



Contribution ID: 177

Type: Session

Bridging the Data Science and Research Data Communities Through Education and Shared Practices

Tuesday, October 14, 2025 11:30 AM (1h 30m)

The data science and research data communities share many common goals and challenges. Despite this, the two communities tend to have separate venues for convening, membership, and educational tracks. This session of short presentations and panel discussion will explore some of the ways that these two worlds can come together in the areas of education, training, and data stewardship practices.

This session explores the evolving intersection of the research data and data science communities through diverse lenses ranging from foundational stewardship to citizen engagement with speakers from across the world. Leo Lahti emphasizes the critical journey from observation to interpretation, underscoring the importance of data throughout the research lifecycle. Daphne Raban highlights how data stewardship serves as a vital bridge between research and data science, ensuring that data is managed, documented, and reused effectively. Phil Bourne further examines this bridge, focusing on practical integration between data science techniques and robust research data infrastructures and suggesting ways the organizations that support both communities can come together. Padmanabhan Seshaiyer brings an educational perspective, advocating for embedding research data principles into K-12 and community college bridge programs to foster inclusive data science literacy. Kelsey Druken will discuss how ACCESS-NRI is embedding FAIR through software and data workflows for Australia's climate modelling infrastructure. Together, these contributions reveal key synergies and a shared commitment to building interoperable, ethical, and impactful data ecosystems.

The session is meant to elicit ideas and suggestions for action assembled from the audience as well. An outcome will be input for a roadmap and a similar session proposal at the next Academic Data Science Alliance (ADSA) meeting. This proposed session, as well as the one at ADSA, are unique and have not been held before. The closest approximation were the prior sessions that sought to bridge the gap between the research data community and high performance computing (HPC). Sessions were held at IDW 2022, 2024, ISC (IDW for HPC people in Europe) and Supercomputing 22 & 23. Establishing baselines of understanding and presenting shared priorities and goals was key for driving progress and creating awareness of the gap. Aside from companion presentations, a desired outcome might be a future CODATA Task Group, an RDA interest group, or an ADSA activity. Such a group would look at a roadmap for shared education - including professional development and training opportunities, as well as ecosystem tools and services, and shared research priorities.

Issues to Be Addressed by the Session

1. Fragmentation Across Communities

Despite overlapping goals, research data and data science communities often operate in siloed venues with different infrastructures, memberships, and training programs. This session will explore how to foster cross-community dialogue and collaboration.

2. Disconnect Between Practice and Infrastructure

Research data practices (e.g., data curation, stewardship, provenance) are not always integrated into the day-to-day workflows of data science, limiting reproducibility, transparency, and reuse. How can we better embed stewardship into data science infrastructures?

3. Educational Gaps and Opportunities

There is a lack of integrated education and training pathways that bridge research data management and data science skills. The session will address how to co-develop curricula, professional development, and early education programs that reflect both domains.

4. Institutional and Organizational Coordination

Many data-related and data science-related consortia (e.g., CODATA, ADSA, RDA, WDS, GO FAIR) operate in overlapping and unique domains. The session will discuss how these entities might coordinate activities,

policies, and collaborate on funding priorities to support shared goals.

5. Roadmapping Shared Goals

There is currently no unified roadmap for aligning the research data and data science ecosystems. The session aims to collect input from the audience to help define shared priorities, pain points, and actions for a future roadmap and collaboration agenda.

6. Sustainability of Shared Ecosystems

Tools, standards, and services that support FAIR, open, and ethical data practices often struggle with sustainability. The session will explore how shared infrastructures and cross-community collaboration can improve the resilience and utility of data ecosystems.

7. Governance, Ethics, and Impact

As both fields intersect with sensitive data (health, environment, education), there's a need for shared approaches to governance, equity, and ethical AI/data practices. How can we co-create responsible frameworks for stewardship across domains?

Speakers

Leo Lahti, University of Turku, Finland, CODATA Executive Committee: From Observation to Interpretation
Daphne Raban, University of Haifa, Chair Israel CODATA NC: The Role of Data Stewardship in Research and Data Science

Phil Bourne, University of Virginia, US National Committee for CODATA, ADSA Board member: Bridging Data Science and Research Data

Padmanabhan Seshaiyer, George Mason University, US National Committee for CODATA: Weaving research data lessons into K-12 and community college data science bridge programs

Kelsey Druken, Australian National University: Embedding FAIR through software and data workflows for Australia's climate modelling infrastructure

Panel moderated by **Christine Kirkpatrick**, San Diego Supercomputer Center, US National Committee for CODATA

Primary authors: CARROLL, Bonnie (Co-Chair, US National Committee for CODATA); KIRKPATRICK, Christine (San Diego Supercomputer Center / CODATA); DRUKEN, Kelsey (ACCESS-NRI); LAHTI, Leo (University of Turku); SESHAIYER, Padmanabhan (George Mason University, US National Committee for CODATA); KOSTI, Rania (National Academies of Sciences, Engineering, Medicine)

Presenters: KIRKPATRICK, Christine (San Diego Supercomputer Center / CODATA); RABAN, Daphne (University of Haifa, CODATA Israel NC); DRUKEN, Kelsey (ACCESS-NRI); LAHTI, Leo (University of Turku); SESHAIYER, Padmanabhan (George Mason University, US National Committee for CODATA); BOURNE, Phil (University of Virginia, US National Committee for CODATA, ADSA Board member)

Track Classification: SciDataCon Persistent Themes: Data Science and Data Analysis