UNIFYING DATA DIVERSITY AND CONVERSION TO COMMON ENGINEERING ANALYSIS TOOLS

H. Reymond, O. O. Andreassen, C. Charrondiere, M. F. Gomez de la Cruz, A. Rijllart, CERN, Geneva, Switzerland

Abstract

The large variety of systems for the measurements of insulation, conductivity, RRR, quench performance, etc. installed at CERN's superconducting magnet test facility generates a diversity of data formats. This mixture causes problems when the measurements need to be correlated. Each measurement application has a dedicated data analysis tool used to validate its results, but there are no generic bridge between the applications that facilitates cross analysis of mixed data and data types.

Since the LHC start-up, the superconducting magnet test facility hosts new R&D measurements on a multitude of superconducting components. These results are analysed by international collaborators, which triggered a greater need to access the raw data from many typical engineering and analysis tools, such as MATLAB®, Mathcad®, DIAdem™, Excel™... This paper describes the technical solutions developed for the data formats unification and reviews the present status.

INTRODUCTION

The 1700 twin-aperture superconducting magnets [1], constituting the LHC accelerator ring, have all been tested at CERN before their installation in the tunnel. These validation and acceptance tests were executed in a dedicated superconducting magnet test facility named SM18.

For these tests campaigns, many dedicated measurement systems [2] and analysis tools were developed. These tools, which are now coming of age, required many development and validation hours. During the 2003-2007 period they have performed thousands of automated tests and analysis, helping the operators to execute all the required magnet tests within the tight schedule of the LHC assembly.

After the LHC start-up, due to the wide range of testing possibilities for superconducting apparatus and magnet technology, the SM18 test facility was converted into a Transnational Access Test Facility [3], a unique test platform for the superconductivity community. It is equipped with ten benches, allowing tests at very low temperatures (lowest being 1.9K) and high currents (up to 20kA).

Because of its flexibility and great range, several R&D projects have been done in SM18. In the last three years, these projects, managed by international collaborations, have created a need for a more flexible and generic data access solution.

These new projects would typically request to do different or extended analysis, correlating data from sources not foreseen when the first applications were designed [4].

THE LEGACY MEASUREMENT SYSTEMS

All the measurement systems designed in SM18 have been built using the same approach. They are following a classical two tiers application where on one side we have a control and data acquisition system, and on the other side a dedicated analysis tool, custom tailored to handle the dedicated data.

This scheme has been applied to:

- Electrical circuit insulation measurements (HV application)
- Quench performance tests (HF/LF DAQ application)
- Measuring physical properties of magnets (RRR, splice resistance, energy loss, etc.)

Today the cross-correlated analysis requested by the new users community are difficult to achieve since these applications where not designed for this purpose from the very beginning The file formats differ between systems, and the specific legacy tools are mandatory to make any good use of the data. In addition, since these tools were custom tailored, the numerous requests for adapting the analysis to new measurement types cannot easily be done. This causes delays since one has to manually analyse the data and is unsatisfying for the users community.

THE SDA PROJECT

In order to offer a suitable solution, the SM18 Data Access (SDA) project was initiated. It aims to open the data access of the legacy measurement systems, enabling raw data access for existing mathematical or engineering analysis tools, and provide a flexible, user customizable environment, that can quickly adapt to new requirements.

In the beginning, we searched for a common data format that we could directly integrate in the legacy measurement systems, storing the raw data in a common industrial format. This idea were deemed unfit due to the necessary efforts needed to adapt all the legacy applications, and also the potential compatibility issues one might have between old and new data.

Instead, we proposed to create a unique and system independent data conversion module.

In addition, to find a flexible replacement for the legacy tools, we launched a survey to find suitable applications from well-established industrial vendors. The tool had to be easy to use, flexible, fully featured with the most

ISBN 978-3-95450-139-7

common mathematical toolkits and cover all the existing analysis operations done in the legacy applications.

Selecting a Common File Format

As mentioned, the key for the whole SDA project is its ability to take a custom file format and convert to a generic, industrially standardized and flexible file format, making it possible for virtually any engineering tool to read and treat the data.

As a result, our first step was to list and identify all the existing file and data formats generated by the legacy measurement systems (Table1). This exercise confirmed that all of them were different, even the structures of the text files were not compatible.

The second step was to select a commonly used file format, which would be generated by the data conversion module. There are many different popular data file formats on the market [5], each with different strengths and weaknesses.

Text based file formats, like eXstensible Markup Language (XML) or Data Interchange Format (DIF) were eliminated due to their large disk footprint. The HF-DAQ produces data at a speed of 200kHz on 160+ channels, which would have led to a huge waste of disk space.

Archiving and compression file formats such as Tape Archive (TAR), Roshal Archive (RAR), SQX and ZIP where neither found suitable. They are not so straightforward to use and they require dedicated and (sometimes) proprietary software, both when writing and reading the files.

