

Requirements for tokamak remote operation: application to JT-60SA

Paolo Innocente^a, Jonathan Farthing^b, Gerardo Giruzzi^c, Shunsuke Ide^d, Emmanuel Joffrin^c, Yutaka Kamada^d, Georg Kühner^e, Osamu Naito^d, Hajime Urano^d, Maiko Yoshida^d

^a*Consorzio RFX, Corso Stati Uniti 4, 35127 Padova, Italy*

^b*CCFE, Culham Science Centre, Abingdon OX14 3DB, United Kingdom*

^c*CEA, IRFM, F-13108 Saint-Paul-lez-Durance, France*

^d*Japan Atomic Energy Agency, Naka, Ibaraki-ken 311-0193, Japan*

^e*Max-Planck-Institute for Plasma Physics, EURATOM Association, Wendelsteinstr. 1, 17491 Greifswald, Germany*

Remote operation and data analysis are becoming key requirements of any fusion devices. In this framework a well-conceived data management system integrated with a suite of analysis and support tools are essential components for an efficient remote exploitation of any fusion device. The following components must be considered: data archiving data model architecture; remote data and computers access; pulse schedule, data analysis software and support tools; remote control room specifications and security issues. The definition of a device-generic Data Model plays also important role in improving the ability to share solution and reducing learning time. As for the remote control room, the implementation of an Operation Request Gateway has been identified as an answer to security issues meanwhile remotely proving all the required features to effectively operate a device.

Previous requirements have been analyzed for the new JT-60SA tokamak device. Remote exploitation is paramount in the JT-60SA case which is expected to be jointly operated between Japan and Europe. Due to the geographical distance of the two parties an optimal remote operation and remote data-analysis is considered as a key requirement in the development of JT-60SA. In this case the effects of network speed and delay have been also evaluated and tests have confirmed that the performance of various connection and data access methods can vary significantly depending on the technology used.

Keywords: JT-60SA, Remote participation, Data acquisition, Control.

1. Introduction

The data management system, with its validation, analysis tools and related documentation, plays a central role in the exploitation of the experimental program of a fusion machine. Modern data analysis makes use of more and more sophisticated computational and visualization software. Simulation and analysis codes produce derived data from the experimental measurements, which then become an essential part of the experimental data base (e.g.: reconstructed equilibrium, heating and current drive source profiles, transport coefficients, bootstrap current profiles, etc.). Therefore, a well-conceived and organized data system is now considered as an essential element for the efficient exploitation of fusion machines, and even for the success of their experimental programs.

Compared to a purely domestic experiment, device utilization is even more challenging for a machine exploited by a large international collaboration. In this case strong effort must be devoted to optimize all components for researcher spending a few times at device site or operating remotely also. In the following sections the main components of a data management system appropriate for international collaboration are analyzed defining the principal requirements for each of them. In the last section their application to the JT-60SA[1] experiment is evaluated, considering in particular the role of network bandwidth and time delay in constraining the choice of remote connection technologies.

2. Data model, data archiving and architecture

Data are the key element of any experiment; there are many different kinds of data depending on source, temporal nature and data type. Examples of different data sources are: pulse configurations, setting-up values, systems (i.e. magnetic system, NBI systems, RF systems, Vacuum and wall conditioning systems), feedback, diagnostics, post-pulse processing codes, simulation codes and human comments. Temporal nature: transient pulse-related, transient non-pulse related (like wall conditioning data), continuously sampled (like data from vacuum system). Data types: structured and unstructured data, integer, real, text data and images.

Due to the complexity and huge size of data expected in future experiments, a careful management of all foreseen data is required; the database management system (DBMS) must be able to answer these basic users' questions: what data are available, how data can be retrieved (or inserted). Furthermore, hidden to the end user, the DBMS must provide efficient and secure storage methods to allow fast data store/retrieve and to guaranty data integrity.

The definition of a "data model" has been considered important to answer the first of the two users' questions. The merit of having a data model is that the API (Application Programming Interface) is standardized independently of used technologies and target data, and the whole system can be made highly

modularized. This also extends the longevity of the system as well as facilitates the integration of large-scale system.

While “data model” can have a much broader meaning, in the present context, it mainly refers to a conceptual view of the data seen by the users that provides them with information on what kinds of data exist and what are their properties. It is similar to the Physics Data Model (PDM) that ITER Organization is presently designing to provide great flexibility and a high degree of self-documentation [2].

They comply with a well-defined set of Design Rules and Guidelines. Examples of the rules and guidelines are:

- Naming conventions;
- Self-documentation conventions;
- Data must contain units, dimension and axes;
- Data must contain creation time and related information;
- Data history must be recorded and readable
- Low-level data shall be accessed by referencing high-level data.

