



Contribution ID: 211

Type: Session

## Coalition building to support sustainable digital data standards

*Tuesday 14 October 2025 16:00 (1h 30m)*

This session will bring together members of the research data community with experience of and interest in developing consortia and coalitions that advance the development and application of practices, principles and standards relating to research data. It will aim to identify common considerations and challenges encountered when building coalitions that can inform those embarking on a similar journey, and identify areas for cross-community collaboration to help ensure the long-term sustainability and success of such activities.

We will consider how proposed ideas and specifications have become accepted by communities as agreed-upon 'standards', what sort of community support, governance, and organization 'ownership' is necessary to maintain and sustain digital resources, and the challenges of sustaining the ongoing development and maintenance of outputs in order to meet evolving circumstances. To do this, we will draw on examples of international initiatives at various stages of maturity from across the chemical, earth and life sciences.

In chemistry, stakeholders from across industry and academia, including publishers, data organizations, scientific unions and research institutes, have been convening in workshops focused on the sustainable development of digital standards for describing chemical information across disciplines. These workshops originated from the WorldFAIR Chemistry project, led by IUPAC, with an aim of establishing a coalition of organizations and individuals contributing to long term, pre-competitive services for discovery, adoption, and validation of standards.

At a national level, NFDI4Chem, the chemistry consortium within the NFDI initiative in Germany, brings together research institutions, information providers and learned societies with the aim of digitizing all key steps in chemistry research. In the UK, the Physical Sciences Data Infrastructure project (PSDI) is aiming to connect across various data systems in use by researchers to support process recording, facilitate data compilation and advance computational analysis. Services and infrastructure being developed in these projects require standards for ontologies, metadata and data, and international collaboration with researchers, RDM specialists, ontology engineers, and data scientists to develop guidelines, data models and strategies that support standardization.

The Earth Sciences are global in scope and many international coalitions have been formed to develop digital data standards that enable integration of data collected around the planet. Some of the oldest are the GeoSciML and EarthResourceML, logical data models developed by the Commission for the Management and Application of Geoscience Information (CGI) of IUGS, which are now governed and maintained by the Open Geospatial Consortium and CGI. Sustaining and promoting these standards without ongoing project support is proving challenging.

A more recent earth sciences initiative, OneGeochemistry, brings together major geochemical data providers and the research community with the objective of establishing global standards and best practices for FAIR laboratory analytical data. Standardization and promulgation of best practices and protocols will be a long process as the global geochemistry community is very heterogeneous and fragmented due to a multitude of analytical techniques applied to a wide variety of materials. Developing sustainable long-term coalitions with other international initiatives to facilitate cross-domain interoperability will be critical.

The Genomic Standards Consortium (GSC) is an international organization focused on the development of genomic and environmental metadata standards to facilitate sharing and reuse of genomic data across databases and between studies. The GSC teams up with domain experts to develop and integrate novel minimum information standards (MIxS) through monthly working group sessions and annual meetings. Striving for true reproducibility of genomic data, the GSC teams up with outreach initiatives run by other national and international microbiome and multi-omics data alliances and consortia.

The Worldwide Protein Data Bank (wwPDB) represents a consortium of international organizations in America, Europe and Asia that collaborate to jointly manage a global repository of 3D biological macromolecular structures. As part of this, the wwPDB oversees the governance of mmCIF, a community standard for the exchange, annotation, validation and archiving of macromolecular structural data. The partnership of organizations in the wwPDB enables funds to be channeled from different regions in support of data standards and curation.

More broadly, the Global Biodata Coalition brings together a range of organizations who fund life science research to jointly address the need for sustainable financial support for global biodata resources and share approaches and strategies for efficient management and growth of infrastructure in this area. The challenge for databases and knowledgebases that archive and add value to research data is that they constitute infrastructure requiring stable long-term support, but are generally funded through short-term grants and contracts.

This session will be designed as a panel discussion with brief presentations from a broad sample of established and emerging coalitions such as those described above, several of whom have agreed to participate. We will share current progress and goals, hear about success stories, how barriers were overcome and ongoing challenges that might be addressed by wider cooperation and collaboration.

Specifically we will ask panelists to reflect on the challenges encountered, strategies adopted and outcomes achieved in the following areas:

- Core goals and target stakeholders
- Leadership and governance structures
- Engaging and coordinating across diverse stakeholders
- Communicating vision and activities to wider communities
- Attracting resources needed to advance aims
- Maintaining momentum of activities over time
- Timescales - initial aspirations vs reality
- Strategies for ensuring sustainability of activities and outputs

We anticipate that the result of this session will be an understanding of strategies that can successfully catalyze the development of initiatives aiming to organize across stakeholders and disciplines, how to avoid common pitfalls, and opportunities for broad community collaboration of benefit across coalition building activities.

**Primary authors:** BRUNO, Ian (CCDC); LEHNERT, Kerstin (Lamont-Doherty Earth Observatory of Columbia University); MCEWEN, Leah (Cornell University); WYBORN, Lesley (Australian National University); RATTENBURY, Mark (GNS Science); KOEPLER, Oliver (TIB - Leibniz Information Centre for Science and Technology); Prof. HARTSHORN, Richard (University of Canterbury); HODSON, Simon (CODATA)

**Presenters:** BRUNO, Ian (CCDC); LEHNERT, Kerstin (Lamont-Doherty Earth Observatory of Columbia University); MCEWEN, Leah (Cornell University); WYBORN, Lesley (Australian National University); RATTENBURY, Mark (GNS Science); KOEPLER, Oliver (TIB - Leibniz Information Centre for Science and Technology)

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