Architecture for the implementation of the Fusion FAIR Data Framework

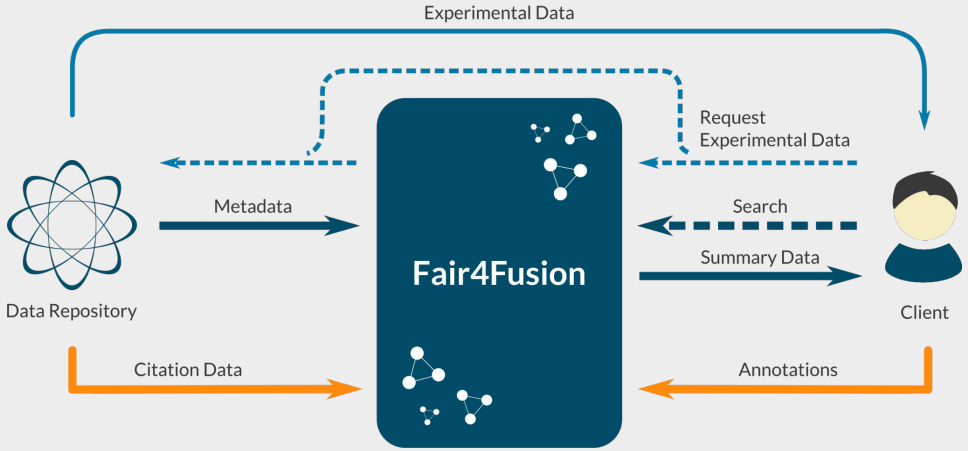
Marcin Plociennik(PSNC IBCh PAS), Bartosz Bosak (PSNC IBCh PAS), Raul Palma (PSNC IBCh PAS), Michal Owsiak (PSNC IBCh PAS), Shaun de Witt (UKAEA), George Gibbons (UKAEA), Nathan Cummings (UKAEA), Iraklis Klampanos (NCSRD), Andreas Ikonomopoulos (NCSRD), Stasinos Konstantopoulos (NCSRD), Pär Strand (Chalmers),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

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

- One of the main outcome of the Fair4Fusion project
- 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
    - Components
    - Relationship, standards, protocols
    - Initial technological options
  - **Costs, roadmap for implementation**
- Available for comments:
  - https://cutt.ly/2mcDHSX



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

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.

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

Use case definition/requirements specification

- Developed with fusion researchers and members of FAIR groups
- 76 user stories identified so far
- Covers
  - Security
  - Data Access
  - GDPR
  - Searching & Querying
  - Experimental Impact
  - Visualisation
  - Who
    - Fusion scientists / data scientists / members of the public
  - What
    - scalar data / 1-d data / 2-d data / structured data
  - Limitations
    - usually from the data providers (GDPR, no impact on operations, cost)
- Requirements have come from the community and external researchers
-

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

Investigation of similar platforms

An investigation of other FAIR platforms and their architectures to look for reuse opportunities.

- CLARIN
- EUROPEANA
- ICOS
- FOODIE
- ROHUB
- EVER-EST

Investigation of Existing Tools in Alternative Domains

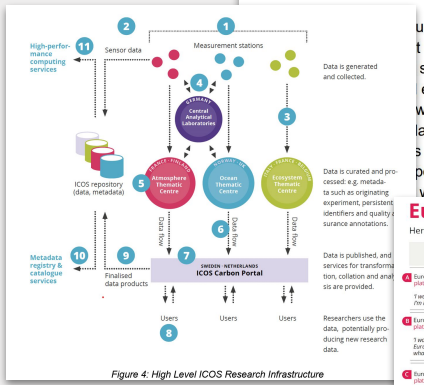


Figure 4: High Level ICOS Research Infrastructure

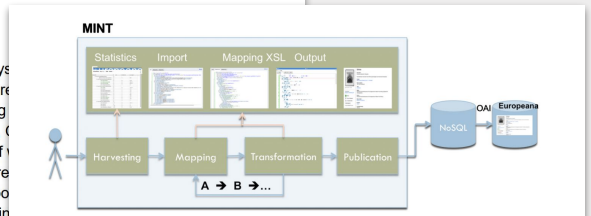
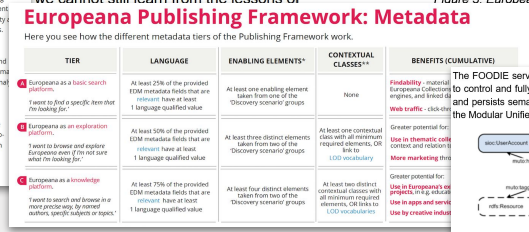


Figure 3: Europeana Ingestion Mechanism (MINT)[4]



