

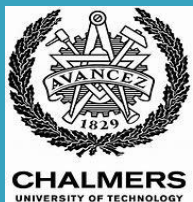


Update of Architecture for the implementation of the Fusion FAIR Data Framework

Marcin Plociennik(PSNC IBCh PAS), Bartosz Bosak (PSNC IBCh PAS), Raul Palma (PSNC IBCh PAS), Michal Owsiak (PSNC IBCh PAS), Shaun de Witt (UKAEA), George Gibbons (UKAEA), Nathan Cummings (UKAEA), Iraklis Klampanos (NCSRD), Andreas Ikononopoulos (NCSRD), Stasinios Konstantopoulos (NCSRD), Pär Strand (Chalmers), David Coster (MPIPP), Frederic Imbeaux (CEA), Joan Decker (EPFL), Yves Martin (EPFL), Olivier Sauter (EPFL)

29.11.2021

4th TM on Fusion Data Processing, Validation and Analysis



This slide set is released under [CC-BY-SA-4.0](https://creativecommons.org/licenses/by-sa/4.0/)



Fair4Fusion - open access for fusion data in Europe

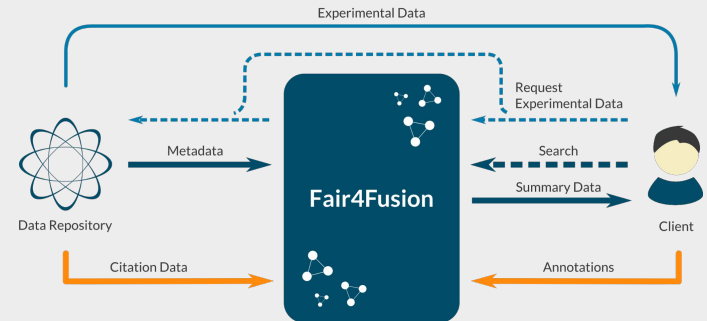


This project has received funding from the Euratom research and training programme 2014-2018 under grant agreement No 847612

Architectural Blueprint for Fusion Open Data Framework

Fair4Fusion project is to demonstrate the impact of making experimental data from fusion devices more easily findable and accessible. The main focus towards achieving this goal is to improve FAIRness of the fusion data to make scientific analysis interoperable across multiple fusion experiments.

- One of the main outcome of the Fair4Fusion project is the architectural blueprint
- Long term Blueprint architecture for Fusion Open Data Framework
- Blueprint content (current status):
 - Current state of the art – policies and data access, **fairness status**
 - User stories , requirements, list of functionalities
 - Architecture
 - Baseline architecture, **Detailed diagrams**, Components,
 - Relationship, standards, protocols, Technological options
 - **Costs, Benefits**
 - **Licensing**
 - **Roadmap for implementation**
- Version 1.4 available for comments:
 - <https://shorturl.at/mAO58>



