CIRM
CALIFORNIA INSTITUTE FOR REGENERATIVE MEDICINE

CIRM Data Sharing and Management

CIRM
CALIFORNIA INSTITUTE FOR REGENERATIVE MEDICINE

Discovery, Preclinical, and Clinical Program Requirements

The sharing of data and knowledge produced from CIRM-funded projects is key to advancing the field of regenerative medicine and accelerating the discovery, validation and development of treatments for patients. CIRM requires awardees to manage and preserve raw data, processed data, and metadata, and make applicable data and metadata available to the broader scientific community. CIRM also requires applicants to allocate funds in their proposed budget for personnel and/or activities related to managing and sharing data produced from the funded project.

To ensure data processing steps can be replicated and data can be reused by other researchers, CIRM requires sharing of data in accordance with FAIR and CARE data principles, using established repositories where possible. CIRM requires that applicants provide a **Data Sharing Overview** in their proposal, and awardees develop and execute a detailed **Data Sharing and Management Plan (DSMP).** The data repositories selected and other information about deposited data must be reported to CIRM during and after the project period. To promote FAIR data sharing and open science, CIRM may publicly share information about CIRM-funded data, including what types of data were generated and where data are deposited.

All application, awarded project, and active award requirements described here are incorporated by reference in Program Announcements and Requests for Applications which reference this resource.

Application Stage: Data Sharing Overview

As part of a Proposal, applicants must provide a general, high-level plan for sharing data produced in the proposed project, the Data Sharing Overview. Applicants must allocate funds in their proposed budget for personnel and/or activities related to managing and sharing data produced from the funded project. For guidelines, please refer to the Data Sharing Budget Justification Guidelines. Instructions for completing the Data Sharing Overview are provided in the application.

Awarded Projects: Data Sharing and Management Plan (DSMP)

If a project is awarded, a DSMP must be submitted to CIRM as Just in Time (JIT) material during Pre-Funding Administrative Review (PFAR). CIRM will review DSMPs and work with awardees to optimize the DSMP, including negotiation of milestones and budget. Awardees must agree with CIRM on a DSMP and associated milestones and budget prior to CIRM issuing a Notice of Award.

Please consult CIRM Data Sharing and Management Plan (DSMP) Guidelines for the purpose and scope of the CIRM DSMP, data terminology, and instructions for submitting a DSMP to CIRM and updating it throughout the duration of the award.

Active award stage

Grantees will report on their data sharing and management activities during regularly scheduled progress reporting and will work with CIRM staff to adjust the DSMP and other data-related milestones as necessary and align data sharing processes with other initiatives at CIRM.

Additional Preclinical Development Program Requirements

CIRM expects that knowledge resulting from PDEV awards will be shared within the CIRM network to drive efficiency and reduce potential roadblocks by leveraging proven processes, study designs, and regulatory pathways to optimize development and eliminate redundant efforts. Sharing learnings with other CIRM awardees will improve product development progression and support a risk-based approach to both planned and unexpected changes throughout the preclinical drug development process while retaining IP

and patient/donor privacy. PDEV awardees are asked in the Data Sharing Overview section of the application to certify to work with CIRM to align with knowledge sharing processes as they are implemented.

Additional Clinical Development Program Requirements

CIRM stipulates that clinical trials funded by CIRM in whole or in part are required to share completed study results with the scientific community. In addition to submitting a DSMP and submitting applicable data to repositories (see CIRM Data Sharing and Management Plan (DSMP) Guidelines), the following requirements apply to Clinical awardees to facilitate clinical trial data sharing:

- Trial is registered at ClinicalTrials.gov no later than 21 calendar days after the enrollment of the first participant
- Informed consent documents for clinical trials are to include a specific statement relating to posting of clinical trial information at ClinicalTrials.gov
- Ensure the responsible entity updates information in the clinical trial record at least once every 12 months
- Results are submitted to ClinicalTrials.gov 12 months from the primary completion date

	Discovery	Preclinical Development	Clinical Development
Application	Data Sharing Overview [CIRM DSMP Guidelines] Budget Justification [Budget Justification Guidelines] Data Project Manager [CIRM DSMP Guidelines, Data Sharing FAQ]		
JIT/PFAR	Metadata Catalog [CIRM Data Explorer Instructions] Questionnaire [CIRM Data Explorer Instructions] Data Use Limitations Institutional Certification [CIRM Data Explorer Instructions]		**Register data sharing statement on ClinicalTrials.gov**
Award	Update Metadata Catalog [CIRM Data Explorer Instructions]		**Informed consent documents include data sharing language, if necessary** **Clinical Trial record is updated at least every 12 months**
Award End	Deposit data in appropriate repository [Data Repository Guidance] Submit final Metadata Catalog to CIRM [CIRM Data Explorer Instructions]		
Post Award			**Results submitted to ClinicalTrials.gov 12 months from completion date**

TABLE 1. Data sharing expectations per program and award stage. A table outlining the different data sharing steps at Application, Just-in-Time, Active Award, Award End, and Post Award stages for Discovery, Preclinical Development, and Clinical Development Programs. Resources for each step are linked in brackets.

