

Tools for Validation and Analysis of a Large Amount of Experimental Data from Magnetic Measurement System: a Case Study at CERN

V. Afonso¹, L. Fiscarelli^{1,2}, J. García Pérez¹, G. Montenero^{1,2}, L. Walckiers¹

¹ *European Organization for Nuclear Research – CERN, Genève, Switzerland*
{*veronica.afonso.rodriguez, lucio.fiscarelli, juan.garcia.perez, giuseppe.montenero, louis.walckiers*}@cern.ch
² *Department of Engineering University of Sannio, Benevento, Italy*

Abstract- A flexible and easy-to-implement management structure for large amount of measurement and test data is proposed. Its main aim is to cope with industrial and research environment where the difficulty to make measurements requires to minimise the number of measurements campaigns and maximize the use of the data taken in these campaigns. This goal is achieved by providing well-defined structures for raw data storage and Data Base storage and clear separation between the analysis processes. A case study on test data of superconducting magnets for the Large Hadron Collider at the European Organization for Nuclear Research (CERN) is illustrated for managing present and future magnetic measurements requests.

I. Introduction

Measurement techniques of magnets have always followed the requirements and developments in the field of particle accelerators.

The main requirement for the measurement systems concerns the improvement of respectively (i) available bandwidth for long measurement run and (ii) accuracy. Therefore, a large amount of data to be processed is unavoidable. The data to handle come from measurement campaigns expensive by nature, as well as hardly to be repeated, thus it is mandatory to review several times raw data with new algorithms, by keeping unchanged the analysis process as a whole. Measurements carried out on magnets under development for accelerators for medical applications, high-intensity synchrotron light sources, or high-energy physics require similar data management layers.

Several proposals are available for the management of large-scale data sets from scientific experiments and simulation. A Web-based architecture can be developed by exploiting Structured Query Language (SQL) and Extensible Markup Language (XML) [1]. Such an architecture is suitable for distributed systems, but requires useless effort for a centralized repository for the data of interest. Optimization access to data [2] and Data Base (DB) solution [3] represent a costly design approach for projects where other resources are already available, i.e. utilities for raw data storage and commercial DB.

This paper deals with the organisation of the off-line analysis work consecutive to the data acquisition. The whole experience accumulated in the magnetic measurement field has to be formalised in order to understand the behaviour of the magnets and to improve the measurement accuracy. A flexible and easy to implement data management structure is requested since only magnet and measurement specialists cope with the programming effort. In particular, in Section II, constraints on the process of data storage and analysis tools are illustrated. In Section III, the architecture of the proposed management structure, and, specifically, the implementation tools, are described. In Sections IV and V, the experimental case study of data gathering and the on-field experience so far, are reported, respectively.

II. Constraints to structure the data storage and analysis tools

In order to guarantee consistency of the large amount of data to the final user whether magnet or beam optics specialist, an appropriate structure of management has to be created. Such a structure must provide an easy and flexible handling of the stored data and the required analysis tools. However, some constraints have to be taken into account:

- The raw data files must be kept accessible with the possibility to sort in a flexible way the “good” measurement files from the doubtful ones and to later repeat that sorting with other criteria.
- The raw data and final results must be separated. It is moreover essential to keep intermediate analysis level files to separately tune the different analysis tools and reduce the data after each analysis step.
- A strict bookkeeping is needed for the tools making these intermediate analysis in order to revisit the results in a systematic and efficient way.
- The programming effort must be kept minimum since tuned analysis tools becomes quickly obsolete for

short duration magnet development projects.

- Easy-to-use programming software must be chosen since the programming effort, in particular for the analysis algorithms, is entirely covered by magnet measurement specialists.
- The need for "informal" tools to document and improve the structure, or easy-to-use formal tools. It is always possible to go for more systematic, but considered over killing for the moment.

