Data-driven operation and maintenance for particle accelerators

A. Apollonio [TE-MPE], L. Serio [EN-ARP].

Introduction and context

Particle accelerators are very complex machines, relying on a number of novel and often tailored key technologies (e.g. Radio Frequency cavities, magnets, vacuum, etc.), complex technical infrastructures and operational procedures (i.e. hardware commissioning, beam commissioning, machine cycles and beam optimization, 24/7 shift work, schedules and unscheduled technical stops for maintenance). Optimizing the maintenance, operation and long-term consolidation of such complex machines is far from trivial. To this end, big research centers rely on the long-lasting experience with previous particle accelerators as a fundamental ingredient. Particle accelerator applications in the industrial or medical domain cannot profit from the same background, although their systems and components are similar and their availability and reliability requirements are equally demanding, if not more, than accelerators for fundamental research. In the medical domain in particular, maximizing machine availability translates into maximizing the number of patients that the facility is able to treat and it might even have an impact on the patient prognosis. It is therefore of fundamental importance to achieve this goal. In the context of developing countries, reaching high availability targets is even more difficult, considering the challenging environments in which accelerators have to be operated, in particular the ageing and potentially obsolete technical infrastructures. This may involve unavailability of basic infrastructures, trained system experts or spare parts for failed equipment. In order to operate within such boundary conditions, accelerators have to be designed for high reliability and maintainability, as well as to minimize downtime required for eventual interventions. In particular, classical corrective or preventive maintenance approaches might be limiting the efficiency and reliability goals due to budget constraints. Predictive maintenance is a valid alternative to reduce interventions to only what is required, but its viability comes at the price of increased monitoring and diagnostics to be implemented. Overall, the ideal compromise should be found based on operability and costs.

The reasons to evaluate the potential of predictive maintenance are manifold: predicting failures allows timely identification of optimal maintenance windows, reducing the costs for storing spares and lead times for procuring new ones. It also allows performing coordinated preventive maintenance on a number of systems at the same time, in order to minimize the total number of interventions and the need for specialized personnel on site.

The ability to perform predictive maintenance relies on a number of technologies and infrastructure requirements. Machines have to be conceived with appropriate sensors to monitor the status of key parameters of each key system. The resulting data needs to be logged and stored continuously in a database with adequate capacity and performance. If sufficiently equipped with sensors, the accelerator will produce large amounts of data, which have to be processed in a timely fashion to preserve the possibility of early failure identification. This requires advanced data processing methods. Machine learning is gaining more and more attention in the field of predictive maintenance, dependability analysis and decision trees for its ability to process large datasets and identify hidden patterns, which could not be revealed with traditional data processing methods.

CERN is currently starting to explore the potential of machine learning to support operation and maintenance activities with different initiatives. In particular, it relies on different technologies for capturing, storing and processing data in an effective way. In the next paragraphs, a brief description of the present CERN experience, as well as the key adopted technologies is proposed. In the concluding paragraph, a discussion on the possible continuation of these activities and extension to new use cases is carried out.

Present Data Sources and Ongoing Studies

Accelerator Fault Tracker (AFT)

The AFT project was launched in February 2014 as a collaboration between the Controls and Operations groups with stakeholders from the LHC Availability Working Group (AWG). The AFT allows to capture information about the faults of CERN machines, specifying the failed systems, their recovery times and potential causality relations among different faults. The AFT system has been used successfully in operation for LHC since 2015, yielding a lot of interest and generating a growing user community for its capability of providing useful summaries and reports of accelerator performances to the operations team and the management in quasi real-time, while still providing detailed descriptions for system experts and fault follow-ups. In 2017 the scope of the AFT was extended to cover CERNs entire Injector Complex following a recommendation from the CERN Machine Advisory Committee.

