



Experimental data from EUROfusion machines

Pär Strand with input from David Coster, Joan Decker, Stephen Dixon, Jonathan Hollocombe, Frederic Imbeaux together with (the old) WPCD team and Fair 4 Fusion project

CHALMERS



This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training programme 2014-2018 under grant agreement No 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission.



- Brief overview of the status of IMAS on different experiments
- Tools and methods for accessing data
- Moving towards FAIR (and why!)
- Future requirements and implementation - Fair4 fusion blueprint
- Summary



Status of IMAS on experiments

- WEST stores all processed data in IMAS and most raw data is available through IMAS interfaces as well. Reference METIS simulations for each pulse are also available in IMAS format.
- AUG has the intention to introduce IMAS and FAIR principles in their data management
 - Currently
 - Trview combines various profile data and produces core_profile, equilibrium etc. for core modelling
 - Have tools that can populate the 1D part of edge_profiles
 - Old mappings exist for some of the diagnostics (see EWE tables)
 - Have a proof-of-principle code for the generation of SUMMARY IDS for metadata
- JET
 - IMAS Strategy is under development
 - Currently
 - Have a proof-of-principle code for the generation of SUMMARY IDS for metadata
 - Tools from WPCD available for data transfers



Status of IMAS on experiments

- TCV
 - Has built an IMAS server that has started storing data relevant for physics studies. The objective is to extend the scope and availability of this data as an effort towards open science.
- MAST/MAST-U
 - MAST data is already "open" and available through UDA services
 - Strategy for MAST-U is under development (following the same path as MAST?).
- Wendelstein
 - IMAS strategy depends on suitability of IMAS for stellarator specific data

Resources and responsibilities for full implementation in the EUROfusion environment needs to be clarified.



Snapshot from F4F catalog Qt demonstrator showing data for AUG shots with $1 < H_{98} < 2$

FAIR4Fusion

Demonstrator Dashboard

Jump to shot: Go

Date:

Machine:

Plasma Current:

Magnetic Field:

q95:

Power Ohm:

global_quantities/h_98/value:

Variable

Open search plots

Shot	Run	Machine	Date	Plot
30373	0	AUG-F4F	2022-04-09	
30380	0	AUG-F4F	2022-04-09	
30265	0	AUG-F4F	2022-04-09	
30291	0	AUG-F4F	2022-04-09	



Snapshot from F4F catalog_qt demonstrator showing data for JET shots with $1 < H_{98} < 2$

FAIR4Fusion

Demonstrator Dashboard

Jump to shot: Go

Date:

Machine:

Plasma Current:

Magnetic Field:

q95:

Power Ohm:

global_quantities/h_98/value:

