

C I R M

CALIFORNIA INSTITUTE FOR
REGENERATIVE MEDICINE

Data Science and Software Engineering Awards (INFR9)

Concept Plan

Janie Byrum, *Sr. Science Officer*
INFR9 Concept Development Team

May 29, 2026

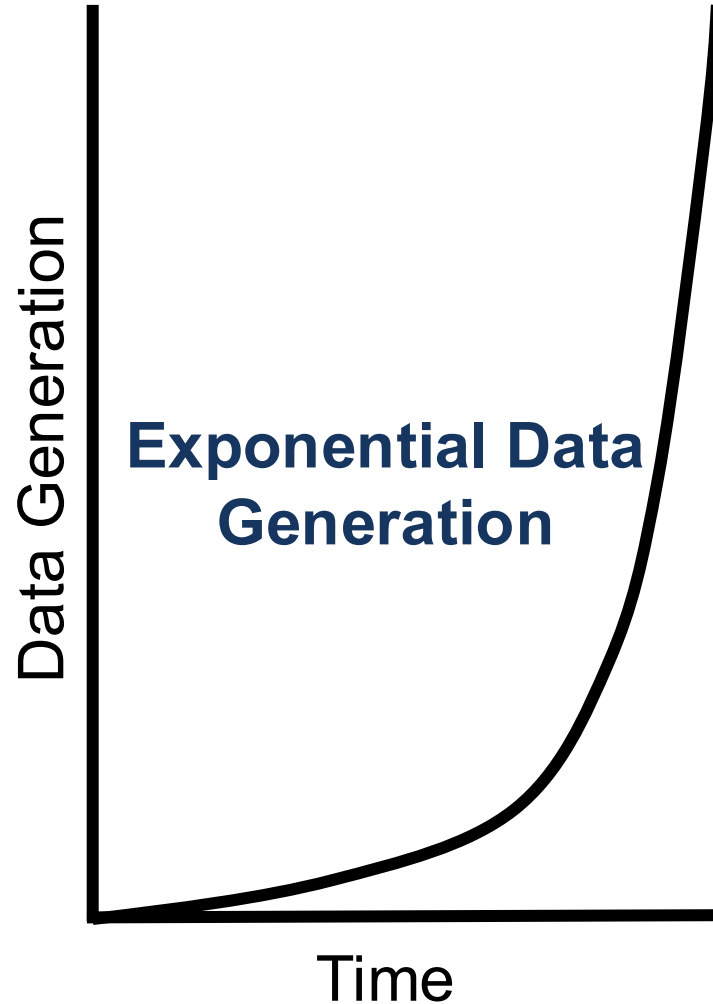
Our Mission

To accelerate world-class science to deliver transformative regenerative medicine treatments in an equitable manner to a diverse California and the world

Concept Outline

- 1. Background**
2. Objective and Approach
3. Scope
4. Award Structure and Budget
5. Eligibility
6. Application and Review
7. Timeline
8. Summary and Request for Approval

