NATIONAL INSTITUTES OF HEALTH GENERALIST REPOSITORY ECOSYSTEM INITIATIVE (GREI) COUNCIL OF COUNCILS WORKING GROUP REPORT

CONTENTS

Executive Summary 3

Introduction 5

- Charge of the Working Group 5
- Working Group Methodology 6
- GREI Goals and Current Status 6

Recommendation 1: Support the expansive potential of data reuse 8

Recommendation 2: Focus community engagement 10

Recommendation 3: Develop metrics to support impact of biomedical research 12

Conclusions 13

Appendix A: Working Group Roster 14

Appendix B: Working Group Timeline 15

EXECUTIVE SUMMARY

NIH-funded researchers are generating and sharing an increasing amount and variety of data. To fully capture these valuable datasets and allow for future reuse, NIH supports an ecosystem of scientific domain-specific data repositories to accommodate the influx of varying data types, storage sizes, and access controls. The focus of domain repositories is either by data type (e.g., sequence data, protein structure, continuous physiological signals) or by biomedical research discipline (e.g., cancer, immunology, or clinical research data associated with a specific NIH institute or center) and often form a nexus of resources for their research communities. These domain-specific data-sharing repositories, whether funded by NIH or other sources, are good first choices for researchers, and NIH encourages their use. However, when no domain-specific repository exists, generalist repositories, i.e., repositories that accept data regardless of data type, format, content, or discipline, may be a good choice to share research data. The Office of Data Science Strategy (ODSS) recognized existing gaps in the repository landscape for data preservation and sharing. After conducting the Generalist Repository pilot and a 2020 workshop on the role of generalist repositories to enhance data discoverability and reuse[1], in 2022, ODSS spearheaded the Generalist Repository Ecosystem Initiative (GREI).

The GREI's mission is twofold: to establish a common set of cohesive and consistent capabilities, services, metrics, and social infrastructure across various generalist repositories using a coopetition[2] model. The secondary goal is to raise general awareness and targeted outreach to facilitate researchers' adoption of the Findable, Accessible, Interoperable, and Reusable (FAIR) principles to manage, share, and reuse NIH-funded data. Thus, the GREI awardees are self-sustaining generalist repositories that receive funds from the NIH to harmonize infrastructure capabilities across entities (e.g., software architecture, data submission, quality assurance, and metadata). Since no individual entity would invest in harmonizing with competitors, NIH has chosen to provide funds to conduct this community-benefiting work.

NIH also has synergistic efforts to support data sharing and data reuse, including but not limited to funding for domain-specific biomedical data repositories, the National Library of Medicine (NLM) Data Catalog to provide a resource for searching for datasets, and the Data Curation Network to support professional development for data curators, data management experts, data repository administrators, and institutions. The GREI repositories are currently being made searchable in the NLM Data Catalog. The GREI initiative has partnered with the Data Curation Network to provide ancillary training in data management for librarians, data curators and repository managers, and researchers generating data.

In May 2024, the NIH appointed the GREI Working Group of the NIH Council of Councils. The charge of the group was to review the GREI's scope and objectives to date and provide recommendations on the future of this initiative to best serve NIH and NIH-funded researchers and highlight potential success criteria for moving forward. Following an overview presentation of the program, the working group members identified four key areas requiring focused discussions: researcher use cases, community engagement, metadata, and metrics. Over five months, the GREI awardees leading these areas co-chairs presented to the working group who asked questions and discussed their progress, resulting in recommendations and considerations.

Overall, the working group felt that the initiative provided opportunities to enhance the data sharing and reuse ecosystem by creating a framework for the generalist repositories to collaborate on shared activities. The findings and recommendations from the working group are:

- The GREI initiative, which 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.

---

[1] https://datascience.nih.gov/data-ecosystem/NIH-data-repository-workshop-summary

[2] Coopetition is a strategic concept to cooperate with competitors to achieve common goals

However, future needs must focus on data reuse and user case studies with clear impacts that convey how biomedical researchers use data in generalist repositories.