Variable

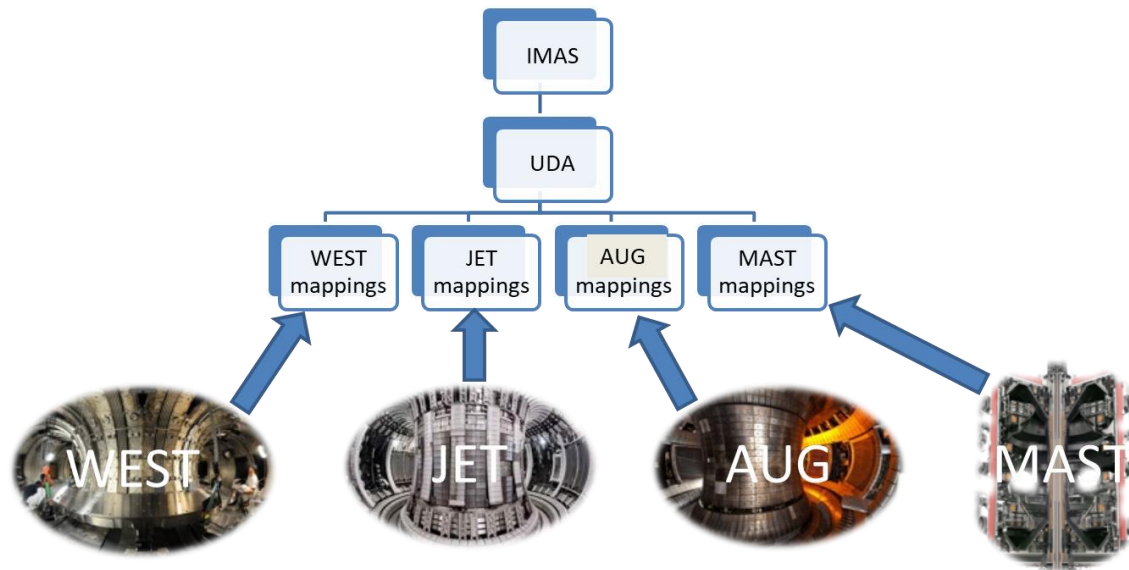
Open search plots

Shot	Run	Machine	Date	Plot
98695	0	JET-F4F	2022-04-09	
98701	0	JET-F4F	2022-04-09	
98652	0	JET-F4F	2022-04-09	
98677	0	JET-F4F	2022-04-09	

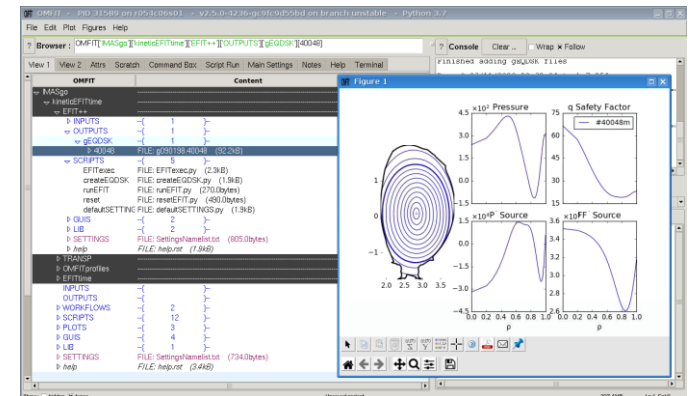
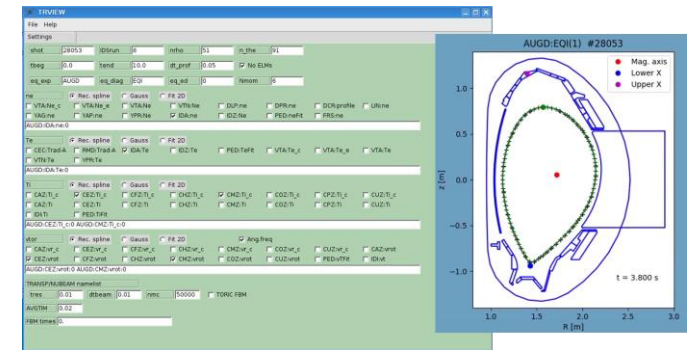
EUROfusion tools to access experimental data in IMAS format



- **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! (Omfite plugin), TCV2IDS,...



With the IMAS/UDA paradigm we gain the additional benefit of ITER being automatically technically integrated in the EUROfusion device ecology.



Data available in IMAS form*



IDS Name	JET	TCV	AUG	MAST	WEST
iron_core	Green	Grey	Grey	Grey	Red
magnetics	Green	Green	Cyan	Green	Cyan
mse	Cyan	Grey	Red	Cyan	Grey
pf_active	Green	Green	Cyan	Green	Green
tf	Green	Green	Cyan	Green	Green
thomson_scattering	Green	Cyan	Red	Green	Grey
wall	Green	Green	Green	Green	Green
core_profiles	Green	Green	Green	Green	Green
equilibrium	Green	Green	Green	Green	Green
nbi	Green	Green	Green	Green	Grey
ic_antennas	Green	Grey	Yellow	Grey	Green
ec_antennas	Grey	Green	Green	Grey	Grey
core_sources	Yellow	Yellow	Yellow	Yellow	Yellow

Colour	Meaning
Grey	Data not relevant for this machine
Red	Data is missing
Yellow	Data mapping in development
Cyan	Data available
Green	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 (AUG)
 - IMASgo (MAST, JET, K-Star, DIII-D)
 - TCV2IDS (TCV)
 - ReadAUG (AUG)
 - exp2ITM (CPOs + converter)
 - UDA (MAST, JET, magnetics)

Tools developed/adapted under WPCD, partially taking over by TSVVs as per individual needs.
No formal continuation in FP9!!!! Internal charge on e-TASC to provide that role (by 2023).

