General E-Tasc Meeting February 2026

(Status and plans for the) Availability of machine data through DMP

Pär Strand, Chalmers for the DMP team

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. Maini[7], J. Morales[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*

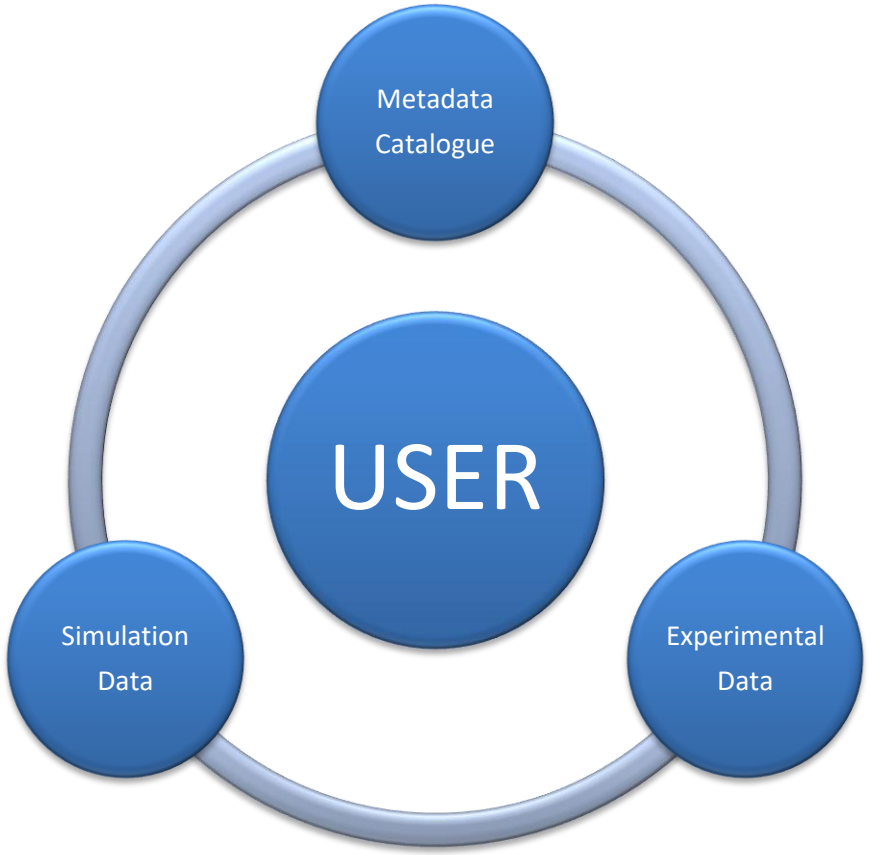
This work has been carried out within the framework of the EUROfusion Consortium, funded by the European Union via the Euratom Research and Training Programme (Grant Agreement No 101052200 — EUROfusion). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect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

Outline

- Background and implementation
- Staged approach towards FAIR access to data
- (2023) 2024-25 activities: prototyping
  - Inventory/catalogue of available data (metadata / waveforms)
  - Remote access to data.
- 2026-27: starting to exploit and open up the data
  - Finalize infrastructure bringing controlled access to AUG, [COMPASS/-U], JET, MAST/MAST-U, TCV and WEST online. (Authe tication and Authorisation extensions to sites)
  - Make the tools available for user (TSVV's, ENR's, DTE's and WP TE modelling)
    - User inputs and requirements needed (task/code specific)
  - Add simulation/modelling data as another facility
  - Enhance provenance and promote the use of Persistent identifiers (PID's).
  - Build a strategy for open data and release a subset of data from participating devices.

Implementation of the Data Management Plan

The longer term aim is to provide a production-ready, federated data management ecosystem for harmonised access to experimental and simulation data across EUROfusion sites and users and in the longer term allow for some elements to be fully open sourced.



Three elements:

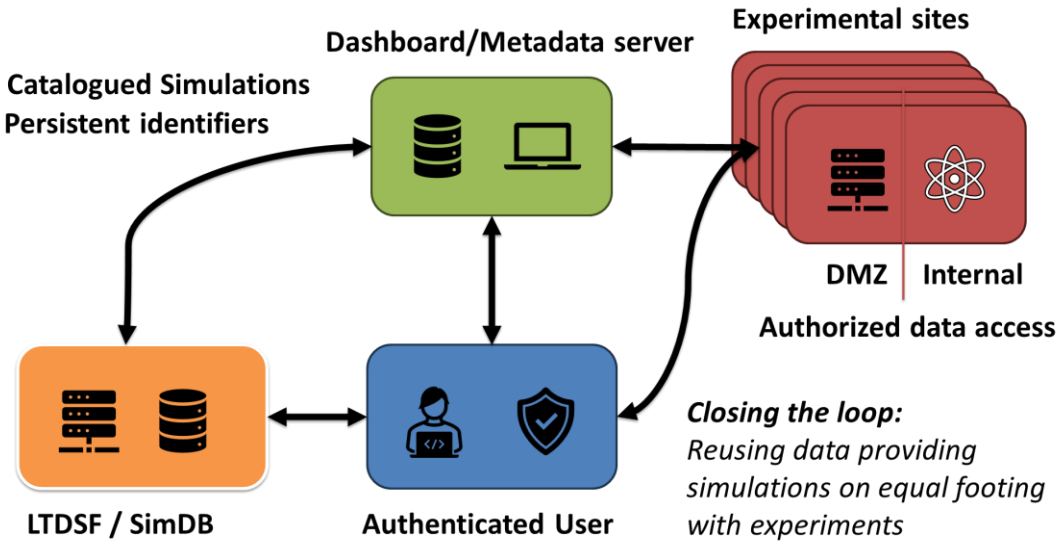
- A metadata catalogue that provide a searchable inventory of the available data (metadata - waveforms)
- (Remote) access to experimental data using a single toolset. (Mindful of not creating secondary repositories)
- Integration of simulation and modelling data in the same infrastructure (access and searchability).

General considerations

- Data availability - policies, embargos and formats
- Data quality, integrity and security – data mappings much more than just mapping namespaces

Implementation of the Data Management Plan

The longer term aim is to provide a production-ready, federated data management ecosystem for harmonised access to experimental and simulation data across EUROfusion sites and users and in the longer term allow for some elements to be fully open sourced.



Three internal activities:

- General Infrastructure components (software, hardware for db's and search engines, Authentication and Authorisation Infrastructure) (PSNC) [12PM]
- IT services at each site (facility) (AUG, COMPASS/ COMPASS-U, JET, MAST/MAST-U, TCV, WEST)
- Data mappings (All sites) [16.9PM + (14PM)]

General considerations

- Provenance capture
- Harmonizing conventions over different devices and tools.

Background - FAIR for Fusion (ww.fair4fusion.eu)

EURATOM Coordinated Support Action 2019-2022 aimed at defining data sharing methodologies and tools for fusion. Aiming at ”as open as possible as closed as necessary”. Open data sharing aim in the EUROfusion Grant Agreement.