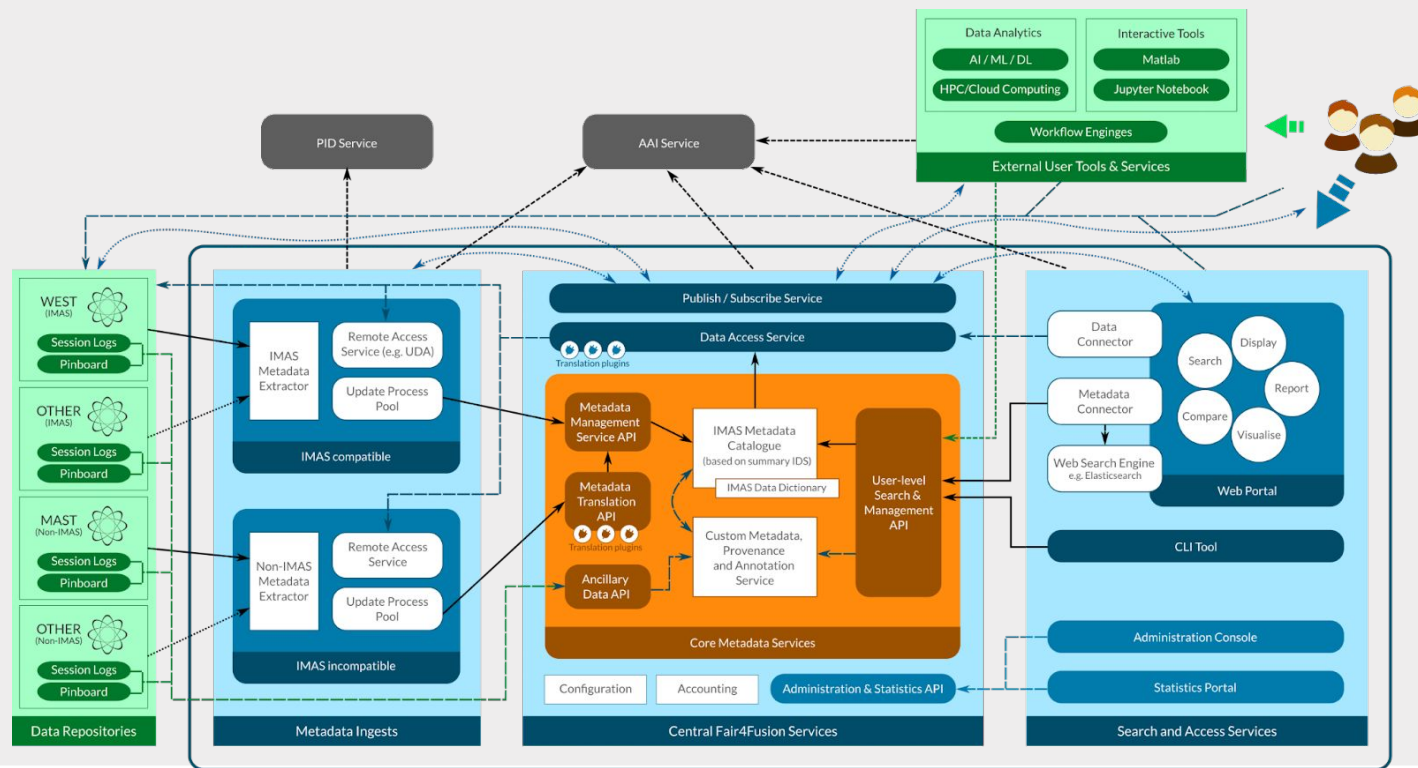
Data Identification and Policy Landscape

- Existing experiments have some internal degree of FAIRness
 - But not across the EU fusion community
- Need a way to make scientific analysis and discoverability interoperable across multiple fusion experiments.
- Benefits not only for manual database queries
 - but would also enable the use of new methods of research with Data Mining and Machine Learning techniques at an unprecedented scale
- The IMAS Data Dictionary is recommended for achieving interoperability within the community.



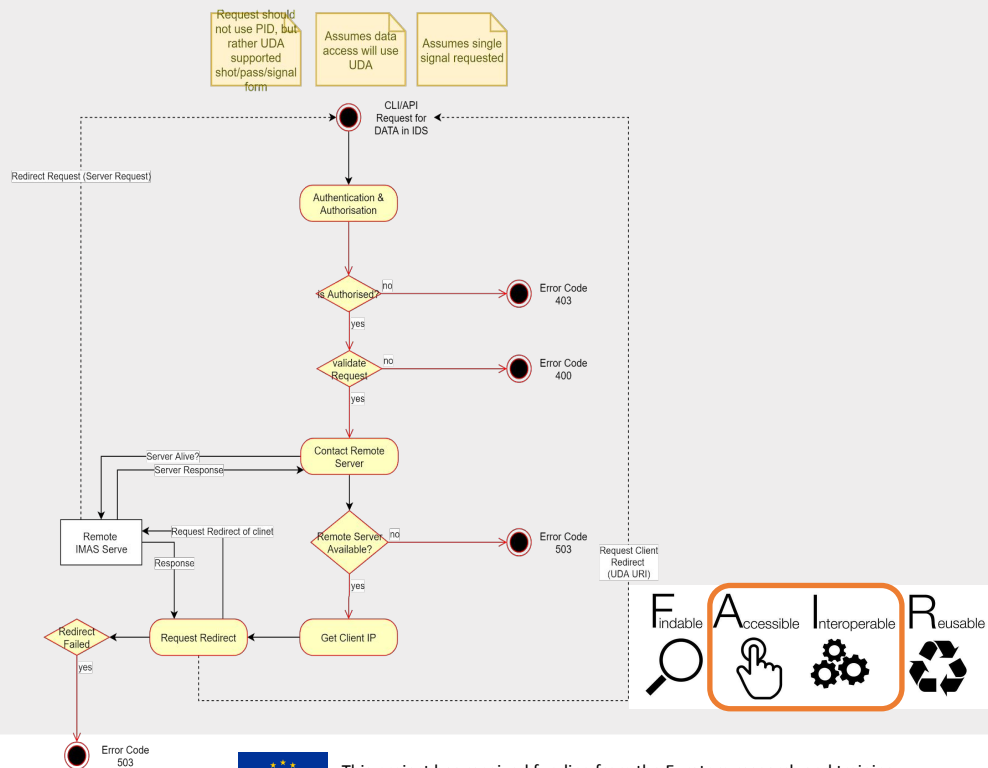
The Magnifying glass, Tap, Gears set, Recycle sign, Storage, Infinity, Discussion, Shield, and Man User icons made by [Freepik](https://www.flaticon.com) from www.flaticon.com are licensed by [CC 3.0 BY](https://creativecommons.org/licenses/by/3.0/). All other icons made by ARDC. Entire FAIR resources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

