Contribution ID: **63** Type: **Oral (Regular)**

# The IAEA Fusion Data Lake Project - Accelerating AI and Big Data Applications through Open Science and FAIR Data

*Friday 12 September 2025 10:40 (15 minutes)*

The application of AI technologies is an ever-growing area of research within the development of fusion energy. The technology plays a key role in the production of surrogate models and digital twins that enable accelerated development by supplementing the use of computationally expensive 2D and 3D simulation codes, defining non-parametric models for insufficiently characterised phenomena, and computationally cheap alternatives for real-time applications [1]. The IAEA is supporting the fusion communities' efforts in this area with the AI for Fusion Coordinated Research Project (CRP), a five-year initiative launched in 2022, which involves 24 institutions across 11 countries [2]. A key goal of the CRP is to support the development of modern, scalable, and accessible data infrastructure that is required to produce diverse and rich data sets, which are required to develop generalised AI models that are machine agnostic and can safely extrapolate into the parameter space of future fusion power plants.

The IAEA is playing an active role in contributing to the data infrastructure with the Fusion Data Lake project. A modern data platform to enable the development of AI workflows in line with FAIR (Findable, Accessible, Interoperable, and Reusable) data principles. The platform comprises three major components:

1. An international data catalogue;
2. A centralised medium-term storage; and
3. A data federation of the various fusion data platforms around the world.

A proof of concept (PoC) has been built, which demonstrates the data cataloguing and federation capacity by integrating with the UKAEA's MAST Data Catalog [3]. Currently, the second phase of the PoC will involve further generalisation of the data pipeline codebase and the ingestion of shot catalogues from two additional experimental fusion devices.

This presentation will present a high-level overview of the Fusion Data Lake project, including:

- Technical architecture and design, collaborations and contributions, and the PoC solution;
- Data and metadata model development and ontological concepts; and
- The approach to data governance and terms of service.

This presentation will illustrate the approach, results, and direction of the work, highlighting the high potential value to the fusion community of increasing the visibility and accessibility of the numerous international experimental data sets.

References

1. P. Brans, "AI ignites innovation in fusion", ITER Organization, 2025, website https://www.iter.org/node/20687/ai-ignites-innovation-fusion (accessed 5th September 2025)
2. AI for Fusion, International Atomic Energy Agency, 2025, website https://nucleus.iaea.org/sites/ai4atoms/ai4fusion/SitePages/AI4F.aspx (accessed 5th September 2025)
3. S. Jackson et al. 2025, IEEE Trans. on Plasma Sci., doi: 10.1109/TPS.2025.3583419. https://ieeexplore.ieee.org/document/11128905

## Speaker's email address

dsgahle1771@outlook.com

## Speaker's Affiliation

## Member State or International Organizations

**Author:** GAHLE, Daljeet (IAEA)

**Co-author:** Mr BARBARINO, Matteo (International Atomic Energy Agency)

**Presenters:** GAHLE, Daljeet (IAEA); GAHLE, Daljeet Singh (International Atomic Energy Agency)

**Session Classification:** Information Retrieval and Visualisation