



Contribution ID: 148

Type: **Session**

An evolving role for Data Scientists in the Age of Intelligent Automation

Wednesday, October 15, 2025 2:00 PM (1h 30m)

Data science has evolved significantly over the past two decades, becoming a force within academic research, public and private sector workplaces, and in government policies and practices. Exponentially increasing volumes of publicly available datasets throughout the social and scientific realms have contributed to an explosion of data science applications, including AI tools and Large Language Models (LLMs) that are being connected to digital technologies of all types, such as personal computers, cell phones, social media, smart devices, and sensor networks. Data science has become a broad term with very different meanings and instantiations, growing far beyond the traditional panoply of techniques applied to derive value (economical, intellectual, or cultural) out of data. A former editor of the CODATA Data Science Journal argued in a recent retrospective essay that, “The scientific data community requires more from data science than other communities” (Rumble, 2023, p. 2), arguing that topics like data preservation, provenance, and traceability cannot be ignored when developing data science applications for scientific research. In this session, we will present broad perspectives on this key question: What are the distinctive aspects of data science for the scholarly and scientific data community, and how should the present day data scientists adapt to address issues like FAIRification, AI-readiness and Ethics?

A diverse panel will engage the CODATA community in a discussion of the role of data science in relation to scientific data initiatives, and how this community is contributing to the growth and evolution of data science. The speakers in this session will address several important questions about the current status and future evolution of data science, including:

- What is data science, and how does it relate to the “science of data”?
- Is data science a scientific domain in its own right?
- What are the main gaps / opportunities that must be addressed in data science going forward?
- What are the key data science trends in relation to the scientific data community?
- Science has always been data driven; what is different now? How have changes in the way that data are or should be shared influenced data science?
- How do AI technologies impact data science and contribute to FAIR research data ecosystems?
- How should the Data Science Journal and SciDataCon respond to changes in the nature of data science?

The talks in this session will inform broader discussions of the implications of these questions for CODATA, the World Data System (WDS), the International Science Council, and the many international scientific associations and unions. How can these organizations both contribute and respond to the changing nature of data science via concrete actions, programmes, and initiatives?

References:

Rumble, J. (2023). Thoughts on starting the CODATA Data Science Journal. *Data Science Journal*, 22, 13. <https://doi.org/10.5334/dsj-2023-013>

Speakers:

- Matt Mayernik (Session co-Moderator) - NSF National Center for Atmospheric Research, USA, Editor-in-Chief, *Data Science Journal*

- Gita Yadav (Session co-Moderator) - Scientist, National Institute of Plant Genome Research (NIPGR), New Delhi India; Member, CODATA IDPC; Professor of Data Science, IISER Bhopal, India Founder, #SemanticClimate; Member Editorial Team, Data Science Journal
- Deb Mohanty - Director, National Institute of Immunology, DBT, Govt of India; Member, CODATA India National Committee; Chairperson, Indian Biological Data Center
- Mark Parsons - Former Editor-in-Chief, Data Science Journal; Member, Arctic Data Committee for the International Arctic Science Committee (IASC) and the Sustaining Arctic Observing Networks (SAON) CODATA representative to the International Polar Year 2032-3 Planning Committee
- Dimitris Symeonidis - University of Tartu, Institute of Computer Science Member, Digital Government Society
- Dr. Lili Zhang - Executive Director, GOSC IPO, member of CODATA IDPC Senior Research Scientist, CSTCloud, CNIC,CAS; Member Editorial Team, Data Science Journal

Primary authors: MAYERNIK, Matthew (NSF National Center for Atmospheric Research); Dr YADAV, Gita (National Institute of Plant Genome Research (NIPGR))

Co-authors: MOHANTY, Debasis (National Institute of Immunology, New Delhi); PARSONS, Mark; SYMEONIDIS, Dimitris (University of Tartu); ZHANG, LILI (COMPUTER NETWORK INFORMATION CENTER, CAS)

Presenters: MAYERNIK, Matthew (NSF National Center for Atmospheric Research); Dr YADAV, Gita (National Institute of Plant Genome Research (NIPGR)); MOHANTY, Debasis (National Institute of Immunology, New Delhi); PARSONS, Mark; SYMEONIDIS, Dimitris (University of Tartu); ZHANG, LILI (COMPUTER NETWORK INFORMATION CENTER, CAS)

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