



INFR9 Awards: Funding Opportunity for Data Science and Software Engineering

Concept Plan: For ICOC Approval June 25, 2026

OVERVIEW	
Objective	Support the development of innovative, open source software (OSS) for integration of disparate datatypes to accelerate research in regenerative medicine. The Expected Outcome is the development and deployment of an OSS tool to solve a multimodal data integration bottleneck in stem cell-based and genetic therapy research that has wide applicability and scalable impact.
Scope	Innovative software engineering led by pairs of investigators to develop, maintain, and/or extend OSS (i.e., computational tools, pipelines, or resources) to address data integration bottlenecks to catalyze data science and accelerate the identification or validation of targets and biomarkers and early candidate discovery
Recurrence	Once per year
AWARD DETAILS	
Max Award Amount	\$500,000 total costs
Max Award Duration	2 years
ELIGIBILITY	
Applicant Organization	Must meet CIRM's definition of a California Organization
Critical Roles	The Principal Investigator (PI) and Co-Investigator (Co-I) must each commit at least 5% effort, and Co-I and PI must be from different labs
SCHEDULE AND DEADLINES	
GWG Selection	Approximately 60 days post submission deadline
GWG Discussion	Approximately 60 days post GWG Selection
Award Approval	Next available Application Review Subcommittee meeting
Start Date	Must be ready to start award activities within 90 days of award approval

Background

The mission of the California Institute for Regenerative Medicine (CIRM) is to accelerate world class science to deliver transformative regenerative medicine treatments in an equitable manner to a diverse California and world. In September of 2024, CIRM's Governing Board, the Independent Citizens' Oversight Committee (ICOC), approved a Strategic Allocation Framework (SAF) to guide and optimize the value of CIRM's current and future investments. This exercise defined an ambitious goal for CIRM to catalyze the discovery and translation of novel targets and biomarkers, integrating them into preclinical or clinical research for diseases in California. Data integration, collaborative data science, and new software tools are necessary to realize this goal.



Translating fundamental research into stem cell-based and genetic therapies is a significant challenge in part due to the complexity and heterogeneity of research and development (R&D) data. Multimodal datasets, from molecular and cellular profiles to preclinical models and patient-level clinical data, must be integrated to be leveraged for biological insight. Data integration and computational analysis have unique technical challenges and are resource-intensive, often requiring bespoke computational tools, robust analytical pipelines, and scalable software infrastructure. Despite their importance spanning basic, translational, and clinical research phases, these tools frequently lack dedicated funding for development and maintenance. OSS engineering and data science can powerfully address these bottlenecks. OSS development fosters collaboration, transparency, and generalizability to build and refine computational methods designed for regenerative medicine. Data science approaches, including machine learning and artificial intelligence, can reveal mechanistic understanding, bridging gaps between multidisciplinary research and patient outcomes.

Advances in data science, computational modeling, and AI have created opportunities to accelerate discovery in regenerative medicine; however, federal funding opportunities often focus on hypothesis-driven research and may not fully support the iterative software development, data integration and computational data infrastructure required to capitalize on emerging data science capabilities. This funding gap is aligned with CIRM’s mission under Prop 14, which recognizes the importance of supporting research that may not receive timely or sufficient federal funding, or that may face limitations that could impede the research. In the context of regenerative medicine, such limitations can include challenges related to multimodal data integration, interoperability, reusable computational infrastructure, and scalable open-source software development. Targeted funding can help bridge the gap between emerging computational capabilities and their practical use by regenerative medicine researchers.

Work to increase data interoperability and data integration is especially underfunded, as these activities (e.g., developing APIs, integrating with foundation models, building reproducible pipelines, etc.) are often considered infrastructure rather than biomedical science and may fall between traditional federal research funding mechanisms and commercial investment.

Through the Data Science and Software Engineering program, CIRM will support the development of OSS tools with applications in regenerative medicine. The Data Science and Software Engineering awards fund collaborations to build and deploy creative OSS solutions to capitalize on recent advances in stem cell and genetic research, digital biology, and computational science, including:

1. Advances in a broad range of technologies, both data-generating and analytical approaches (e.g. multiomics, massively parallel or automated approaches, ML/AI, medical devices and wearables)
2. Advances in disease models and mechanistic understanding (e.g., stem cell models, preclinical animal models, microtissues)
3. Advances in data sharing (e.g., federated learning, privacy-preserving technologies, data volume available for reuse)

The Data Science and Software Engineering program aims to integrate these advances with CIRM’s support for discovery-stage research to accelerate data science across a wide range of disease areas. This program focuses on enabling the integration of diverse data sources (e.g., clinical specimens, electronic health records, *in vivo* models, stem cell models, multiomics, computational modeling), to break down data silos, enable meaningful linkages, and strengthen the validity and reproducibility of research findings.

