SciDataCon 2025



Contribution ID: 175

Type: Presentation

FAIRifying at Scale: Lessons from NIAID's Ecosystem-Wide Approach to Repository Interoperability

Wednesday 15 October 2025 15:06 (11 minutes)

As global communities continue to adopt the FAIR Principles, many organizations face the challenge of not just FAIRifying individual datasets, but entire ecosystems of data repositories and services. The many tools and assessments developed for FAIRtend to be customized to a particular data architecture, especially data organized in a file with a DOI and listed on a website. The National Institute of Allergy and Infectious Diseases (NIAID), an arm of the U.S. National Institutes of Health, manages a data landscape that includes infectious, allergic, and immunologic data. NIAID data has been at the forefront of research and therapeutics for today' s biggest public health challenges, and this project targets increasing the discovery and reuse of these data. With this diversity of topic, use, and audience, assessing NIAID data resources for FAIRness brings a host of challenges: diverse data architectures, varying levels of entity management, declaration, and resolution, and data that has been collected over many time scales—decades in some cases. This session will share insights and actionable strategies from the NIAID Data Landscaping and FAIRification project undertaken by GO FAIR US and partners including:

-Framing a data resource landscape for an organization

-Customized FAIR assessment instruments

-Approaching record-based repositories (as opposed to file-based ones)

-Constructing a FAIR common strategy for your organization

-Assessing individual data resources' progress towards a common

approach

-Using Impact assessment to drive adoption of community- and

standards-based metadata and PIDs

-Collecting and extending assessment data in a FAIR, AI-Ready way for

further exploration

-Increasing the value and use of citations for all types of data products

The session will delve into a novel, stakeholder-informed method for assessing how well data repositories meet the FAIR Principles—developed and applied by GO FAIR US with support from GO FAIR Foundation with speakers from GO FAIR US, NIAID, and National Center for Atmospheric Research. The method builds on and extends prior impactful work1,3 from and including FAIRsFAIR2, the RDA Data Maturity Model Working Group4, the National Science Foundation's EarthCube5, GO FAIR Foundation's FAIR Implementation Profiles, and others. We will also present a number of novel tools that have been developed to maximize our understanding, and discuss how these have supported our progress on the project.

For this project we designed tailored questionnaires and interviews for different repository roles (managers, technical staff, data depositors, data users) to provide a multi-dimensional view of FAIR practices. The assessment products combine desk-based research and structured templates to create a detailed FAIR baseline for each repository—including metadata practices, technical infrastructure, and governance. This addresses the limitations of automated tools when assessing complex or secure repositories, and offers a qualitative, human-centered alternative. The assessment is then used to produce targeted FAIRification strategies that align with a repository's goals, constraints, and domain-specific practices. This method enables comparative analysis across multiple repositories and informs broader strategic planning for FAIR implementation, while supporting repository ownership and buy-in. We will discuss several common challenges including inconsistent understanding of metadata, the need for trust and buy-in from repository staff, and difficulty in addressing

interoperability goals while supporting unique repository approaches. In addition, our API strategy considers a landscape with hundreds of diverse data systems that implement web-based Application Programming Interfaces (APIs) in many different ways. The sociotechnical aspects of top-down vs. bottom up change and how these impacted the FAIRification process also will be discussed.

We will present a toolkit of the instruments, processes, and strategies developed by the GO FAIR US team. The results support collecting, analyzing, and communicating FAIR knowledge of the ecosystem, and motivating desirable changes. Specific advances include a detailed framework for FAIR data collection (on-line and in-person role-based questionnaires), an analysis of methods to evaluate record-based repositories (i.e., that don't contain data files), development of repository profiles and FAIR summaries, and the use of the FAIR Implementation Profile (FIP) and FIP Wizard to collect raw repository information that is then analyzed by a FAIR Evaluator tool.

Lastly, the session will illuminate the steps and challenges related to FAIRifying the information. That is, how does one collect FAIR assessment data that is AI Ready and can be converted into a knowledge graph for further exploration? Audience questions, feedback, and discussion will be encouraged. The session is designed for data producers, data repository or resource owners and technical managers, data stewards, and those interested in gaining further data stewardship skills.

References:

- 1. European Commission: Directorate-General for Research and Innovation, European Research Data Landscape Final report, Publications Office of the European Union, 2022, https://data.europa.eu/doi/10.2777/3648
- 2. FAIRsFAIR's "Fostering FAIR Data Practices in Europe" project documentation
- 3. Mathers, B.J., L'Hours, H., Increasing the Reuse of Data through FAIR-enabling the Certification of Trustworthy Digital Repositories, https://doi.org/10.5334/dsj-2020-041 which explores the alignment of the FAIR Data Principles with the CoreTrustSeal Trustworthy Digital Requirements
- 4. Recommendations from a Research Data Alliance Data Maturity Model Working Group, which identifies FAIR data maturity model indicators (Bahim et al., 2020)
- 5. Questions that EarthCube Office staff used to interview EarthCube project leads about the organization' s impact (Stocks and Evans, 2022)
- 6. "Data Science Dispatch."NIAID NIH. 2025. https://www.niaid.nih.gov/research/data-science-dispatch
- 7. "GO FAIR US awarded a NIAID FAIR data and ecosystem contract by Frederick National Laboratory for Cancer Research." GO FAIR US, Dec.
 - (a) Press release. https://www.gofair.us/post/go-fair-us-awarded-a-niaid-fair-data-and-ecosystemcontract-by-frederick-national-laboratory-for-can

Primary authors: KIRKPATRICK, Christine (San Diego Supercomputer Center / CODATA); GRAYBEAL, John (San Diego Supercomputer Center / GO FAIR US)

Presenters: KIRKPATRICK, Christine (San Diego Supercomputer Center / CODATA); POKUTNAYA, Dasha (NIAID/NIH); FILS, Doug (San Diego Supercomputer Center / GO FAIR US); GRAYBEAL, John (San Diego Supercomputer Center / GO FAIR US); MAYER, Lisa (NIAID/NIH); MAYERNIK, Matthew (NCAR); HOEBELHEINRICH, Nany (Knowledge Motifs LLC / GO FAIR US); SHABMAN, Reed (NIAID/NIH); VAN PANHUIS, Wilbert (Previously NIAID/NIH)

Session Classification: Presentations Session 7: Open research through Interconnected, Interoperable, and Interdisciplinary Data

Track Classification: SciDataCon Persistent Themes: Open Data, FAIR Data, Innovation, Industry and Development