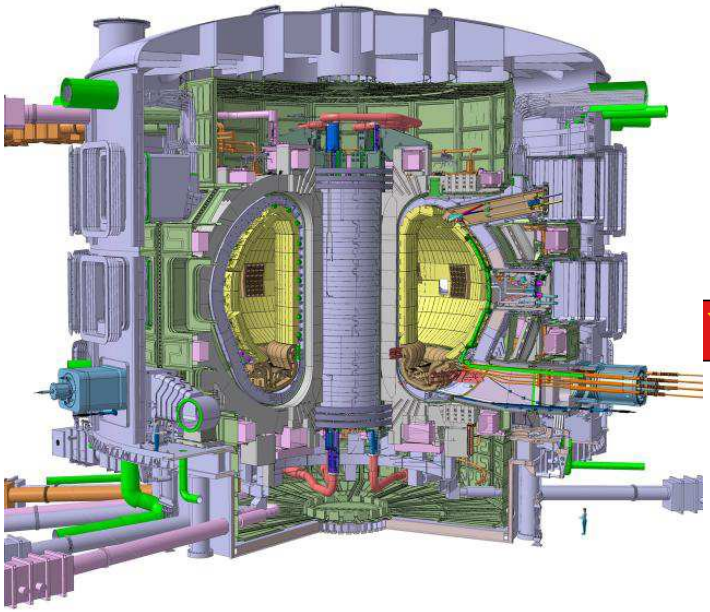
Simulation and data analysis: Data and data access requirements for ITER Analysis and Modelling Suite

P. Strand

ITERIS IM Design Team, 7-8 March 2012

1st EUDAT User Forum

"ITER aim is to demonstrate that it is possible to produce commercial energy from fusion."



First plasma 2019, full operation 2026 .
-Modelling framework being defined now.
-Simulation and modelling important high profile activity in the running up for operations.

Experimental facility(*):

- 10Gbit/s during discharges, 500-1000s
- 20-100PB/year

*lower bound estimates

International partners:

- Data replication several offsite repositories
- (Near) real time data streaming, inline
- modelling data to/from remote centers
- 5000+ worldwide users

IMAS Infrastructure, Framework

By ITER Modelling and Analysis Suite (IMAS) Infrastructure, or Framework, we mean a set of tools including:

- Data Model
- Component Model
- Workflow orchestration tool
- Data Bases, Catalogues
- Data Access (API), Querying tools
- Monitoring
- Web-based Portal to access various applications
- Lifecycle management of all the above

•The ITER users are global and we are projecting towards a global use and sharing of data

- Extending the above to a distributed global user community is a new challenge only partially tackled → EUDAT dialogue?!

IMAS Infrastructure, Framework

By ITER Modelling and Analysis Software (IMAS) Infrastructure, or Framework, we mean a set of tools including:

- Data Model
- Component Model
- Workflow orchestration tool
- Data Bases, Catalogues
- Data Access (API), Querying tools
- Monitoring
- Web-based Portal to access various applications
- Lifecycle management of all the above

Prototyped in EFDA ITM-TF and tested in the EUFORIA project. Strong fusion tech flavor!

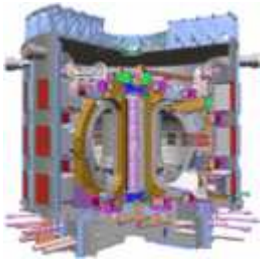
Being adapted and extended for ITER usage –"mature"

•The ITER users are global and we are projecting towards a global use and sharing of data

- Extending the above to a distributed global user community is a new challenge only partially tackled → EUDAT dialogue?!

Sketch of data flows and modelling activities

Plant Operation Zone



-"Raw" data
- Short term storage

-Control systems
-Safety systems

- Data Sanity checks

Nuclear Licensed environment

ITER IO and Control Room

-"Raw" data
- "processed" data
- Persistent storage

-Real time displays
-Real time modelling
- Physics support

- Modelling and analysis (Tier -2)
- Scenario development
- Expt Planning & support

Validated experiments
Detailed waveforms

ITER Partner facilities

- Replicated data
- New simulation data

-Real time displays
- Physics support

-Extended modelling and analysis (Tier-1, Tier-0)
-Scenario development
-Experimental design
-Physics exploitation

Data model (ontology) key concept

- Detailed data model for systems, diagnostics, experimental and simulation data
  - Abstracted in XSD schemas
  - Logically linked elements are structured in compound objects
  - Compund objects: standard structure for data access
  - Self-described with provenance data embedded