* Current as of a year ago...



- Data access for an end user is currently use case dependent:
- Different access tools depending on
 - What analysis is planned
 - Which device(s) are to be used
 - And what internal processing has been done/needs to be done
- Significant issues in
 - Managing and recording provenance
 - Sharing data back to community/experiments
 - ..
- A lot to be gained if we can provide/shift to a single access method with validated data mappings
 - UDA can provide this but some performance issues need to be solved
 - IMAS based “profile maker” would allow for more streamlined data management
- Simulation data needs to be provided with a home (a data facility in its own right)
 - A proposal is to implement a simulation data base system (SimDB) with an associated Long term simulation storage facility (LTSSF), proposal of the gateway operation group.

IMAS data access through UDA



- UDA is a tool to retrieve data from heterogeneous sources through a common interface and provide a common data-object, no matter where it was read from or from what original format.
- In the context of IMAS mappings, a new UDA plugin is added to handle the data-access interface for each experimental device.
- The data can be processed by these plugins to implement mappings between the native data formats and the IDS data structure. The mappings are generally controlled by configuration files in human-readable and easy to edit formats such as XML.
- Data access has previously been established in this way for WEST, MAST, JET, AUG, and TCV. What data has been mapped to IDS from each machine is then principally managed by teams at each site.
- The current status on the EUROfusion Gateway is that data is available from WEST and MAST; recent breaking changes to the JET MDS+ server are being investigated to return access to JET data. TCV are looking at setting up an UDA interface to their IMAS data.
- Access to data from locations outside of the Gateway is complicated by firewalls and access controls implemented by each experimental site. Often sites whitelist the individual IP addresses they will allow to connect. Any new users would then have to register with each site individually.
 - The planned EUROfusion Authentication and Authorization Infrastructure (AAI) between sites could fix this.

UDA plans as part of IMAS



- Some performance bottlenecks have been identified in retrieving whole IDS's from remote locations as the IMAS/UDA interface sends out a very large number of small requests for individual IDS fields.
- The third deliverable of the ITER simulation databases management project involves the improvement of access to remote IMAS data – both mapped experimental data and native IDS files.
- Improving the data transfer rate over UDA is still in progress but will likely involve the transfer of richer data objects containing whole or partial IDS structures rather than individual data nodes – which is what is transferred currently.
- As part of this work various other improvements to UDA will be made including cleaning up the code, improving documentation and adding training material – documentation/training material for creating UDA plugins is currently being worked on and should be available soon.
- Existing UDA plugins will continue to work with any changes to UDA made, though some changes to the plugins may be required to make use of the improved data transfer functionality.
- This work will incorporate the IMAS URI syntax currently being finalised, and should allow for seamless access to data whether the data is available locally (i.e. through the existing MDS+ backend) or the data is on a remote server (i.e. through the UDA backend).
- This should then allow for data located via SimDB to be accessed using the Access Layer from any (authorised) location.

Background - open data as a scientific driver



Opening up EUROfusion data so that it is more accessible to fusion researchers should (if done through a FAIR based approach)

- Enhance scientific quality and output and foster new collaborations
- Allow for novel AI based analysis approaches
- Enhance re-use and secondary use of existing data
- Significantly improve provenance capture for published data

EUROfusion formally need to address some EU driven requirements that are new for FP9.

