

MICADO – THE DATA IMPORT ENGINE OF THE CERN ENGINEERING AND EQUIPMENT DATA MANAGEMENT SYSTEM

D. Widegren, T. Ladzinski, R. Lyzwa, S. Mallon Amerigo, S. Petit, B. Rousseau, CERN, Geneva, Switzerland

Abstract

The CERN Engineering and Equipment Data Management System (EDMS) is one of the largest and most complex data management systems of its kind. For the last 10 years, it has received huge quantities of data generated in the different LHC project phases. Capturing all this information would not have been possible without a robust, failsafe, yet flexible and user-friendly data import engine. For this purpose the Micado toolkit was developed at CERN using XML standards, providing multi-level data verifications, an advanced queuing mechanism and batch processing of large amounts of import requests. Whereas Micado originally was developed to facilitate the capturing of LHC manufacturing data, its modular architecture has allowed a cost-effective extension to also cater for the LHC installation and hardware commissioning data import processes. Recently Micado has broadened its scope even further, including also imports of safety inspections, and at the same time the next generation of the tool providing Web Services is already being developed.

INTRODUCTION

CERN EDMS – the service

The project of implementing the CERN Engineering and Equipment Data Management System (EDMS) was launched in 1997 [1] and has over the years grown into a full-scale service comprising not only the system itself but also a user support, consulting activities and a complete training programme. Initially put in place for the LHC Project, the scope has today been widened and the service is currently provided throughout the whole Organization, including all its projects and experiments as well as the existing accelerator injector chain.

CERN EDMS – the system

The system provides today complete Product Lifecycle Management (PLM) capabilities allowing full traceability of equipment from the cradle to the grave. This includes support for the design work, follow-up of the manufacturing, installation and commissioning processes as well as maintenance management [2]. To achieve this, the chosen strategy has since the beginning been to interface existing off-the-shelf tools and to complement them when required with custom-made tools and web interfaces. The CERN EDMS is thus a compound system based on two commercial systems as its main pillars: Oracle Agile PLM for documentation and design data and Infor EAM for asset tracking and maintenance management. These systems have been linked at the

database level on top of which CERN has developed a set of web interfaces. Two of the most commonly used interfaces are the EDMS Web used for document and drawing management and MTF used for equipment follow-up [3].

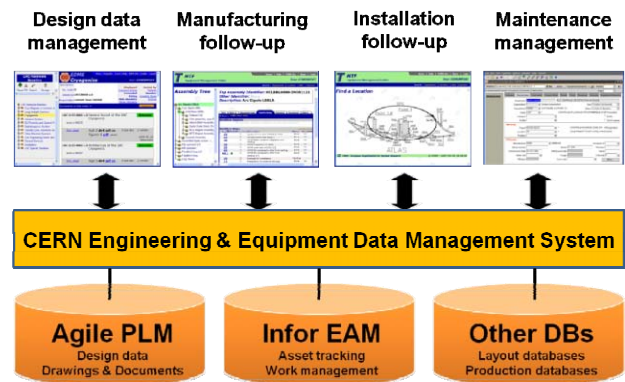


Figure 1: Overview of the CERN EDMS architecture

Initial requirements for automated imports

During the LHC manufacturing phase, MTF was used extensively for capturing results and progress of fabrication and test operations. Almost 600.000 individual LHC components were traced with MTF together with technical parameters and documentation. The general strategy was to register the information as close as possible to where it was generated, often using the MTF web interface directly at the manufacturing sites. However, with monthly rates of more than 12.000 equipment descriptions and even more related documents to register in MTF, it became obvious that a new automated import mechanism for large amounts of data was required. A new import system called MICADO (MTF Import Chain to Avoid Data Overdose) was therefore developed and put into production during the second half of 2004 [4].

OVERVIEW OF MICADO

Basics of the import engine

Micado was developed to provide a simple tool where users could submit data import requests that would be placed in a queue and processed automatically every night. By performing such massive, and thus CPU-consuming, imports in batch processing during night-time it was possible to minimize the potential impact of reduced system performance for normal users. After having submitted an import request, the user