Objective

The objective of this program is to support cutting-edge OSS engineering of data science tools to accelerate research in regenerative medicine. The Expected Outcome is the development and deployment of an OSS tool to solve a multimodal data integration bottleneck in stem cell-based and genetic therapy research that has wide applicability and scalable impact.



This vision is realized through an interdisciplinary funding opportunity supporting software development, with an emphasis on technical innovation, open science, and leveraging the integration of multimodal data to accelerate the discovery and translation of potential therapeutics in regenerative medicine.

Beyond the unique contributions of individual software tools, CIRM recognizes the value of collaborative data science in applying novel computational methods to make scientific breakthroughs by leveraging available regenerative medicine datasets. The Data Science and Software Engineering program is designed to provide dedicated support for OSS that can harness the rich and diverse data generated across CIRM's R&D portfolio to address bottlenecks in the field of regenerative medicine. Projects supported by this program are expected to develop adaptable and extensible software tools and resources that have broad utility in the regenerative medicine community.

Through this dry lab-only funding opportunity, CIRM intends to bolster computational support in regenerative medicine labs by funding awards for engineers in research. Data scientists and software engineers in research environments are often supported via piecemeal, project-specific funding, which limits their opportunities to lead research programs. By recognizing and elevating the contributions of these engineers, the Data Science and Software Engineering awards seek to expand their opportunities for leadership within the research enterprise through dedicated support for the development of high quality, reusable software that follows engineering best practices.

Program Guiding Principles

Guiding Principles are how CIRM translates the SAF recommendations into portfolio outcomes. Guiding principles shape program objectives, inform the criteria by which projects are selected for merit review and evaluated, and inform the recommendations that CIRM teams bring to the Application Review Subcommittee (ARS) to support funding decisions.

The INFR9 Program is designed to develop high quality, user-friendly software tools that address a data integration bottleneck in regenerative medicine research. To achieve this objective, the INFR9 portfolio will prioritize projects that will:

- Address barriers to integrating multimodal data in regenerative medicine
- Innovate through open and collaborative software development to accelerate research
- Create broadly-applicable, validated tools that maximize the value of CIRM-funded data

Scope and Structure

Data Science and Software Engineering awards will support OSS development led by pairs of interdisciplinary investigators to address bottlenecks in regenerative medicine. This funding opportunity has a focus on development, maintenance, and/or extension of software to integrate data from different sources (e.g., data from disparate data acquisition modalities and model systems, and data across scales) to enable computational research in target, biomarker, and candidate discovery in regenerative medicine.

The Data Science and Software Engineering awards will develop generalizable computational tools, pipelines, or resources that will integrate and expand the impact of existing datasets to make an impact in the field of regenerative medicine. This program is product-focused and projects will produce a stand-alone tool or application with open licensing. Successful projects will build a solution to overcome a bottleneck in stem cell-based and genetic therapy research that has wide applicability and scalable impact.

The Data Science and Software Engineering awards are intended to equip regenerative medicine research labs with computational expertise through collaboration and catalyze data science research. By enabling data integration, meta-analyses, and collaboration, it will facilitate the generation of new knowledge about human biology (e.g., disease mechanisms or drug responses) and create new data linkages across research domains.

Data Science and Software Engineering awards must propose projects designed to achieve the Expected Outcome defined in the Objective. Projects must follow transparent development and OSS best practices,



develop and execute a Deployment, Adoption, and Sustainability plan, and incorporate community/contributor engagement activities to inform the development of the software project.

To maximize the impact of the project outcome, applicant teams must appropriately manage and share data generated during a Data Science and Software Engineering award. All data used in meta-analyses or as proof-of-concept must be shared in established repositories. The codebase must be in a public repository approved by CIRM and follow OSS development principles. Refer to “Open Source Software Requirements” on pg. 6 for more information.

The Data Science and Software Engineering program awards are open to all applications that fulfill the eligibility requirements detailed below, without restrictions in research topics or disease indications.

This program is not designed to support proposals that include wet-lab experimental biology. Proposals centered on developing tools with limited applicability beyond a single research question do not align with this program’s focus on OSS development with broad utility across the regenerative medicine research community. Applicants are encouraged to explore the full range of CIRM Funding Programs to identify the opportunity that best matches their research objectives.