- Data is open unless explicitly excluded (on commercial, security or similar grounds)
- Data should be made available under FAIR principles (limited access for the public embargos and metadata only, with full access for EUROfusion members)
- We have most of the individual building blocks to provide a full (two way) integration of EUROfusion experiments with data analysis and simulation modelling tools from the TSVV activities as well as the beneficiaries

Fair4Fusion received 2 MEuros from EURATOM to look at FAIR issues. The continuation of this work should be embedded in EUROfusion in FP9 with a seamless handover of responsibilities

Target Audience

The main users of the (open) data access and tools are the active fusion researchers within Europe (= EUROfusion which has [some] rules for joint collaboration and knowledge sharing already established):

- Promote broad internal and new/novel collaborations
 - Facilitate cross device research initiatives
 - Integration of the horizontal eTASC (theory, simulation, V&V) activities with experiments
- Build tools facilitating data sharing between experiments, modelling data and researchers on a higher level than what is now typically available.
- Reduce thresholds for large scale data mining and AI/ML approaches
 - Increase secondary use of data
 - Consolidate knowledge and tools towards a consistent “data and software” ecology for exploitation of ITER and DEMO.
- Explore/exploit a FAIR based approach to data management and access!

How do we fare with FAIR?

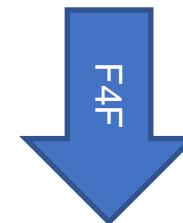


The Magnifying glass, Tap, Gears set, Recycle sign, Storage, Infinity, Discussion, Shield, and Man User icons made by [Freepik](https://www.flaticon.com) from www.flaticon.com are licensed by [CC 3.0 BY](https://creativecommons.org/licenses/by/3.0/). All other icons made by ARDC. Entire FAIR resources graphic is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

On top of this: open data access for the public (limit to metadata and select data sets for educational purposes)



Open data: how and for whom?



Target Audience

The main users of the (open) data access and tools are the active fusion researchers within Europe (= EUROfusion which has [some] rules for joint collaboration and knowledge sharing already established):

- Promote broad internal and new/novel collaborations
 - Facilitate cross device research initiatives
 - Integration of the horizontal eTASC (theory, simulation, V&V) activities with experiments
- Build tools facilitating data sharing between experiments, modelling data and researchers on a higher level than what is now typically available.
- Reduce thresholds for large scale data mining and AI/ML approaches
 - Increase secondary use of data
 - Consolidate knowledge and tools towards a consistent “data and software” ecology for exploitation of ITER and DEMO.
- Explore/exploit a FAIR based approach to data management and access!

How do we fare with FAIR?

Findable 	Persistent Identifiers (PIDs) 	Rich metadata 	Indexed data repositories 	PIDs in metadata
Accessible 	Standard communications protocol 	Open, free protocol 	Authentication, where necessary 	Metadata is always available
Interoperable 	Vocabularies 	Vocabularies are FAIR 	Linked metadata 	
Reusable 	Metadata have multiple attributes 	Usage license 	Provenance 	Community standards

On top of this: open data access for the public (limit to metadata and select data sets for educational purposes)



