Contribution ID: **53** Type: **Session**

Stronger together: Advancing the data repository ecosystem through strategic coopetition

Monday, October 13, 2025 2:30 PM (1h 30m)

Overview
The NIH Office of Data Science Strategy launched the Generalist Repository Ecosystem Initiative (GREI) in February 2022, in recognition of the key role generalist repositories play in the NIH data sharing landscape to support the FAIR sharing of data and other research outputs. The GREI program represents a groundbreaking collaborative model that brings together seven repositories (Dataverse, Dryad, Figshare, Mendeley Data, Open Science Framework, Vivli, and Zenodo) to enhance data sharing and reuse in alignment with NIH's mission. By prioritizing user needs, adopting shared standards, and creating flexible governance, GREI has modeled an innovative approach to balancing competition and cooperation among diverse repository partners.

In this session, we will build off a previous GREI presentation at SciDataCon in Salzburg in 2023 highlighting work in progress at the midpoint of the program. As we approach the next phase of the program, we look forward to sharing the impact of the work we have accomplished together, as well as the challenges and successes of the coopetition model itself.

Over the last four years, the initiative has focused on improving data accessibility, standardizing metadata, and facilitating comprehensive metrics, all in an effort to create a more interoperable research data ecosystem, both within and beyond generalist repositories. By working collaboratively across repositories and through strategic partnerships with organizations and community initiatives like DataCite, the Carpentries, and ROR, GREI has managed to create collective progress that would not have been possible through individual repository efforts. Critically, GREI's success lies in its ability to leverage each repository's unique strengths while maintaining a shared vision of open and FAIR data sharing infrastructure.

As we look to the future of our initiative, we are seeking feedback from the research data community on our work thus far and suggestions for where we might collectively turn our focus next. In this session, we will seek audience input to gather feedback on the work we have done on interoperable metadata, consistent metrics, and community engagement, and will mediate a forward-looking audience discussion on future GREI priorities to serve the research data ecosystem.

Session Agenda
Our proposed session agenda includes short presentations from GREI team members highlighting GREI accomplishments, impact, and future directions. We will thread interactive polls throughout the session, and follow this with a facilitated discussion (~30 minutes) and Q&A focused on the SciDataCon audience's suggestions for future GREI efforts.

Overview of the GREI program - the NIH perspective
Proposed speaker: Ishwar Chandramouliswaran, NIH Office of Data Science Strategy
This presentation will expand on the program history and overview shared in the Session Overview above, and give the NIH perspective on both the role of generalist repositories in the NIH data sharing landscape and the rationale behind the GREI program.

Coopetition way of working - challenges & successes
Proposed speaker: Kristi Holmes, Northwestern, Zenodo
This presentation will describe how GREI demonstrates the potential of coopetition to advance open science and collectively address complex challenges to create a more interoperable research data infrastructure. By working together, the seven GREI repositories have produced collective improvements that would have been unattainable by siloed individual repositories.

Highlights & impacts from 4 years of GREI

Metadata & Search
Proposed speaker: Mark Hahnel, Figshare
This presentation will showcase the development of our refined core metadata recommendation based on the DataCite Metadata Schema, through which we have strategically enhanced data discoverability and compliance with NIH data sharing policies. We will detail how we have standardized our handling of optional fields, introduced controlled vocabularies, and integrated persistent identifiers to streamline research data workflows and improve data interoperability across repositories.

Common Metrics
Proposed speaker: John Chodacki, California Digital Library, Zenodo
This presentation will explore how GREI has developed and continues to implement common metrics through the Make Data Count principles, creating a standardized approach to tracking data usage, citation, and impact across repositories. We will show how these harmonized metrics provide actionable insights for researchers, funders, and institutions, ultimately recognizing and reinforcing the value of open data sharing practices.

Community Outreach
Proposed speaker: Traci Snowden, Mendeley Data
This presentation will trace the evolution of GREI's extensive community engagement efforts, revealing how we have built capacity and support for researchers navigating data management and sharing requirements via targeted webinars, workshops, training materials, and collaborative partnerships. Our proactive approach to outreach has ensured we address emerging needs, engage diverse research communities, and foster trust through transparent feedback mechanisms.

Real-world user stories highlighting program impact
Proposed speaker: Ana Van Gulick, Figshare
This presentation will share compelling real-world user stories that illustrate how GREI's collaborative efforts have made NIH data sharing and reuse more impactful. By highlighting concrete examples of researchers benefiting from improved metadata standards, enhanced discoverability, and streamlined repository workflows, we will demonstrate the tangible value of GREI's innovative approach to open science.

Future directions for collaboration
Proposed speaker: Ana Van Gulick, Figshare
This presentation will explore future directions for GREI, focusing on improving data value and interoperability across repositories through enhanced metadata, connected digital objects, and thoughtful AI integration for data curation and metadata enhancement. We will also prioritize user engagement and incentives for transparent data sharing. Potential future tasks include developing a repository maturity model, facilitating federated search, and fostering a culture of data sharing and reuse to advance the research data ecosystem.

Facilitated audience discussion and Q&A

Primary author: VAN GULICK, Ana (Figshare, Digital Science)

Co-authors: CHANDRAMOULISWARAN, Ishwar (NIH Office of Data Science Strategy); HOLMES, Kristi (Northwestern University); Dr HAHNEL, Mark (Figshare, Digital Science); BUYS, Matt (DataCite)

Presenters: VAN GULICK, Ana (Figshare, Digital Science); CHANDRAMOULISWARAN, Ishwar (NIH Office of Data Science Strategy); HOLMES, Kristi (Northwestern University); Dr HAHNEL, Mark (Figshare, Digital Science); BUYS, Matt (DataCite)

Track Classification: SciDataCon2025 Specific Themes: Open research through Interconnected, Interoperable, and Interdisciplinary Data