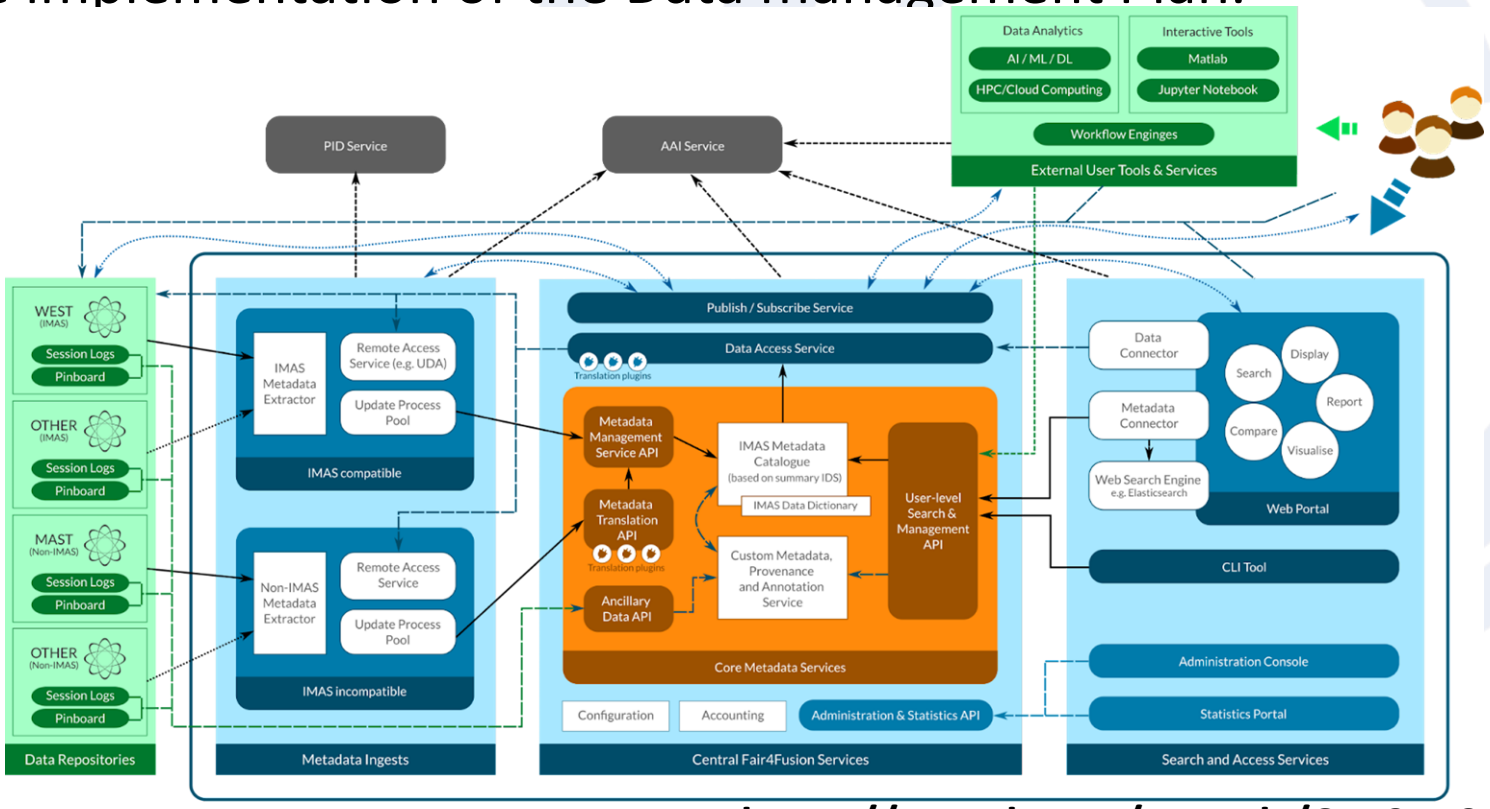
FAIR4Fusion defined a blueprint architecture based on detailed requirements and analysis
→ Being used as the basis for the implementation of the Data management Plan.

Findable

Accessible

Interoperable

Reusable



[Fair for Fusion - Cordis reports etc]

https://zenodo.org/records/6759119

Implementation of the DMP resourced in stages

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

WP 2023-2025

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

  Infrastructure in place (performance enhances needed), demo data available. User inputs/requirements needed!

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

  Activity prototyped 2025. Some further developments needed in view of access authorisation in 2026. USER INPUT/REQUIREMENTS KEY ! Leads into Scenario C.

WP 2026-2027

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

  Extension of the data mappings to cover provenance and start to put data in the hands of users! Input on data mappings and specific data for validation. LTDSF

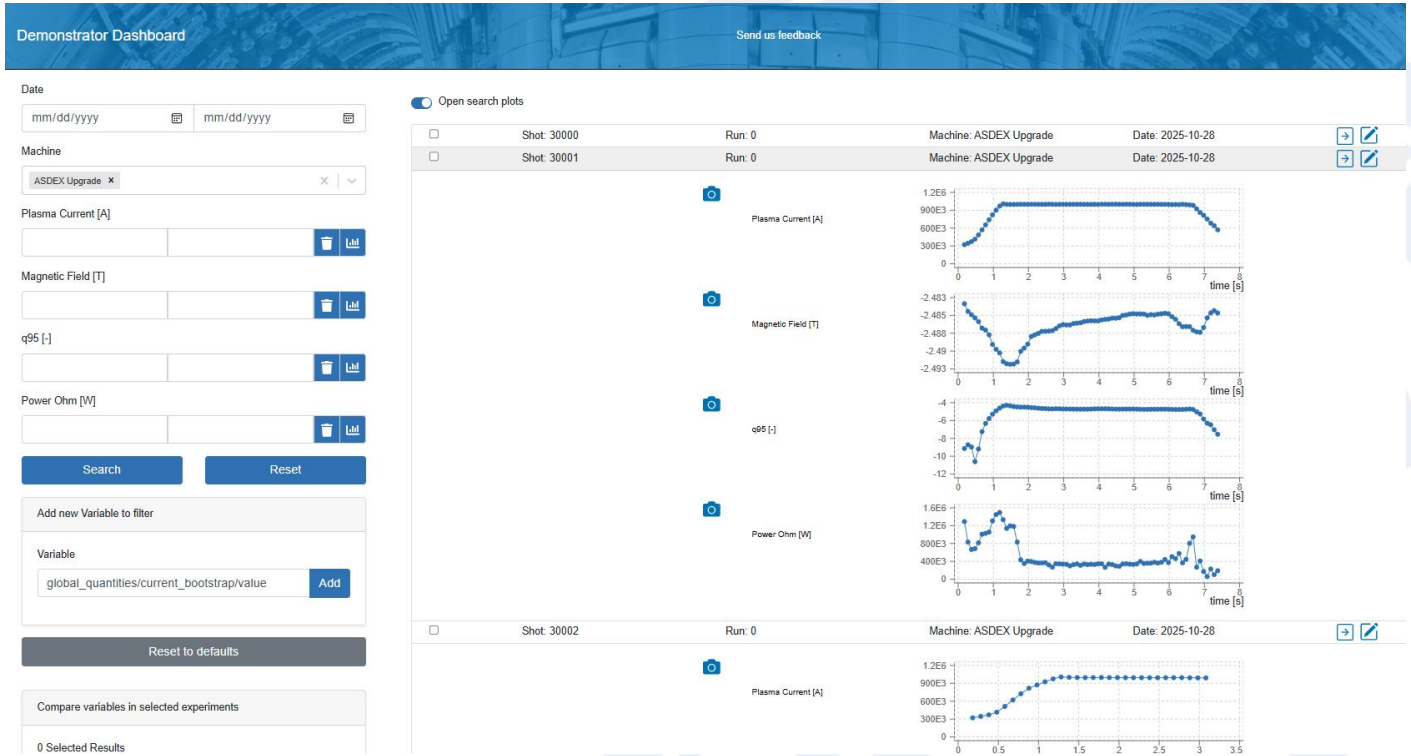
- Scenario D: adds a lightweight layer for open access to non-embargoed metadata and, where allowed by the facilities, also data access for export in human readable formats (CSV files) [F,A,I,R] and open.