No specific requirements have been set for archiving methods and system architecture as long as the functional requirements are satisfied. But each system should satisfy the following principal requirements:

1. The system should allow the data listed below. The data should be accompanied with the descriptions of the data (i.e. metadata) that provide users with the information necessary to use them, their provenance and their quality;
2. The system should allow 0D parameter, 1D/2D signals (data array) and images/movies in RAW or standard formats. It should be flexible enough to allow custom format data also;
3. The system should manage continuous data recording by the ability to append data and retrieve subsampled data without retrieving the entire signal. APIs to manage continuous/long recording should be the same (may be with additional parameters) used for standard data;
4. The system must track the creation, insertion, modification times and the user name. It should also notify users when stored data were subject to modifications. External information such as in the form of HTML language, referring to the status of facilities and diagnostic systems at the time of pulse execution, can be linked;
5. The system should be equipped with search tool facilities.

3. Remote computers and data access

A data analysis environment is necessary to visualize and analyze the experimental data in JT-60SA. Three top level requirements have been identified for remote data access and remote computer access:

1. Participant teams should have identical access regardless of their geographic proximity to the device physical location;
2. Equivalent functionality with equivalent efficiency should be available to all participating remote researchers;

3. Data retrieval should be supported in both local and remote computers with the same API.

There are two complementary architectures allowing users to have remote access to data known as Remote Computer Access (RCA) and Remote Data Access (RDA). Both RCA and RDA depend on high speed and high availability network connections.

- RCA involves the secure remote login to computers located at the facility (local) and allows running of local applications. RCA also provides access to local clusters to allow code development and code execution;
- RDA allows applications and tools to run on remote computer clusters and access the data from the facility. It thus allows users to work in the environment they are familiar with, without having to port their codes to the local cluster environment. Transfer of the data would be on a signal-by-signal basis using client-server technology. When data analysis is performed remotely the processed data should be written back to the master data archive. This implies that the client-server RDA interface must provide the ability to write as well as to read data.

The choice of RCA or RDA depends on the application and on the amount of data to be transferred, for this reason both technologies should be provided. In this context in the last chapter network implication on the two methods will be analyzed.

4. Support tools

Support Tools are all the facilities that can provide an easier and more effective exploitation of the experiment by both off-site and on-site researchers. For this reason Support Tools are not limited to those available from remote (relatively to device location) or used from researchers located in geographically separated laboratory.

4.1 Requirements

Support Tools are required for the whole experimental cycle, which can be split in phases each one requiring its tools:

1. Pulse preparation;
2. Pulse execution;
3. Data analysis;
4. Results discussion;
5. Publication preparation and internal presentation.

Pulse preparation requires a high number of discussions between all potentially involved researchers. Summarizing in this phase support tools should provide:

- Interactive many-to-many and one-to-one (or few-to-few) discussion;
- Written/Recorded many-to-many discussion and planning functionality;
- Complete device/diagnostic information;
- Discharge parameter editor.

Pulse execution can be performed with some or many researchers remotely located. The quality of remote participation can be enhanced by providing remotely most of the information and the feeling available in the

local control room. In this phase support tools should provide:

- Remote participation;
- Textual comments (with/without keywords);
- Shared control room terminal display;
- Remote control room.

Data Analysis usually requires a preliminary analysis performed without any data processing; in this case a powerful and flexible data visualization tool is important. Since a big part of data analysis is performed on post-processed data which can be modified, it is useful to have an easy way to manage and know which, why and when data are modified. Summarizing in this phase support tools should provide:

- Data visualization;
- Data version information;
- Code developing, code managing tools.

Results discussion requires some of the tools already presented for the preparation phase:

- Interactive many-to-many and one-to-one discussion;
- Written/Recorded many-to-many discussion (BLOG).

Publication preparation and internal presentation are important elements to guarantee the quality of the scientific production, papers before submission to journals or conferences should be available for review by designated reviewers and for comment by all researchers. All published papers (or posters) should be available for reference. For this reason support tools should provide:

- Abstract/papers/poster presentation/comment;
- Abstract/papers/poster archiving.

4.2 Implementation

Previous requirement can be satisfied by three different approaches:

- Commercially or freely available software;
- Customization of WEB based tools;
- Dedicated developed applications.

Commercially or freely available software can be used for interactive many-to-many and one-to-one discussion. Several nearly equivalent solutions are available. No specific choice has been made but the selected products must provide desktop sharing and interactive white board integrated in the usual conference capability.

Customization of WEB based tools can manage experiment proposal, device information, textual comments and publication process. Web based blogging tool and content management systems tools like WordPress[3] and Drupal[4] provide a convenient way for display experiment proposal and discussion.

Dedicated developed applications are the only solution to provide the features necessary to the two very important applications: the discharge parameter editor and data visualization. They can be developed in the form of web-applications or standard desktop/client-server application. The ability to run on any client without distributing and installing software make web-based application the optimal solution for remote

analysis but presently web-based applications provide fewer features than standard applications. For this reason, at least for data visualization, a standard or client/server application must be also developed.