- The GREI initiative developed a baseline of consistent metadata that, at a high level, provided data discoverability and data use capabilities; however, future efforts should support the technologies, capabilities, and relevant metadata driver by user communities to fill gaps not already covered by specialist biomedical repositories.
- The GREI initiative provided significant outreach and training; however, future development efforts need an engagement strategy tailored to meet the needs of different communities, with a higher priority on collaborative projects with the larger data and resource ecosystem, including biomedical researchers, and domain-specific repositories.

The recommendations in this report are intended to provide areas of focus for the upcoming years regarding activities and objectives. The “Future Work and Considerations” section highlights the group’s next steps in considering longer-term coopetition objectives for future harmonization and usability of generalist repositories. The final recommendations are to:

1. Support the expansive potential of data reuse.
2. Tailor future development via stronger and broader community engagement.
3. Develop metrics to quantify frequency and impact of biomedical research data reuse.

INTRODUCTION

The Office of Data Science Strategy (ODSS) leads and coordinates the implementation of the NIH Strategic Plan for Data Science. This Strategic Plan provides a roadmap for modernizing the NIH-funded biomedical data science ecosystem as follows:

- supporting a highly efficient and effective biomedical research data infrastructure,
- promoting the modernization of the data resources ecosystem,
- supporting the development and dissemination of advanced data management, analytics, and visualization tools,
- enhancing workforce development for biomedical data science, and
- enacting appropriate policies to promote stewardship and sustainability.

In support of the NIH Final Data Management and Sharing Policy, in January of 2022, ODSS launched the Generalist Repository Ecosystem Initiative (GREI) based on a Generalist Repositories Pilot. From this pilot, NIH learned three key lessons:

- Generalist repositories are growing as more researchers deposit and link their publications to data in generalist repositories,
- Researchers require additional education and guidance on where to publish data and, how to describe datasets, and
- Metadata enhancement enables greater discoverability.

This initiative is part of a larger NIH and ODSS data ecosystem strategy to support established domain-specific data repositories, develop new domain-specific data repositories, improve access to controlled-access clinical data repositories, enhance healthcare data collection and harmonization, and enrich training in data management and data sharing.

While supporting specialized repositories, NIH recognized that generalist repositories fill a gap where domain or discipline-specific resources do not exist and can contribute to supporting researchers' data-sharing needs, positively impacting data sharing and discovery. Therefore, the GREI initiative was launched with seven generalist repositories to work together to establish consistent metadata, develop use cases for data sharing, and train and educate researchers on FAIR data and the importance of sharing of research objects.

The GREI's mission is twofold. The primary mission is to establish a common set of cohesive and consistent capabilities, services, metrics, and social infrastructure across the participating generalist repositories. The secondary mission is to raise general awareness and targeted outreach to encourage researchers to adopt the Findable, Accessible, Interoperable, and Reusable (FAIR) principles to better share and reuse NIH-funded data.

CHARGE OF THE WORKING GROUP

The GREI Working Group of the Council of Councils is charged with providing an assessment of the GREI's progress to date and recommendations for its future. NIH expects to use these recommendations to guide the initiative's future focus.

Specifically, the charge for the Working Group will be to:

- 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 needs of the biomedical research community.
- Provide recommendations on future success measures for the GREI initiative, accounting for a diverse community of researchers.

WORKING GROUP METHODOLOGY

Appointed in May 2024, the international members of the Working Group brought valuable expertise in FAIR data management, sharing, (re)-use, and community building (see Appendix A: Working Group Roster). The working group met virtually from August 2024 to March 2025. The initial meeting was a one-hour meeting, followed by monthly two-hour meetings. During the initial meeting, the working group was introduced to the GREI program, it's goals, and objectives. From that initial meeting, four key topics were identified for further discussion: metrics, use cases, community engagement, and metadata. The GREI awardee co-chairs were invited to present on each of these topics throughout four meetings. These key topics sessions were followed up with a session featuring three researchers who use GREI repositories. The recording and notes were made available if WG members could not attend. Each session was summarized, and the working group recommendations were recorded for further refinement. From December 2024 to January 2025, the working group drafted and edited the report asynchronously and virtually. The meeting timeline and agendas are in Appendix B: Working Group Timeline and Appendix C: Meeting Agendas.

GOALS AND CURRENT STATUS

The GREI has the following goals:

1. Make it easier for NIH researchers to share data that is not suitable for existing specialized repositories.
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 generalist repositories.
5. Encourage NIH-funded researchers to be both contributors and consumers to increase the reuse of data.

