1 st Sites and Core Services (SiCo) meeting

Intention of the meeting is to provide a forum for technical discussion on the interfaces and integration of hte core and site services. Today was intended to be the first meeting with detailed discussions.

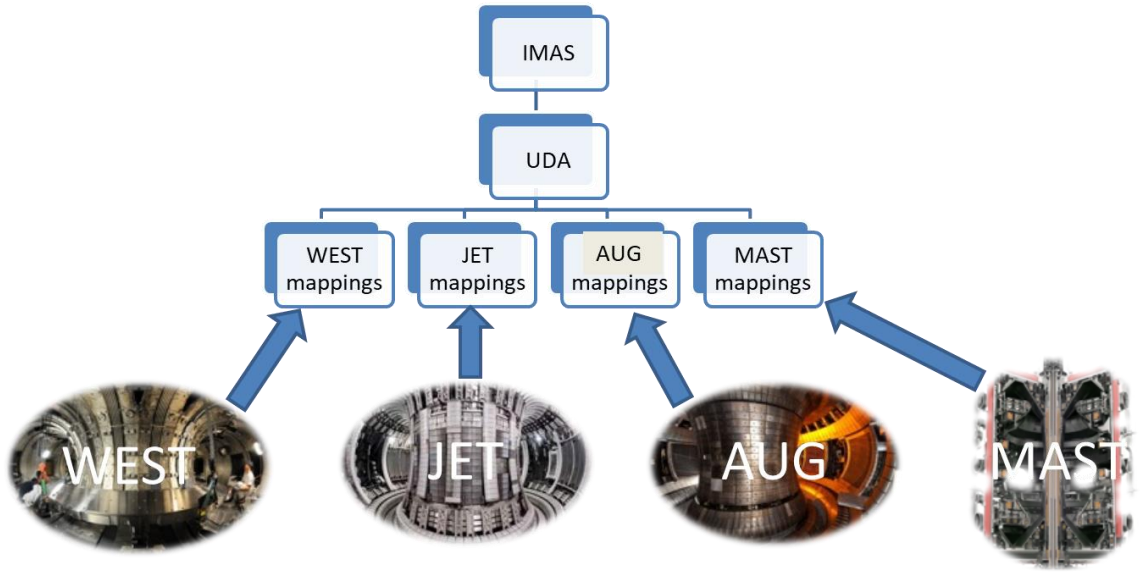
- Aim: providing a structured approach for implementing scenario A an to start prototyping scenario B. However a slight disruption for today…
- New developments on the EUROfusion level need to be urgently adressed.
  - About 2PPY/yr has been set aside to provide Imasification of machine data for AUG, TCV, WEST, JET, MAST, MAST-U, W7-X, COMPASS, [COMPASS-U], JT-60SA
  - Volker, Denis, Sara and Duarte participating to get input on how to move forward.
- We will schedule a second SiCo meeting soon if needed to make up for lost time.

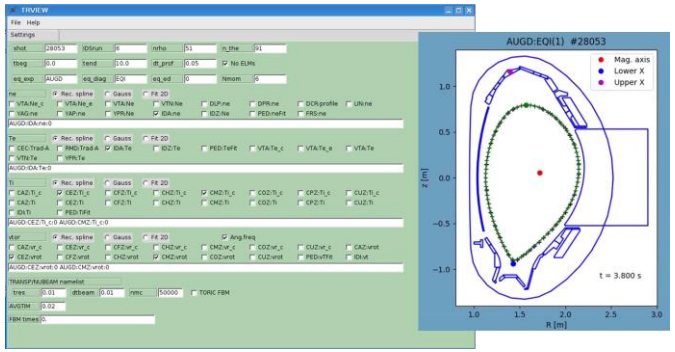
Agenda (for approval)

