

A. Pau (PI), B. Labit, Y. Martin, P. Zhang, M. van Rossem, V. Cristina, the TCV team and the EUROfusion Team.

¹ EPFL (École Polytechnique Fédérale de Lausanne), Swiss Plasma Center (SPC), CH-1015 Lausanne, Switzerland

DATA: Scalable Production & Collaborative Environment for Multi-Machine Databases

Contact: alessandro.pau@epfl.ch | DEFUSE Release 1.0 | Apache 2.0 - LGPL License | gitlab.epfl.ch

Introduction & Motivation

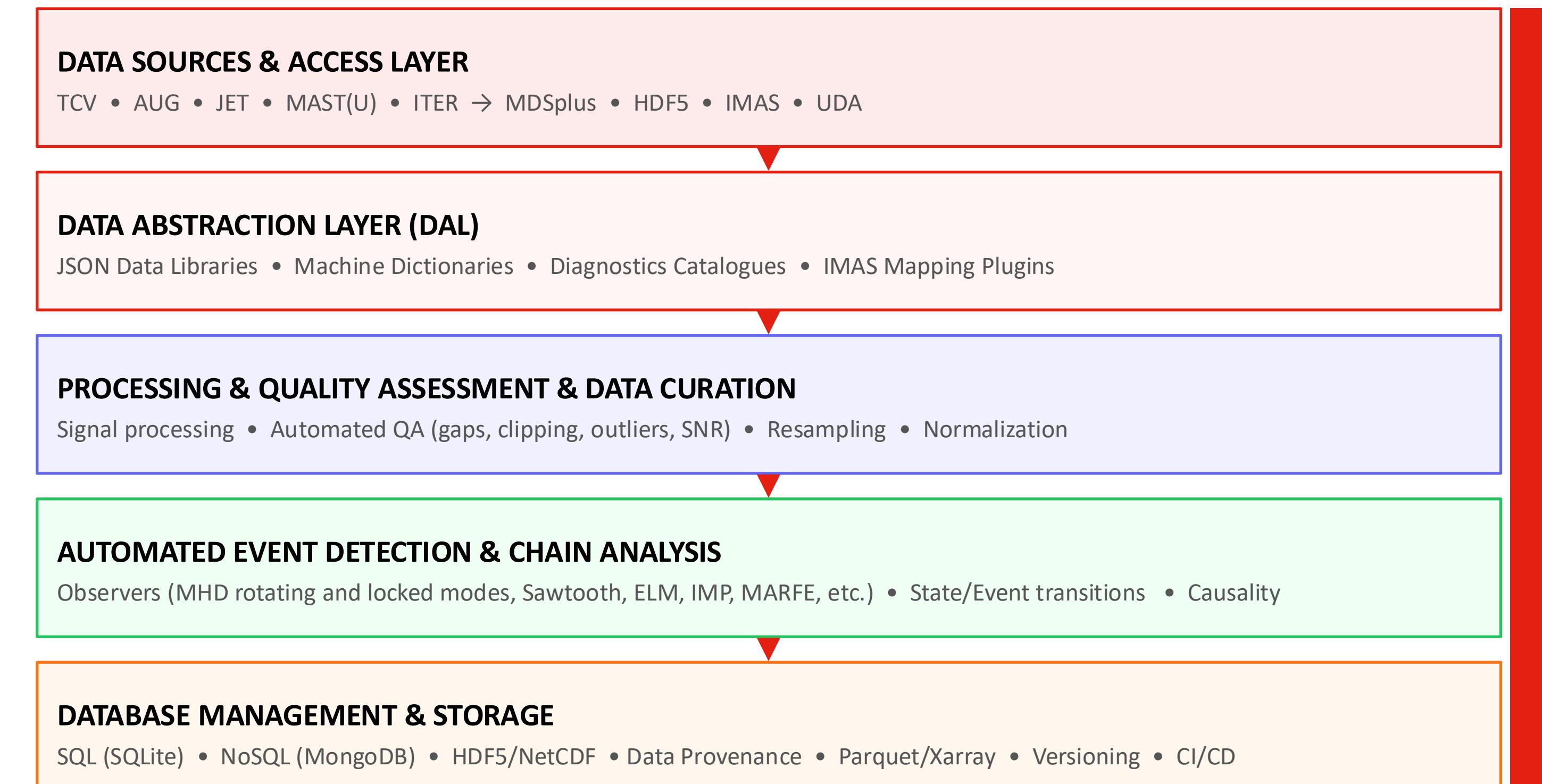
The European fusion programme requires standardized, validated multi-machine databases to support disruption avoidance strategies, AI/ML model training, and extrapolation to ITER and DEMO. Managing heterogeneous data from multiple tokamaks – each with different diagnostics, naming conventions, and data systems – poses a major infrastructure challenge.