Massive data growth necessitates transformative tools

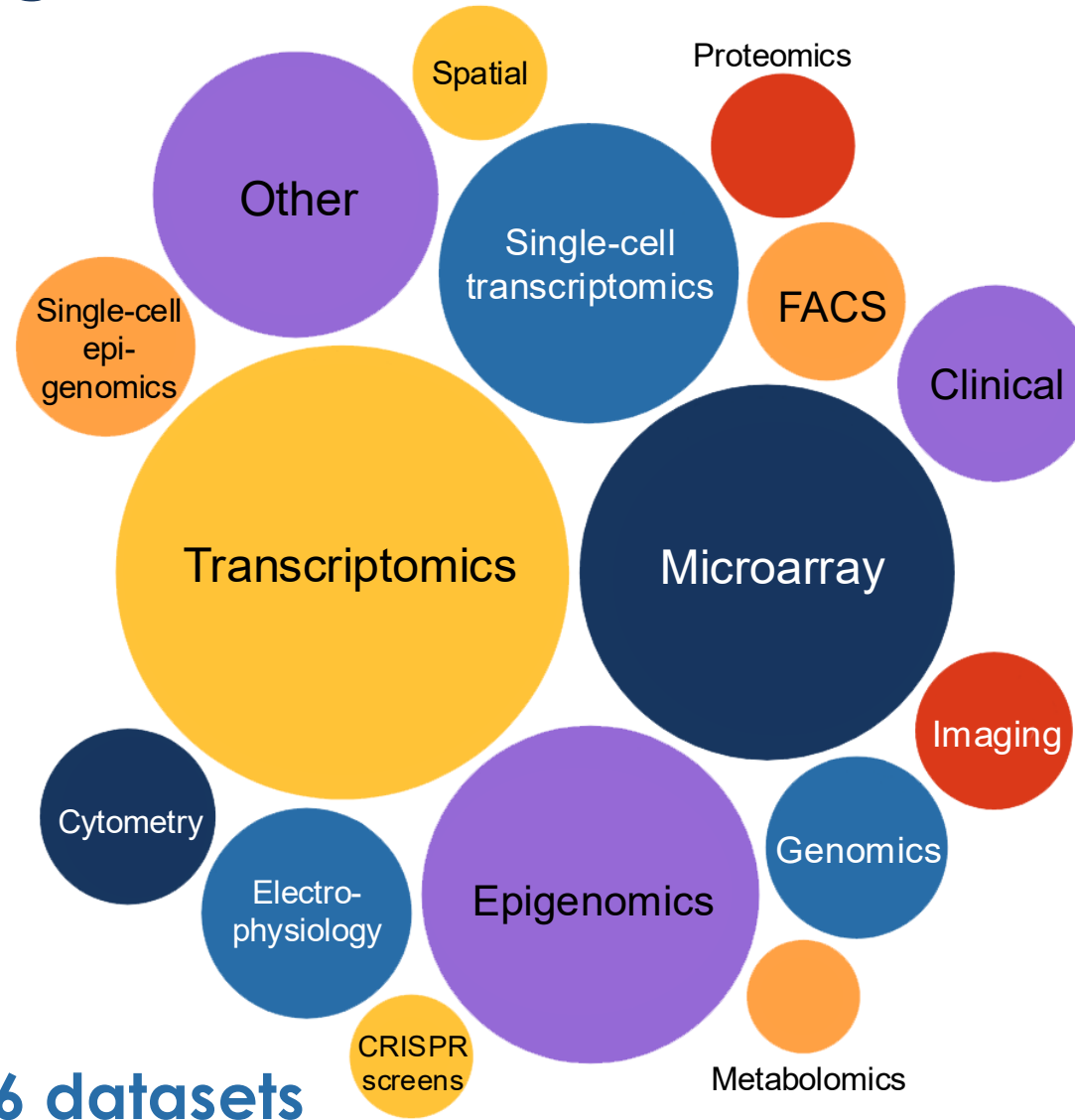


Scalable tools are essential for integrating heterogeneous data

**The challenge...
is how to store,