CIRM
CALIFORNIA INSTITUTE FOR REGENERATIVE MEDICINE

CIRM Data Sharing and Management Plan (DSMP) Guidelines

Purpose of DSMP

To leverage CIRM-funded data and enable reuse of data by other researchers, CIRM awardees are expected to share their data consistent with FAIR (Findable, Accessible, Interoperable, and Reusable) and CARE (Collective Benefit, Authority to Control, Responsibility, and Ethics) data principles and reflective of practices within specific research communities. Development and execution of the CIRM **Data Sharing and Management Plan (DSMP)** is intended to facilitate:

- **F**indability of data through a public dashboard, the CIRM Data Explorer,
- **A**ccessibility of data by deposition in data repositories accessible to other researchers, and
- **I**nteroperability and
- **R**eusability of data by associating deposited data with necessary and sufficient metadata.

The DSMP intends to capture information (metadata) about the biological samples used for data generation, the methods and data analysis pipelines used during the funded studies and the Data Use Limitations that apply to the data. The information assembled in the DSMP should provide 1) sufficient detail for another researcher to repeat the data processing stages (replicate results), 2) sufficient context to use the data in new ways with confidence in their interpretation of the data and its provenance (reuse data), and 3) sufficient information on data use limitations, enabling adherence to CARE principles.

The expectation is that information captured in the DSMP will be included when data is deposited in a repository, with the goal of making the data FAIR. Data submission rules of data repositories must be followed.

CIRM appreciates your careful attention to this matter and your support of these aims.

Scope of DSMP

CIRM requires DISC, PDEV, and CDEV awardees to manage and preserve raw data, processed data and metadata, and share applicable data[1] and metadata, i.e. make applicable data and metadata available to the broader scientific community through data repositories accessible to other researchers. CIRM expects all applicable data generated under CIRM DISC and PDEV awards to be shared no later than the time of publication or by the end of the award period, whichever comes first. CIRM expects clinical trial data and results generated under a CLIN2 award to be shared no later than 12 months after the study's primary completion date. Even data not used to support a publication, including null or negative findings, are considered data.

For some programs and data types, CIRM has developed specific data sharing expectations (e.g., data types to share, relevant standards, repository selection, timelines) that should be reflected in a DSMP. When no specific CIRM data sharing expectations apply, researchers should propose their own approaches to data sharing and management.

CIRM requires that anyone deriving data from living humans must be prepared to ensure privacy and confidentiality protections (i.e., de-identification, Certificates of Confidentiality, and other protective measures), in accordance with applicable federal, Tribal, state, and local laws and regulations.

[1] Definition of *Applicable Data* in 'Data Terminology'

CIRM
CALIFORNIA INSTITUTE FOR REGENERATIVE MEDICINE

	DISC	PDEV	CLIN
DSMP required	Yes	Yes	Yes
Data sharing of applicable data required via deposit into data repository	Yes—at time of publication or by end of award	Yes—at time of publication or by end of award	Yes—no later than 12 months after the study's primary completion date

Data Terminology

Data: The Intellectual Property Policy for CIRM Awards defines "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).

Data generation: generation of raw data

Data processing: all data processing steps (dry lab) following generation of raw data

Data production: overarching term, referring to both data generation and data processing

Data products: the result of each data generation step and each data processing step (Each data product should be listed in the DSMP Metadata Catalog)

Raw data: data produced by an instrument (e.g., raw sequence data) or by other methods, such as measurements and surveys, or obtained from a data repository

Processed data: data produced from raw data and from subsequent processing steps (e.g., quantification files, alignment files, etc.)

Final processed data: data produced from last processing step (e.g., aggregated quantification, etc.), on which conclusions are based

Metadata: data that provide additional information needed to make shared raw and processed data findable, interpretable and reusable. Metadata information is requested in the DSMP.

Metadata categories in CIRM Data Explorer

- **Data Product Details:** methods used for data generation (machine, instrument), data processing (software toolkits, pipelines) and data sharing (data repositories).
- **Biological Material Details:** information about the source and modifications of the biospecimens and the final cell product used for data generation
- **Goal of Experiment:** information about diseases studied and/or biological questions addressed
- **Sample Preparation:** information about experimental approaches used to prepare the sample for data generation
- **Protocols & Publications**

Additional metadata

- **Map of unique identifiers:** a document that details the persistent unique identifiers or other standard indexing tools, assigned by data repositories and used to track projects and samples, enabling other researchers to find related data deposited in different repositories.

- **Data Dictionary:** a document that defines field names, such as male/female is represented by 0/1 or 1/2 or m/f etc. (only needed if not using an existing Data Standard, such as this LOINC code for sex at birth)

Data standards: guidelines or formal rules for producing, structuring, naming, and describing data.

CIRM expects that an awardee will apply data standards that are common to their field of study in the production of data and to metadata that are deposited in a Data Repository. Examples of data standards can be found at CDISC or LOINC.