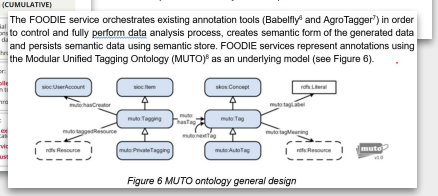
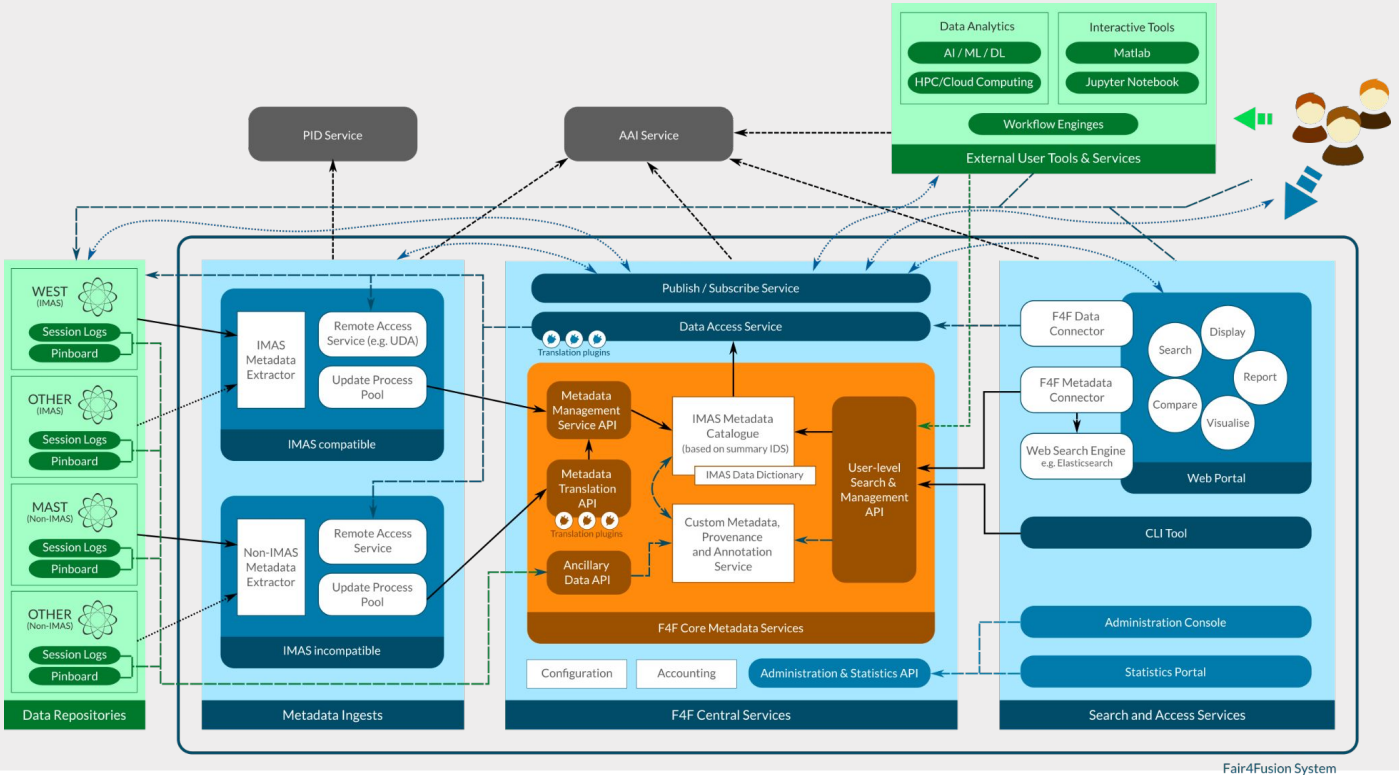


Figure 6 MUTO ontology general design

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

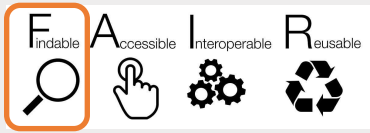
Proposed architecture overview



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

Metadata Standards and Ontology Mapping

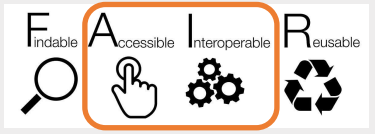
- Agreed to use Summary IDS as basis for physics metadata schema
  - IDS structures already based on extensive discussions with the community
  - Augmented with elements from Dublin Core to support FAIR
- Gathered list of existing fusion glossaries
  - Which sort of highlights the issue with developing a Ontology
- Investigating use of OAI-PMH for harvesting
  - …but looks like a ‘push’ model from experimental sites is more suitable