  Activity starting now by defining use cases: TCV, WEST experimental data, data collections for code validation,…?

Inventory of EUROfusion data

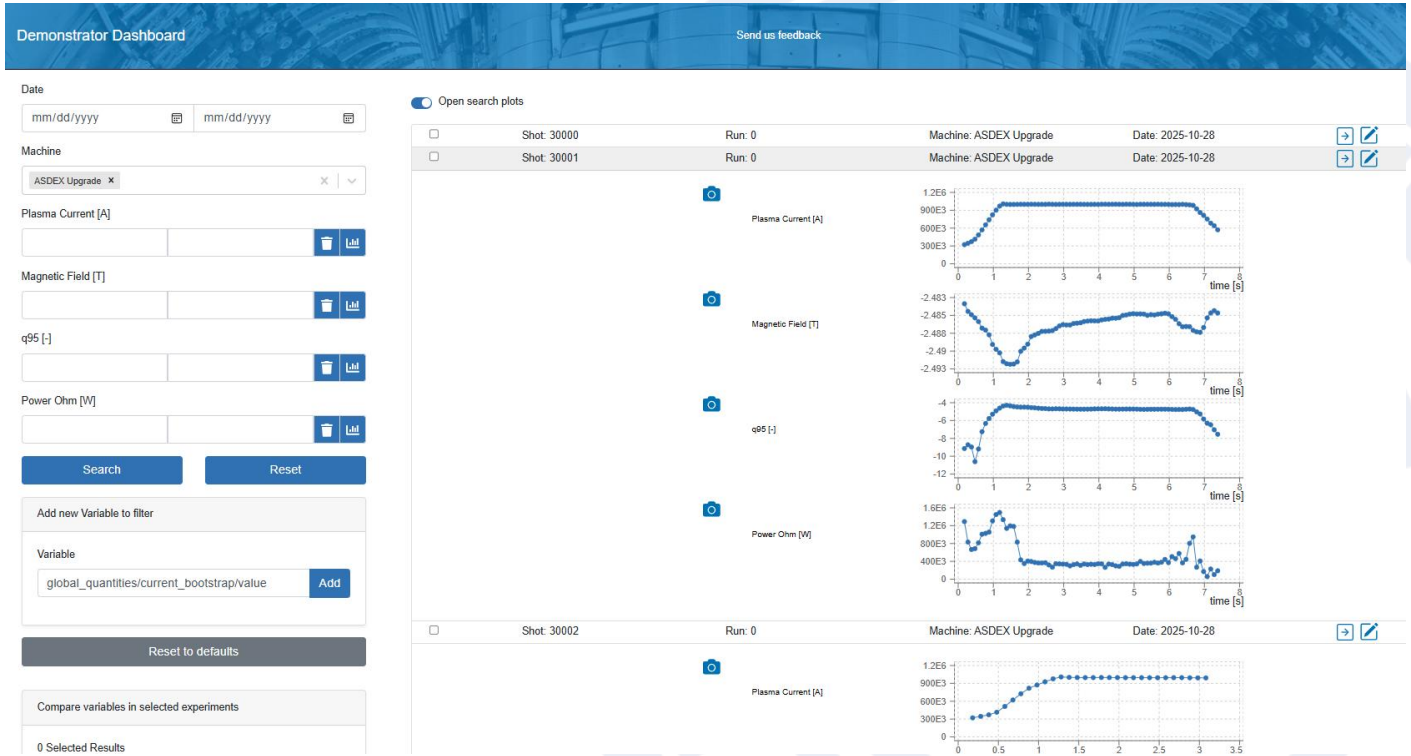
A. Searchable Inventory of experimental data

- Provide a searchable catalogue of experimental data from EUROfusion.
  - IMAS based conventions (consistent conventions over devices)
  - 1st version of Dashboard/Catalogue demonstrated (SOFT 2025)
  - Moving towards production on the new Gateway. (dmp.eufus.eu)
  - BUT also moving to new backend infrastructure for robustness and performance
- Overview data (IDS_summary)
  - Time dependent waveforms
    - IP, BT, q95, vloop,…
    - Extendable on user requests
    - Link to actual data sources



A. Searchable Inventory of experimental data

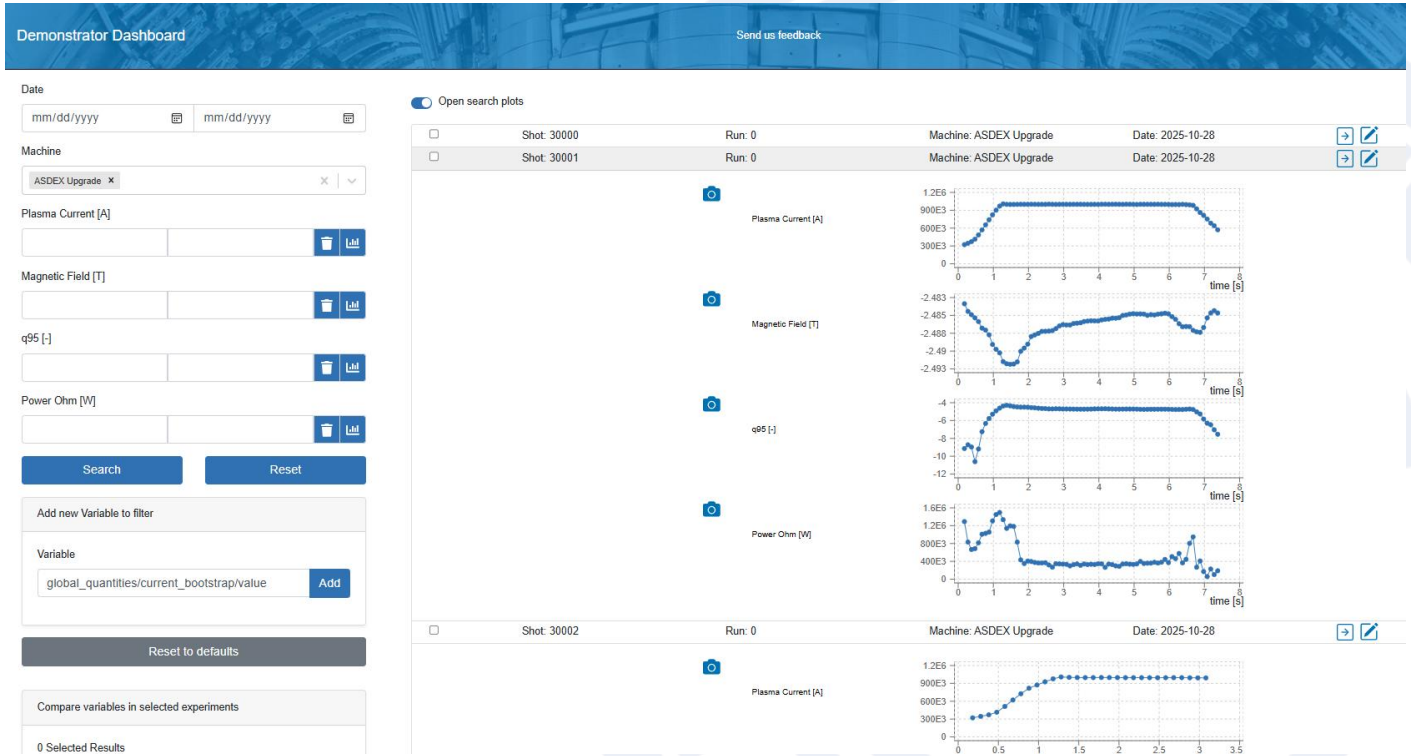
- Provide a searchable catalogue of experimental data from EUROfusion.
  - IMAS based conventions (consistent conventions over devices)
  - 1st version of Dashboard/Catalogue demonstrated (SOFT 2025)
  - Moving towards production on the new Gateway. (dmp.eufus.eu)
  - BUT also moving to new backend infrastructure for robustness and performance
- Overview data (IDS_summary)
  - Time dependent waveforms
    - IP, BT, q95, vloop,…
    - Extendable on user requests
    - Link to actual data sources



Metadata is context dependent – someones metadata is someone else's data and viceversa:
Additional meaning here: "metadata is also the waveforms needed for a researcher to not only know that a shot was performed but also to assess the shot for insights and future use".