III. Management Structure Architecture

These constrains call for informal documentation tools to describe the relations between the processes, the stored raw data files, and the databases [4]. More systematic software development tools [1-3] are for the moment considered excessively time consuming. The proposed data management structure has been developed for the most demanding need, i.e. detailed measurement of the dynamic effects present in the Large Hadron Collider superconducting dipoles by means of harmonic coil method [5]-[6]. The architecture is general in order to be adapted to other harmonic coil measurement systems and all the other measurement methods foreseen for the future magnet projects at CERN should be easily integrated in this software layer.

In the following, (i) in Section III.1, the tools selected to respect the aforementioned constrains are described, and, (ii) in Section III.2, the architectural view of the proposed Management Structure, specifying the relationship with the measurement system taken into account, are shown.

III.1 Data Management Structure: Proposed Tools

During the design phase of the data Management Structure, we looked for a replica management for cached and archived copies, a container technology to optimize archival storage performance and co-locate small data sets, and tools for uniform collection management across file systems, databases, and archives [7]. The conclusions are:

- XML is considered not adapted for the proposed approach since the team involved and data to exchange over networks are reduced. The emphasis is on flexibility of calculation more than on data storage.
- MATLAB[®] applications are adapted to our needs, with possibility to transfer them in the measurement systems as on line algorithms once they are validated.
- The Engineering Data Management System (EDMS, [8]) for storing the data because is preferred for the flat files storage since well supported at CERN.
- The measurement conditions are changing and can be characterized by lists of different parameters. The metadata are written in flat files (called [Testdata] in Fig. 2) as their structure is likely to change across measurement systems, even from different specifications according to the magnet types to measure (i.e. more or less dynamic concern, more or less stability in the calibration data). These flat files are replicated in a database to collect consistence information with the data analysis results.
- The intermediate and final results of the analyses are stored in Oracle[®] database widely used at CERN. The end users, teams involved in the magnet qualification process or in the accelerator beam physics, request a general database storage.

III.2 Data Management Structure: Architectural view

In Fig. 1, the top diagram of processes to obtain the results and have them accessible for the users in a format convenient for further exploitation is shown. (Right) The data flow inside the acquisition process, and (Left) the architectural view of the proposed "Data Management Structure" are highlighted.

In the acquisition system a metadata file [TestData] is created with all the information relevant to the measurement and its further analysis. Then the operator launches the measurement process and the acquisition starts. Data are then generated from different sources: the magnetic measurement data, the magnet excitation currents, and the mechanical data read by other sensors when needed.

Once the data acquisition is finished a first validation process is made. It includes graphical interfaces and validation steps in order for the measurement specialist to check the consistency and integrity of the data. This avoids storing data, which are later invalidated by the analysis. This checking can go up to evaluate the "generic" harmonics, a pre-analysis stage made with nominal calibration coefficients of the measuring equipment. Then the data are uploaded to the EDMS structure where they are stored for the detailed analysis.

In the following steps the analysis is made with the appropriate Matlab[®] programs implementing the harmonic analysis algorithms. The results of the analysis are then stored into a database. These results are then made available for a 'customers' in a standard format [Multipole].