Key challenges:

- Heterogeneous data formats and access methods across fusion devices (MDSplus, HDF5, IMAS, UDA)
- Automated, reproducible validation pipelines for thousands of discharges
- FAIR-compliant data curation with full provenance and traceability
- Scalable infrastructure for large-scale statistical analysis and AI-ready datasets

DEFUSE (Disruption & Event analysis framework for FUSion Experiments) provides the production stack underpinning EUROfusion multi-machine databases, combining advanced data analysis, optimization, and scalability for off-normal event and disruption characterization.

DEFUSE Framework Architecture



MATLAB + Python dual-language OOP framework. JSON data libraries for machine-agnostic interoperability. Automated setup with cross-platform configuration management.

Scalability & Optimization

CI/CD Automated Build & Deployment

GitLab CI/CD pipelines with Jenkins integration. Automated building, testing (~80% coverage target), and deployment. Version-controlled on EPFL GitLab with GitHub mirror.

SQL Multi-Layer Database Management

SQL (SQLite) relational model for IMAS mapping with role-based authentication. NoSQL (MongoDB) for off-normal events and plasma state. JSON extension for complex nested structures.

Parallel Batch Processing

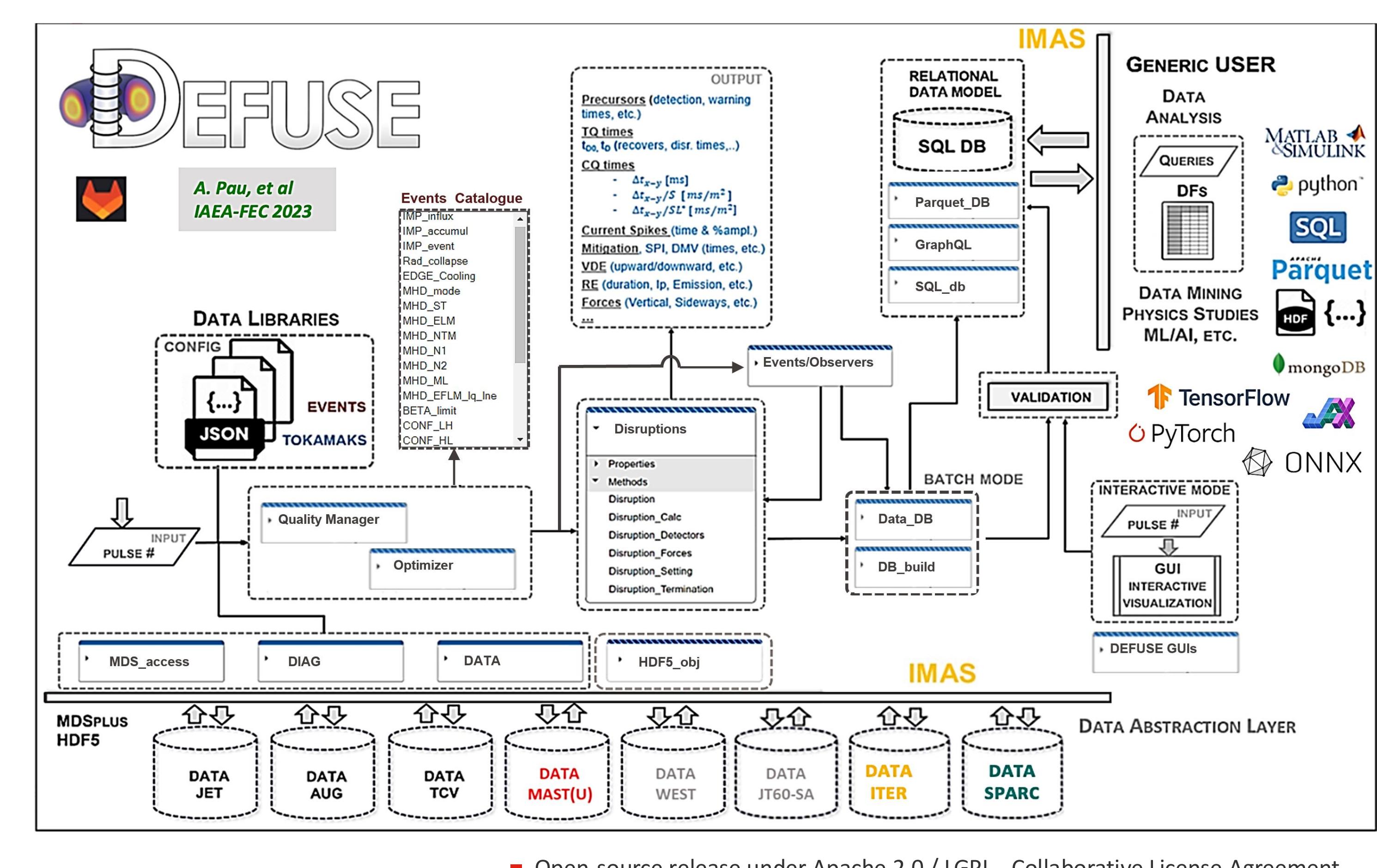
Configurable parallel workers for large-scale analyses. Checkpoint-based batch execution with merge & recovery. Performance benchmarking: <30s per shot for full disruption characterization.