Program Activities

CIRM **will fund** the following activities under this opportunity:

REQUIRED ACTIVITIES	
✓	Development and deployment of OSS
✓	Development and execution of a Deployment, Adoption, and Sustainability Plan
✓	Activities associated with managing, preserving, and sharing data and knowledge from the study
✓	Community engagement activities, e.g., workshop, hackathon, user feedback channels or focus groups to gather feedback to inform development of the software tool and encourage community contributions
ALLOWABLE ACTIVITIES	
✓	Activities contributing to meeting the Data Science and Software Engineering objective to support cutting-edge OSS engineering of data science tools to accelerate research in regenerative medicine
✓	Activities that address a data science research question in the identification or validation of targets or biomarkers, or candidate discovery (through lead optimization)
✓	Partnering activities with patient-centered organizations or other project-relevant community groups to incorporate patient, provider, or user perspectives into the development of the software project
✓	Activities to support outreach or communication of research plans or outcomes with the wider public
✓	Travel and accommodation expenditures associated with attendance of CIRM organized meetings and conferences
✓	Engagement activities with trainees supported through CIRM’s EDUC or INFR programs

CIRM **will not fund** the following activities under this opportunity:

UNALLOWABLE ACTIVITIES	
✗	Wet-lab activities, including new laboratory-based data generation from biological samples



✘	Activities to build new data repositories (i.e., creating a database for the purpose of storing and sharing datasets)
✘	Costs incurred on or before the date of ICOC approval
✘	Activities already budgeted or paid for under a prior, existing or pending CIRM award or which are already supported by another funder
✘	Building a knowledge platform or data coordinating center (i.e., digital infrastructure that serves as a combined repository and collaborative workspace to store, manage, and/or analyze data) to support CIRM-funded programs
✘	Construction or renovation of physical facilities

Award amount and duration

The maximum amount of funding that may be requested for a Data Science and Software Engineering award is \$500,000. The maximum award duration is 2 years. The requested amount must be adequately justified and is subject to adjustments prior to the issuance of an award based on assessments by the Grants Working Group, the CIRM team, or by the Application Review Subcommittee of the ICOC.

Award Management

All awards made under this opportunity are subject to CIRM’s standard award terms, including requirements related to intellectual property, data sharing, reporting, and award administration. These terms are set forth in CIRM’s applicable regulations including but not limited to the CIRM Grants Administration Policy for Discovery, Translation, and Education Projects (and any successor policy governing CIRM awards) and the CIRM Intellectual Property Policy for Non-Profit and For-Profit Organizations.

GWG Application Selection

If the number of applications submitted exceeds the volume that can reasonably be reviewed by the GWG panel, the GWG will conduct an initial screening to identify those applications that best align with the INFR9 program objectives. The basis for selection will be defined in the INFR9 Program Announcement.

Applications receiving the highest overall reviewer assessments, consistent with the target number established for GWG Merit Review, will advance to full panel discussion.

The CIRM President and CEO, upon recommendation from CIRM scientific staff, may advance a limited number of additional applications not selected by the GWG if there is a compelling programmatic justification.

Reviewers may provide brief comments or categorical assessments at this stage; however, detailed feedback to applicants is not guaranteed.



Provisional timetable

The Data Science and Software Engineering funding opportunity will follow the provisional timeline below:

PROVISIONAL TIMETABLE	
GWG Selection	Approximately 60 days after application due date
GWG Discussion	Approximately 120 days after application due date
Application Review Subcommittee Award Approval	Next available Application Review Subcommittee meeting
Award Start	Approximately 90 days after award approval

Eligibility

All the following requirements must be fully satisfied for an application to be accepted and considered for funding by CIRM. Requirements marked with a * incorporate by reference the requirements and definitions described in **CIRM Funding Opportunities: Common Requirements and Definitions**, which may be amended from CIRM from time to time.

ELIGIBILITY REQUIREMENTS	
1	The proposal must be in scope and aim to develop OSS that will lead to the expected outcome defined in the Objective
2	The Principal Investigator (PI) must commit a minimum of 5% effort and adhere to CIRM's requirements
3	The Co-Investigator (Co-I) must commit a minimum of 5% effort and not currently be part of the same lab as the PI
4	An individual may not serve as PI on more than one application or serve as a Co-I on more than two Data Science and Software Engineering applications per funding cycle
5	The PI and/or Co-I must not currently have another application that is substantially similar or has overlapping activities pending review or approval under any CIRM opportunity
6	The applicant must be ready to initiate work on the funded project within 90 days of award approval
7	The application must be complete and accurate
8	The applicant organization must meet CIRM's definition of a California Organization*
9	For-profit organizations must demonstrate solvency
10	The applicant must be in "good standing"*

Open Source Software Requirements

To support transparency, broad usability, and long-term impact of software developed under the INFR9 Data Science and Software Engineering Awards, all projects must meet open source software requirements, which will be further described in the Program Announcement and briefly summarized here:

- Be released under a recognized open source license that allows unrestricted reuse,



- Make the full codebase publicly accessible in a repository approved by CIRM throughout the duration of the award,
- Include (in the approved repository) clear, up-to-date documentation (including communications related to project development) sufficient for users to use, contribute to, and evaluate the software,
- Ensure that all datasets used for benchmarking, validation, and demonstration adhere to CIRM's data sharing requirements, and
- Include a plan for community engagement and sustainability.