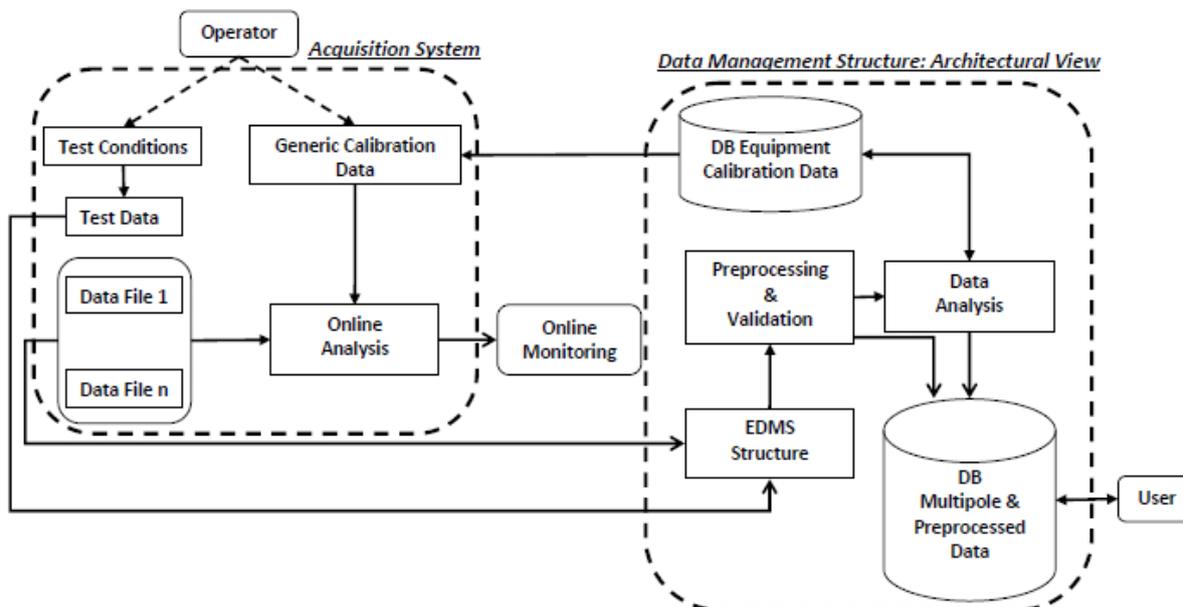


Figure 1 Diagram showing the architecture of the Data Management Structure from the Acquisition System to DB.

IV. Data Gathering: Experimental case study

As described in Section III the proposed “Data Management Structure” was mainly developed to fulfil the requirements from a high-performance rotating coil based measurement system, i.e. the Fast Acquisition Measurement Equipment (FAME) [9]-[10]. The voltage across a coil rotating in a 2D field of the slim aperture of an accelerator magnet is integrated over time to give the magnetic flux as a function of the angular position. This periodic signal is Fourier analysed to give the multipole or harmonic content of the field [11]. The excitation current of the magnet is measured with the same time resolution.

The resulting data files “raw data” from the measurement process are (Fig.2 Acquisition System):

1. Integrated Flux file: this is the biggest one (several Gigabytes per hour) and contains the integrated flux and its associated UTC time for every angular encoder tick (trigger).
2. Current file. It contains the current value read from the power supply, and has as well the time information in the form of UTC time with μs precision and a frequency of 50 Hz.
3. Test data file. It contains the metadata associated to the measurement in particular the reference of the measurement head, the data acquisition system settings and the power supply settings (current cycles for the different power supplies used, magnets being composed of main and corrector windings).

MeasureName: MB2524_2010_2_22_10_45_tFT_standard_precycle_50As_300sflatTop						
LOGIN_DATE: 4/6/2010 10:37						
B3	B5	B7	Field	Current	Time	
-252.51	20.26	-8.81	-0.51	50.00	40231.41	
-252.70	20.27	-8.94	-0.51	50.00	40231.41	
-252.66	20.29	-8.91	-0.51	50.00	40231.41	
-252.76	20.31	-8.95	-0.51	50.00	40231.41	
-252.72	20.29	-9.02	-0.51	50.00	40231.41	
-252.85	20.26	-9.02	-0.51	50.00	40231.41	
-252.70	20.28	-9.03	-0.51	50.00	40231.41	
-252.87	20.28	-9.07	-0.51	50.00	40231.41	
-252.74	20.29	-9.04	-0.51	50.00	40231.41	
-252.75	20.31	-9.04	-0.51	50.00	40231.41	
-252.83	20.27	-9.01	-0.51	50.00	40231.41	
-252.76	20.27	-8.96	-0.51	50.00	40231.41	
-252.71	20.27	-8.95	-0.51	50.00	40231.41	
-252.68	20.28	-8.94	-0.51	50.00	40231.41	

Table 1 Excel file content with formatting defined by the user i.e. the measured quantities of interest.