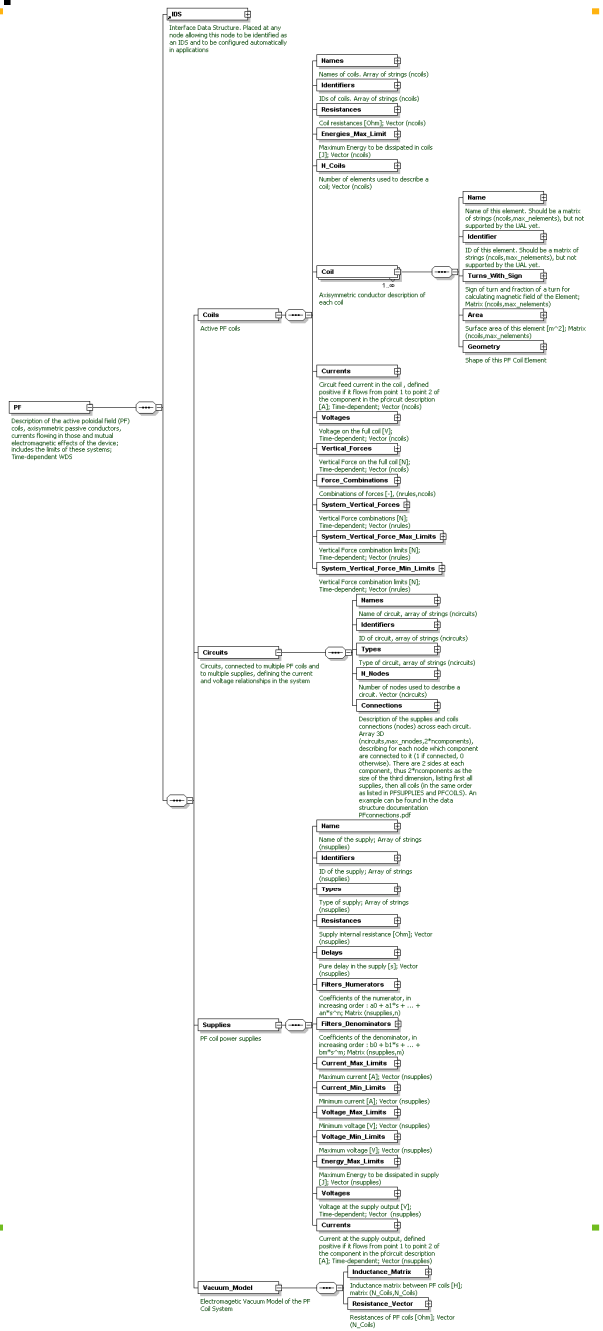
→Defines data storage trees

→Defines low level data access methodology

→Defines user APIs (multiple languages)

→Defines data transport layer for analysis and simulation workflows

→ challenges in distributed access to compute resources and scale of some data (studied in EU MAPPER project)



Data access

- ITER will be set up to run discharges (pulses) and most data will be identifiable through
  - Pulse number (discharge) + run number (version) [unique!]
    - In general each discharge will be more or less frequently reprocessed and updated – PID access for published subset of data important! No real master for some large scale data!
- Simulation data (joint available with experimental data) are largely
  - Simple types: Integers, Strings, Reals, Complex types and multi-dimensional arrays of these types
  - Complex types: structured objects composed of the above Simple Types, as well as one-dimensional arrays of such hierarchical objects – in specific magnetic geometry
  - Data reference
- Fusion community , most if not all current experiments, have adapted MDSplus (www.mdsplus.org)
  - A client-server storage and data access system with powerful server side scripting facilities
  - Defining features need to be maintained in ITER production env.

EUDAT dialogue

- Local Structure and tools (albeit fusion specific) reasonable well established and in principle fully defined for simulation and modelling activities
- Interface with EUDAT: Dialogues on the distributed components
  - PID and relevant infrastructure for persistent access for published data (not in general working level data)
  - Incorporation (federation/replication and collection) of remote facilities
    - May be huge drivers/producers of simulation data – data residing at remote facilities
  - Common services and exploitation:
    - Explore what the common subset /needs are –Need to align with ITER policies /infrastructure evolving through its own logic
- Framework should be defined and at least locally prototyped within two years. Provisions for distributed access and integration of partner facilities should be included. EUDAT as sounding board for technical solutions, tools and best practices. For sure room for future dialogue!