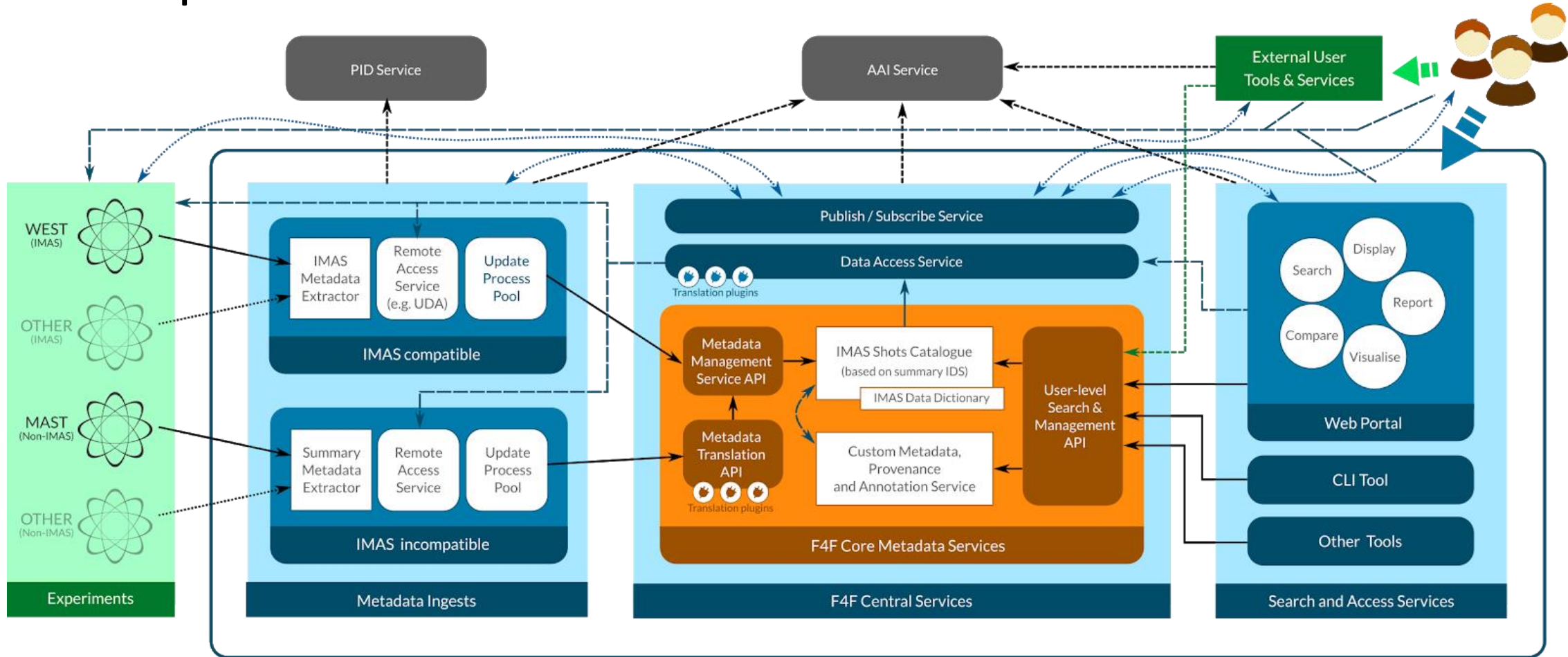
FAIR paradigm gives a structure for addressing this

What is shared, who has access and how as well as safeguarding →

- Findable: (F4F central services/portal and dashboard)
 - Published metadata (IDS: ids_summary, fair_data) [“waveforms”],
 - Link with persistent identifiers
 - Searchable and machine actionable
 - Points to data resources or landing pages
- Accessible: (UDA, AAI)
 - Data (and metadata) are available through standardized protocols
 - Allows for authorized access whenever needed
- Interoperable: (IMAS)
 - Shared and broadly applicable language for knowledge representation.
- Re-usable: (F4F extensions on provenance, CC-BY-NC-SA license)
 - Clear and available licenses for data usage
 - Provenance capture for all stages

For more details see e.g., <https://www.force11.org/group/fairgroup/fairprinciples>

Blueprint Architecture



Fair4Fusion System

Current public version available for download on project web site <https://fair4fusion.eu/>