FMT Multi-Format Data Pipeline

HDF5 (unstructured data, IMAS backend), Parquet/Xarray (analytics), NetCDF (compatibility). Automated parquet table generation: DATA, COORD, LABELS, NORM variants per shot.

QA Validation & Provenance

Shot-level validation with versioning, rollback, and logging. Field-level provenance tracking. Pre-commit testing framework with regression detection.



■ Open-source release under Apache 2.0 / LGPL - Collaborative License Agreement

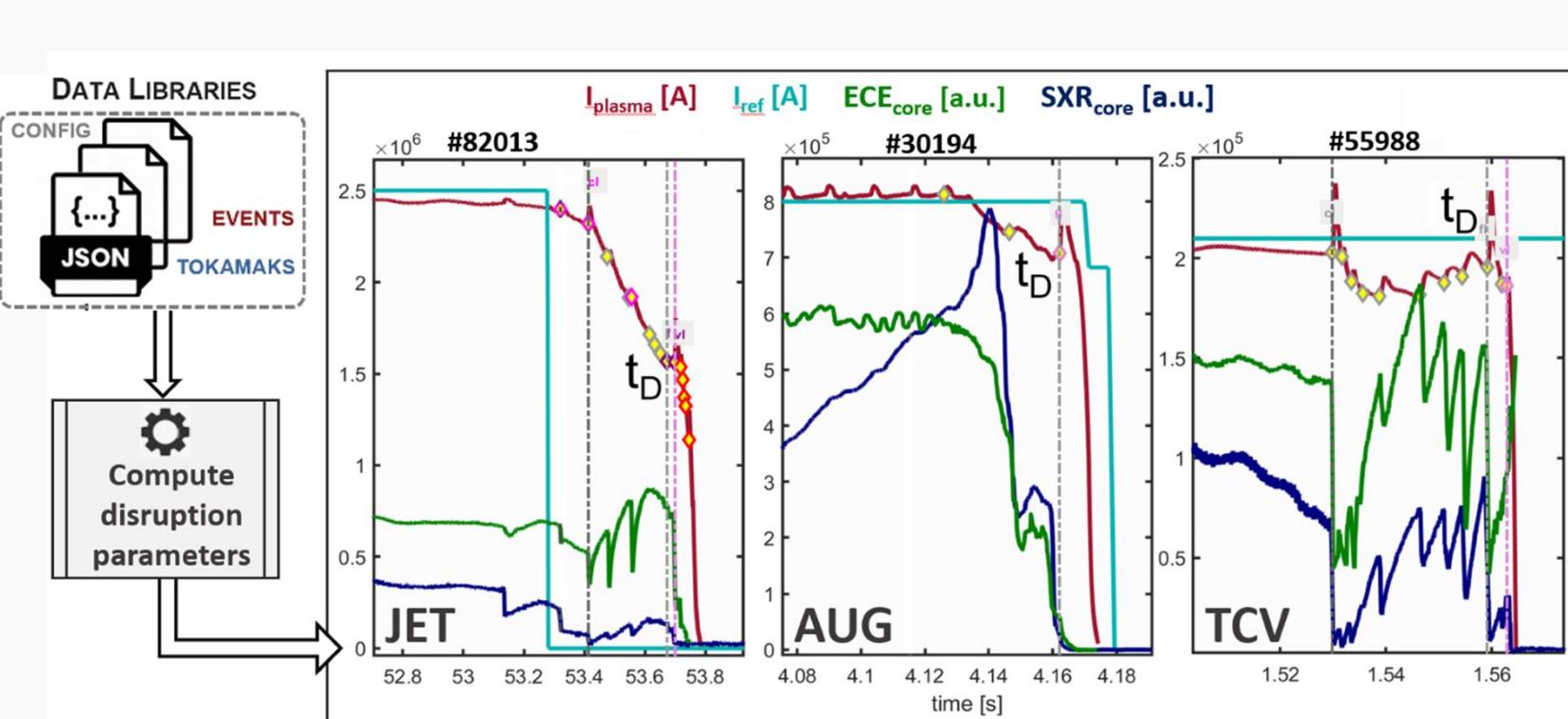