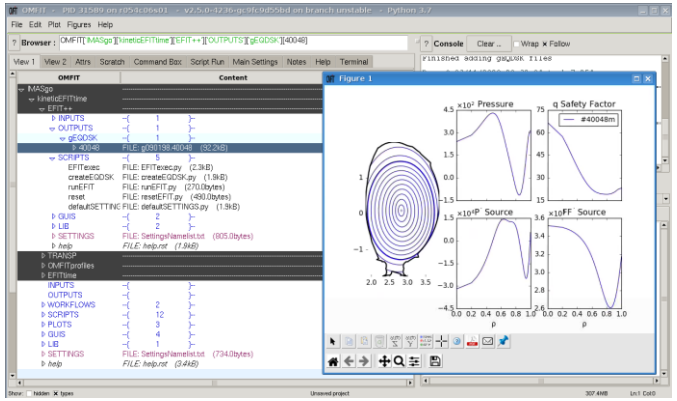
- Background info (PS)
- Discussion on possible implementation for imasification of machine data (I)
- Technical presentation from services
- [Revisit: Discussion on possible implementation for Imasification of machine data (II)]
- Recommendation/advice on Imasification implementation
- Next steps for implementation of DMP – scheduled activities

EUROfusion tools to access experimental data in IMAS format from WPCD to TSVVs

- **Bring data to users:** Implementing and supporting access methodologies for EUROfusion devices (Machine descriptions, data mappings and data access)
  - UDA (In principle available but yet to be fully established as the general tool)
  - Current practices: Bespoke toolsets used to map data from experiments
    - Exp2itm, Trview (for AUG data), readAUG, IMASgo! (Omfit plugin), TCV2IDS, ex2GK…







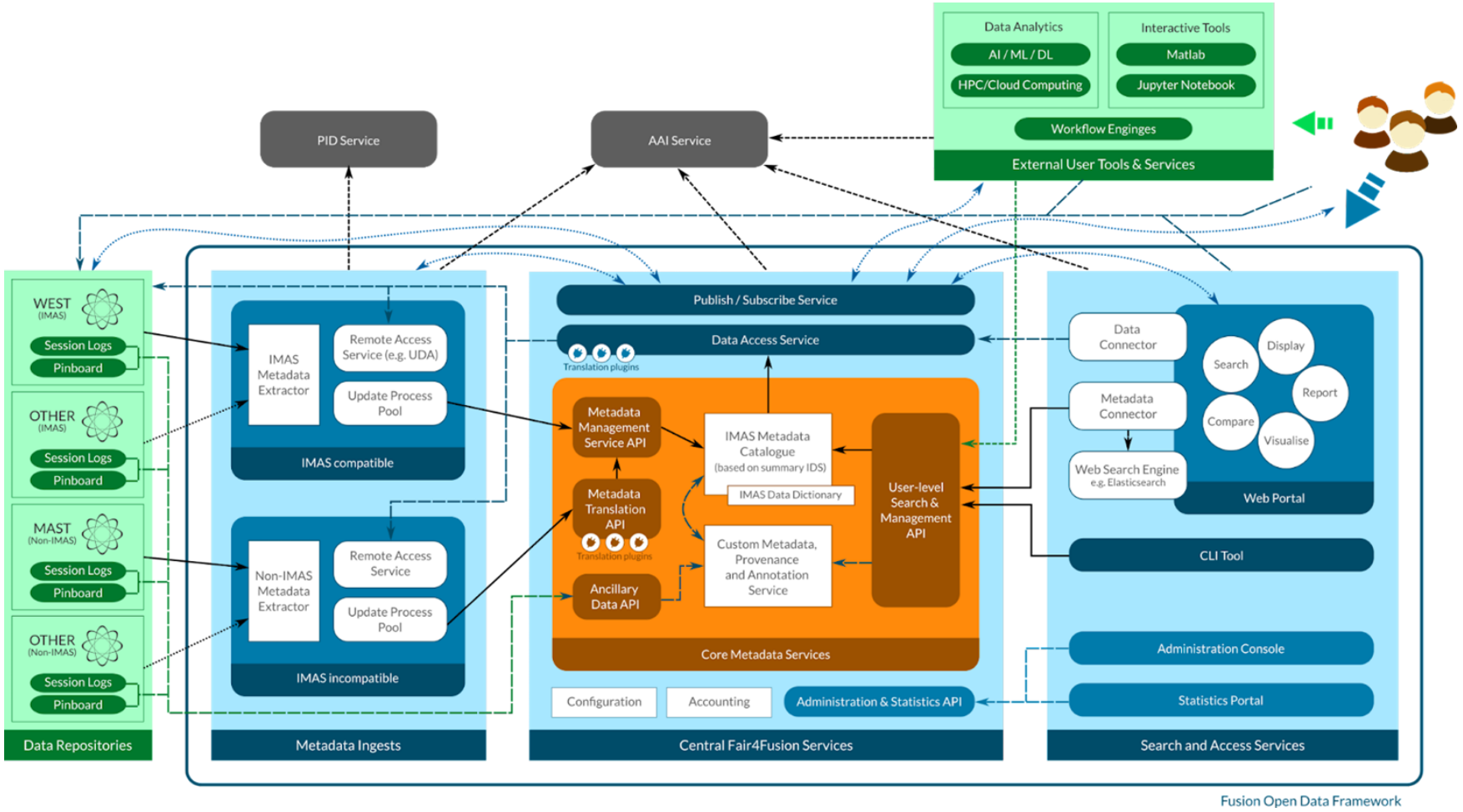
With the IMAS/UDA paradigm we gain the additional benefit of ITER being automagically technically integrated in the EUROfusion device ecology. Robustness and ease of implementation?