A. Searchable Inventory of experimental data

- Provide a searchable catalogue of experimental data from EUROfusion.
  - IMAS based conventions (consistent conventions over devices)
  - 1st version of Dashboard/Catalogue demonstrated (SOFT 2025)
  - Moving towards production on the new Gateway. (dmp.eufus.eu)
  - BUT also moving to new backend infrastructure for robustness and performance
- Overview data (IDS_summary)
  - Time dependent waveforms
    - IP, BT, q95, vloop,…
    - Extendable on user requests
    - Link to actual data sources



The catalogue will serve as an inventory of available data and host for EF db's BUT it can evolve with different capabilities if needed. E.g.,

- Adding different data summaries (scatterplots, …)
- Automatic data labelling on import, …
- Serve as a data repository for AI/ML tools

Suggestions welcome…

A. Searchable Inventory of experimental data (adding simulation data)

Reasonable # of shots in demonstration database

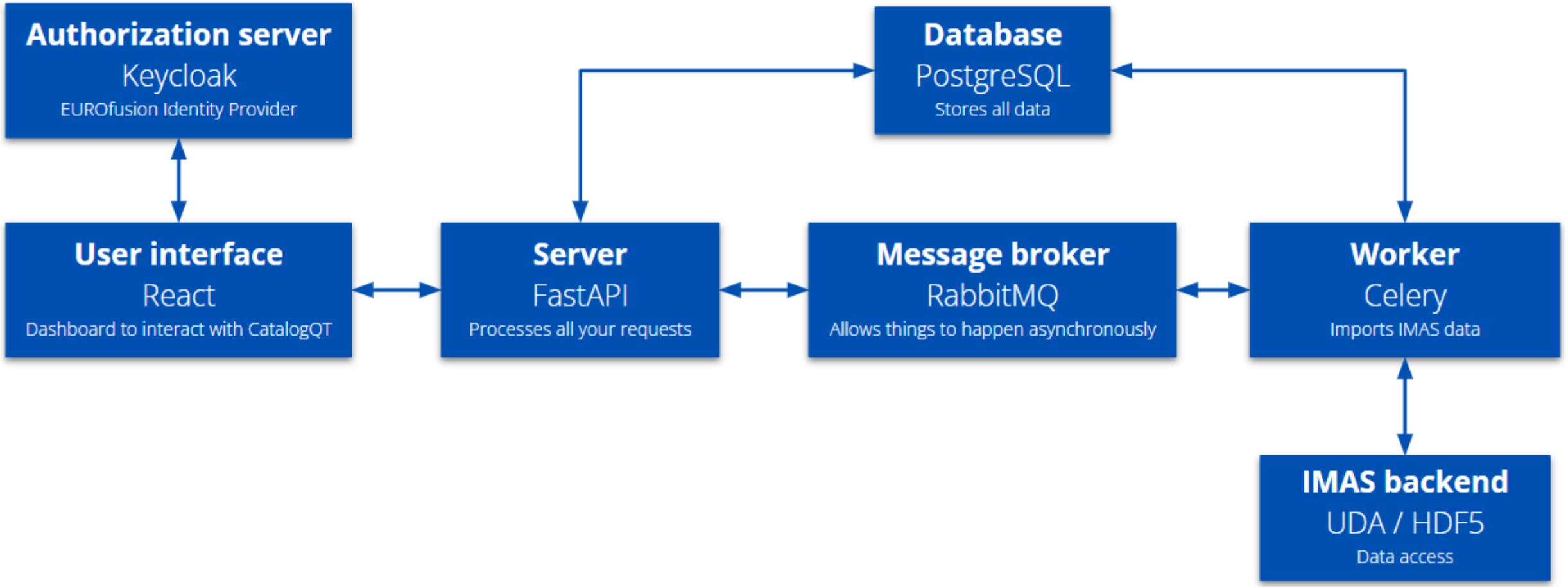
Performance upgrade needed and planned with version 2 backend (java → python)

For 2026 (release and promote)

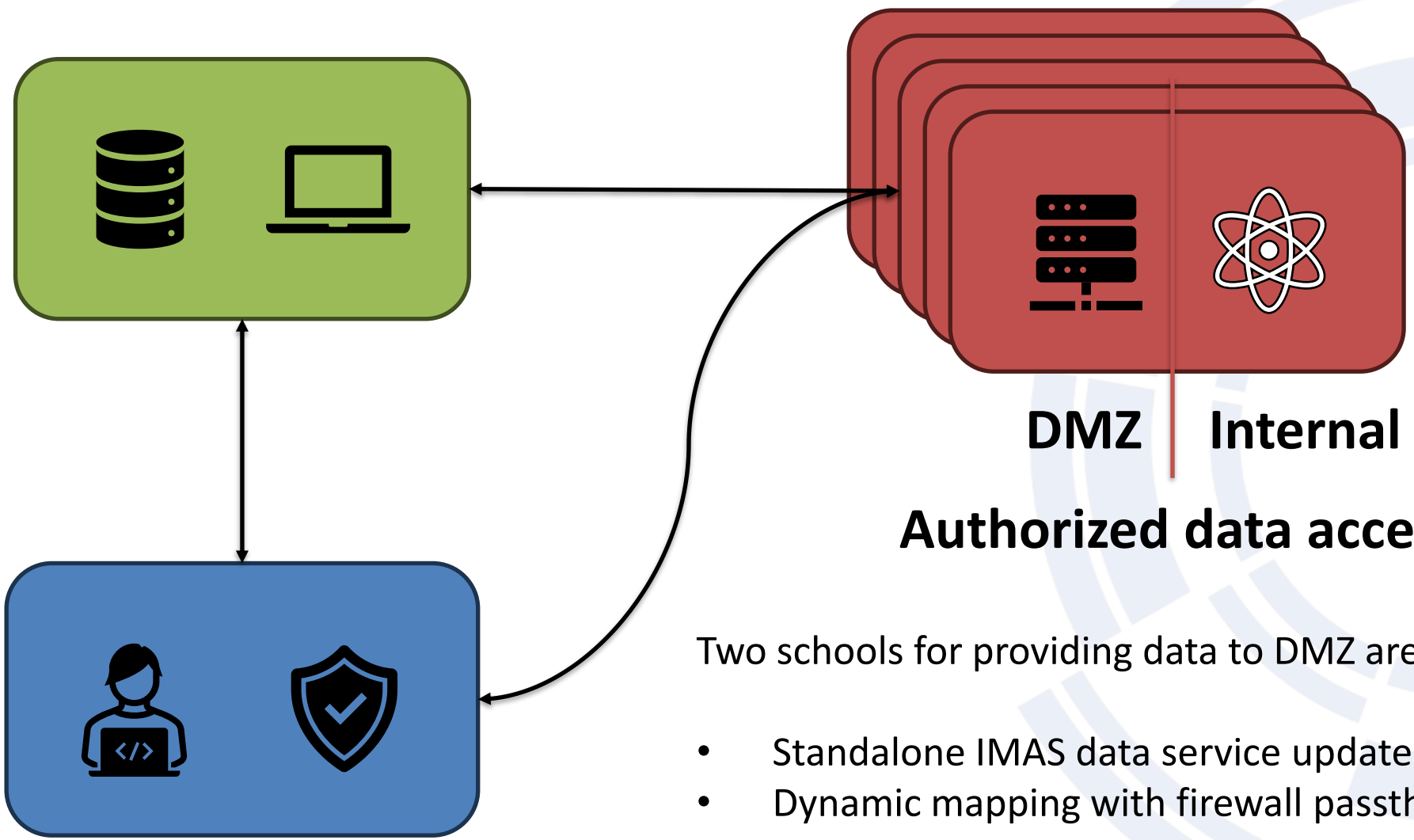
- Extend data coverage: with users (TSVVs etc) define minimum list of required quantities (all sites)
- Collect user feedback to improve usability, interfaces and identify performance bottlenecks.
- Integrate site data with push requests to core services for automated ingestion and updates of released data.

