



National Institutes of Health  
*Office of Data Science Strategy*

# Generalist Repository Ecosystem Initiative (GREI) Council of Councils Working Group Recommendations

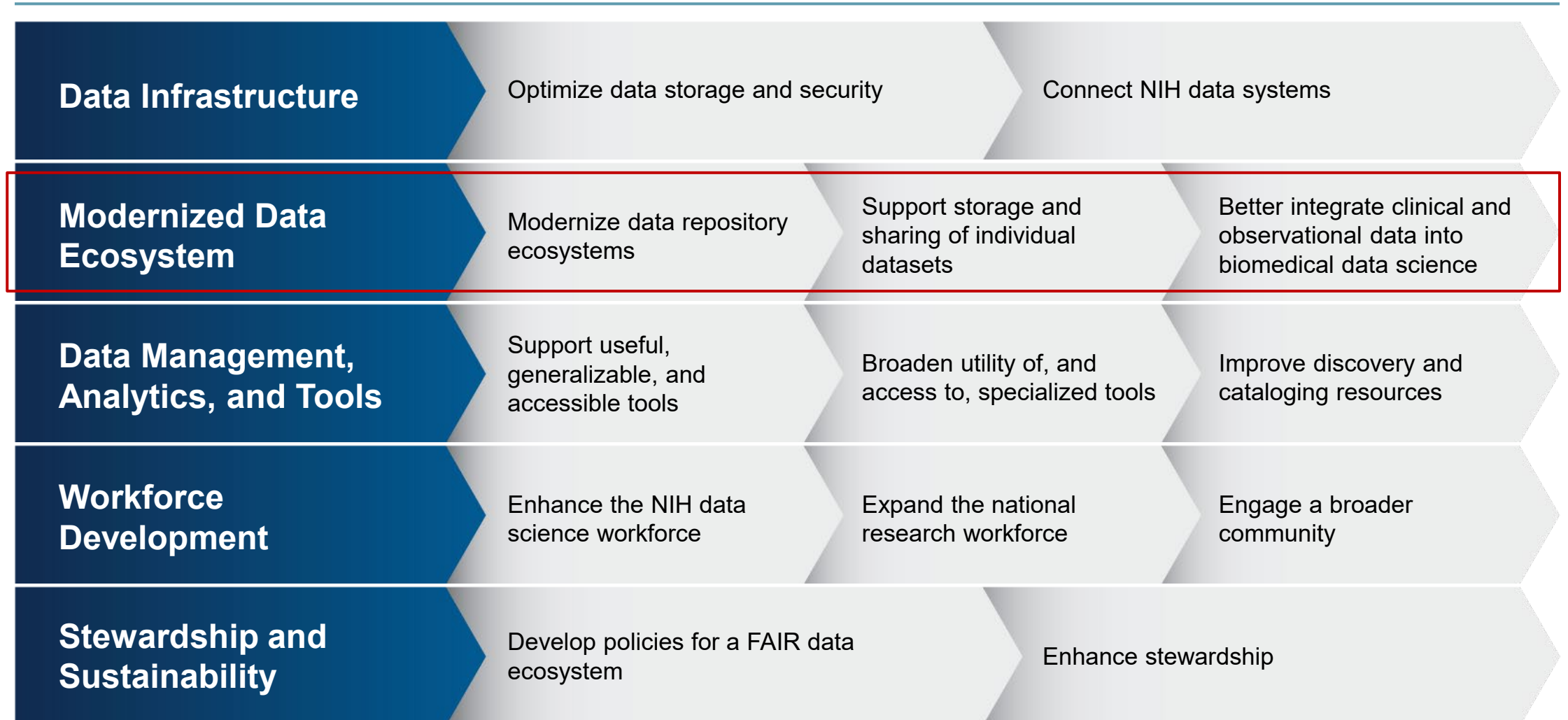
Dr. Susan Gregurick  
NIH Associate Director for Data Science  
Director, Office of Data Science Strategy

Dr. Rafael Irizarry  
Professor and Chair, Department of Data Science, Dana-Farber Cancer Institute  
Department of Biostatistics, Harvard T.F. Chan School of Public Health

Council of Councils Meeting

May 29, 2025

# NIH Strategic Plan for Data Science: Goals and Objectives



# The Generalist Repository Ecosystem Initiative (GREI) Program

## Generalist Repositories' Activities:

## Expected Outcomes and Impact:



**Implement consistent capabilities (NOT-OD-21-016)**



**Make data sharing easier**



**Increase access to and discovery of NIH-funded data**



**Improve discoverability**



**Conduct outreach and training on FAIR data practices**



**Increase reproducibility of research**



**Engage the research community**



**Encourage secondary use of data**



**MENDELEY DATA**



**OSF**



# GREI's Goals and Objectives

---

## Goals

1. Make it easier to for researchers to share data
2. Enable the discoverability of NIH-funded data across generalist repositories
3. Support greater reproducibility of NIH-funded research by ensuring data associated with publications are readily available
4. Avoid duplication of data across repositories
5. Encourage NIH-funded researchers to be both contributors and consumers to increase the reuse of data

## Objectives

1. Commit to “coopetition”
2. Implement best practices for data repositories
3. Support discovery of NIH funded data
4. Adopt consistent metadata models
5. Facilitate quality assurance and quality control
6. Connect digital objects
7. Catalog use cases supported by GREI
8. Implement open metrics
9. Prepare training materials
10. Conduct outreach



MENDELEY DATA



OSF



# GREI Work Rationale

---

The initiative has resulted in outcomes resulting in tailored community training, a common core metadata schema, and best practices for sharing data.