retrieve, integrate,
and analyze this
mountain of
complex data
—Francis Collins**



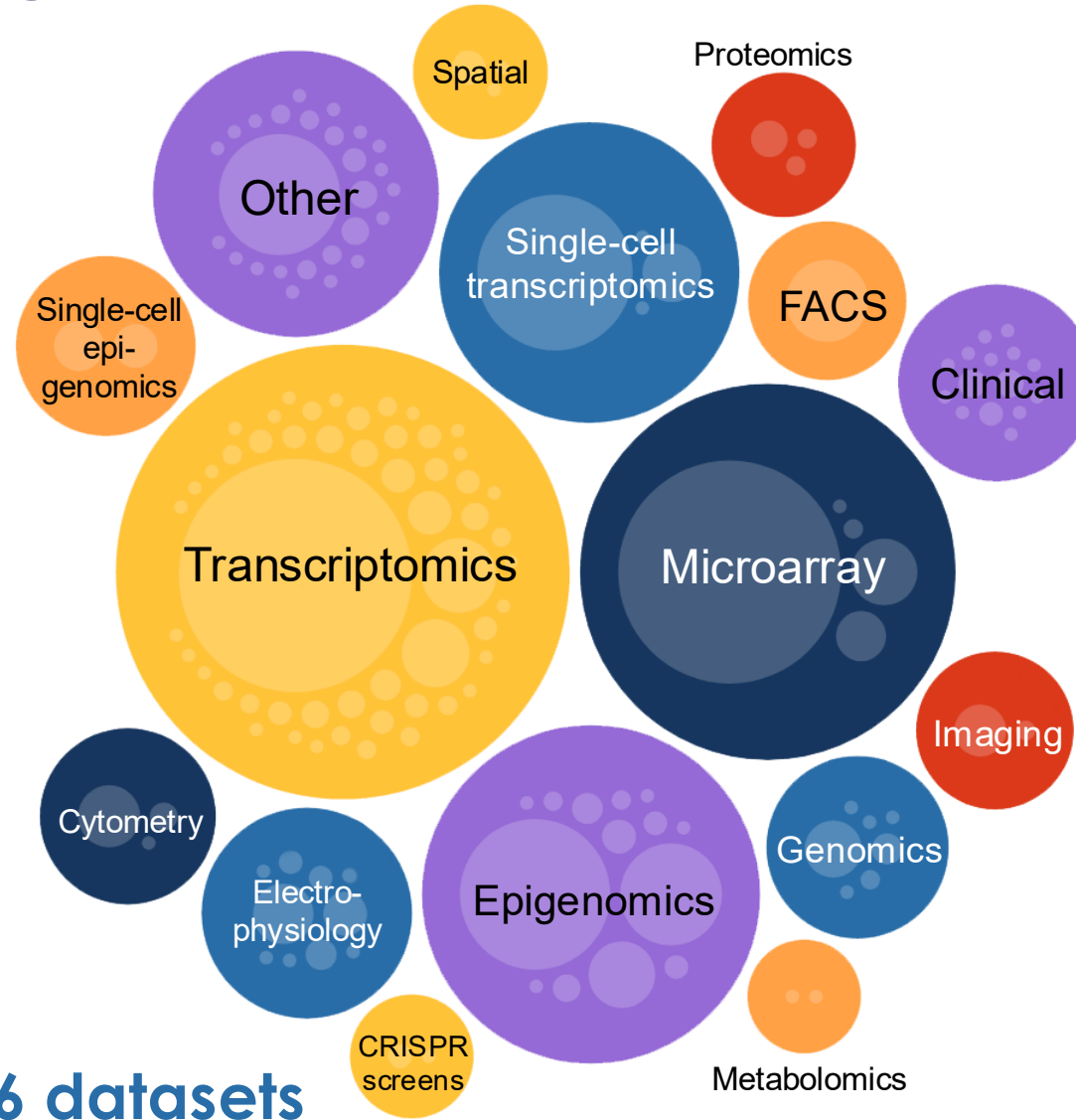
CIRM researchers generate vast amounts of diverse data