Device	Description	# of shots	Comment
AUG	Summary and dataset_description IDSes	11150	Currently with DD 3.39.0 and AL 5.0.0
COMPASS/COMPASS-U	Summary and dataset_description IDSes		Currently with DD 3.39.0 and available only internally. Selected COMPASS tokamak discharges will be provided, mapped to IMAS format via the UDA server. The infrastructure will be ready for the operation of the COMPASS Upgrade tokamak.
JET	Summary and dataset_description IDSs	DTE2: 99294 - 99982 DTE3: 103987- 104697	DTE2 and DTE3 3.42.0 Subset of pulses selected for demonstration but can be extended to include all pulses
MAST-U	Summary and dataset_description	40000 onwards (current shot is 52634)	Shot 40000 - current 3.42.0 Not publicly available due to embargo and data policy
MAST	Summary and dataset_description	M4 - M9	3.42.0 Not publicly available due to embargo and data policy
TCV	Summary and dataset_description	More than 10000 shots	Mostly DD 3.41 and AL 5
WEST	Summary and dataset_description IDSes	8000	All shots from campaigns C4 to C11

A. New Backend for waveform/metadata is being upgraded



Prototyping remote data access



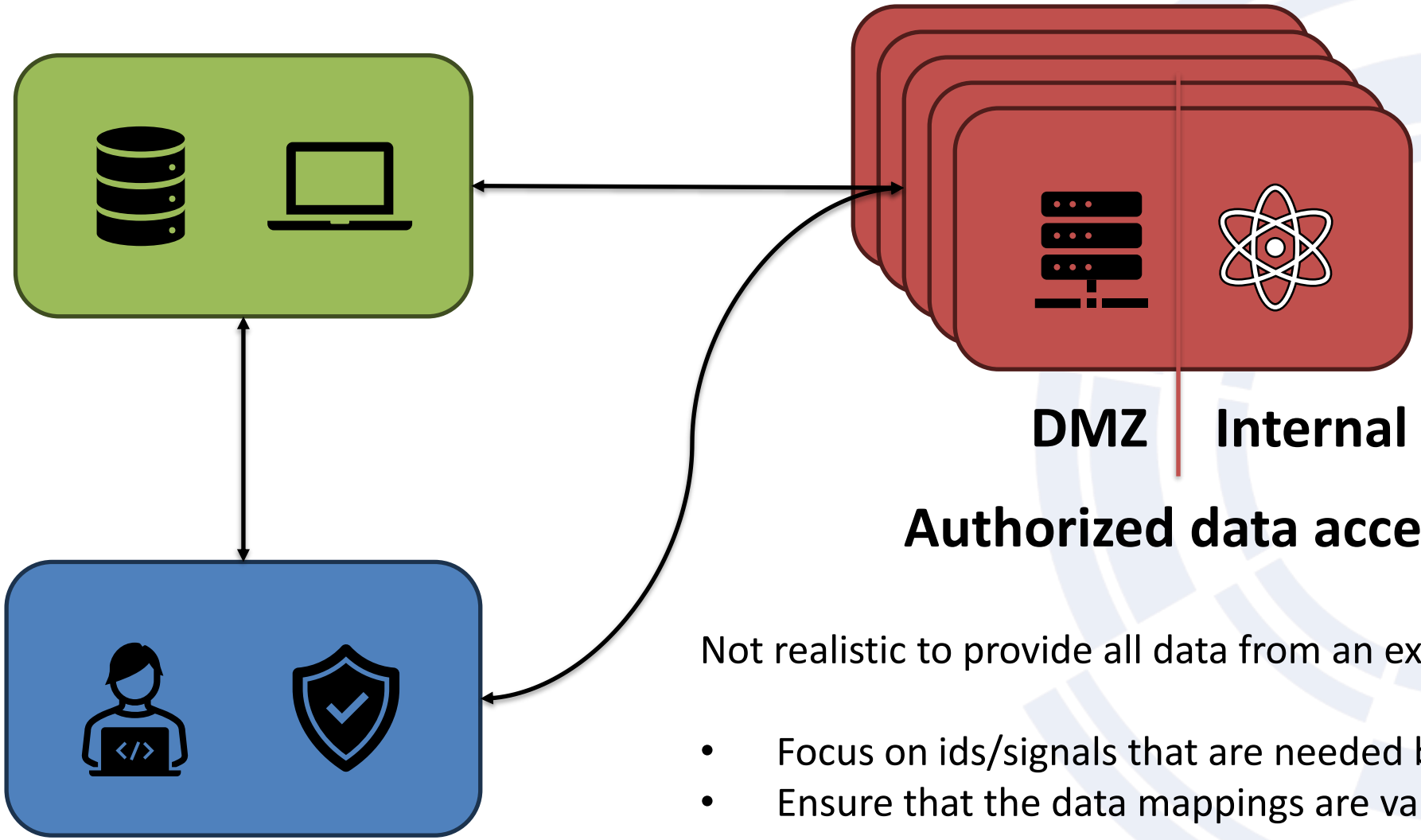
Two schools for providing data to DMZ area:

- Standalone IMAS data service updated on need/request
- Dynamic mapping with firewall passthrough

Availability of data resides with the devices. NB there is no Magic Bullet - data need to be mapped to IMAS unless native

Dashboard/Metadata server

Experimental sites



DMZ | Internal

Authorized data access

Not realistic to provide all data from an experiment:

- Focus on ids/signals that are needed by users/codes.
- Ensure that the data mappings are validated

Authenticated User on GW

Close collaboration with TSVV's are needed!

