



Contribution ID: 188

Type: **Session**

Early Career Researcher perspectives on data repositories across disciplines, geographies and cultures

Tuesday, October 14, 2025 4:00 PM (1h 30m)

Data is of ever increasing value to the global research ecosystem, as underlined by recent emphasis on the FAIR principles and Open Science. Research infrastructures and data repositories are key to enabling Open Science and implementing the FAIR principles within the research ecosystem. Early Career Researchers (ECR) play an essential role in shaping and evolving new data practices and bringing fresh perspectives to well established data domains and research infrastructures.

Reflecting the importance of the ECR community, the two hosting organisations of SciDataCon have their own ECR networks, CODATA Connect and WDS ECR Network. They will be joined by local ECRs to host a session on research infrastructures and repositories across disciplines and geographies, highlighting the uneven landscape in their use and access.

The aim of the session is to showcase excellent work carried out by ECRs and to explore perspectives on use, accessibility and value of research infrastructures and repositories across a range of disciplines, geographies and cultural contexts. Extending to infrastructural, financial and policy related challenges, particularly in the Global South which often contrast sharply with the Global North. Aligning strongly with the conference themes, with international representation of speakers, the session will focus on open research, equity, global collaboration and in one example, CAREful Indigenous Data Governance.

We propose research presentations, sharing best practices and lessons learnt while underscoring where advancement is needed. Highlighting resources and thinking needed to improve the data repository structures, to enable more complete adoption of the FAIR and CARE principles. Followed by discussion with the audience.

Ntsundeni Louis Mapatagane, Uneven Data, Unequal Futures: Climate Change Data Disparities Between the Global North and South

Presenting the disparities in climate change data infrastructure on a global scale, highlighting how the predominance of data repositories situated in the Global North constrains the integration of localised data, particularly from entities in the Global South. Drawing on research conducted within South African universities, a microcosm for fostering sustainable innovation, this discourse will underscore the critical necessity for establishing regional data repositories. Such repositories are essential for enhancing climate resilience, facilitating education, and tracking progress toward the Sustainable Development Goals (SDGs). With the UN's 2024 SDG Report and COP29's forthcoming emphasis on education, the presentation will advocate for the pivotal role of universities in producing reliable, localised climate data that both enriches global datasets but also promotes equitable climate action.

Aдриanna Eufrosina Bora, AI Against Modern Slavery: The Role of Open Data and Infrastructure
Sharing insights from the speaker's experience leading Project AIMS (Artificial Intelligence against Modern Slavery), an initiative that leverages AI to analyze corporate modern slavery statements for compliance with legislation in the UK and Australia. The project addresses the challenge of processing thousands of corporate disclosures by developing machine learning tools capable of reading and benchmarking these reports. Drawing from this work, the talk will explore the critical importance of high-quality, accessible, and well-structured data in developing trustworthy AI systems for high-risk applications. Highlighting how inconsistencies in data formats, limited machine readability, and gaps in data availability can impede the effectiveness of AI tools, particularly in sensitive areas. The session will also discuss the infrastructural and policy-related challenges encountered in building and maintaining open-source AI initiatives for social good.

Pragya Chaube, Who Gets to Share? Understanding the Challenges of Open Research Data in India

Sharing perspectives on the author's experience and drawing from a qualitative study conducted at a leading research institution in India. This study examined attitudes toward open research data across academic disciplines and faculty career stages—from ECRs to senior professors. The study mapped the perceived value of open data, as well as the challenges faced in its adoption and implementation. These include infrastructural and institutional limitations, disciplinary norms, and varying levels of awareness and motivation to engage with open data practices. The findings offer insights into how openness is negotiated within the Indian research environment context, and how faculty perspectives differ based on career stage and discipline.

Claire Rye, Learnings from helping to build data repositories

Comparing and contrasting two experiences working to build data infrastructures, this talk will reflect on the journey of Ingestion service of the Human Cell Atlas Data Coordination Platform. The software infrastructure and metadata standards that support data sharing across the Human Cell Atlas project, while based at the European Bioinformatics Institute. Next on the development of Aotearoa Genomic Data Repository, an Aotearoa-based resource, that enables researchers and Māori communities to fulfil their obligations relating to the guardianship, management, sharing and use of genomic data from biological samples that are taonga (treasured).

Cyrus Walther, What do we do with all the data? Insights on Data Repositories in Large-Scale Multinational Collaborations in Physics

Addressing the challenges and respective approaches in high-energy experimental physics, a data-driven field of research, relying on large data volumes while requiring collaboration across borders and continents. The talk will introduce several of these international collaborations, MAGIC, CERN, or SKAO, providing context and showcasing their necessity in advancing research. Furthermore, critical challenges in data storage, stewardship, and usage that originate from the nature of data acquisition are discussed, utilizing these research use cases. While respecting their individualities, the talk will highlight examples of such approaches for data repositories, demonstrating the opportunity for interdisciplinary transfer of these methods. Overarching best practices in high-energy experimental physics collaborations will be presented and perspectives towards implementation in other areas of the research will be discussed with the audience.

Primary authors: Dr RYE, Claire (University of Auckland/WDS ECR co-chair); WALTHER, Cyrus (TU Dortmund University/CODATA Executive Committee); EUFROSINA BORA, Adrianna (Queensland University of Technology); MAPATAGANE, Ntsundeni Louis (Walter Sisulu University/CODATA Connect); Dr CHAUBE, Pragya (UPES/CODATA Connect)

Presenters: Dr RYE, Claire (University of Auckland/WDS ECR co-chair); WALTHER, Cyrus (TU Dortmund University/CODATA Executive Committee); EUFROSINA BORA, Adrianna (Queensland University of Technology); MAPATAGANE, Ntsundeni Louis (Walter Sisulu University/CODATA Connect); Dr CHAUBE, Pragya (UPES/CODATA Connect)

Track Classification: SciDataCon Persistent Themes: Data and Research