Contribution ID: **54** Type: **Oral (Regular)**

JDDB: A Flexible and Extensible Data Processing Framework to Accelerate AI4Fusion Research

Wednesday 10 September 2025 14:50 (25 minutes)

The increasing integration of artificial intelligence (AI) into fusion research demands scalable, standardized, and traceable data infrastructures. To address this need, we introduce JDDB (J-TEXT Disruption Database), a flexible and extensible most of all, a light weight data processing framework designed to streamline the management and transformation of fusion data across tokamak experiments.

As the name suggests, JDDB started as a database for data driven disruption prediction research. As it evolves, it now has the potential to play a foundational role in supporting AI4Fusion research by enabling robust data handling practices, reproducible workflows, and cross-device compatibility.

The design philosophy of JDDB is to be simple and lightweight. As long as the data and metadata is kept complete and organized, the rest is not core concern of JDDB. The key feature of JDDB is the data is split into two repositories. One named FileRepo is a bunch of HDF5 files, simple and straight forward for data processing jobs to handle. The other is the MetaDB, which is a MongoDB database contain all the metadata used for searching and tracing the data. For the FileRepo storage format, the ITER Integrated Modelling & Analysis Suite (IMAS) data dictionary is used as it's the most accepted standard for fusion scientific data. For metadata, the MongoDB is used because if offers flexible scheme so different metadata for different signals can be stored without modify the database scheme. For data processing, JDDB provides a parallel, modular workflow engine that ensures efficiency and reproducibility. A core principle of JDDB is data immutability: raw data remain unchanged, while all processed data are versioned with full provenance tracking. This guarantees traceability and transparency in complex analysis pipelines, which is critical for verification and reproducibility, collaboration, and regulatory compliance in fusion research.

By unifying data structure, metadata, and workflow management into a coherent framework, JDDB significantly lowers the barrier for developing and deploying AI-driven models in fusion applications. It enables researchers to focus on physics-informed modeling and data analysis, while ensuring that their pipelines are robust, scalable, and transparent. JDDB is currently in active use within the J-TEXT tokamak and is being extended to support cross-device applications and collaborative development. Its architecture is well-suited for integration into future AI foundation models, intelligent simulations, and control strategy optimization for magnetic confinement fusion devices.

Speaker's email address

zhengwei@hust.edu.cn

Speaker's Affiliation

Huazhong University of Science and Technology

Member State or International Organizations

China

Authors: Mr XUE, Fengming (International Joint Research Laboratory of Magnetic Confinement Fusion and Plasma Physics, Huazhong University of Science and Technology); SHEN, Chengshuo (Huazhong University of Science and Technology); Ms LUO, Runyu (Huazhong University of Science and Technology); AI, Xinkun (International Joint Research Laboratory of Magnetic Confinement Fusion and Plasma Physics, School of Electrical and Electronic Engineering, Huazhong University of Science and Technology); Mr YI, Yukai (Huazhong University of Science and Technology)

Co-author: ZHENG, Wei (International Joint Research Laboratory of Magnetic Confinement Fusion and Plasma Physics, Huazhong University of Science and Technology)

Presenter: ZHENG, Wei (International Joint Research Laboratory of Magnetic Confinement Fusion and Plasma Physics, Huazhong University of Science and Technology)

Session Classification: Information Retrieval and Visualisation

Track Classification: Information Retrieval and Visualisation