- ❑ Implemented common metadata fields – funder and grant ID
- ❑ Included new metadata fields in faceted search
- ❑ Normalized and comparable counting for usage metrics (views & downloads)
- ❑ Counts by institution/funder
- ❑ Connect research and researchers via PIDs
- ❑ Teaching & learning materials at <https://about.zenodo.org/projects/grei/>

NIH sought Council recommendations to guide the future focus of the initiative.

# GREI Working Group Charge

---

The charge of the GREI Working Group of the Council of Councils is to provide an assessment of the GREI's progress to date and to provide recommendations for the future of this initiative, and specifically:

- ❑ Review the current scope and goals of the GREI as well as progress to date;
- ❑ Provide recommendations on future GREI objectives and goals, based on progress and the biomedical research community's needs;
- ❑ Provide recommendations on future success measures for the GREI initiative, accounting for a diverse community of researchers

# Working Group Formed: Feb-May

---

## **Co-chairs**

- Rafael Irizarry, Ph.D.
- Susan Gregurick, Ph.D.

## **Ex-officio Member**

- Ishwar Chandramouliswaran

## **Executive Secretary**

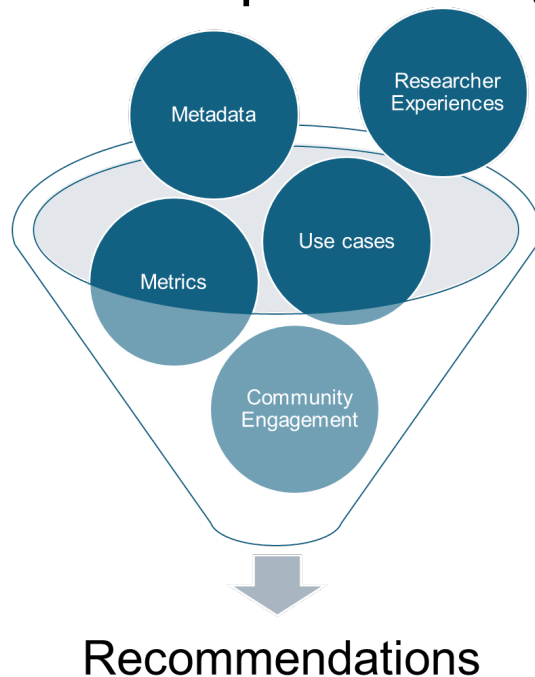
- Amanda Skarlupka, Ph.D.

## **Members:**

- Kasper Hansen, Ph.D.
- Rick Horwitz, Ph.D.
- Kari L. Jordan, Ph.D
- Jeff Leek, Ph.D.
- Susanna-Assunta Sansone, Ph.D.
- Andrew Su, Ph.D.
- Jason Williams

# Working Group Sessions

- 10 Sessions
  - Kick-off meeting
  - 5 Thematic Discussions
  - 4 Report Writing



Date & Time	Topic	Contents	Presenter(s)
June 27 <sup>th</sup> 3-4pm ET	Kickoff	Introductions, GREI program overview, Logistics	Susan Gregurick and Rafael Irizarry
August 15th 2-4pm ET	Community Engagement	Community engagement efforts taken for years 1 and 2, and the direction for year 3	Ana Van Gulick
September 5th 1-3pm ET	Use-cases	Identification of use cases and the progress toward improving	Julie Wood
September 30th 1-3pm ET	Metadata	Identification of metadata and changes to infrastructure to allow for cross platform searches	Nici Pfeiffer
October 21st 1-3pm ET	Metrics	Harmonization of metrics used by GREI	John Chodacki
November 5 2-4pm ET	Researcher Stories	Invited speakers provide data reuse research stories	Data Reuse Researchers
December 12th 2-4pm ET	Report Writing	Discuss the report and recommendations	Rafael and Susan
January 8th 1-3 pm ET	Report Writing	Discuss the report and recommendations	Rafael and Susan
January 14 2-4 pm ET	Report Writing	Discuss the report and recommendations	Rafael and Susan
March 20 3-5 pm ET	Report Finalization	Prepare presentation	Rafael and Susan
May	Council Presentation	Present to Council and publish report	Rafael and Susan



# Findings & Recommendations

---

# Findings

---

- The GREI initiative focused on supporting use cases for data sharing and mainly directed at supporting the Final NIH Data Management and Sharing Policy, was appropriate for an initial phase.
- The GREI initiative developed a baseline of consistent metadata that, at a high level, provided data discoverability and data use capabilities.
- The GREI initiative provided significant outreach and generalized training.

# Recommendations

---

WG provided areas of focus for co-opetition towards further harmonization and usability of generalist repositories.

- Support the expansive potential of **data reuse**
- Tailor future development via **community engagement**
- Develop metrics to **quantify impact**

# Support the Expansive Potential of Data Reuse

---

Focus on:

- Data reuse by biomedical researchers
- Data without a home and approaches to collect, enrich, & reuse
- Capabilities to enable faceted search

# Tailor Future Development via Community Engagement

---

Focus on:

- Expand engagement with biomedical researchers
- Engage with others in repository landscape (domain-specific repos)
- Contribute to FAIR-enabling methods and approaches
- Establish a steering committee from the community

# Develop Metrics to Quantify Impact of Biomedical Data (NIH-funded)

---

Focus on:

- Impact of data reuse and case studies
- NIH-funded researcher use cases
- Universal adoption of data citations & linking to publications
- Identify new partners to develop metrics (for biomedical data)

# Council Action

---

Vote to accept the Working Group Report on the Recommendations for the Generalist Repository Ecosystem Initiative.

# Q & A

---