Proposed architecture overview



Detailed diagrams: search and access services

- Example diagram
- Web/CLI/REST API access
- Authorised users can gain access to full resolution experimental data
- Experiment themselves and access is provided by the existing UDA or MDSplus protocol
- This requires HTTP redirection based on the existing standards



Licensing (1)

- We recommend to follow use Creative Common licenses for opening validated fusion data and metadata to the public and wider research community.
- An embargo period of a few years for data to give sufficient time for the Institutes running the experiment or collaborating with it to exploit the data first
- Data used for publication should be released co-incident
- The precise combination of CC flavour to be chosen is left to each data owner (experiments, modellers,...), but we recommend using CC-BY-NC-SA (e.g. MAST has already moved to this license):
 - “this license allows users to distribute, remix, adapt, and build upon the material in any medium or format for noncommercial purposes only, and only so long as attribution is given to the creator. If you remix, adapt, or build upon the material, you must license the modified material under identical terms”

Licensing (2)

- A good trade-off in terms of openness and fair usage of the data produced by a huge effort based on public funds
 - BY (attribution): requires citing the data creator.
 - SA (share-alike): requires that, if derived data is produced and exploited, it has to be published with the same license as the original data. This would also allow the experiments to benefit from any adaptation of their own data under the same license.
 - NC (non-commercial): prevents commercial usage of the data and of its adaptations. This leaves the possibility for public Institutes to benefit from a potential commercial exploitation of their data with e.g. a private partner, if there is such an opportunity.

Licensing (3)

- It should also be noted that if different data producers use different licenses it will create difficulty when trying to combine data for the benefit of the community.
- For strategic, commercial aspects, this license does not preclude any site from partnering with industry and sharing data with them under a bespoke license or even charging for data access requested by commercial entities

Benefits

• Financial Benefits

- Reduced training costs for new generation of scientists
- Common software infrastructure means support and development can be shared across sites
- A common security infrastructure means practices can be shared across organizational boundaries

• Non-Financial Benefits

- Better ability to validate simulation by allowing simpler comparison with a wider range of devices and configurations
- Closer collaborations between experiments may be needed to more rapidly make progress towards realisation for commercial fusion
- Possible improved collaboration with industry leading to faster technology breakthrough, for example in materials science,
- Collaboration on algorithms and infrastructure with other communities

Roadmap

- Get blueprint/implementation of proposed architecture embedded into EUROfusion system
- Acceptance from data providers that the proposed system will be adopted

For EUROfusion:

- Decide on policy for data publication, licenses (and make its DMP consistent with the policy)
- Define an implementation project : scope, organisation, resources, timeline
- Organise dissemination of the services and user feedback, evolutions of the project

Central Services:

- Host the Services - (the recommendation is to start with the prototypes developed by F4F)
- Provide AAI, PIDs
- Set up a helpdesk system with first line support
- Set up a monitoring service
- Guarantee data preservation and archiving

Experimental sites:

- Provide Summary physics metadata
- Implement mechanisms to submit the data to Central Services (including remote data access)
- Progressively map more physics data to IMAS
- Progressively add provenance information in the output of experiments-related workflows
- Maintain FAIR information related to the submitted data (dataset validity, publications, ...)
- Extend the data submission to simulations related to experiments



Summary

- Big step forward for the community towards FAIR and Open Data
- Looking forward to feedback from Fusion community on the Blueprint Architecture to improve the final recommendations
- Still to do...
 - Complete blueprint architecture
 - To detail the data access
 - Gaps analysis,
 - Recommendations
 - Include non-experimental data



Questions?

For more please follow us on social media and check the web site for project updates

<https://www.fair4fusion.eu>



[@fair4fusion](https://twitter.com/fair4fusion)



<https://www.facebook.com/Fair4fusion>



<https://www.linkedin.com/company/fair4fusion>



Fair4Fusion - open access for fusion data in Europe



This project has received funding from the Euratom research and training programme 2014-2018 under grant agreement No 847612