Table 1: The Measurements Data Formats

Application	**Data Format**	**Data Analysis**
HV	Text	Included in App.
RRR	Text	Dedicated tool/Excel
SPLICE	Text	Dedicated tool/Excel
LOSS	Text	Dedicated tool/Excel
ACT	Text	Included in App.
HF-DAQ	Binary	Dedicated tool
LF-DAQ	Binary	Dedicated tool

Custom Binary files, in contrast to text files, have a significantly smaller disk footprint. Their main drawback is that they are not human-readable and cannot immediately be opened by standard software. To share these files, one must ship them with an application that interprets the specific binary file correctly. Custom binary files where also removed from our list of potential formats.

After evaluating several hierarchical data formats such as HDF5 or BHM, we ended up selecting the Technical Data Management and Streaming format (TDMS) [6]. It has been developed several years ago by National Instruments, specifically to meet the needs of engineers and scientists collecting large quantities of test data. TDMS mixes the small disk footprint of the binary file and the self-description of the XML with a comprehensive header and an attributes list.

Some attributes such as file name, date, and many others, common to signals and channels properties, are stored automatically when creating TDMS data files. In addition, the user can easily add his own custom attributes as well.

Another advantage of this file format is the built-in three-level hierarchy: file, group, and channel levels (Fig. 1). A file can contain an unlimited number of groups, and each group can contain an unlimited number of channels. One can add attributes to any of these levels, better describing and documenting the test data. This hierarchy creates an inherent organization of the test data.

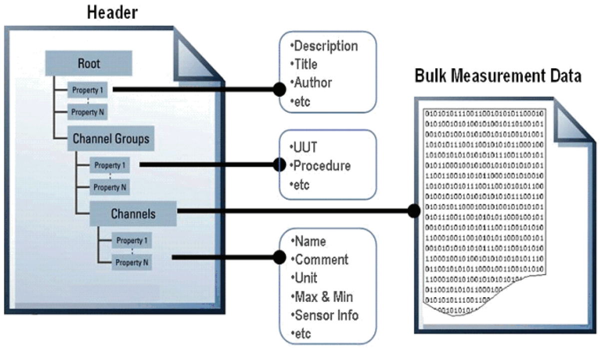

Figure 1: Overview of the TDMS file structure.

When saving data in a TDMS file, a complimentary index file is automatically generated. It provides consolidated information on all attributes and pointers in the bulk data file. This is done to speed up read access on larger data sets.

Finally, although TDMS files are binary, they can be opened in many common applications, such as Microsoft Excel™ or OpenOffice™ with a free plug-in. Even if these applications are not suited for deeper mathematical analysis, they are often used either for first level analysis (based on user developed macros) or simply to present results for large audience. In the same way, MATLAB® can easily handle TDMS data files, with an appropriate existing plug-in.

The Implementation of the Conversion Module

As the legacy analysis tools have been built with the LabVIEW™ programming language, and we had the knowledge and expertise in the team, it was a natural choice for the data conversion module (Fig. 2). By re-using the existing LabVIEW™ based files handling modules, from the legacy applications, we drastically reduced the development time, compared to a new development from scratch, in C or JAVA.

A valuable advantage of this language is the presence of a built-in palette of functions for TDMS file creation.

They have speeded-up the coding of the TDMS file generation and its implementation.

A major effort was put in the definition of the TDMS structures and the hierarchy of the properties over the three levels (file, group, channel).

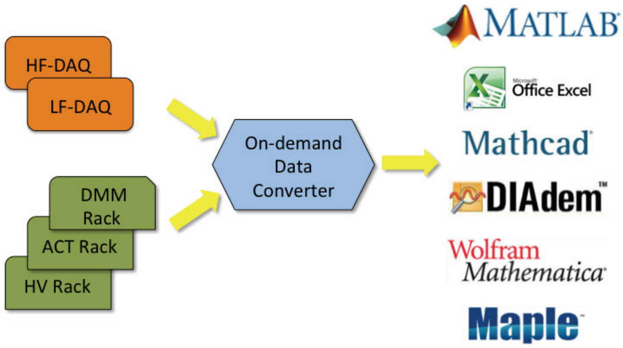

Figure 2: Principle of the conversion module.

One of the most difficult tasks, when making the data conversion module, was merging the seven different files and data types from the HF and LF DAQ systems. Originally these data are produced at different frequencies (from 250Hz to 200kHz), through many triggers, time scales and channels configurations. Until now, all these DAQ files were only viewable from two separated analysis applications. These one exploiting some related configuration files, to apply if needed, some formulas and conversion factors, as specified by the users, before launching the acquisitions. We have been able to embed the whole HF and LF characteristics (data, frequencies, time patterns, formulas...) into one TDMS file, taking advantage of its native hierarchical structure.

For all the others legacy systems based on mobile racks, the coding and integration into the conversion module was done without particular difficulties. Their text-formatted data files, were much easier to convert into TDMS files.