The data from the above files are the input to the analysis chain defined by the structure on the left of Fig. 2. Once the data are uploaded into the DB they can be accessed by a graphical interface enable to choose the resulting content of the final user file. Table 1 shows, as an example, a spot into such a file. Each column represents the measured value for the quantities of interest, in particular the most significant integral multipole field component for a superconducting dipole magnet measurements.

The reported results concerns a measurement carried out during a so called standard LHC cycle measurement [12]. Figure 2 depicts the graphical representation of the analysed data: the integral b_3 field harmonic vs the measured current (along the whole measurement cycle).

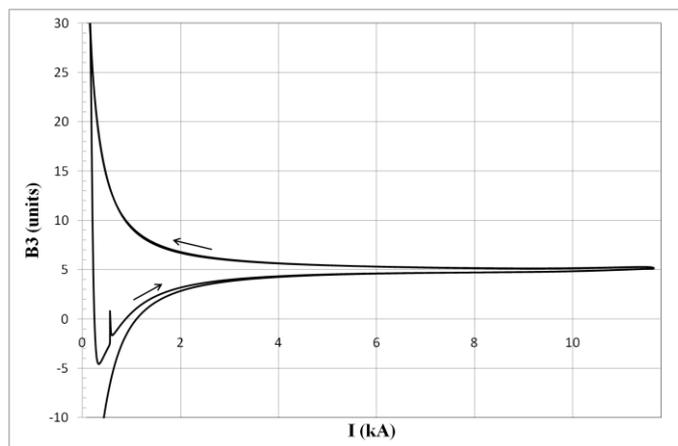


Figure 2 Measured hysteresis loop of the sextupolar field content in a superconducting magnet during a powering cycle: the analysis includes more than 3000 DFT.

V. Experience so far with this structure

The proposed structure already proved to help and give the following advantages:

- separation of the tasks of analysis and corrections that allows to split the programming work between the various physicists of the team,
- the possibility to come back to calibration data obtained after the measurements (already very useful for series measurement in the past),
- the possibility to go back in systematic way to the full analysis process,
- enough flexibility to adapt the structure as long as the measurements progress, and use that structure for new instrumentation or measurement systems.

V.1 EDMS Structure for storing raw data and test data from magnets measurements

The Engineering Data Management System, EDMS[®], allows the storing of data and documents in a very flexible way. A structure is created for having together the data coming from the FAME system as well as the documentation associated to its treatment. The structure is composed of folders grouping all documents associated to each magnetic measurement cycle. These folders and internal documents have properties as keywords, access rights, history management of the updates, and they are publicly accessible by standard web tools.

V.2 Data Base Structure for the measuring coil calibration data

In the Data Base the results after partial or completed analysis of the data are stored. There are many different tables created. A first type is the information describing the coils, basic elements in most of the magnetic measurement devices. During the last years a lot of different coils geometries and sizes have been produced. These coils are assembled into supports, or segments, which finally constitute the instruments. Calibration data of each long shaft, assembled with segments (Fig. 1, Equipment Calibration Data), are structured in detail. Some characteristics of the instruments are regularly calibrated into a reference magnet (coil surface, rotation radius) are also updated into the database. Using these parameters one can obtain the coil factors of the instrument, which are later use to obtain the harmonics from the raw measurement data.

V.3 Tools to process data

The set of tools that interact with the data files and the database tables and that process these data are described here below for the measurement systems already included and tentatively for the systems to include later. These tools are modules internal to the processing blocks of right of Fig 1. In order to have a good traceability of the results the various versions of codes of this analysis are saved into EDMS.

The two main tasks to analyse the data are, the calculation of the calibration coefficients of the instrument, and the analysis programs to obtain the pre-processed data [FTn] and harmonic components [Multipoles] from the FAME data.