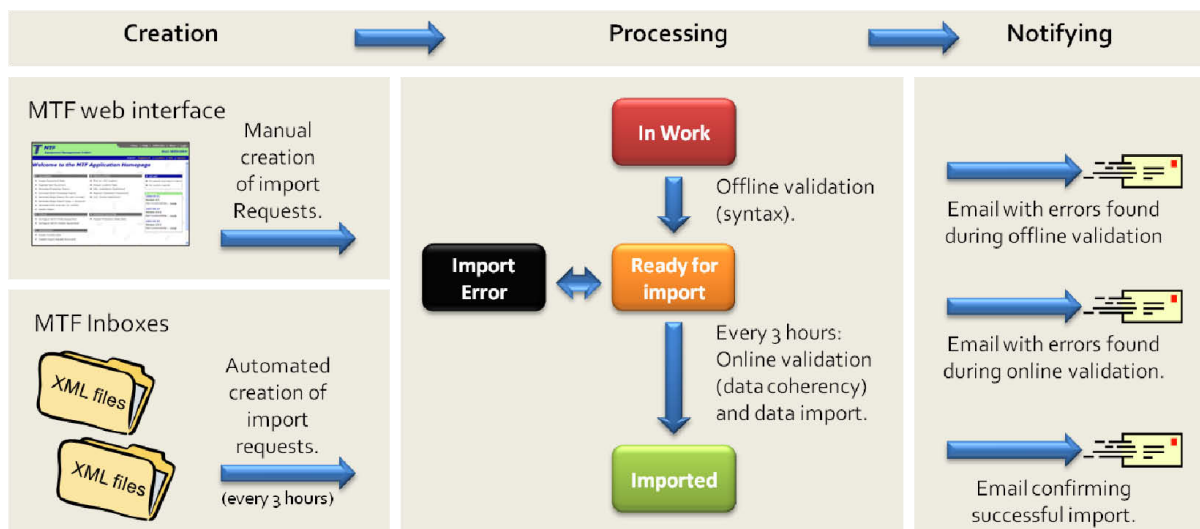


Figure 2: Overview of the creation and processing of Micado import requests.

would get a notification via email indicating progress or errors found in the request. Even though initially specified and developed for facilitating the import of manufacturing data, it was later discovered that the Micado functionality was required as well for the LHC installation and commissioning phases, for which MTF has been heavily used too.

Strategies and technologies used

When starting to develop Micado, one of the goals was to carry out an as cost-effective implementation as possible by relying on the existing infrastructure. It was therefore decided to generate the import requests as a special document type in EDMS. This allowed reusing of existing EDMS functionality such as the customizable document workflows and the flexible access rights module. Whenever possible, existing data validation functions written in Oracle PL/SQL were also reused. The actual Micado application was however entirely written in JAVA, which provided a higher degree of flexibility and access to many public libraries; such as JAVA classes for Excel data parsing. Such functionality was most important since pre-formatted Excel templates had been selected as the standardized format for input of data to Micado. These pre-formatted Excel templates were later complemented with XML files as a second import format allowed for more specialized types of imports.

The chosen implementation strategy showed to be very successful which was also proven by the fact that the first version of Micado, only targeting manufacturing data, could easily be extended to also cover installation and hardware commissioning data. It also allowed increasing the frequency of imports from once per night to once every three hours as the requirements changed over time.

IMPORT WORKFLOW

Import request creation

Micado is currently supporting two different ways of creating import requests; manual creation from the MTF

web interface or automated creation by using the MTF Inboxes. The manual creation of import requests is performed from the MTF web interface where the user has the possibility to submit a filled-in Excel template or an XML file formatted according to the Micado rules. In addition to this, the user can upload a zip file containing job result documents related to the objects referred to in the Excel or XML file. Using the MTF inboxes will have exactly the same result, but can only be performed by users with access to the specific folders on the CERN computer network that have been defined as MTF Inboxes. These special folders can be seen as “drop boxes” where well-defined users can place files that Micado will fetch every three hours and automatically upload into new import requests. To indicate that the files have been fetched by Micado the file extension is automatically changed to “.imp”. Both the downloading and renaming of files is technically done using the SAMBA protocol providing access to networked folders independently of operating system. The MTF Inboxes have proven to be a simple but very efficient solution when the data to import is the output of another application. An example of this are the drop boxes created for the sequencer application used for the LHC hardware commissioning where the produced test results are stored in MTF. A current limitation with the inbox functionality is however that only XML files can be uploaded, meaning that Excel templates are not accepted.