Selecting a Common Analysis Tool

There are several engineering analysis tools available on the market. Since we had already settled for the TDMS file format, we focused our attention on finding a tool, which came with TDMS support, out of the box.

In addition, we wanted to find a flexible, easy to use and fully featured application, in the sense of math and analysis modules. They would have to cover the capabilities of the legacy analysis, and be adaptable to the new calculations, requested by the collaborations users.

Some of the tools we looked at, were MATLAB®, Mathcad® and Mathematica®. These softwares are maintained by large companies and they offer a whole mathematical environment. However, they suffer of a complex interface and even if proven, the flexibility is mainly founded on scripting language.

We finally turned our choice on the DIAdem™ software, proposed by National Instruments. It met all our needs, such as the manipulation of large volumes of scattered data, the intuitive analysis interface and the aptitude to handle TDMS files. This software is designed to quickly locate, load, visualize, analyse, and report measurement data, collected during data acquisition or generated from simulations. New bricks for specific algorithms can easily be built, by sequencing appropriate functions.

Using of the Conversion Module

The conversion module has been designed with ease of use in mind. The user selects the tests to convert, through a simple list, choose the output format (TDMS or CSV for now) and the tool will propose a location where the output file should be stored. At this point, the converted data can be analysed by any engineering tool (Fig. 3).

Behind the scene many things are going on. The data source can be located on Windows or Linux storage (DFS or NFS), in a transparent way for the users. In addition, for some data types, formulas and units conversion factors can be applied. The user can then choose to retrieve scaled or raw data.

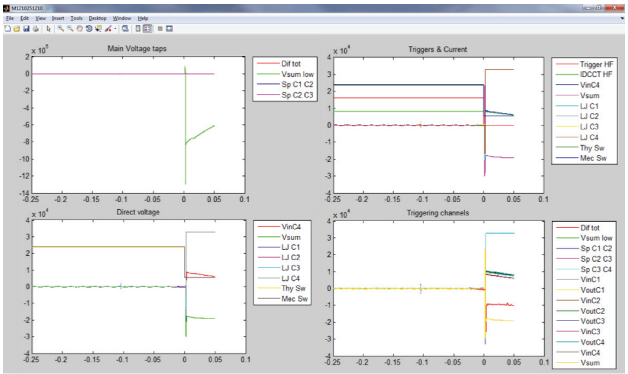

Figure 3: HF-DAQ data visualized in MATLAB®, after conversion of the test referenced: HCMQXCM0001-CR000001_2/M1210251210.a02.

CONCLUSIONS AND PERSPECTIVES

A data conversion tool has been developed to offer the users of the CERN superconducting magnet test facility, the possibility to analyse the measurement data, across several common engineering tools.

Until now, the raw data generated during these tests, were only readable with the legacy and fully customised analysis tools. The data opening that we put in place, gives a total freedom to the users, in term of application selection from those existing on the market, but also in term of developing their own mathematical toolbox.

We have proposed the use of both a new unifying data format (TDMS) and a powerful engineering tool named DIAdem™. Almost twenty persons working around the magnet test facility followed the training course, to get the most out of this application.

Using this new data format allows, for example, the correlation and viewing of all the measurements types of one dedicated magnet. This was not possible with the legacy analysis tools, due to the incompatibility of the file formats and the need of several independent viewers.

The TDMS format that we have selected after a survey of the main storage formats, gives a good compromise in term of file size. As example, the magnet test showed in the figure 3 has a size of 60MB, in the legacy binary format. It decreases to 18.9MB in ZIP format, but takes only 26MB in TDMS and can be opened directly with MATLAB®.

Even if we take into account the cost of the training, our approach was the most efficient, compared to a new development in C or JAVA. The use of LabVIEW™ allowed us to reuse existing code and facilitated the TDMS file generation. Also the license cost of the DIAdem™ software was not an issue, as we benefit at CERN of a full site license for the National Instruments products.

A modular approach has been used to program the data conversion module. This will simplify the future integration of raw data formats. It will also ease the capacity to accept other legacy DAQ application data as input.

Finally this method could be applied to other CERN facility areas or in any environment hosting a wide range of legacy file formats.

REFERENCES

[1] G. de Rijk et all, "Status Report on the LHC main magnet production", ASC 2004, Jacksonville, USA.
[2] A. Raimondo, P. Coutinho-Ferreira, G. D'Angelo, H. Franca-Santos, M. Gateau, M. Peryt, H. Reymond, A. Rijllart, A. Siemko, "The hardware recognition system for the SM18 LHC cryomagnet test benches", ICALEPCS 2003, Gyeongju, Korea.
[3] http:/eucard.web.cern.ch/EuCARD/activities/research/WP7
[4] A. Raimondo, H. Reymond, A. Rijllart, "A distributed architecture for the LHC dipole series measurements", IECON 2000, Nagoya, Japan.
[5] http://en.wikipedia.org/wiki/List_of_file_formats
[6] http://ni.com/white-paper/5696/en

ISBN 978-3-95450-139-7