Discharge parameter editor is a human-machine interface for the experimental team to prepare an ensemble of plant system data to be loaded onto the plant after validation by the relevant officers. It should include specific algorithms, learning tools and machine systems or even plasma simulators for preparing the pulse. High level requirements for the discharge parameter editor are:

- Discharge condition could be prepared from a remote site with the exact same interface and validation tools than on the local site;
- The discharge condition editor should be able to use the database of previous discharges and previous discharge condition or simulation data from any source;
- Pre-check tool of discharge parameter against operating instructions of the machine and subsystems must be available in the discharge parameter editor;
- The discharge parameter editor concerns all subsystems of the machine including heating systems and diagnostics set-up;
- The discharge parameter editor should be able to view the availability of each system;
- The discharge parameter editor should be able to record/access post-discharge information such as alarms from the plant or from the plasma, warnings, or any details relevant for the analysis of an incident;
- The discharge parameter editor should be able to use and set-up all the control tools available on the machine library and the plasma and machine simulation tools available;
- Loading onto the plant must be made on the local site only by the local responsible officer.

Data visualization tool should provide an easy way to visualize and interact with all kind of stored data. Most important requirements are:

- 2D/3D, polar, contour and image plots;
- Multiple waveforms (traces/pulses) in one window;
- Annotations, drawing tools;
- Axes rescaling, interactive zoom, pan, pointer;
- Pointer cursor can follow represented traces, obtain numerical values, intervals;
- When pointer position is on one plot, its position is replicated on all other at the same xy coordinates;
- Linked plot to synchronize 3D signals or images to 2D time bases when moving the pointer cursor;
- Interactive browsing or query to search signals in data-files;
- Signal manipulation before visualization like simple linear transformation, derivative, FFT, smoothing or more sophisticated processing custom-written macros, or functions with interpreted languages (IDL, Cling, Python);
- Local and remote data access.



Fig. 1: RFX to JAEA Network connection.

5. Remote control room.

Remote Control Room (RCR) or Remote Experiment Centre means a site playing a role in the experimentation of the device, having the following functionalities:

- Specifying pulses design;
- Participation in their execution;
- Monitoring their progress;
- Analyzing the experimental results.

RCR can play a specific role in remote experiment exploitation thank to its ability to nearly completely remove the need of having researchers tokamak site at the time of experiment execution. It can significantly reduce the cost of moving researcher over long distances.

Remote control should allow the functioning of specific hardware from outside the plant operation zone by operation request. The operation request: (based on existing experience, JT-60U, JET and others) relayed through operation request gateway which interprets the request and relays them to the system if appropriate depending whether they are authorized or not. The functionality of the remote and local site should be equivalent from the user point of view. This implies that he can design pulses, participate to their execution, monitor the experiment progress and analyze the experiment data as if he was sitting in the local site.

6. Application to the JT-60SA experiment

Previous requirements apply to any fusion device but they are very important for JT-60SA experiment. In the JT-60SA case stringent requirements are related to the long pulse operation, the large distance between EU and Japan and the corresponding time lag, which makes even telephone communication difficult.

Since network is the underlying components that make possible remote exploitation a preliminary survey of network connection has been carried out performing some simple tests between JAEA (Naka) and Consorzio RFX (Padova) to evaluate if it is suitable for remote participation and remote data-analysis.

Firstly has been measured the time delay between Padova and Naka. By the ping command an average delay of about 280 ms (round trip) has been found. Such delay is mainly due to trans-oceanic hop because no direct connection exists between Europe and Japan but all high speed connections are through the USA.

To analyze the effect of such time delay on remote participation some computer connections have been established using the tools available at JAEA. A VPN connection between an RFX desktop computer and JAEA network has been established using RSA SecurID strong authentication system. Remote computer access has been tested starting a session on a

JAEA data analysis server based on RedHat. The two available connection methods have been used: SSH+X11[5] and NX[6]:

- SSH+X11 performs very badly: typed character takes about one second to be display, a signal visualization tool (daisyx) takes about 30 seconds to display the main window and menus do not shown their content before pointer moves over them.
- NX performs very well: only small delay have been observed in typing character and running the daisyx visualization tool. Editing a file remotely is only slightly slower than locally.

Remote data access has been simulated running on a browser of the local desktop computer a Java Applet based visualization code and transferring a file with scp (Secure Copy):

- Java Applet is very slow in data loading: to display about 10 waveforms with 150 data points the Java Applet takes about 20 seconds, after this delay data manipulation is quite fast because data and code are inside the local computer.
- File transfer with Secure copy uses a fraction of the bandwidth: about 140 seconds are required to transfer 30 MB file of, with a mean a transfer speed of 207 KB/s (or 1.7 Mbits).

Preliminary tests on Europe to Japan network connection show a satisfactory experience with RCA properly selecting the connection tool; oppositely RDA seems to require some optimization to provide an acceptable response. Though big improvement on network performance can be possible from now to the JT-60SA operation starting time it is manifest that technologies must be carefully selected and optimized.

Acknowledgments

This work was supported by EURATOM and carried out within the framework of the European Fusion Development Agreement. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

References

- [1] P. Barabaschi, Y. Kamada, S. Ishida, Fusion Engineering and Design **86** (2011) 484
- [2] F. Imbeaux et al, *Design and first applications of the ITER Integrated Modelling & Analysis Suite*, Proc. 25th Fusion Energy Conference, Saint Petersburg (Russian Federation), TH/P3-41.
- [3] <http://wordpress.org>
- [4] <http://drupal.org>
- [5] http://en.wikipedia.org/wiki/Secure_Shell
- [6] <http://www.nomachine.com/>