Examples of IDSs from codes being IMASified by ACH-04

TSVV	CODE	IDS list
1	GENE	equilibrium, core_profiles, core_transport
3	SOLEDGE3X	equilibrium, edge_profiles, edge_transport, turbulence, wall
5	EIRENE	Equilibrium, edge_profiles,…
7	ERO2.0	edge_profiles, wall
8	JOREK	equilibrium, mhd, wall?
10	HYMAGYC	equilibrium, mhd_linear, mhd + HFPS like
11	HFPS, ETS	Equilibrium, core_profiles, core_sources, edge_profiles. Edge_sources, wall, nbi, ec_launchers, ic_antennas, tf, pellets, interferometer, pulse_schedule, radiation, core_transport, edge_transport, …
11	DYON	core_profiles, core_sources, equilibrium, ec_launchers, em_coupling, gas_injection, magnetics, pf_active, pf_passive, plasma_initiation, pulse_schedule, radiation, wall
12	ASCOT5	equilibrium, nbi, distributions,…

There is a lot of complexity hidden in these: how is processed ids obtained (core_profiles, core_sources) – needs for specific formats ggd structured 3-d data etc…, interpretative or predictive ….

Design driven by User Stories → Blueprint architecture



Scenarios

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

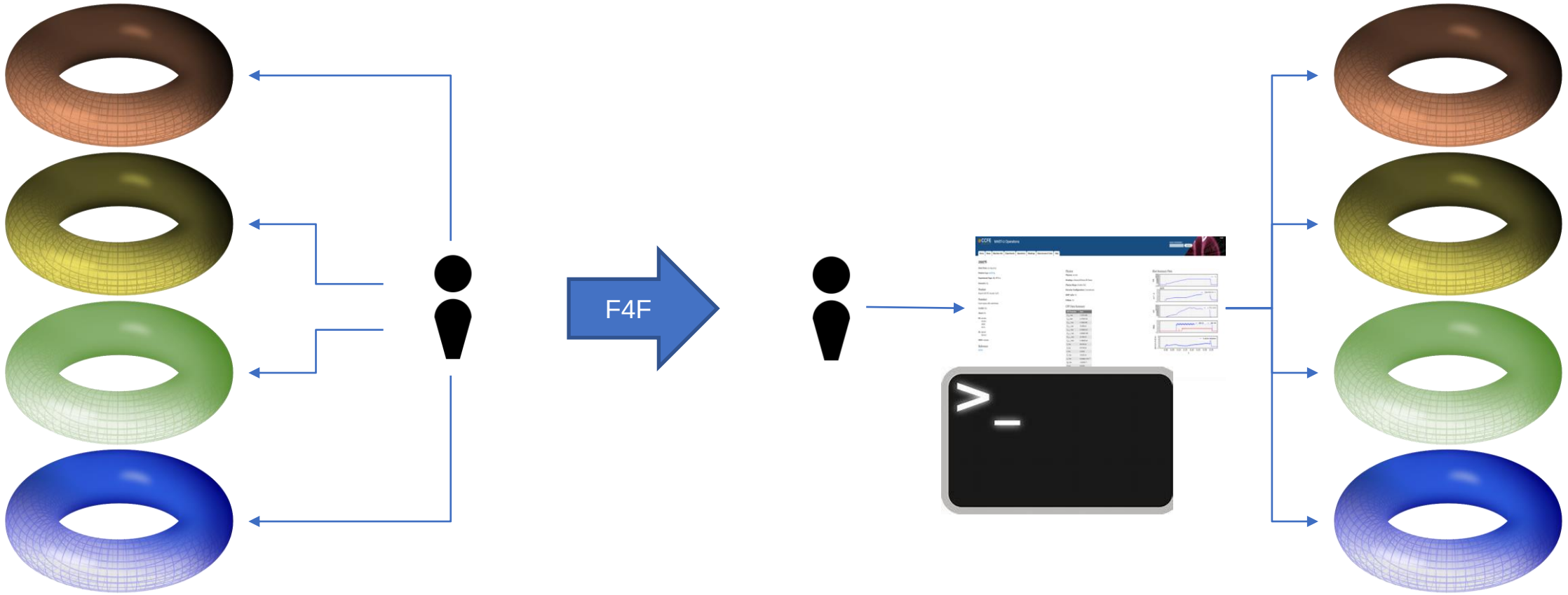
Scenario B: adds to Scenario A by allowing a subset of the data to be accessed using common tools (for example UDA). Facilities are responsible for the access level and qualification of data through the data mappings [F,A,I,(R)] prototype → implement (FSD proposal)

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

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. defer!

Only Scenario D implies open data outside of EUROfusion members.

Streamlining data access for end users



Design well aligned with user requriements

Most advanced and well expressed needs from TSVV11 (F. Casson, C. Bourdelle)