Data sharing: making data available to the broader scientific community by depositing in a data repository accessible to other researchers

Applicable data:

- All data that are needed for another researcher to replicate results and to reuse data. Minimally this includes raw data, final processed data and metadata.
- CIRM does not anticipate that researchers will preserve and share all data produced in a study. Researchers should decide which data to preserve and share based on ethical, legal, and technical factors that may affect the extent to which data are preserved and shared. The rationale for these decisions must be provided in the DSMP Questionnaire.
- Data not used to support a publication, including null or negative findings, can be considered applicable data.

Replicate results: another researcher uses shared data and same code/software as original researcher to obtain the same results

Reuse data: another researcher uses shared data and different tools / software to obtain new results, or uses shared data in combination with their own data

CIRM
CALIFORNIA INSTITUTE FOR REGENERATIVE MEDICINE

Instructions for Submitting and Updating DSMP

The CIRM DSMP has 3 components:

- Metadata Catalog
- Data Use Limitations (DUL) Institutional Certification
- Questionnaire

Together, these components outline how the data for the funded project will be shared with the scientific community.

For all data you propose to generate, please prepare a Data Sharing and Management Plan (DSMP):

1. Join/log into **CIRM Data Explorer**
2. Complete the **Metadata Catalog** for expected data
3. Complete, sign, and submit the **Data Use Limitations (DUL) Institutional Certification form**
4. Complete the **DSMP Questionnaire**

The Metadata Catalog will be a living record:

- **Initial Metadata Catalog:** Prior to CIRM issuing the Notice of Award (NoA), the initial Metadata Catalog is submitted to CIRM. It contains minimal information about the anticipated data types and experimental design of the project.
- **In progress Metadata Catalog:** Throughout the project, the Metadata Catalog is continually updated as data is produced and metadata is collected. This ensures timely and progressive assembly of all information necessary for data deposition at the end of the project. The most up to date version of the Metadata Catalog is submitted to CIRM as part of each scientific progress report and is subject to CIRM review and approval.
- **Final Metadata Catalog:** At the end of the award, the Metadata Catalog, as well as the DUL form, Map of Unique Identifiers and Data Dictionary, are finalized and serve as a record of metadata that is shared with raw and processed data.

Once data have been deposited by the awardee, the metadata provided in the Metadata Catalog and the DUL information will be made public and displayed in the CIRM Data Explorer, a dashboard that scientists can use to discover CIRM-funded data and determine where they are deposited.

CIRM
CALIFORNIA INSTITUTE FOR REGENERATIVE MEDICINE

CIRM Data Sharing and Management Budget Justification Guidelines

Applicants should consider and justify any resources needed to adhere to their plans for data sharing and management, as described in the **Data Sharing Overview** in the application, and as anticipated in their more detailed Data Sharing and Management Plans **(DSMPs)** which will be submitted to and negotiated with CIRM prior to issuance of a Notice of Award (NOA), if awarded. For more information on data sharing and management requirements, please see CIRM Data Sharing and Management.

Costs required for implementing the DSMP must be requested and justified in the submitted application as part of the overall Budget Justification for the project and incurred during the funded 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 data management and sharing budgets. Costs may not be double charged or inconsistently charged as both direct and indirect costs.

Reasonable, allowable costs for managing and sharing data may be included in CIRM budget requests. Resources to cover total cost of data sharing or adhering to the DSMP do not solely have to be funded by CIRM.

Examples of data sharing and management 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 data sharing plan proposes preserving and sharing 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 data sharing plan proposes deposition to multiple repositories, costs associated with each proposed repository may be included.

- Personnel costs required to perform data management and sharing activities. Provide effort, annual salary and personnel cost for this project.

CIRM
CALIFORNIA INSTITUTE FOR REGENERATIVE MEDICINE

CIRM Guidance for Data Repositories and other Resources

This document provides a non-comprehensive set of resources for identifying and selecting domain-specific or generalist data sharing repositories for discovery, preclinical, and clinical data. Researchers are advised to reference the repository policies.

Springer Nature Data Repository Guidance

This resource provides listing and guidance on specialized and generalist repositories. The specialized repositories are categorized by scientific discipline or data types and the notes include suggestions or recommendations for repository selection. The listing also includes links to the repository entry on FAIRsharing.org where researchers can obtain more information on the repositories.

https://www.nature.com/sdata/policies/repositories

NIH Data Resources

The NIH maintains a non-comprehensive listing of NIH-supported domain-specific data repositories as well a list of external generalist repositories. The lists are organized as tables and include descriptions as well as links to data submission and data access policies.

NIH-supported open domain-specific data repositories—The 148 repositories listed in this table are generally open to domain-specific data submission and user access.

Generalist Repositories—The 9 generalist repositories in this listing accept data regardless of type, format, content, or discipline.

Repository Search Tools

Fairsharing.org—A community driven resource that promotes FAIR principles by providing a searchable database of repository profiles, data standards, and journal and funding sharing policies.

External Resources Related to Sharing Protected Health Information (PHI)

Informed Consent for Secondary Research with Data and Biospecimens: Points to Consider and Sample Language for Future Use and/or Sharing –NIH Office of Science Policy resource for drafting informed consent language for data sharing

Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule –US Department of Health and Human Services guidance on approaches to achieve de-identification