Contribution ID: **56**   Type: **Oral (Regular)**

Development of the IMAS Data Model Towards an Open Standard for Fusion Data

Friday 12 September 2025 09:50 (25 minutes)

The cornerstone of the Integrated Modelling & Analysis Suite (IMAS) developed by ITER is its machine-agnostic data model, called the Data Dictionary (DD). It was recently made open access [1] to permit a broader adoption outside of the ITER Members and by private fusion ventures. Discussions and contributions are welcome and held publicly in GitHub, either for new extensions, clarifications or changes to the Interface Data Structures (IDS) that compose the DD.

While it aggregates decades of effort to standardize the description of fusion data for both simulations and experiments, it has been found to introduce a maintenance burden on adapted codes through the frequent introduction of non-backward compatible changes introduced in IDS that are still in an alpha development stage. The major release of DD version 4 in 2024 improved the situation with a 50% increase of the number of stable IDSs, focusing on diagnostic and sub-system IDSs, while in version 3 the focus was on IDSs describing simulations and processed data. To go further, ongoing efforts are exploring more disruptive approaches based on lessons learned from the historical design choices.

A first effort focuses on a better separation of the Data Model and the IDS interfaces designed in support of integrated modelling. It relies on the definition of Standard Names that can be used as metadata to describe physical quantities and their units without restricting the dimensionality or coordinates associated with the data. The Standard Names approach is adapted from the CF-Conventions [2] which is successfully applied to climate and geoscience data over 20+ years. This approach aims at reducing the complexity of the Data Model and improving backward compatibility and is intended to be complementary to the IDS definitions.

A second effort focuses on improving the storage format for IMAS data with the intention to lower maintenance needs by simplifying the software stack and improve performance, particularly for remote data access. The main objective is to define a self-describing data format that keeps metadata together with data and to use open-source tools and libraries for access and manipulation. Early comparisons are made between netCDF, Zarr and the legacy solutions based on MDSplus, HDF5 and UDA for remote access.

[1] https://github.com/iterorganization/IMAS-Data-Dictionary
[2] https://cfconventions.org/

Speaker's email address

Olivier.Hoenen@iter.org

Speaker's Affiliation

ITER Organization

Member State or International Organizations

ITER Organization

Author: HOENEN, Olivier (ITER Organization)

Co-authors: Dr VAN VUGT, Daan (Ignition Computing); Mr SEBREGTS, Maarten (Ignition Computing); MCINTOSH, Simon (ITER Organization); PINCHES, Simon (ITER Organization)

Presenter: HOENEN, Olivier (ITER Organization)

Session Classification: Information Retrieval and Visualisation

Track Classification: Information Retrieval and Visualisation