The seven awardees in GREI work together toward ten primary objectives to accomplish their goals:

- Commit to coopetition.
- Implement best practices for data repositories.
- Support the discovery of NIH-funded data.
- Adopt consistent metadata models.
- Facilitate quality assurance and quality control.
- Connect digital objects.

- Catalog use cases supported by GREI.
- Implement open metrics.
- Prepare training materials.
- Conduct outreach.

Since January 2022, GREI has accomplished the following key objectives:

- Engaged with key audiences, such as librarians, academic institutions, and biomedical research communities, to ensure training and outreach are appropriately tailored and impactful.
- Developed a common core metadata schema[3] for each generalist repository to enhance the interoperability and discoverability of datasets across repositories and allow NIH to track the impact of NIH-funded research data.
- Published community resources[4], including best practices for sharing data, guidance on including generalist repositories in NIH data management and sharing plans, and use cases for sharing and searching data in each of the GREI repositories.

---

[3] 10.5281/zenodo.8101956

[4] https://zenodo.org/communities/grei

RECOMMENDATION 1: SUPPORT THE EXPANSIVE POTENTIAL OF DATA REUSE

To date, the GREI initiative has identified and focused on six use cases:

1. As an NIH-funded **researcher**, I want to **share my data** to comply with the data management and sharing plan and the conditions of my NIH grant, but there are no discipline-specific repositories for my data type.
2. As a **researcher**, I want to **find research data** of interest so that I can validate findings, reuse data, and build on work within my discipline. Still, there are no discipline-specific repositories for the data type that interests me.
3. As a **funder**, from a specific NIH institute or in general, I want to **find datasets** that we have funded for which there are no specialized repositories so that I can report on compliance with policies and track the impact of research funding and data usage.
4. As an **institution**, I want to **report on all datasets** from my institution so that I can ensure compliance with research data sharing and management plan commitments by our researchers.
5. As an **institution**, I want to capture and **preserve research data** from my institution by using several repositories (including a generalist repository).
6. As an **institution**, I want to **use a generalist repository infrastructure** for my institutional data repository.

To support these use cases, GREI awardees have spent extensive time on metadata enhancements and user interface capabilities to improve the data uploader's experience and provided use case information guides that contain key information and features for each user.

1.1: DATA REUSE BY BIOMEDICAL RESEARCHERS MUST BE THE HIGHEST PRIORITY.

The working group acknowledged that although each use case listed above has value, the most impactful use case requires investigators to effectively search for, find, and reuse data to support their research objectives. The other use cases are important but should support or demonstrate the researcher's ability to find and reuse data. In addition, the working group noted that the current list of use cases provided did not capture publishers' role in the data-sharing environment. Likewise, for some data, instrument manufacturers deliver data with proprietary formats that are not easy to deal with, hindering the development of analytic tools. Use cases for these key stakeholders need to be included to support the biomedical research working group's recommendation to prioritize data reuse.

While the prior use cases were useful organizing principles in the earlier stages of GREI's development, the working group strongly desires to see them evolve into actual user stories and case studies. Although the case studies presented during the meetings had compelling elements, the working group felt that the GREI initiative had reached a level of maturity where extensive interactions with NIH-funded investigators is critical to developing practical solutions others are likely to adopt.

Since all use cases should support efficient data reuse, the connections between the differing use cases and their outputs should be organized and measured by their assessments to ensure that they will positively impact data reuse. The development and implementation of use case solutions can then be prioritized by their potential benefit to NIH and NIH researchers. Therefore, the working group also recommends supporting the development of a framework that illustrates how each use case impacts and supports biomedical data reuse, ensuring cohesive connections and promoting data sharing. The working group recommends that as part of this framework, the NIH consider creating a logic model to illustrate how secondary use cases mutually strengthen primary use cases in biomedical data reuse.

1.2: ENGAGE WITH BIOMEDICAL RESEARCHERS TO IDENTIFY DATA THAT HAS NO HOME AND IS CONSIDERED A PRIORITY FOR REUSE AND DEVELOP A METHODOLOGY TO CAPTURE, STORE, AND SERVE IT POWERED BY RICHER DESCRIPTIONS.