n = 1276 datasets

CIRM Data Explorer
datahub.cirm.ca.gov

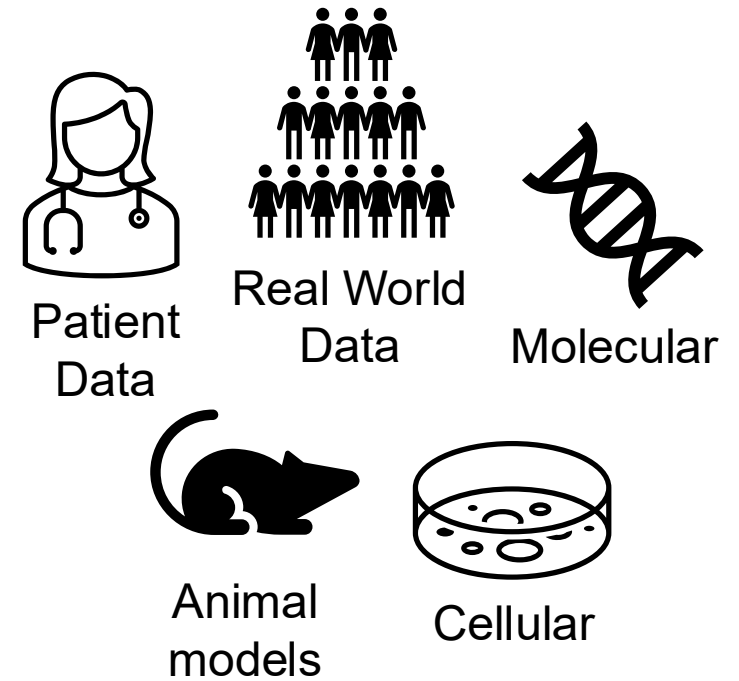
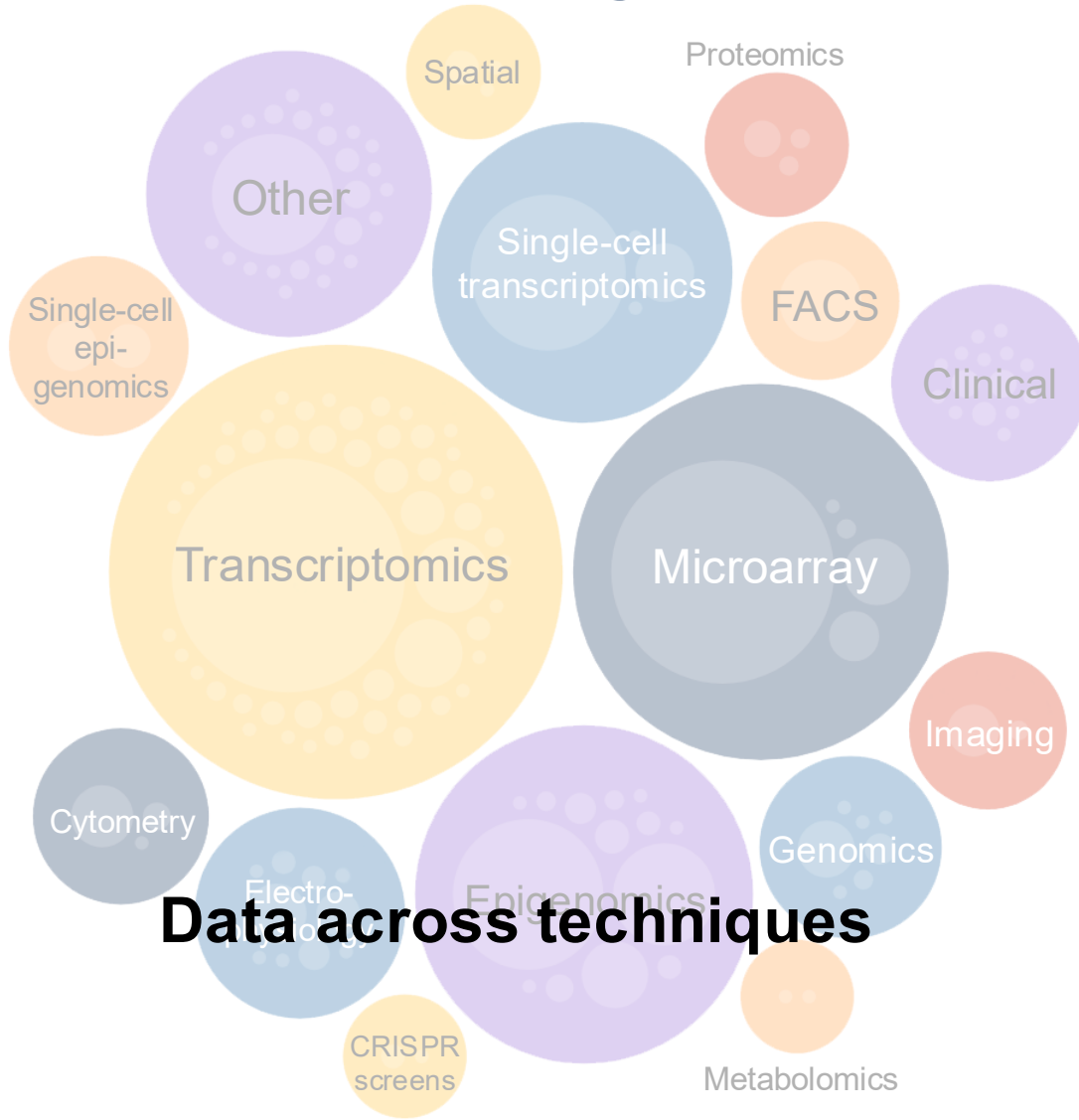
CIRM researchers generate vast amounts of diverse data



n = 1276 datasets

CIRM Data Explorer
datahub.cirm.ca.gov

CIRM researchers generate vast amounts of diverse data



Accelerate research with accessible software tools

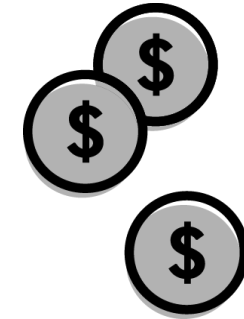
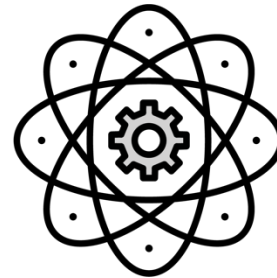
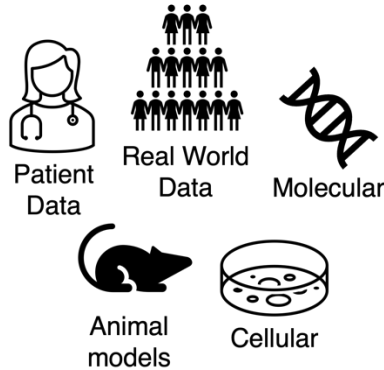
SAF Goal 1

Catalyze the identification and validation of at least 4 novel targets and biomarkers, ensuring integration into preclinical or clinical research for diseases in California

Recommendation:

Enable data science collaborative efforts via dedicated grants

Bottlenecks in Regenerative Medicine Research



Multimodal data (across scales, diseases, systems) need to be integrated

Software tools are needed to accelerate stem cell-based and genetic research

Lack of dedicated funding for developing and maintaining critical open source software

