

ITERDB – The Data Archiving System for ITER

G. Abl¹, G. Heber², D.P. Schissel¹, L. Abadie³, A. Wallander³, D. Robinson²

¹General Atomics, PO Box 85608, San Diego, CA 92186-5608, USA

²The HDF Group, 1800 South Oak St., Suite 230, Champaign, IL 61820

³ITER Organization, Route de Vinon-sur-Verdon, 13115, St. Paul-lez-Durance
email: abla@fusion.gat.com, Phone: (858) 455-3103; Fax (858) 455-4156

For ITER, acquiring, managing and archiving its data is an essential task. ITER is foreseen to produce up to one terabyte of data per pulse and several petabytes of data per year. All the produced data needs to be stored and managed. The stored data is expected to serve the data access needs of ITER researchers located both on the ITER premises as well as worldwide during ITER's lifetime and beyond.

The ITER data archiving system has multiple requirements and design challenges. The aggregate rate of data to be recorded at ITER is large. This requires the system to balance its capacity to record the steady stream from many data sources. The recorded data must also be progressively replicated from short-term data storage to permanent data storage, while concurrently being served to users. This requires a robust and efficient data replication mechanism that can insure data integrity while conducting a fast data transfer. The permanent data holdings will be required to be stored in an efficient scientific file format, while the system is also responsible for managing and serving a detailed provenance and metadata catalog which is required for searching and discovering datasets. All collected data, along with the metadata catalog, needs to be delivered to end users via a unified data access interface. The number of data sources and the size of datasets are expected to grow over the years, thus the data system must be scalable to meet the growth demand. Finally, ITERDB needs to provide monitoring and management capabilities, and is expected to be integrated with other ITER CODAC tools and network infrastructure.

ITERDB is a work-in-progress being designed for centralized ITER data archival and data access. It will be designed to manage and serve both unprocessed and processed data from the ITER plant systems and data analysis workflows. Major features of the current ITERDB design are: 1) The continuous recording of high-volume and high-rate streams of scientific data. 2) The scalable infrastructure for both data recording and serving. 3) An integrated data access interface for scientific and engineering, and configuration data. 4) Collecting and managing the metadata, including provenance. 5) Continuous migration of ITER data into a long-term archive. 6) The capability for system administration, and health monitoring.

In this paper, we report on the ITER Data Archiving System software requirements and priorities that have been identified by working with ITER staff and a large number of stakeholders. We will describe the design challenges as well as our solutions. We will also present the complete software architecture of the ITERDB.

Keywords: ITER, CODAC, Data Archiving, Requirements, System Architecture

TOPICS: Database Techniques for Information Storage and Retrieval

Preference: Oral

Journal Publication: Yes

Internet Connection: Yes

*This work was supported by the US Department of Energy under ITER/CT/12/430000571.