



## Tokamak data mirror for JET and MAST data

Our data pilot will provide a mirror of experimental data from two magnetic confinement nuclear fusion devices (Tokamaks) at the Culham Centre for Fusion Energy (CCFE): the Joint European Torus (JET) and the Mega Amp Spherical Tokamak (MAST). The research community will be plasma physics and fusion researchers, engineers and technologists from the 29 members of the EUROfusion consortium and around 100 associated organisations, including those delivering the next generation nuclear fusion device (ITER) in southern France, namely ITER-IO (France) and Fusion 4 Energy (Spain).

## The Scientific Challenge

Data from the JET and MAST experiments has been collected over many years (JET has been operating since 1984). It is hosted at CCFE and made available via bespoke APIs and visualisation tools. We would like to make more use of standard data infrastructure including object-storage platforms and modern APIs.

The challenges in making use of a third party platform include:

- Maintaining the native data versioning and validation status information;
- Maintaining the link to local identifiers for data items;
- Not losing information from the native hierarchical structure of the data;
- Complying with UK government and EU policies on hosting and access restrictions;
- Keeping mirrored data in sync as new versions of individual data items supersede old ones.

There is scope for EUROfusion members to make more use of each other's data. We intend to make it simpler to access JET and MAST data remotely. Data volumes are ever increasing - both the total per experiment and the size of individual signals such as high-resolution camera data. It's necessary to plan ahead and evolve our data infrastructure to cope with this continued growth. We are also keen to develop and pilot data management approaches for the next generation nuclear fusion device, ITER, which is currently being constructed in southern France. ITER's individual experimental runs will have a much longer duration than the current generation of tokamaks and will generate up to 0.4PB of data per day. There is lots of potential for researchers to make more use of HPC facilities and we aim to provide more convenient ways to make data available for this purpose. We estimate that several hundred users might initially make use of the EUDAT data mirror once it's fully tested and publicised.

## Who benefits and how?

The research community for this EUDAT data pilot will be members of EUROfusion consortium and associated organisations. EUROfusion consists of 29 research organisations from 26 European countries plus Switzerland. In addition about 100 third parties contribute to research activities through consortium members. EUROfusion also collaborates with the organisations delivering the next generation nuclear fusion device (ITER) in southern France, namely ITER-IO (France) and Fusion 4 Energy (Spain).

Researchers within the EUROfusion community currently have controlled access to JET and MAST data and are discouraged from creating local copies of the data sets. The EUDAT data pilot will provide researchers with access to alternative sources of JET and MAST data which are trusted, carefully synchronised with the master data set and include better meta-data for finding data of interest. There will also be an opportunity to transfer and

EUDAT receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 654065.



stage sub-sets of data to other high performance computing (HPC) services than those provided at Culham (UK). This will expand the HPC capacity available to scientists and allow more direct use of new technologies, for example the new nuclear fusion HPC service being developed by CINECA (Italy).

The impact of these changes is expected to be twofold. Firstly, facilitation of data access will encourage more researchers to access and use JET and MAST data. This will be achieved with appropriate publicity and marketing and further encouraged by the wider variety of access tools provided by EUDAT. The second impact will be development of new ways to find and analyse the data, through improved meta-data and data discovery interfaces and server-side big data analytics

The organisation, data types and access patterns for JET and MAST data are well aligned with the concept of object rather than file based storage technologies. CCFE is working with STFC GRIDPP to develop CEPH and Swift OpenStack plug-ins for IDAM and also as part of the H2020 funded SAGE project (led by Seagate) to develop an Exascale, big-data centric deeply tiered percipient data storage platform. The EUDAT data-pilot will allow us to compare and contrast file-based EUDAT services with these object storage based infrastructures.

In summary, we will leverage related project work with STFC together with work in collaboration with European data centres (e.g. Julich), within Universities (York, ANU etc.), European fusion community collaborators (IPP-CAS, LECAD) and industrial partners (Seagate, Bull etc.) to help strengthen the project and ensure that the infrastructure delivers a step change in capability.

## Technical Implementation

The first phase of our project ran from April to October and was internal work (described below). Phase 2 started at the beginning of October and full collaboration with our EUDAT partners was established from this point.

### **Phase 1 (complete): New fusion data interface**

The first part of our project was to design and implement a new generalised data interface for fusion data (Simple Access Layer) for use with both the internal data system and later with the EUDAT data mirror. It has been designed to allow reading and writing fusion data of various types from multiple sources but the initial implementation supports read-only access to JET processed time-series data stored in our existing data system.

The main components of the system that have been built are:

1. An abstraction layer or Virtual File System (VFS) which can support multiple storage options (persistence providers).
2. A persistence provider interfacing with the existing JET processed data system.
3. An HTTP REST API.
4. A new Python API for reading JET data implemented as a thin wrapper for the REST API.

### **Phase 2 (ongoing) - Answer key design and policy questions.**

Upload of data samples:

- Initial test upload of MAST-U open data B2SHARE training instance.
- Second upload with revised data model.

Exploratory work:

- Design and analysis to establish a common data and metadata model for JET and MAST-U.
- Mapping this fusion data model onto EUDAT services.

---

EUDAT receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 654065.



- 
- How our business logic layer could use B2FIND to allow access to data items by their metadata.

## Preliminary Results

We have agreed a common data model for JET and MAST-U data covering various classes of fusion data. Samples of data from each experiment have been produced in this form.

A small sample of MAST-U open-access data mapped to this common data model has been successfully uploaded to the B2SHARE training instance.

An internal beta version of the new data access interface (Simple Access Layer) and Python API has been released for user testing. This lays the groundwork for the planned user interface to the EUDAT data mirror via a second instance of the same system with an alternative persistence provider using EUDAT services.

## Contacts

- Alys Brett, CCFE, [alys.brett@ccfe.ac.uk](mailto:alys.brett@ccfe.ac.uk)

## Furhter Information

[Scientific Challenges behind the Pilot](#)

Learn more about the EUDAT/Tokamak pilot collaboration [here](#)

[Read more](#)