CIRM has an opportunity to support **open source software development** that **drives research in regenerative medicine**

Successful open source software in life sciences

scikit-learn
Machine Learning in Python

Kaggle State of Data Science and Machine Learning 2022 survey (24k people) found that it's the most popular ML framework, over 80% of respondents use it

<https://www.kaggle.com/kaggle-survey-2022>

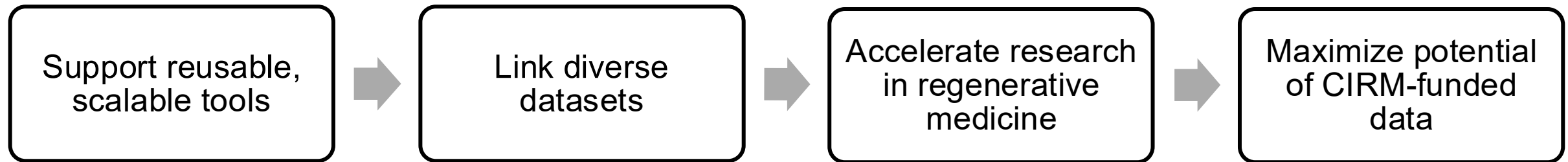
65.5k stars on Github

Scikit-learn learns from data to help with:

- Prediction, what will happen next?
- Classification, is this diseased or healthy?
- Clustering, can we group the similar cells?

Data Science and Software Engineering Awards (INFR9)

Support **open source software** that **resolves data integration barriers** in regenerative medicine



Distinctions between INFR9 and what we already support

DISC0/4/5 Supported	INFR9 Focus
<ul style="list-style-type: none">Ancillary to research projectsNo requirements for software engineering standards and best practices	<ul style="list-style-type: none">Enhanced scalability, reusability, and reproducibility of open-source research software
<ul style="list-style-type: none">Highly specialized for a specific research project	<ul style="list-style-type: none">Expanded use cases, built for community use, maintenance

INFR9

Data Science and Software Engineering Concept

**Catalyze data science research
through data integration**

Guiding Principles: INFR9 (proposed)

Guiding Principles help CIRM shape its portfolio

The INFR9 portfolio will:

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

Objective and Approach

C I R M



INFR9 | Data Science and Software Engineering

Objective

Support the development, maintenance, and/or extension of innovative, open-source software for data integration of disparate data types to accelerate research in regenerative medicine for the purpose of target identification and validation

Approach

Investigator pairs leading interdisciplinary software engineering projects

Scope

C I R M



INFR9 | Scope

INFR9	
Scope	Software tools for data integration of disparate data types for regenerative medicine research
Examples	<ul style="list-style-type: none">• AI/ML approaches for <i>in silico</i> screening, protein design, molecule generation, cell and gene engineering• Knowledge graphs for bridging structures, behavior assays, functional genomics, pathways, patient data and EHRs, etc.

Award Structure and Budget

C I R M



INFR9 | Award Structure and Budget

INFR9	
Recurrence	Annual
Max Duration	Up to 2 years
Award Amount	Up to \$500k total costs
Awards/Year	15-20
Total Funds/Year	\$10M

Eligibility

C I R M



INFR9 | Eligibility

INFR9	
Applicant	Must meet CIRM's definition of a California Organization
Critical Roles	2 CA-based Investigators: <ul style="list-style-type: none">• 1 Principal Investigator (PI)• 1 Co-Investigator (Co-I), must not currently be part of the same lab as the PI
Investigator Effort	<ul style="list-style-type: none">• PI - 5% min• Co-I - 5% min
Project	The Proposal must be in scope and develop open-source software that could lead to the expected outcome defined in the INFR9 Objective

Application and Review

C I R M



INFR9 | Application and Review

INFR9 will adopt a GWG selection process to:

- Allow GWG to select applications that are most responsive to the funding opportunity and hold the greatest potential for impact
- Allow CIRM pre-planning for improved scientific and technical review

INFR9 will adopt a 1-100 numerical GWG scoring system to:

- Align across CIRM programs
- Improve granularity and visibility for score driving decisions

Timeline & Outreach Plan

C I R M



INFR9 | Proposed Program Timeline



Outreach Plan

Targeting outreach to:

- Non-UCs, Community Colleges and CSUs
- Professional organizations, consortia/networks, and/or computer science and data science university departments

Tools:

- Email campaigns and social media posts
- Landing page
- Webinar or video guide
- Events



More to come!

Requested Subcommittee Action

The CIRM team requests that the Science Subcommittee recommends approval of the Data Science and Software Engineering (INFR9) concept plan to the full board.