EUROfusion Multi-Machine Databases

EuDDB – Disruption Database

Validated multi-machine disruption database covering TCV, JET-ILW, AUG, MAST.

- JET-ILW entire lifecycle: ~6k disruptions (Ip > 750 kA).
- TCV: ~4k disruptions (Ip > 50 kA).

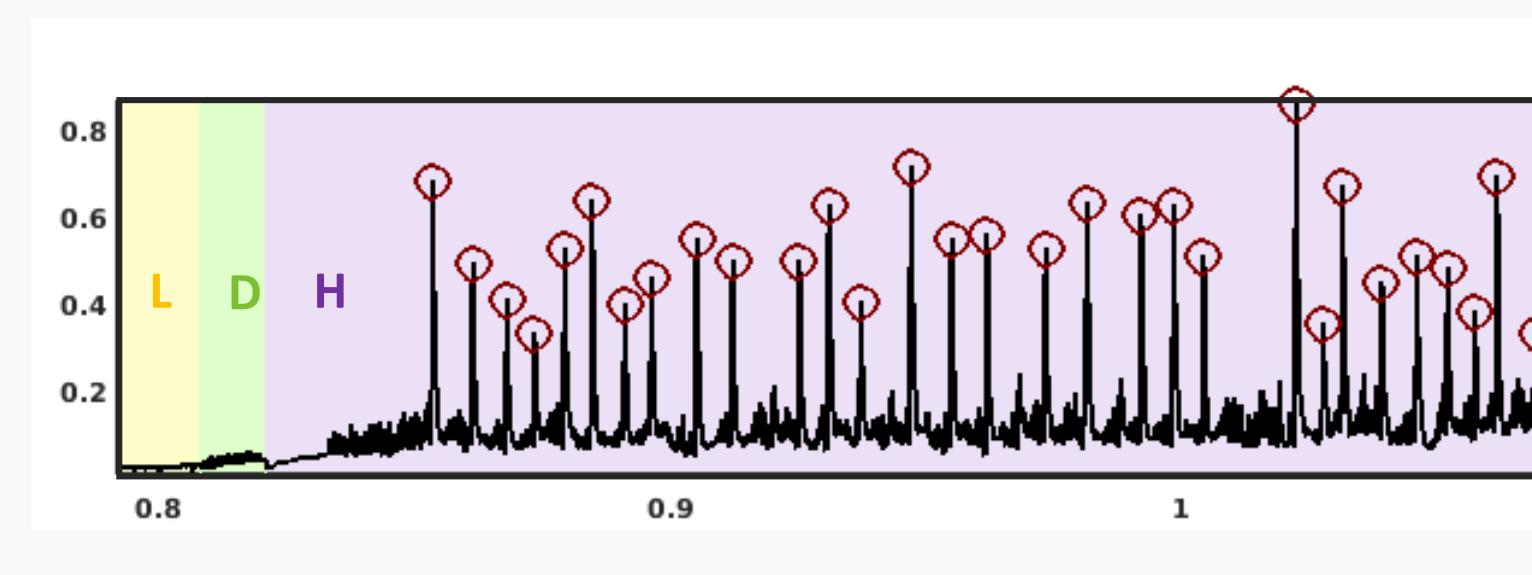
Automated chain-of-events analysis with physics precursors, off-normal events, and control system alarms & reactions.



L-H Transition Database

TCV public database undergoing 'IMASification' to support various EUROfusion activities.