Generalist repositories intend to accept all data types with minimal limitations. Therefore, required metadata for generalist repositories is collected as domain-agnostic study-level information instead of file/data-level metadata. By its very nature, generalist repositories need to allow arbitrary metadata. However, this increases the complexity of data search and validation. The metadata for the GREI initiative has focused on dataset-level descriptors (e.g., authors, funders, research organizations). It uses the community driven DataCite Metadata Schema[5], for persistent identifier (PID) assigning data repositories, to identify and recommend metadata elements for the GREI repositories to use (https://zenodo.org/records/8101957). The initiative is harmonizing metadata across the generalist repositories to allow for data cataloging and cross-data repository search.

However, GREI partners should also consider enabling search by data type, more granular concepts, and their relationships and cater to biomedical data types where no domain-specific repositories exist. To do this, the GREI members should also be aware of the dynamic landscape of thousands of community-driven data and metadata standards (https://doi.org/10.1038/s41587-019-0080-8) that have identified the most highly sought-after characteristics to describe and share a given data or study type. Whilst some standards are well established and already implemented by a repository, other standardization efforts are at various stages of their life cycle (e.g., design or testing), which is a complex combination of technical and social engineering processes (https://doi.org/10.6084/m9.figshare.3795816.v2). To navigate such a dynamic landscape, the GREI members should also collaborate with resources and networks, such as FAIRsharing[6], that can help them track and identify those communities who are at an early stage of the standardization process and, ideally, have already identified their metadata needs. NIH should also support the initiative to engage with relevant experts (e.g., data scientists and analysts, data reuse researchers, and database maintainers in neighboring domains) and establish sustained collaborations. Working with and for these identified communities will enable GREI partners to demonstrate how their core repository infrastructure can actually (a) capture this richer domain-specific metadata, (b) store it along any relationships among the concepts, and (c) serve via their search interfaces to drive queries by humans and machines (via API). To ensure metadata is FAIR and actionable, GREI members should leverage tools for templating, such as CEDAR[7], reuse existing standards (such as controlled vocabularies or ontologies for certain established fields), and consider validation tools to check for compliance to the agreed minimal set of descriptors. Ultimately, if successful, this, in turn, will provide the biomedical community with a framework and exemplars (i) showcasing storage, sharing, and search capabilities for data types or studies that had no home and (ii) improving the ability of biomedical researchers to search, find and reuse this data based on sought-after characteristics and more granular concepts rather than simply by its primary publication, authors, funders, and organizations.

1.3: ADD CAPABILITIES THAT ALLOW DATASETS TO BE SEARCHABLE AND FINDABLE BY DIFFERENT PARAMETERS.

Data reuse is a broad concept (see https://datascience.codata.org/articles/910/files/submission/proof/910-1-5817-1-10-20190620.pdf and https://pmc.ncbi.nlm.nih.gov/articles/PMC7518187/pdf/peerj-08-9954.pdf), and the GREI initiative must specifically define the data reuse cases they are prioritizing in

---

[5] https://schema.datacite.org/

[6] https://doi.org/10.1038/s41587-019-0080-8

[7] https://doi.org/10.1093/jamia/ocv048

alignment with investigator's current and future needs. For instance, biomedical researchers may wish to search for a specific dataset associated with a specific source (e.g., a primary research article). They will track this data down through citation or author communications. On the other hand, biomedical researchers may wish to locate many datasets of a specific class and type that fits their research question (e.g., lung images). Because biomedical researchers are interested in data by various parameters (type, study, scientists, etc.), the NIH should consider flexible capabilities that enable data discoverability within the GREI initiative and more broadly. The research communities' need for data discoverability will likely include generalist, institutional, and domain-specific data repositories, particularly for new and emerging areas in artificial intelligence, such as multimodal AI that involve combining multiple data types to draw more accurate insights. As AI plays a more significant role, the GREI initiative must also prepare for the reuse potential enabled by increasingly AI-capable approaches to extracting information from possibly unstructured data. Therefore, specific definitions and use cases for reuse can assist NIH in supporting the prioritized cases and providing appropriate metrics. Considering this, NIH could explore and encourage potential partnerships with programs such as NAIRR (National Artificial Intelligence Research Resource) Pilot and B2AI (Bridge to AI).