First offline data validation

A first offline validation is launched immediately after the submission of an import request. This is done automatically for MTF Inbox requests and triggered by a status change to “Ready for Import” for the manually created import requests. The purpose of this first validation is to identify possible errors in the syntax of the data and to find other general formatting problems without requiring any database interaction. Typical errors detected during this check are invalid dates or badly formatted equipment identifiers. If problems are detected,

the import request remains in the “In Work” status and an email notification is sent to the requester indicating in which cells of the Excel file or on which row in the XML file the errors are found. The requester must then correct that data and try to submit it a second time by changing the status once more to “Ready for Import”. If there are no errors detected by the first validation, the status of the import request will be changed to “Ready for Import” and it will be put in the import queue.

Second online validation and data import

All import requests that pass the first validation are put in the import queue which is scanned and processed every three hours. This final import process starts by running a second validation, this time by performing a full data coherency check with the database. If a problem is detected, no data will be imported and the whole import request will be rejected. The status will be automatically changed to “Import Error” and a notification will be sent to the requester indicating what data to be corrected before re-submitting it to the “Read for Approval” status. In case of no errors, the data will be imported and a final email notification will be sent to confirm this.

RECENT IMPROVEMENTS

During the last year a major effort was put into the consolidation of the Micado application by both streamlining the code and improving the maintainability of the software, notably by implementing the JAVA Spring framework. Due to the growing complexity and wider use of the tool, stricter configuration management procedures have also been applied including for example full version management of the software code in CVS.

Other major achievements carried out during the last 12 months were the extensions made to Micado to also cater for imports of safety inspections. EDMS-MTF is the Organization’s official tool for managing safety inspections and following the recent reorganization of the execution of such inspections at CERN, new batch import functionalities became urgently needed. Thanks to the modular architecture of Micado it was possible to provide these new import capabilities both rapidly and with limited resources.

IMPORT STATISTICS

In order to be able to efficiently follow the performance of Micado, a dedicated tool was developed to generate statistics based on the Micado log files. This tool provided important insights in how the tool was used and made it easier to identify where improvements would have the most significant impact. With peak months of more than 12.000 equipment registered, 35.000 jobs results recorded and almost 150.000 technical property values imported, Micado has clearly justified its existence and has proven to be a highly performing system. Since the first version in 2004 Micado has in total imported

some 400.000 equipment descriptions, 150.000 job results and 1.4 million property values.

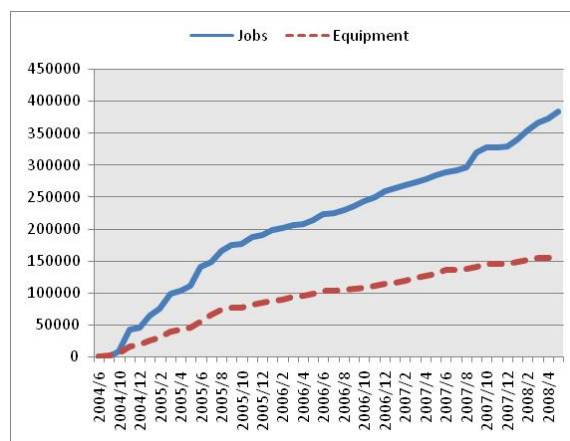


Figure 3: Total numbers of imported jobs and equipment

CONCLUSIONS AND OUTLOOK

The CERN EDMS has been an essential tool ensuring proper documentation and quality follow-up for the LHC manufacturing, installation and hardware commissioning. The system has received and managed to store securely huge amounts of technical data vital for the operation and maintenance phases of the LHC machine. Capturing all this information would simply not have been possible without the Micado import engine and its associated user support. In addition to this Micado has been a strategic tool for the support team allowing import tasks to be done by the end-users instead of performing such tasks internally in the team.

Micado has proven to be robust and failsafe yet flexible enough to allow easily implemented extensions. This combined with its user-friendly interface and email notifications has resulted in the decision that the import engine be extended to cover all object types in EDMS. Some first developments have been launched in order to link Micado to the EDMS core web services which would widen the scope and use of the import engine even further. The first results of this development project are expected by the end of summer 2008.

REFERENCES

- [1] T. Pettersson *et al.*, “The CERN EDMS – Engineering and Equipment Data Management System”, EPAC’02, Paris, June 2002
- [2] D. Widegren, “PLM at CERN - A True Challenge”, PDT Europe 2007, Geneva, September 2007
- [3] P. Martel *et al.*, “Manufacturing and Test Folder; MTF”, EPAC’02, Paris, June 2002
- [4] T. Pettersson *et al.*, “Equipment Manufacturing and Test Data Tracking for the LHC”, EPAC’04, Lucerne, July 2004