Implementation of the Data Management Plan

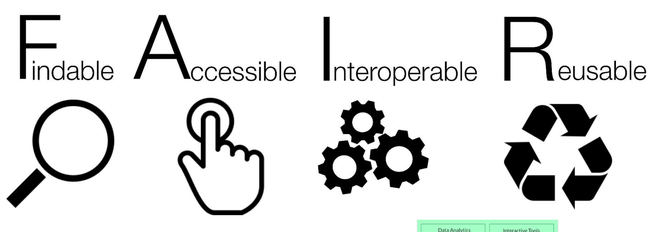
P, Strand[1], D. Coster[2], L. Kripner[3], I. Hanák[3], D. Kománek[3], A. Gonzalez Beltran[4], N. Cummings[4], S. Dixon[4], T. Farmer[4], J. Hollocombe[4], L. Kogan[4], A. Parker[4], C. Powell[4], G. Szepesi[4], R: Coelho[5], J. Decker[6], L. Simons[6], S.C Yildiz[6], L. Fleury[7], F. Imbeaux[7], P. Grabowski[8], K. Niznik[8], M. Owsiak[8], B, Pogodzinski[8]

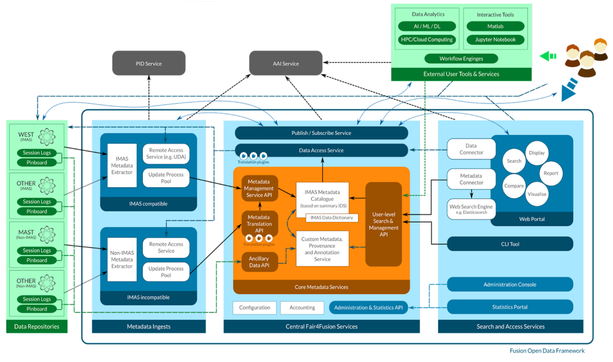
[1]Chalmers University of Technology, Sweden, [2]Max-Planck-Institut für Plasmaphysik, Germany, *[3]Institut for Plasma physics, CAS, Czech Republic, [4]UKAEA, United Kingdom, [5]IST, Portugal, [6]SPC/EPFL Switzerland, [7]CEA, France, [8]IPPLM (PSNC IBCH PAS), Poland*

The aim is to provide a production-ready, federated data management ecosystem enabling harmonised access to experimental and simulation data across EUROfusion sites and users.

Built on established tools and frameworks (IMAS, UDA, SimDB,…)
Maintain data integrity and quality
- **Authentication and Authorisation protocols for controlled data access**
- **Data mappings developed by data producers/owners**

FAIR based approach to data sharing
Stepwise implementation in different scenarios
Prioritisation to be based on users' needs and capabilities





Fusion Open Data Framework

Blueprint implementation from Fair4fusion (www.fair4fusion.eu)

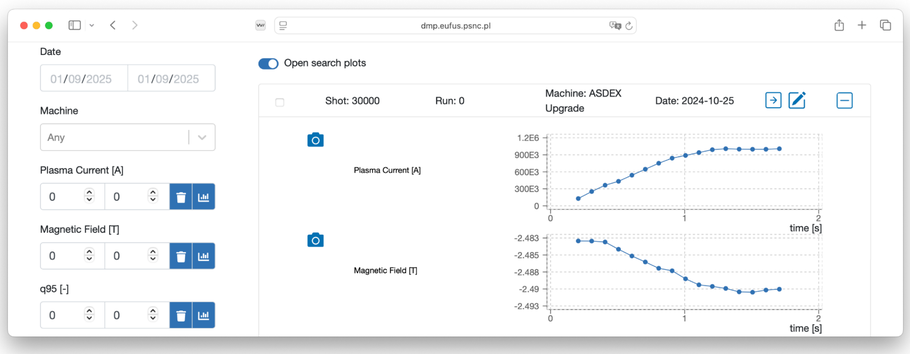
Divided into 3 technology activities:

- **central core services ACH support**
- **IT/infrastructure services on each site**
- **Data mappings**

The data management plan defines 4 scenarios/stages of increasing ambition

Scenario A: **making metadata only available and searchable using IMAS data subsets for interoperable definitions of quantities [F,(I)].**

Developed , going into production 2026-27 (upgrading for performance improvements)



Metadata in this context is waveforms from select IDS_summary signals downsampled to ~100 timepoints (if needed) together with data from dataset_description. Link back to original data sources. Catalog also used for EF databases. (dmp.eufus.eu)

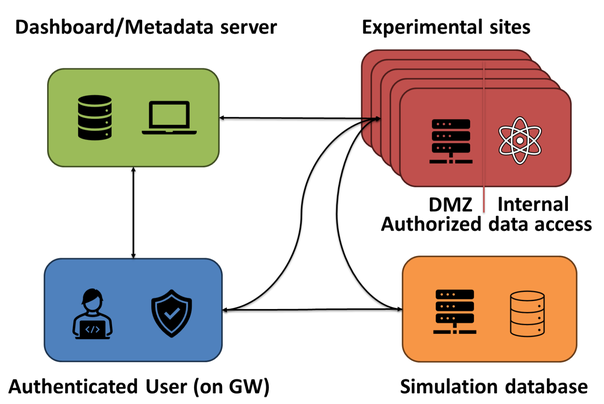
Current data availability:

AUG 11150 shots, JET select discharges from DTE2 and DTE3 campaigns for demonstration (extendable) , TCV > 10000 shots, WEST ~8000 (all shots from C4 to C11),

MAST (data from M4-M9)/MAST-U (data from shot 40000 and onwards) NOT available due to embargo and data policies. Selected Compass discharges will be made available testing and future Compass-U infrastructure is being integrated.