- Remote access to experimental data from EF devices, ideally with a single tool to access expt. data in IMAS form from all the devices! This would be waveforms, core profiles, and equilibrium…. Scenario B
- Access to searchable metadata across devices, ideally on a single dashboard Scenario A
- Single sign-on credentials across Eurofusion data access and services!
- A long term simulation data store to support SimDB and IMAS datasets

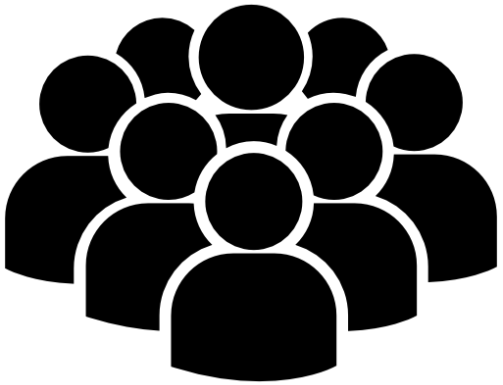
Additional needs from AI/ML community (conclusions from FSD Science Coordination Meeting - AI & ML)

- A framework for annotated data would help accelerate the application of AI/ML, and storage is also an important subject to consider for the future. Scenario A +B
- DMP initiative is important: to accelerate the progress the implementation of full scenario **B** is important (preferably in FP9, WEST and AUG can be pilot projects) . Scenario B
- Establish a mechanism to support local data experts (for experimental machine data) to help on data validation / annotation / mapping to IMAS Site services in Scenario A+B
- In future, introduction of version control for data sources is important (*discussions within DMP might be needed, Scenario C!*)

Coordination:

- Regular open meetings with reports on(status of implementation, services, response to request and new data proposals);
- Implementation of a DVCM like structure for user request and proposals; and
- User trainings and demonstration (Demonstration sessions and training, Maintenance of Asynchronous information archives).
- In dialogue with the PMU, foster exchanges with external entities: infrastructure components (mainly other data and data service providers) and applications (non-Eurofusion) devices

Stakeholders meetings



Site services:

Scenario A.