IDS name	Type	AUG	COMPASS	JET	MAST	MAST-U	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

IDS name	Type	AUG	COMPASS	JET	MAST	MAST-U	TCV	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
wall	Expt.	trview magnetics tool		tokamap		tokamap	gdat	WEST
waves	Derived	trview						

Color coding:

IDS – active, alpha

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

core_profiles

- core_profiles/profiles_1d(itime)/q(:)
- core_profiles/profiles_1d(itime)/electrons/density_thermal(:)
- core_profiles/profiles_1d(itime)/electrons/temperature(:)
- core_profiles/profiles_1d(itime)/t_i_average(:)
- core_profiles/profiles_1d(itime)/zeff(:)

Equilibrium

- equilibrium/time_slice(itime)/boundary/outline/r(:)
- equilibrium/time_slice(itime)/boundary/outline/z(:)
- equilibrium/time_slice(itime)/global_quantities/ip
- equilibrium/vacuum_toroidal_field/r0
- equilibrium/vacuum_toroidal_field/b0(itime)
- equilibrium/time_slice(itime)/global_quantities/magnetic_axis/r
- equilibrium/time_slice(itime)/global_quantities/magnetic_axis/z
- equilibrium/time_slice(itime)/profiles_2d(1)/grid_type/name = 'rec
- equilibrium/time_slice(itime)/profiles_2d(1)/grid_type/index = 1
- equilibrium/time_slice(itime)/profiles_2d(1)/grid/dim1(:)
- equilibrium/time_slice(itime)/profiles_2d(1)/grid/dim2(:)
- equilibrium/time_slice(itime)/profiles_2d(1)/psi(:,:)
- equilibrium/time_slice(itime)/profiles_1d/psi(:)
- equilibrium/time_slice(itime)/profiles_1d/f(:)
- equilibrium/time_slice(itime)/profiles_1d/q(:)
- equilibrium/time_slice(itime)/profiles_1d/pressure(:)
- equilibrium/time_slice(itime)/boundary/psi
- equilibrium/time_slice(itime)/profiles_1d/rho_tor_norm
- equilibrium/time_slice(itime)/profiles_1d/j_tor
- equilibrium/time_slice(itime)/global_quantities/psi_axis
- equilibrium/time_slice(itime)/global_quantities/psi_boundary

core_sources

- core_sources/source(inbi)/profiles_1d(itime)/electrons/energy(:)
- core_sources/source(inbi)/profiles_1d(itime)/total_ion_energy(:)
- core_sources/source(inbi)/profiles_1d(itime)/j_parallel(:)
- core_sources/source(inbi)/profiles_1d(itime)/ion(iion)/element(1)/a (ion(iion) (needs to be set for the main ion species that are set up in JETTO)
- core_sources/source(inbi)/profiles_1d(itime)/ion(iion)/element(1)/z_n
- core_sources/source(inbi)/profiles_1d(itime)/ion(iion)/atoms_n
- core_sources/source(inbi)/profiles_1d(itime)/ion(iion)/particles(:)
- core_sources/source(inbi)/profiles_1d(itime)/momentum_tor(:)
- core_sources/source(iic)/profiles_1d(itime)/electrons/energy(:)
- core_sources/source(iic)/profiles_1d(itime)/total_ion_energy(:)
- core_sources/source(iic)/profiles_1d(itime)/j_parallel(:)
- core_sources/source(iec)/profiles_1d(itime)/electrons/energy(:)
- core_sources/source(iec)/profiles_1d(itime)/j_parallel(:)
- core_sources/source(ilh)/profiles_1d(itime)/electrons/energy(:)
- core_sources/source(ilh)/profiles_1d(itime)/total_ion_energy(:)
- core_sources/source(ilh)/profiles_1d(itime)/j_parallel(:)

We want to expand the data mappings to cover the different use cases in EUROfusion.

A list of needed IDS's is a good start but need more details - down to signal level -to be really useful.

core_profiles

- core_profiles/profiles_1d(itime)/q(:)
- core_profiles/profiles_1d(itime)/electrons/density_thermal(:)
- core_profiles/profiles_1d(itime)/electrons/temperature(:)
- core_profiles/profiles_1d(itime)/t_i_average(:)
- core_profiles/profiles_1d(itime)/zeff(:)

Equilibrium

- equilibrium/time_slice(itime)/boundary/outline/r(:)
- equilibrium/time_slice(itime)/boundary/outline/z(:)
- equilibrium/time_slice(itime)/global_quantities/ip
- equilibrium/vacuum_toroidal_field/r0
- equilibrium/vacuum_toroidal_field/b0(itime)
- equilibrium/time_slice(itime)/global_quantities/magnetic_axis/r
- equilibrium/time_slice(itime)/global_quantities/magnetic_axis/z
- equilibrium/time_slice(itime)/profiles_2d(1)/grid_type/name = 'rec
- equilibrium/time_slice(itime)/profiles_2d(1)/grid_type/index = 1
- equilibrium/time_slice(itime)/profiles_2d(1)/grid/dim1(:)
- equilibrium/time_slice(itime)/profiles_2d(1)/grid/dim2(:)
- equilibrium/time_slice(itime)/profiles_2d(1)/psi(:,:)
- equilibrium/time_slice(itime)/profiles_1d/psi(:)
- equilibrium/time_slice(itime)/profiles_1d/f(:)
- equilibrium/time_slice(itime)/profiles_1d/q(:)
- equilibrium/time_slice(itime)/profiles_1d/pressure(:)
- equilibrium/time_slice(itime)/boundary/psi
- equilibrium/time_slice(itime)/profiles_1d/rho_tor_norm
- equilibrium/time_slice(itime)/profiles_1d/j_tor
- equilibrium/time_slice(itime)/global_quantities/psi_axis
- equilibrium/time_slice(itime)/global_quantities/psi_boundary

