CIRM Data Sharing and Management Plan (DSMP) for Omics / Flow Cytometry Data

Guidelines for Discovery Awards

DO NOT SUBMIT DSMP with APPLICATION.
If funded, submit DSMP as Just in Time (JIT) material.

Content:

1. Background

As articulated in CIRM's strategic plan, CIRM is committed to building infrastructure that organizes and democratizes data through knowledge networks and fostering open science. Data Sharing is therefore a vital component of CIRM's commitment to California researchers.

CIRM requires DISC awardees to manage and preserve raw data, processed data and metadata, and make applicable* data and metadata available to the broader scientific community. CIRM expects all applicable data generated under a CIRM award to be shared no later than the time of publication or by the end of the award, whichever comes first. Even data not used to support a publication, including null or negative findings, are considered data.

With the goal of leveraging CIRM-funded data, and enabling reuse of data by other researchers, CIRM DISC awardees are expected to develop and execute a **Data Sharing and Management Plan (DSMP)** that is consistent with FAIR (Findable, Accessible, Interoperable, and Reusable) data principles and reflective of practices within specific research communities.

- If a project includes generation of **omics and /or flow cytometry data**, the DSMP must be developed using templates provided by CIRM, and an Initial[#] DSMP must be submitted to CIRM prior to issuing a Notice of Award.
- For **data from other types of experiments** (e.g., imaging, electrophysiology, behavioral, etc.), CIRM may work with the awardee to develop a DSMP and establish data sharing milestones prior to CIRM issuing a Notice of Award.

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.

*Definition of 'applicable data' in 'Data Terminology' section
[#] Definition of 'Initial DSMP' in 'Instructions' section'

2. 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 processing) 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 Data 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 interpretable and reusable

  Types of metadata
    - Data production metadata refers to methods used for data generation (machine, instrument), data processing (software toolkits, pipelines) and data sharing (data repositories). This information is requested in the DSMP Data Catalog.
    - Tissue donor metadata includes clinical, phenotypic, demographic data.
    - Sample handling metadata includes quality of the sample, preparation of sample, and protocols used in the course of the research.
    - Map of unique identifiers refers to 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 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 Dictionary refers to 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 sharing:** make data available to the broader scientific community by deposit in a data repository accessible to other researchers

- **Applicable data**:
  CIRM expects all applicable data generated under a CIRM award to be shared no later than the time of publication or by the end of the award, whichever comes first. Even data not used to support a publication, including null or negative findings, are considered data.
    - CIRM requires that all data that are needed for another researcher to replicate results and to reuse data be deposited in a data repository. 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. Provide the rationale for these decisions in the DSMP Questionnaire.
    - Data derived 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.
    - Data submission rules of data repositories must be followed.

- **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

3. Instructions - DSMP for Omics and Flow Cytometry Data

DO NOT SUBMIT DSMP with APPLICATION.
If funded, submit DSMP as Just in Time (JIT) material.

For all **omics and flow cytometry data** you propose to generate, please prepare a Data Sharing and Management Plan (DSMP) using the following 2 templates:

Part A - DSMP for Omics / Flow Cytometry Data - Data Catalog **('DSMP Data Catalog')**
Part B - DSMP for Omics / Flow Cytometry Data - Questionnaire **('DSMP Questionnaire')**

The 2 templates to complete the DSMP for Omics / Flow Cytometry data can be found [here].

The DSMP will be a living document:

- **Initial DSMP**: At the start of the project, the initial DSMP, submitted to CIRM as JIT material, contains minimal information about the anticipated data types and experimental design of the project. The Initial DSMP will become part of the Notice of Award.
- **In Progress DSMP**: Throughout the project, the DSMP is continually updated as data is produced and metadata is collected. The goal is to avoid last minute scrambles to assemble all information needed for data deposit at the end of the project. The most up to date version of the DSMP is provided to CIRM as part of each scientific progress report, and is subject to CIRM approval.
- **Final DSMP**: At the end of the project, the DSMP is finalized and serves as a record of metadata that is shared with raw and processed data. Data not yet deposited is deposited.

The DSMP Data Catalog intends to capture the methods and analysis pipelines used during the course of the funded studies. The information assembled in the DSMP Data Catalog and the DSMP Questionnaire collectively aims to provide (1) sufficient detail for another researcher to repeat the data processing stages (replicate results) and (2) sufficient context to use the data in new ways with confidence in their interpretation of the data and its provenance (reuse data).

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 findable, accessible, interoperable and reusable (FAIR). CIRM appreciates your careful attention to this matter and your support of these aims.