The [Calculate Kn] tool makes the calculation of the coil factors from the detailed information of the coils [13]. These coefficients are then stored in the [CoilFactors] DB table will be used later on in the analysis process to find the multipoles of the measured magnets.

The second task is split up into two sub programs. The first program [Flux to FTn] uses as input the integrated fluxes of every turn measured with the rotating coil for every tick of the angular encoder, the UTC time of this ticks. It processes the magnet excitation current for accurate synchronisation. Several versions of internal calculations took already place demonstrating the soundness of the split approach for the analysis algorithms. These results are written into the table [FTn]. The second program [Offline analysis and validation] calculates the final multipoles and makes them available for both the magnet and accelerator physicists in the [Multipole] database.

VI. Conclusions

This work describes a data management structure to store and process data coming from a magnetic measurement system and finally save it into the Database. In particular, a case study on the FAME measurement equipment at CERN is reported. The structure is simple and flexible enough to evolve with the requests of magnetic measurements coming to the community of superconducting magnets. Tools have been developed to treat the raw data in order to feed the accelerator community with the results. For the moment all these programs are written in Matlab[®] but some critical objects can be transferred into online software for efficiency. The goal is to facilitate this data manipulation (storage and analysis) for the owners of this data first and eventually the users of the results.

References

- [1] M. Papiani, J. L. Wason, A. N. Dunlop, D. A. Nicole, "A Distributed Scientific Data Archive Using the Web, XML and SQL/MED", SIGMOD Record, Vol. 28, No. 3, September 1999.
- [2] B. Reiner, K. Hahn, "Optimized Management of Large-Scale Data Sets Stored on Tertiary Storage Systems", IEEE Computer Society, Vol. 5, No. 5; May 2004.
- [3] A. Valassi and all, "COOL, LCG Conditions Database for the LHC Experiments: Development and Deployment Status", Nuclear Science Symposium Conference Record, pp. 3021 – 3028, 2008.
- [4] URL: http://folkworm.ceri.memphis.edu/ew/SCHEMA_DOC/comparison/erd.htm.
- [5] P. Arpaia, M. Buzio, V. Inglese, G. Spiezia, "Experimental qualification of the platform for magnetic measurements at CERN", 16th Symposium on Electrical Measurements and Instrumentation & 13th Workshop on ADC Modelling and Testing (TC4), Florence, Italy, 2008.
- [6] L. Bottura, L. Walckiers, R. Wolf, "Field Error Decay and Snap-back in LHC Model Dipole", IEEE Transactions on Applied Superconductivity, Jun 1997, Vol. 7, No. 2 , pp. 62-605.
- [7] URL: http://www.npaci.edu/DICE/Pubs/Data-management_moore.pdf.
- [8] The EDMS Service Web site, URL: <https://edms.cern.ch>.
- [9] P. Arpaia, M. Buzio, L. Fiscarelli, G. Montenero, J. Garcia Perez, L. Walckiers, "Compensation of Third-Harmonic Field Error in LHC Main Dipole Magnets", IEEE Instrumentation and Measurement Technology Conference, Austin, US, May 2010.
- [10] P. Arpaia and all, "Measurement and Scaling Laws of the Sextupole Component in LCH Dipole Magnets", International Particle Accelerator Conference, Kyoto, Japan, 2010.
- [11] L. Bottura, "Standard Analysis Procedures for Field Quality Measurement of LHC Magnets Part I: Harmonics", Nov. 2001, CERN EDMS number 313621.
- [12] S. Sanfilippo, L. Bottura, M. Buzio, E. Effinger / LHC-MTA, "Magnetic Measurements for 15-m Long Dipoles -Extended Program of Tests", Internal notes LHC-MTA-IN-2002-183, CERN, Geneva, Switzerland, 2002.
- [13] L. Deniau, 'Coils Calibration Computation and storage of sensitivity factors', CERN Internal Note, MTA-IN-98-021, October 1998.