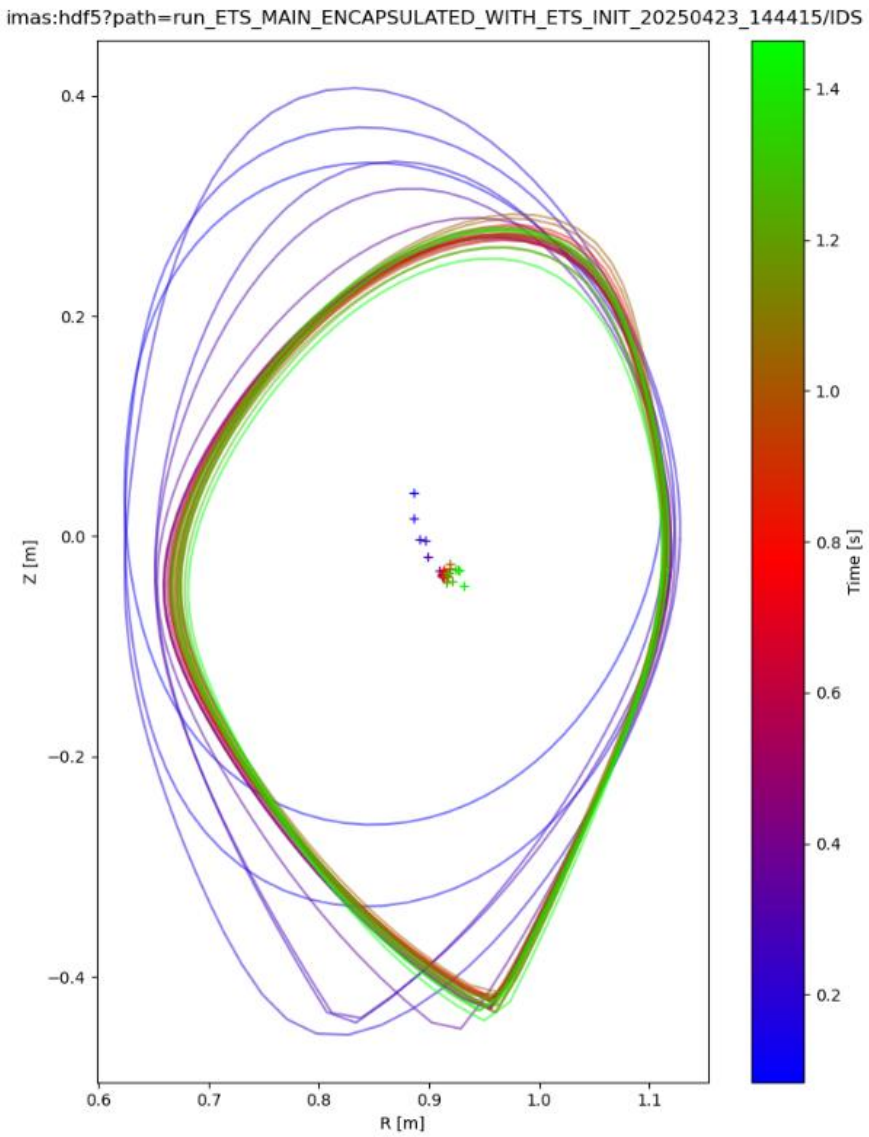
core_sources

- core_sources/source(inbi)/profiles_1d(itime)/electrons/energy(:)
- core_sources/source(inbi)/profiles_1d(itime)/total_ion_energy(:)
- core_sources/source(inbi)/profiles_1d(itime)/j_parallel(:)
- core_sources/source(inbi)/profiles_1d(itime)/ion(iion)/element(1)/a (ion(iion) (needs to be set for the main ion species that are set up in JETTO)
- core_sources/source(inbi)/profiles_1d(itime)/ion(iion)/element(1)/z_n
- core_sources/source(inbi)/profiles_1d(itime)/ion(iion)/atoms_n
- core_sources/source(inbi)/profiles_1d(itime)/ion(iion)/particles(:)
- core_sources/source(inbi)/profiles_1d(itime)/momentum_tor(:)
- core_sources/source(iic)/profiles_1d(itime)/electrons/energy(:)
- core_sources/source(iic)/profiles_1d(itime)/total_ion_energy(:)
- core_sources/source(iic)/profiles_1d(itime)/j_parallel(:)
- core_sources/source(iec)/profiles_1d(itime)/electrons/energy(:)
- core_sources/source(iec)/profiles_1d(itime)/j_parallel(:)
- core_sources/source(ilh)/profiles_1d(itime)/electrons/energy(:)
- core_sources/source(ilh)/profiles_1d(itime)/total_ion_energy(:)
- core_sources/source(ilh)/profiles_1d(itime)/j_parallel(:)

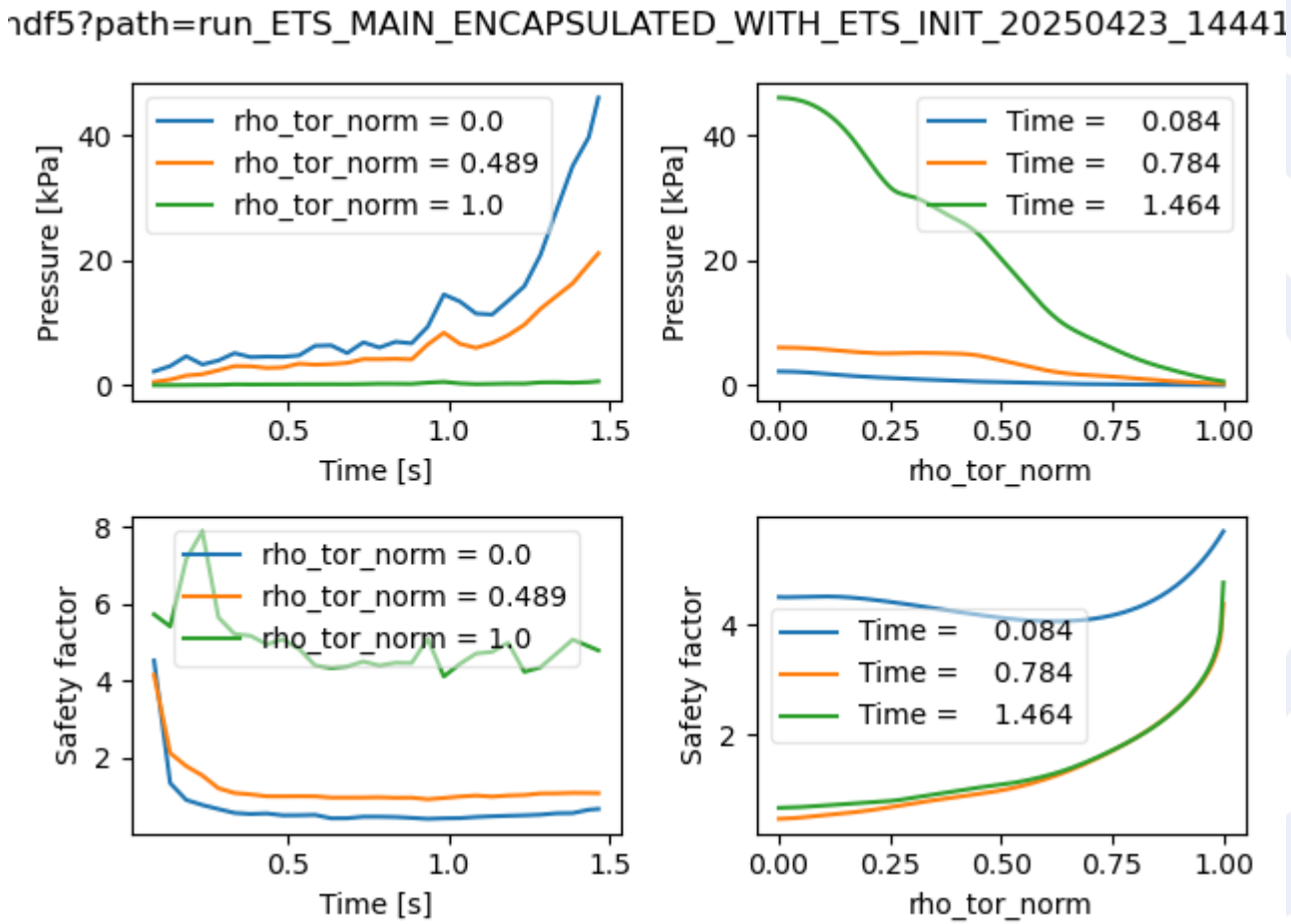
Minimum list of inputs for HFPS (run case dependent): core_profiles, equilibrium, core_sources, pulse_schedule, nbi, pellets,...

BUT for complete and working data mappings specification down to signal level is needed.

Validation of TCV data - interpretative run with ETS



Boundary evolution for the TCV 68340 interpretative run



The workflow has been saved in simdb as **ecf65ca0-210c-11f0-8eb7-0894ef442712** with the alias **g2dpc/tcv/68340/run_ETS_MAIN_ENCAPSULATED_WITH_ETS_INIT_20250423_144415**.

Moving forward

Tasks for 2026

Technical Specification:

1. Data Delivery & Catalogue

- Adopt a **minimum required signal set** and implement it in validation workflows.
- Enable **automated data push and update integration** with central services.
- Promote catalogue tools and access to users.

2. Data Mappings & Use Cases

- Extend and maintain **data mappings** driven by modelling needs.
- Work directly with code owners to validate **use-case–specific mappings**.
- Include derived quantities, pulse-specific settings, and device-specific needs.
- Apply DMP-aligned **validation rules**.

3. IMAS Data Dictionary Migration

- Support **migration to IMAS DDv4**, including code impact.
- Support mapping updates focused on **mandatory signals within recommended IDS list**.

4. Modelling Data Integration (Scenario C)

- Support integration of **SimDB modelling data** in IMAS formats.
- Contribute to joint experiment–modelling data workflows.