RECOMMENDATION 2: FOCUS COMMUNITY ENGAGEMENT

Community engagement is a core component of two GREI objectives: conducting outreach and preparing training materials. The primary audiences included data librarians and data stewards at NIH-funded institutions, researchers either with or applying for NIH funding, and scientific research communities. Additional stakeholders are journals, publishers, research infrastructure providers, specialized repositories, NIH staff, and other federal agencies and funders.

Over the first two years of the initiative, the GREI awardees have provided more than fifteen webinars and workshops, presented at more than twelve conference presentations, established an email subscriber community of over 400 individuals, published twelve blogs, and shared twenty-eight public outputs that have been collectively downloaded more than 19,500 times. Further, through these activities, GREI awardees collected user and attendee feedback to inform community needs and help set new priorities and outputs for the upcoming years.

2.1: EXPAND ENGAGEMENT WITH BIOMEDICAL RESEARCHERS.

The researcher category is composed of a diverse group of individuals: trainees and staff scientists generating and collecting data, post-doctoral and principal investigators identifying potential repositories for DMS plan compliance, manuscript authors and research group data managers responsible for uploading and sharing data to repositories, researchers at institutions with and without robust library infrastructure and resources, researchers looking for data mentioned in a manuscript, and researchers looking for certain data types without a motivating primary source. While there is overlap between these audiences, the outreach approach for audiences will need specific target outcomes, which may not all have the same priority or urgency. To maximize the impact of the GREI initiative, the working group recommends that the NIH adopt a diversified engagement strategy tailored to the needs of different communities, with a higher priority placed on engaging biomedical researchers.

Moreover, the GREI initiative has done substantial work to ensure researchers can comply with NIH policies. Most of the training and outreach focused on engaging the librarian community, given their responsiveness. The working group acknowledged that researchers are not utilizing institution librarians and information management specialists to find data or select appropriate repositories for their DMS plan or final datasets. However, the working group felt that the GREI initiative may be over-emphasizing the train-the-trainer approach and outreach to librarians of these institutions, as researchers are not willingly engaging with the information. The NIH should acknowledge a potential communication gap and plan appropriately to minimize the effects. Steps, such as in researcher-focused workshops, could remediate this gap. Since GREI awardees have a significant footprint in the data-sharing community, they should be attentive to new strategies and flexible in adopting what works and shedding what does not.

Finally, the working group felt that the GREI activities have been communicated to a wide range of audiences through webinars and presentations. Instead of broad content, the working group recommends that workshops or sessions should be more tailored to a specific audience and include targeted instructional videos to have the greatest impact. For example, trainees, students, and individuals responsible for data collection need more in-depth training on how to properly collect metadata required for GREI repositories. In contrast, librarians need information regarding search, discovery, and cataloging data holdings.

2.2: STRENGTHEN THE ENGAGEMENT PROCESS, IDENTIFY COLLABORATIONS, AND STREAMLINE COMMUNICATIONS BETWEEN GENERALIST AND DOMAIN-SPECIFIC REPOSITORIES, CLARIFYING THEIR DIFFERING ROLES.

Generalist repositories are critical in data warehousing and dissemination for data types for which no domain-specific repository exists. Domain-specific repositories have a clear scope and types of data they accept and serve. There is also a wealth of other scholarly research objects (also required by funders, publishers, and institutional data policies) that are not in scope for some domain-specific but are suitable for generalist repositories. However, when these research objects are related, they should be linked, (e.g., cross-referencing a grant report deposited in OSF for example with the corresponding datasets accessible in the Biological Magnetic Resonance Data Bank). Link-related research objects seamlessly and at a scale, and with minimal burden to users, requires continued communications with NIH's Data Repositories and Knowledgebases (DRKB) network as a starting core within the NIH to build partnerships, strengthen relationships and identify solutions that are potentially applicable to domain-specific, and institutional repositories as appropriate.

As the biomedical data landscape continually grows and evolves, it will be important to clarify the specific mechanisms through which generalist, institutional, and domain-specific repositories synergize to promote data submissions and reuse. Recognizing the emergence of new challenges, the working group encourages NIH to provide a forum for the generalist repositories to strengthen cooperation with the domain repositories. This will require clear roles and responsibilities.

