Contribution ID: **156** Type: **Presentation**

FAIR Challenges when using AI to Tailor Data for Climate Change Risks Applications

Tuesday, October 14, 2025 4:55 PM (11 minutes)

Conducting high-quality research increasingly involves complex workflows and the generation of numerous intermediate datasets. Achieving reproducibility requires the availability of extensive and well-structured information. To enable this, researchers need interfaces and tools that are not only user-friendly but also FAIR-aware.

Data analysis workflows typically span multiple tools and platforms, making interoperability of metadata a critical requirement for reproducibility. Ensuring that datasets and software comply with FAIR principles, particularly over time, remains a significant challenge. Such resources must be preserved in permanent, citable repositories to support long-term reuse and citation in scientific publications. This is especially relevant in the context of "long-tail research data," which encompasses not only datasets but also software and workflows.

The emergence of AI-powered tools has further amplified the need for transparency and reproducibility, as AI models are highly sensitive to training data, configuration parameters, and implementation details. Supporting the full lifecycle of research data—especially in domains like climate science—requires sustained and coordinated efforts.

In the climate research infrastructure community, various initiatives have emerged to meet these challenges. The Research Data Alliance (RDA) has been instrumental in providing guidelines and best practices to enhance FAIR compliance across data and software. Key RDA groups contributing to this effort include the FAIR Digital Object Fabric Interest Group (IG), the FAIR Data Maturity Model Working Group (WG), FAIR for Machine Learning (FAIR4ML) IG, and FAIR for Research Software (FAIR4RS) WG.

This presentation will explore the specific challenges of achieving FAIR compliance in the context of tailoring future climate simulation data for climate change risk applications. These efforts are part of the Horizon Europe IRISCC project, which involves multiple Demonstrators and Service Design Labs that rely on such data to support their objectives.

This project (IRISCC) is funded by the European Union under @HorizonEU research and innovation programm under grant agreement N°101131261.

Primary author: PAGÉ, Christian (CERFACS)

Co-author: Mr SPINUSO, Alessandro (KNMI)

Presenter: PAGÉ, Christian (CERFACS)

Session Classification: Presentations Session 5: Rigorous, responsible and reproducible science in the era of FAIR data and AI / Infrastructures to Support Data-Intensive Research

Track Classification: SciDataCon2025 Specific Themes: Rigorous, responsible and reproducible science in the era of FAIR data and AI