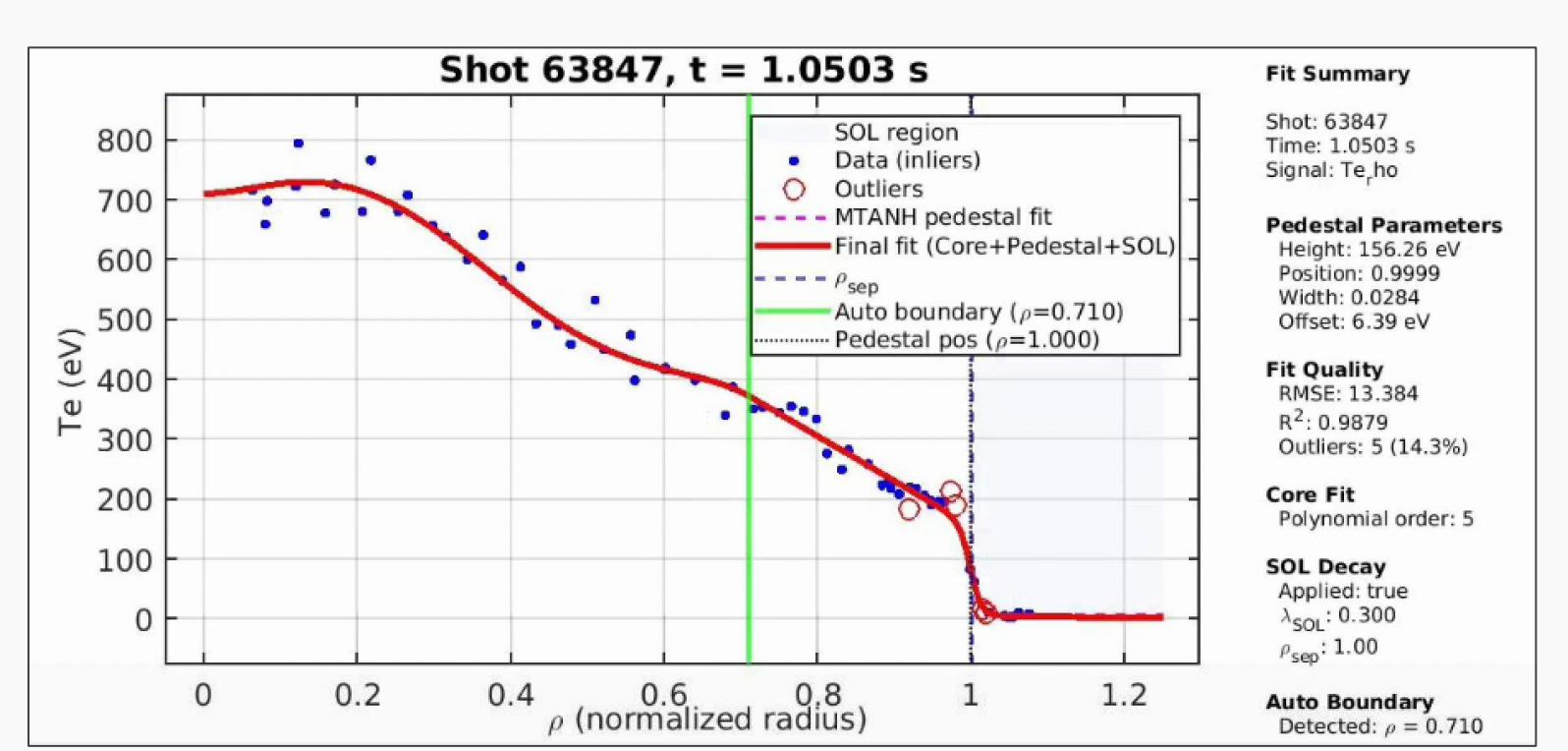
Unified JET database consolidating predictions from various models. Standardized L-H transition characterization across devices and automated confinement state detection