2.3: PLACE THE GENERALIST REPOSITORIES AS AN ACTIVE PARTICIPANT OF AND CONTRIBUTORS TO THE BROADER FAIR-ENABLING ECOSYSTEM OF TOOLS, SERVICES, AND INFRASTRUCTURES.

The current GREI's secondary mission is to facilitate researchers' adoption of the FAIR principles to better share and reuse data. This also requires the GREI repositories to be FAIR, within the limits of their scope. However, the number of existing community-developed FAIR assessment/assistance tools do not deliver comparable results, mainly due to the fact that they do not test the same thing in the same way. Therefore, there is no way to objectively prove any claims of FAIRness from GREI or other parties as a matter of fact. Some GREI repositories are already working to implement a community-driven approach (https://doi.org/10.5281/zenodo.7463421), also approved by the European Open Science Cloud (EOSC), to address this challenge. The strategy involves exposing the same metadata pattern to the FAIR evaluation and assessment tools and creating a uniform metadata harvesting environment for the FAIR tools (as well as other machines and agents) to ultimately deliver comparable results. However, this is just the first step to ensure generalist repositories, and their data promote improved FAIRness, and the results should be tested and showcased in practice. Furthermore, GREI partners should work more closely with the growing ecosystem of FAIR-enabling tools and services, especially numerous under the EOSC umbrella and its funded infrastructure projects (e.g., OSTrails), as well as relevant biomedically-focused research infrastructures (e.g., ELIXIR).

2.4: ESTABLISH A STEERING COMMITTEE OF KEY STAKEHOLDERS TO ENSURE THE DATA ECOSYSTEM IS ALIGNED AND INFORMED.

The NIH should consider including the broader research community's needs and input into future initiatives, including domain-specific repositories, data-stewards, data generators, both large-scale consortia and individual researchers, and researchers who are experts in data sharing and reuse, as well as from the FAIR-enabling ecosystem of tools and resources. The working group believes an external steering committee would satisfy this gap.

RECOMMENDATION 3: DEVELOP METRICS TO SUPPORT IMPACT OF BIOMEDICAL RESEARCH

The open metric development for the GREI initiative intersects with four of the GREI objectives: adopt consistent metadata models, conduct outreach, catalog use cases supported, and implement open metrics. In year one, the initiative conducted a landscape analysis of the open data metrics proposed by the data reuse community. Working with the community, GREI awardees used DataCite and MakeData Count as a standard to focus on repository usage (i.e., page views and downloads) and data citations (i.e., connecting data to publications). By year two, the awardees normalized their approaches across the repositories and began implementing tools for metric collection in year three. The working group realizes it is important to collect similar metrics across the generalist repositories participating in the GREI initiative. However, they also noted that dataset download metrics differ from others, highlighting data reuse in the biomedical community. The working group feels the initiative would benefit from flexibility in developing and collecting metrics, acknowledging that metrics may mean different things in different communities.

3.1: FOCUS METRICS DEVELOPMENT AND IMPLEMENTATION TO ILLUSTRATE IMPACT OF DATA REUSE AND USER STORIES

Building on the early work in the GREI initiative, the working group recommends that NIH focus on illustrating the impact of data reuse from generalist repositories and tie these metrics back to real-world user stories and case studies. Further, new and existing partners should be flexible in developing metrics, which will need to be integrated within the larger community of data sharing and reuse. To decrease the reliance on one entity, it is important to ensure that other partners are evaluated and brought into the initiative to ensure this flexibility.

The working group recommends that metrics identification, development, and implementation should be prioritized by the magnitude of their impact on NIH-funded researchers. Current metrics development is broad in scope, and the impact is too encompassing, covering the entire data reuse community. Metrics and their development should be more related and tied to biomedical use cases specifically beneficial to NIH researchers and biomedical areas funded by the NIH, not just to the data reuse community. The working group also recommends that NIH consider a mechanism for how datasets are cited and referenced in publications.