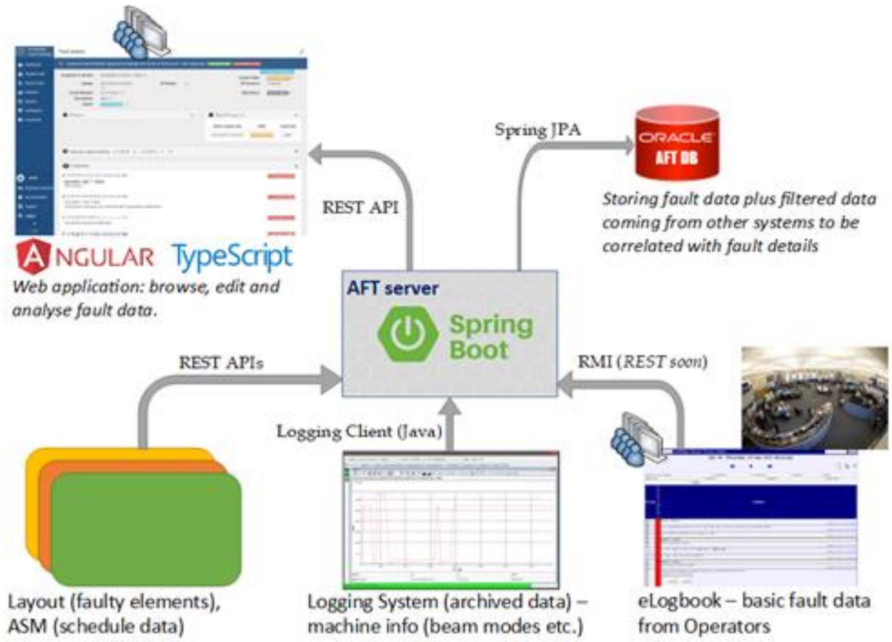

Figure 1 - Reference architecture of the AFT Framework

The AFT currently stores the fault history of the entire CERN complex. It also links to the CERN asset management system INFOR EAM via the Layout Database (see Fig.1). When specifying the 'faulty element' in AFT during a fault creation, the faulty equipment is automatically linked to a functional position. The latter is referenced in the INFOR EAM database to store the history of the related equipment, including past interventions, spare management and storage.