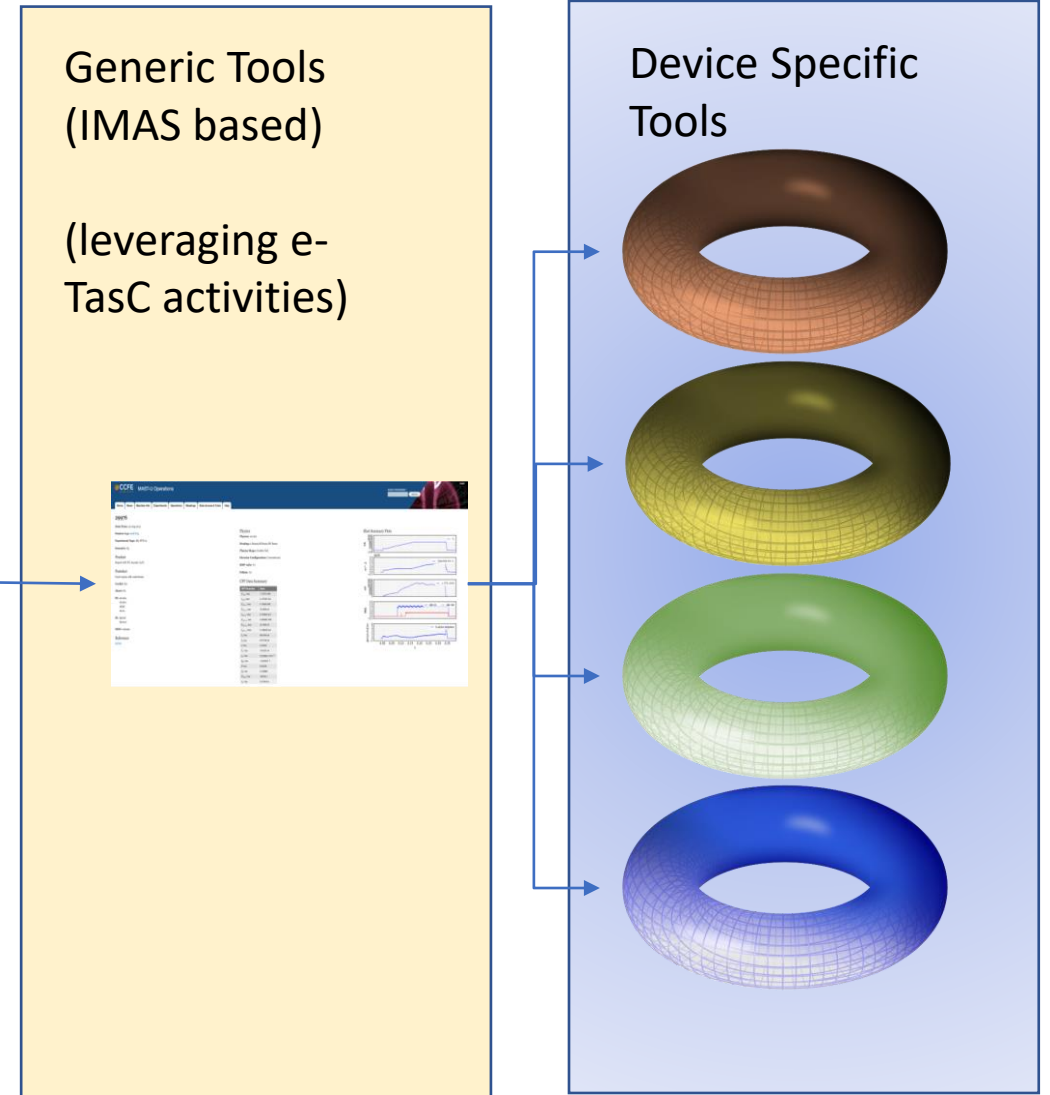
Access paradigm

- Central metadata "portal" for searches, summaries
- Remote data access (through portal) for verified and authorized users
- Data license (credit, peer review,...)
- Improved Provenance capture
- Data standard (IMAS)
- Embargo periods
- Quality assurance through extension to pinboard

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





Native access to experiments is still somewhat limited: Facilities are planning and/or exploring to move towards IMAS and FAIR based data sharing.

- Need to harmonize on tools and methodologies for bringing data to end users (EUROfusion and others)

IMAS provides the means of storing experimental and simulation/modelling data on equal footing

- A great opportunity to create a joint data ecology between experiments and simulations (TSVVs)
 - Build on successful model from CERN, Astrophysics, Climate,...

We have all the ingredients to build and benefit from a FAIR data environment: making data findable, accessible, interoperable and re-usable and we propose to be aligning with the Fair4Fusion (F4F) Blueprint

- F4F Dashboard for metadata and search facilities, pid (Findable)
- EUROfusion AAI and UDA installations with data mappings (Accessible)
- IMAS as the common data dictionary and format (Interoperable)
- Adding provenance capture and persistent identifiers to data workflow (IMAS/IDS already extended) (Reusable)

A brave new world is possible!

- but challenges a little bit the current work practices and
- will require some resources to implement.

EUROfusion will present a resource loaded Data Management Plan where at least part of this is clarified by mid July.

