFICATION

(PART OF ONLINE APPLICATION BUDGET, NOT PART OF DSMP TEMPLATE)

Provide a narrative justification that describes the total costs (both CIRM funded and co-funded) of resources required to prepare the data for sharing/preservation.

Applicants should consider and justify any resources needed to adhere to the DSMP. Examples could include personnel time, curating data and developing documentation, infrastructure necessary to provide local management and preservation, and data deposit fees. Resources to cover total cost of the DSMP do not solely have to be funded by CIRM.

Guidance:

All allowable Data Sharing and Management costs submitted in the final budget request must be incurred during the project period. Consistent with CIRM's Grants Administration policy (GAP), budget requests must not include infrastructure costs that are included in institutional overhead (e.g., awardee's facilities & indirect costs) or costs associated with the routine conduct of research. Costs associated with collecting or otherwise gaining access to research data (e.g., data access fees) are considered costs of doing research and should not be included in scientific data management and sharing budgets. Costs may not be double charged or inconsistently charged as both direct and indirect costs.

Reasonable, allowable Data Sharing and Management costs may be included in CIRM budget requests, including personnel costs required to perform data management and sharing activities. Examples of costs include:

- Curating data and developing supporting documentation, including formatting data according to accepted community standards; de-identifying data; preparing metadata to foster discoverability, interpretation, and reuse; and formatting data for transmission to and storage at a selected repository for long-term preservation and access.
- Local data management considerations, such as unique and specialized information infrastructure (only those not covered by awardee's facilities & indirect costs), necessary to provide local management and preservation (e.g., before deposit into an established repository)
- Preserving and sharing data through established repositories, such as data deposit fees necessary for making data available and accessible. For example, if a DSMP proposes preserving and sharing scientific data for 10 years in an established repository with a deposition fee, the cost for the entire 10-year period must be paid prior to the end of the project period. If the DSMP proposes deposition to multiple repositories, costs associated with each proposed repository may be included.

***Example Answer*:**

The following individuals [or just the position titles if unknown] will be responsible for data collection, curation, developing documentation, management, storage, retention, and dissemination of project data, including updating and revising the DSMP when necessary. Provide their effort, annual salary and personnel cost for this project.

[_____] will be utilized to provide local data management and data preservation.

Anticipated fees for preserving and sharing data are [_____]

Anticipated data deposit fees for [repository] are [_____]