5. Provenance & PIDs

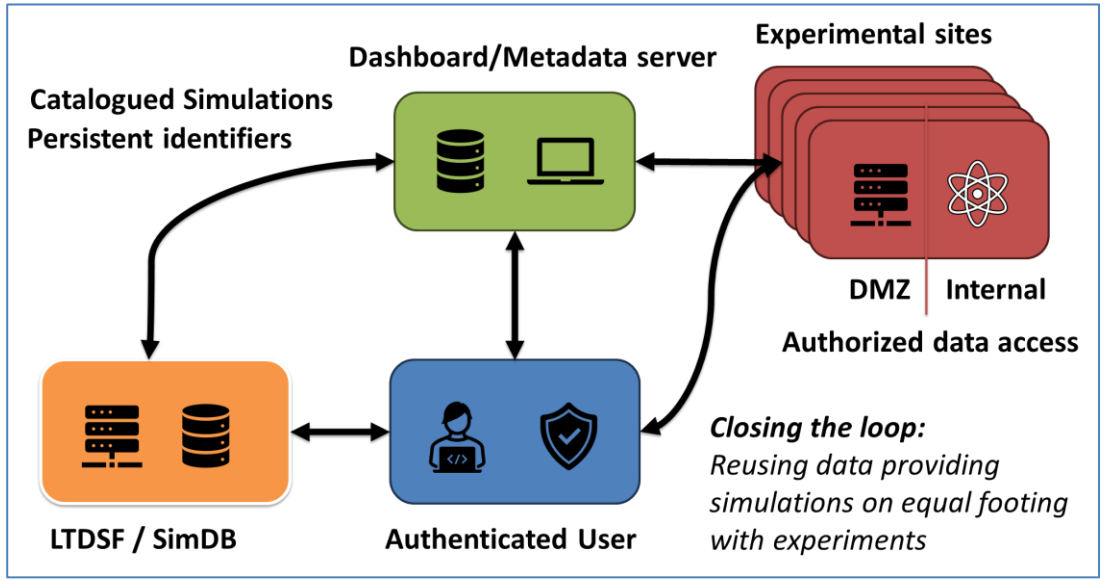
- Extend **provenance metadata** and update datasets.
- Contribute to and apply a **common PID minting strategy**.

Need to base work on the needs from different users and use cases.(Some identified and contacts taken.)
Room for more - get in touch if interested

Formalise requirements capture through WP TM , WP TE and DSO lead.

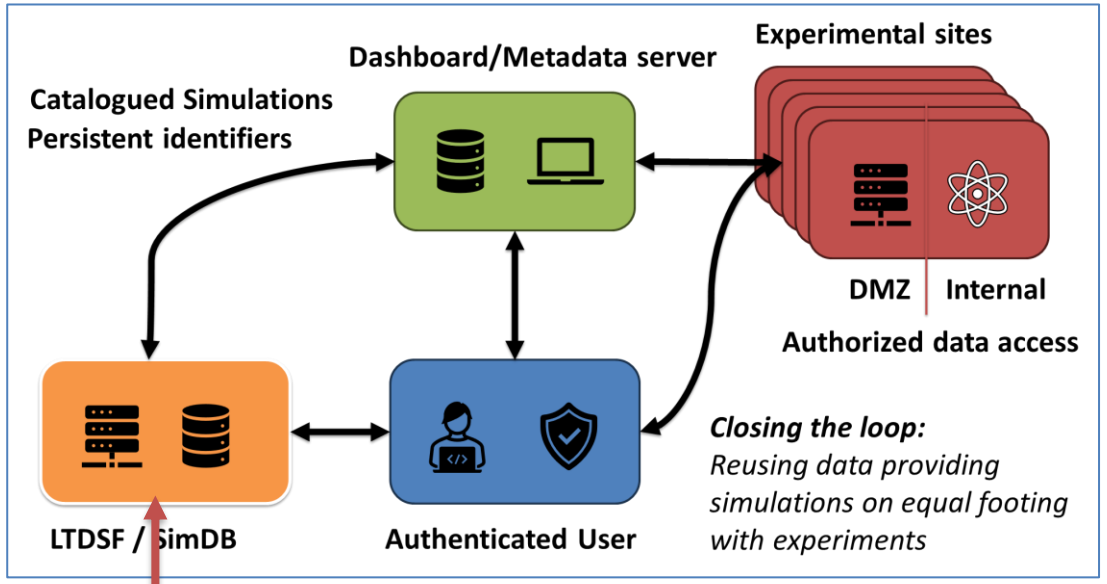
2026-27 moving to exploitation including simulation

- Adding simulation data as another facility.
- Proposal to integrate SimDB in the data workflow
- SimDB available on GW and can link to several levels of simulations input/outputs



2026-27 moving to exploitation including simulation

- Adding simulation data as another facility.
- Proposal to integrate SimDB in the data workflow
- SimDB available on GW and can link to several levels of simulations input



Ingest simulation results back into catalogue ”to close the circle”

Low hanging fruit: Open data access

Make the UDA server in the DMZ Publicly available. Open access but not necessary FAIR

Grant Agreement Clause:

” Within Horizon Europe open data for fusion research is required, and the initial format for data sharing to be based on IMAS data dictionary will be proposed. “

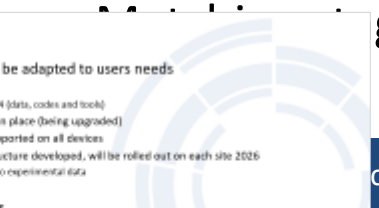
Delivering scenario C is a grant deliverable.

As part of the DMP activity, we will also provide some data from several devices as open source and propose a structure for broader use in 2027 - supporting candidate sites (TCV, WEST) in opening their data.

Milestones overview

No	Title	Description	Expected date
1	MA. 1	Provide scenario A data using DD 3	2026/Q1
2	MA. 2	Provide a list of recommended signals for scenario A	2026/Q2
3	MA. 3	Provide updated scenario A data using DD 4	2026/Q4
4	MA. 4	Include modelling and simulation data in metadata Catalogue	2027/Q2
5	MB. 1	Provide a list of recommended signals from first use cases (live document)	2026/Q1
6	MB. 2	Provide scenario B data using DD 3	2026/Q2
7	MB. 3	Strategy for Migration to DD 4	2026/Q2
8	MB. 4	Implement AAI access to data (on all sites)	2026/Q4
9	MB. 5	Provide scenario B data using DD 4	2027/Q2
10	MC. 1	Strategy for PID integrations	2026/Q3
11	MC. 2	Release of data with enhanced provenance	2026/Q4
12	MC. 3	Release of data with PIDs	2027/Q2
13	MD. 1	Plan for open data releases	2026/Q4
14	MD. 2	Release of open data from at least one device	2027/Q4

ges for core infrastructure - continuous validation with users



Summary

- Infrastructure largely in place – to be adapted to users needs
  - IMAS based data structure and toolsets
    - Need to manage transition from DD3 to DD4 (data, codes and tools)
  - Dashboard/Catalogue for data inventory in place (being upgraded)
  - UDA based client server infrastructure supported on all devices
  - Authentication and Authorisation Infrastructure developed, will be rolled out on each site 2026
    - Needed to allow controlled remote access to experimental data

- 2026-27 extensions/developments
  - Moving from prototyping to production use
    - Enhanced provenance and inclusion of persistent identifiers
  - Integrating simulation /modelling data output
    - Proposed through SimDB building on ITER and TSVV-11, TSVV-H experiences
  - Demonstrating / Supporting Open data access. Infrastructure
    - Candidate devices (TCV, WEST)
    - WP TE repository (emerged this meeting)
- USER input needed: Data requirements, Validation of data mappings,