Scenario B: **a subset of the data is accessed using common tools (UDA). Facilities are responsible for the access level and qualification of data through the data mappings [F,A,I,(R)].**

Prototyped, continued towards production in 2026-27 (supporting user codes)



WEST, AUG and TCV provide good data coverage including some derived quantities.
- **Used to demonstrate remote data access for different use cases.**
- **Different levels of automation in the workflows!**

Other devices need further investments in data mappings!

Authorisation and Authentication infrastructure key to maintain data integrity and security. Being deployed on all sites 2026-27.

IDS name	Type	AUG	COMPASS	JET	MAST	MASTU	TCV	WEST
bolometer	Expt.						gdat	WEST
bremsstrahlung_visible	Expt.						N/A	WEST
calorimetry	Expt.							WEST
camera_ir	Expt.							WEST
camera_x_rays	Expt.							WEST
core_profiles	Derived	trview		TRANSP		TRANSP	gdat	WEST
core_sources	Derived	trview		TRANSP		TRANSP	gdat	
core_transport	Derived			TRANSP (TBC)			gdat	
dataset_description	Derived	trview		tokamap			gdat	
distributions	Derived	trview		*				
ec_launchers	Expt.	trview		N/A			gdat	
ece	Expt.	trview						WEST
equilibrium	Derived	trview magnetics tool		tokamap		tokamap	gdat	WEST
gas_injection	Expt.							WEST
hard_x_rays	Expt.							WEST
ic_antennas	Expt.	trview					N/A	WEST
interferometer	Expt.							WEST
iron_core	Expt.	N/A		tokamap	N/A	N/A		
langmuir_probes	Expt.							WEST
lh_antennas	Expt.							WEST
magnetics	Expt.	magnetics tool		tokamap		tokamap	gdat	WEST
mse	Expt.						N/A	
nbi	Expt.	trview					gdat	
neutron_diagnostic	Expt.							WEST
pellets	Expt.						N/A	
pf_active	Expt.	trview		tokamap		tokamap	gdat	WEST
pf_passive	Expt.					tokamap	gdat	WEST
polarimeter	Expt.							WEST
pulse_schedule	Control	trview				tokamap		WEST
reflectometer_profile	Expt.							WEST
soft_x_rays	Expt.						gdat	WEST
spectrometer_visible	Expt.							WEST
spectrometer_x_ray_crystal	Expt.							WEST
summary	Derived	trview summary tool	cdb2imas WIP	tokamap	tokamap	tokamap	gdat	WEST
thomson_scattering	Expt.			tokamap			gdat	
tf	Expt.	trview magnetics tool		tokamap		tokamap	gdat	WEST
w all	Expt.	trview magnetics tool		tokamap		tokamap	gdat	WEST
waves	Derived	trview						

Data mappings:

Key to future developments - should be fully driven by user requirements - connections to WP TM and WP TE.

--Plan to make detailed requirements capture together with TSVV, ENRs and DTE projects with IMASified codes and workflows. Need to define list of IDS and signals needed to run specific workflows

- Data mappings and validation key to data integrity and scientific quality – requires code owners/users input.

Issues and gaps:

- IMAS update DD3->DD4 (backwards incompatible) - need to define a strategy and schedule for change. Affects data mappings and all IMASified codes!

- Common use cases require access to derived IDSs (core_profiles,…) that may need manual intervention to create. Need input on alternatives and needs.

- resources for data mappings is somewhat scarce.

Color coding:

IDS – active, alpha

Mappings – (partially) validated, available needs work, No data is available, Not applicable (N/A)

Scenario C: **builds on the previous stages and allows for enhanced data provenance and referencing through PID's [F,A,I,R].**

Development and deployment in 2026-27.

Moving from development to implementation: supporting users' data needs

- Support specific use cases from the TSVVs and ENR¨s - validate data mappings together with owners of IMASIfied codes. (requirements capture and implementation)
- Expand and enrich provenance metadata and export together with data
- Develop a strategy for PID minting on generated data (PSNC minting, Zenodo, etc)
- With WP TM, develop a strategy for inclusion of simulation and modelling data in the catalogue.
- Deploy an overall infrastructure supporting full AAI implementation and data access.

Scenario D: **adds a lightweight layer for open access to non-embargoed metadata and where allowed by the facilities also data access for export in human readable formats (CSV files) [F,A,I,R] and open.**

To be demonstrated in 2027.

This scenario will utilize the full range of the toolset to export data sets as open source/open access. Starting with at least two candidates (WEST, TCV) we will promote released data on a public platform under CC(4.0) BY-CA-NC.
Internal use of PID based data references for verification and validation exercises will be explored

Summary and outlook:

A federated data infrastructure have been developed and is being put in operation – still need full inclusion of AAI on the participating sites and some performance enhancements. Major effort of the infrastructure teams!

A metadata catalogue (containing waveforms) is available as a searchable inventory of available data. Currently a small subset of data from IDS summary is available. Need your input on required/useful signals!

Remote data access for the purpose of running selected analysis and physics codes has been demonstrated. 2025 scope was to prototype this facility Need your input on required/useful signals and data!

In 2026-27 the remote data access will move to production phase with enhanced data provenance and inclusion simulation and modelling data on equal footing as experimental data. This requires collaboration with WP TE, WP TM and the code owners and users.

Some of the data will by end of 2027 be released as open data (TCV, WEST are candidates due to national policies). Need to agree platforms for this and also review if data collections for validation can be opened up.

The project activities are co-financed by Polish Ministry of Science and Higher Education from financial resources of the programme entitled "PMW" in 2021-2026 and grants from the Swedish Research Council.

This work has been carried out within the framework of the EUROfusion Consortium, funded by the European Union via the Euratom Research and Training Programme (Grant Agreement No 101052200 — EUROfusion). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.