CONCLUSIONS

The working group acknowledges the NIH's forward-thinking strategy to engage and support both domain-specific data repositories and to create a coopetition model for generalist repositories that ensures collaboration when appropriate but also respects the need to allow for individual innovations. The initiative's first three years laid a foundation that could be built upon for future work. By focusing on supporting the final NIH Data Management and Sharing Policy, the initiative made progress in defining use cases, developing consistent, albeit surface-level, metadata, as well as outreach and training targeting communities of librarians, data curators, and, to a lesser extent, researchers. In some cases, the working group felt that it was difficult to determine the actual impact on the biomedical research community from the initiative, in particular claims of FAIRness.

Therefore, the recommendations outlined in this report aim to focus the initiative on biomedical researchers, current or potentially NIH-funded biomedical researchers, who need to share data within generalist repositories. The focus on biomedical data reuse, in terms of user stories, community engagement, and metrics, is a high priority for the working group. The working group also recommends that the initiative integrate and collaborate more strongly with the domain-specific repositories and other activities in the larger biomedical data-sharing ecosystem. Finally, the working group recommends that the initiative work to support evolving metadata with technologies that enhance data discovery, not only within the generalist repositories but also across other data resources.

APPENDIX A: WORKING GROUP ROSTER

Co-Chairs

Rafael Irizarry, Ph.D.
Council of Councils Member (2028)
Professor
Department of Data Science, Dana-Farber Cancer Institute
Department of Biostatistics, Harvard T.F. Chan School of Public Health

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

Members

Kasper Hansen, Ph.D.
Associate Professor
McKusick-Nathans Institute of Genetic Medicine
Department of Biostatistics
Johns Hopkins University

Rick Horwitz, Ph.D.
Council of Councils Member (4/30/2024)
Executive Director, Emeritus and Senior Advisor
Allen Institute for Cell Science

Kari L. Jordan, Ph.D.
Executive Director
The Carpentries

Jeff Leek, Ph.D.
Vice President and Chief Data Officer
Professor, Biostatistics Program
Public Health Sciences Division
Fred Hutchinson Cancer Center

Susanna-Assunta Sansone, Ph.D.
Director, Oxford e-Research Centre
Professor of Data Readiness, Department of Engineering Science
University's Academic Lead for Research Practice
University of Oxford, UK

Andrew Su, Ph.D.
Professor
Elden and Verna Strahm Chair for Medical Research
Department of Integrative Structural and Computational Biology
The Scripps Research Institute

Jason Williams
Assistant Director, Diversity and Research Readiness, External Collaborations
Cold Spring Harbor Laboratory, DNA Learning Center

Ex-officio Member

Ishwar Chandramouliswaran
Program Director for Generalist Repository Initiative Ecosystem (GREI)
Office of Data Science Strategy, NIH

Executive Secretary

Amanda Skarlupka, Ph.D.
Program Officer
Division of Cancer Prevention, NCI

APPENDIX B: WORKING GROUP TIMELINE

Date & Time	Topic	Contents	Presenter(s)
Jun 27, 2024 3-4pm ET	Kickoff	Introductions, GREI program overview, Logistics	Susan Gregurick and Rafael Irizarry
Aug 15, 2024 2-4pm ET	Community Engagement	Community engagement efforts taken for years 1 and 2, and the direction for year 3	Ana Van Gulick
Sep 5, 2024 1-3pm ET	Use-cases	Identification of use cases and the progress toward improving	Julie Wood
Sep 30, 2024 1-3pm ET	Metadata	Identification of metadata and changes to infrastructure to allow for cross platform searches	Nici Pfeiffer
Oct 21, 2024 1-3pm ET	Metrics	Harmonization of metrics used by GREI	John Chodacki
Nov 5, 2024 2-4pm ET	Researcher Stories	Invited speakers provide data reuse research stories	Data Reuse Researchers
Dec 12, 2024 2-4pm ET	Report Writing	Discuss the report and recommendations	Rafael and Susan
Jan 8, 2025 1-3 pm ET	Report Writing	Discuss the report and recommendations	Rafael and Susan
Jan 14, 2025 2-4 pm ET	Report Writing	Discuss the report and recommendations	Rafael and Susan
Mar 20, 2025 3-5 pm ET	Report Finalization	Prepare presentation	Rafael and Susan
May 29, 2025	Council Presentation	Present to Council and publish report	Rafael and Susan