- a local IMAS installation is maintained;
- existing metadata is mapped to IMAS formats (i.e., filling elements of summary_ids and providing dataset descriptions); and
- metadata is ingested (pushed) to the central metadata catalog.
- The site team is expected to develop the methodology for metadata ingestion together with the Central team.

Scenario B

- extended data mappings (e.g., towards supporting interpretative, integrated data analysis)
- support for remote access through UDA or equivalent tools, for authenticated users.

SiCo Meetings

Core Services

Scenario A + B

- an initial deployment of the Fair4Fusion portal is provided on the gateway;
- the portal interfaces, protocols and backend technologies are extended (over time) to provide a hardened production facility for end users;
- support for metadata ingestion and if needed extension of the access methodologies in the portal is provided to the site;
- a structured support system is put in place for the end users;
- integration with the EUROfusion wide AAI system is implemented; and
- further extensions towards direct data access is promoted.

Implementation proposal

Scenarios are cumulative adn we have customers/needs for both scenarios. Need to review the resource balance between the activities relating to each to come up with a realistic timeline.

Embedded in the site services

- Additional resources towards site services ”extended data mappings (e.g., towards supporting interpretative, integrated data analysis)”
- Develop a strategy for processed ids: core_profiles

Core services:

- Review priorities between A and B resources

If needed,

- Ensure technology support for UDA and other tools in ACH – can funding be channeled from 2ppy/yr if needed?

SiCO 2

- Continued technology discussions
- Technical input
    - UDA status (Jonatahan Hollocombe, UKAEA)
    - Experience of developing MAST UDA plugin (Adam Parker, UKAEA)
- When 202305-12 1400CEST

Data available in IMAS form*

IDS Name	JET	TCV	AUG	MAST	WEST
iron_core					
magnetics					
mse					
pf_active					
tf					
thomson_scattering					
wall					
core_profiles					
equilibrium					
nbi					
ic_antennnas					
ec_antennnas					
core_sources					

Colour	Meaning
	Data not relevant for this machine
	Data is missing
	Data mapping in development
	Data available
	Data validated as input of EWE-2 and EWE-3 workflows

- Initial experimental input datasets provided for “all” EUROfusion machines
- - Iterative process with workflow owners to test / extend the datasets as required
- Alternates to UDA to process native data and map them in IMAS/IDS have been developed to target specific workflows:
  - TRVIEW/GUI (AUG)
  - IMASgo (MAST, JET, K-Star, DIII-D)
  - TCV2IDS (TCV)
  - ReadAUG (AUG)
  - exp2ITM (CPOs + converter), ex2GK (cmp A. Ho)
  - UDA (MAST, JET, magnetics)

Tools developed/adapted under WPCD, partially taking over by TSVVs as per individual needs.

* Current as of a year ago..

No formal continuation in FP9!!!! Not clear who is (if anyone) responsible for providing this capacity for the community.

Opiniated view of the presenter: key issue need to have strong mgmt with clear responsibilities and resourced for the longer term

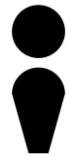
Access paradigm

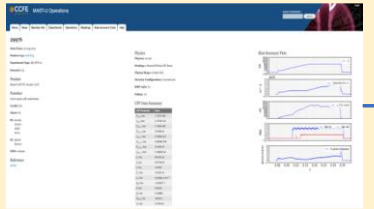
- Central metadata ”portal” for searches, summaries
- Remote data access (through portal) for verified and authorized users – validated data mappings
- Data license (credit, peer review,…)
- Improved Provenance capture
- Data standard (IMAS)
- Embargo periods

Implementation – EUROFUSION,

Several technology options - limit impact/cost on existing activities (experiments):

- Metadata (pushed from experiments to metadata server (as IMAS IDS summary)
- Remote data access (UDA client) serving subset of available data as IMAS based datasets
- Portal + federated AAI, central development





Generic Tools (IMAS based)

(leveraging e-TasC activities)

Metadata resides on Portal.

- Remote access to Expt data in IMAS
- Simulation DB?

Device Specific Tools