Pedestal Database

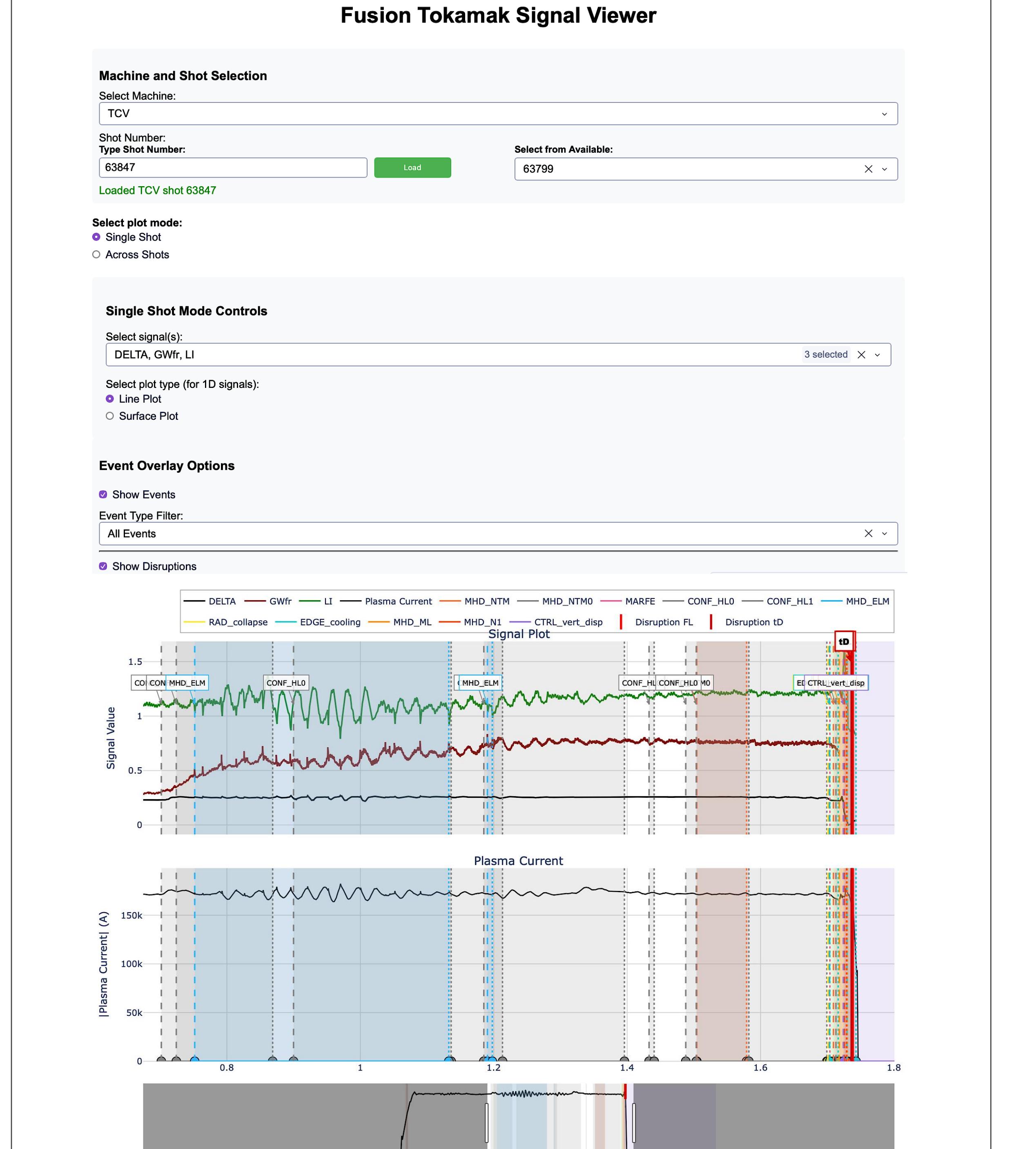
Updated with data from EUROfusion WPTE campaigns (RT01, RT02, RT05, RT07, RT08). IMAS-ready format for Gateway storage.

- Transformed and mapped to the IMAS data model for cross-machine compatibility.
- Automated initial fit and outlier rejection based on Gaussian Process Regression (GPR) and principled Uncertainty Quantification (UQ)



IMAS Mapping: All databases are being mapped to the IMAS data model via Python APIs. Gateway migration planned with 20–50 TB storage in HDF5/Parquet formats. SQL tables with JSON metadata for full IMAS/IDS compatibility.

DEFUSE Disruptions and Events Dashboard



References

- [1] A. Pau et al., IAEA FEC (2023) • [2] DEFUSE Documentation – SPC-EPFL GitLab (2024) • [3] PrIO-2.DB.01-T008-D001 (2024)

Supported by the Swiss National Science Foundation and the EUROfusion Consortium.

Synergy with the AI/ML Grant 'AI-assisted Plasma State Monitoring for Control and Disruption-free Op. in Tokamaks'.