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

Data Interoperability 4 FAIR and Open Data

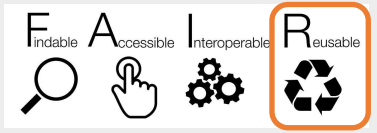
- Investigated different PID solutions and obtained relevant costings
  - Demonstrated interoperability of DataCite and Handle systems
- Proposed recommendations on granularity at which they could and should be applied
  - Shot/Pass level strongly recommended for interoperability
  - Other levels at site discretion
  - Investigating long term funding sources



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

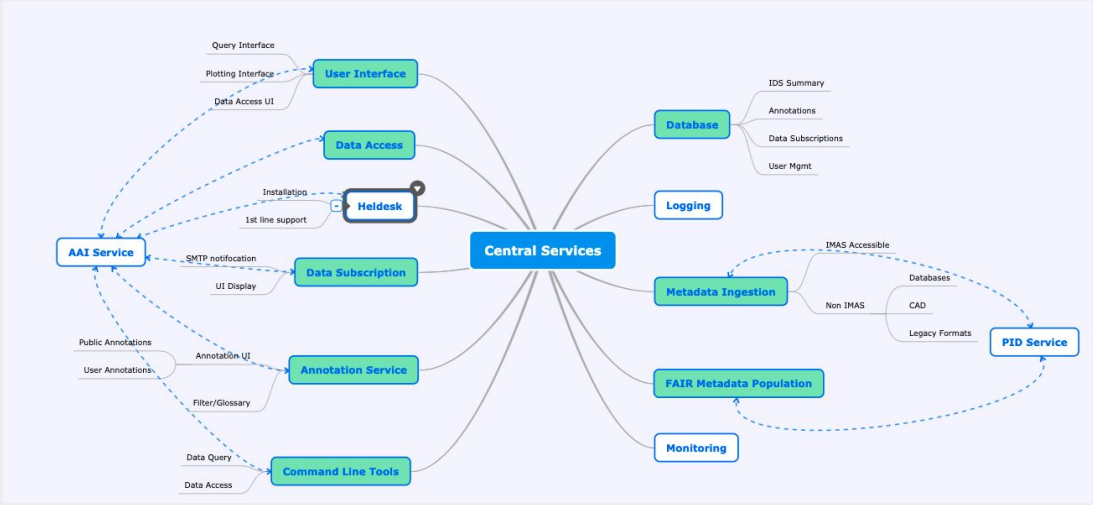
Standardised Models of Provenance

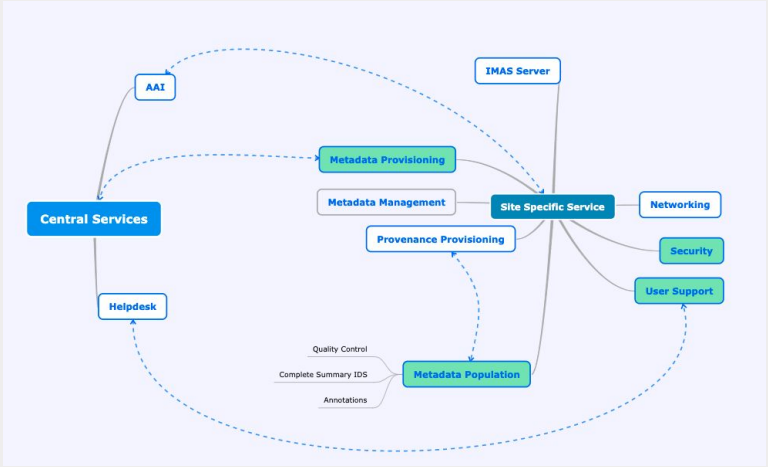
- Proposed provenance model based on W3C PROV
  - Prototype implementation in place
  - Tested on MAST and JET
  - Issues around ownership - recommendation to use roles rather than personal identification
- Proposal to extend IDS structures to incorporate minimal provenance on data access
  - e.g. if, for instance, bolometric data is accessed - include the shot/pass (and/or PID) in the bolometer IDS



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

Initial evaluation of implementation costs





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

Initial components of 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

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

Summary

- Big step forward for the community towards FAIR and Open Data
  - Detailed identification of current state of art
  - Recommendation on the architecture and roadmap implementation
  - Cost analysis
- Looking forward to feedback from Fusion community on the Blueprint Architecture to validate correctness of approaches
- Still to do…
  - Actually get a PID service up and running for the community
  - Propose data access methods
  - Complete blueprint architecture (gaps analysis, evaluation of demonstrators, roadmap, benefit vs costs analysis)

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

Questions?

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

https://www.fair4fusion.eu

@fair4fusion

https://www.facebook.com/Fair4fusion

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

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