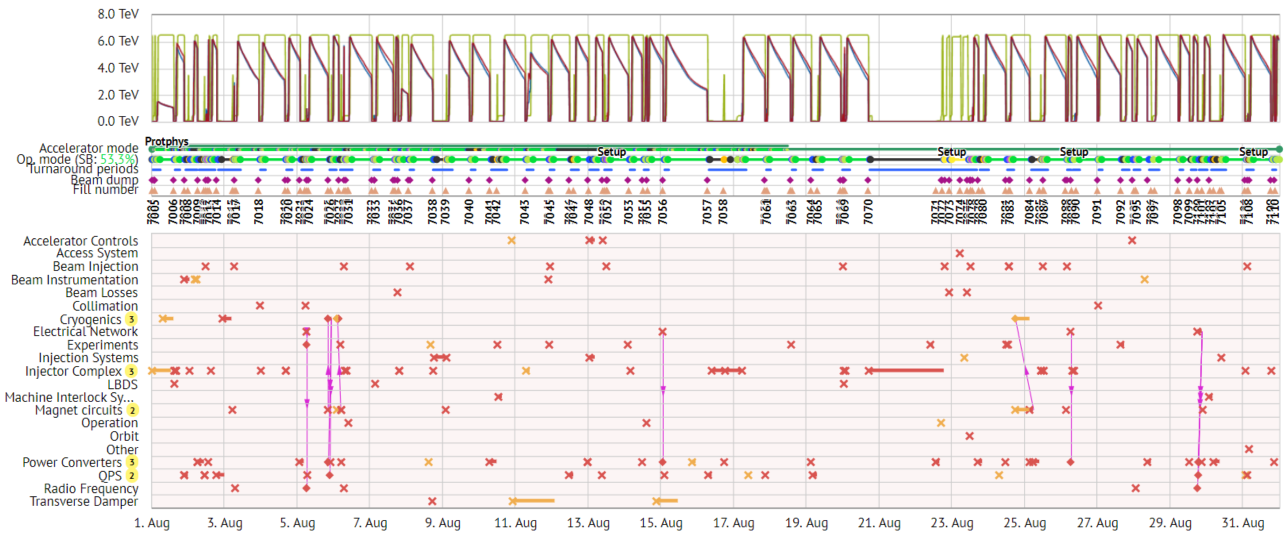

Figure 2 – ‘Cardiogram’ of accelerator operation

In addition, the AFT is already including information coming from the New Generation CERN Accelerator Data Logging Service (NXCALS). This is for the moment limited to variables used to display the so-called ‘cardiogram’ of accelerator operation (Fig. 2). This view combines the fault state of the various accelerator systems with the operating conditions of the machine, expressed in terms of beam energy and intensity. Given this interface, it would be easy to imagine an extension to include all necessary monitoring signals for individual systems in AFT (e.g. vacuum pressures, cryogenic temperatures and gas flows, voltages and currents in magnet circuits, etc.). With this in place, failure data stored in the AFT could be used as a reference to train possible supervised machine learning algorithms looking for precursors of relevant failures in the equipment data stored in NXCALS.
In the long term, the AFT could become a smart platform to automatically plan and manage system interventions, based on operating conditions and measured system trends. Thinking about new-generation accelerators, a tool with these functionalities will be essential to optimize the availability of large-scale machines, considering the increasing system complexity and more challenging maintainability requirements.

CASO Project (based on a H2020 call and currently being developed and tested within a collaboration CERN EN-ARP, Politecnico di Milano, Ecole des Mines Paristech).

The project has been designed on a modular development of individual ML algorithms (‘bricks’) to implement a multipurpose Computer-Aided System for critical infrastructures Operation (CASO), capable of inferring and interpreting data coming from different and heterogeneous sources and systems, to extract descriptive and predictive models and to suggest mitigation and consolidation strategies for the performance enhancement of complex systems and infrastructures.

CASO has demonstrated, on a proof of concept of a representative portion of the LHC, the ability to discover and track hidden dependencies of components and subsystems, to predict failures, to analyze and guide maintenance intervals and operation activities. The final goal of such a generic, overall tool will be to reduce cost and downtimes, and to increase quality, safety, availability and reliability of services and processes.

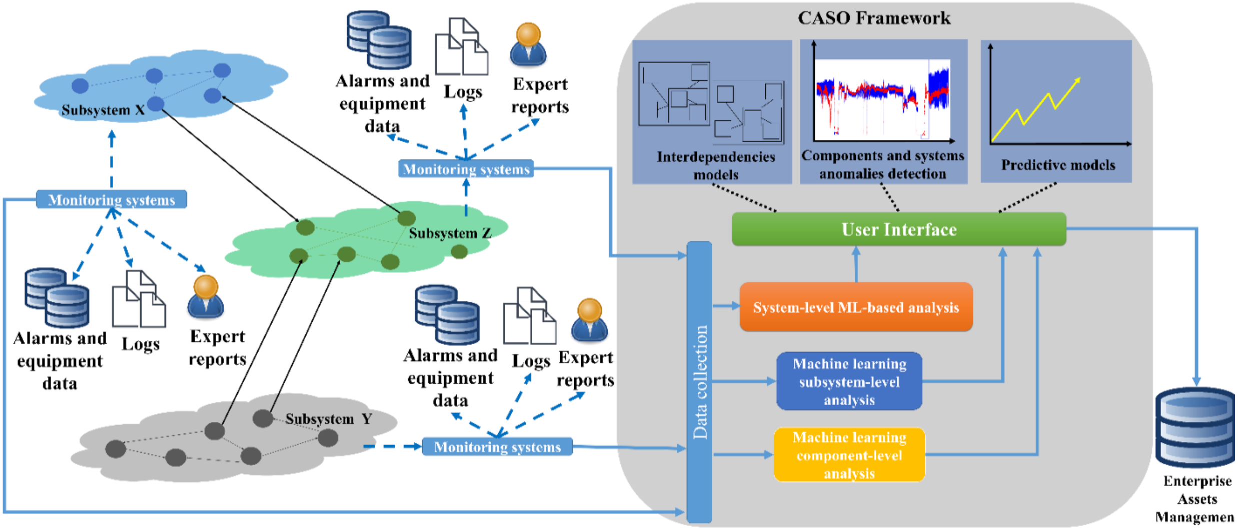

Figure 3 - Reference architecture of the Smart Framework

According to the high-level schema of the approach shown in Fig. 3, the critical infrastructure chosen for the design and the development of the framework is the CERN technical infrastructure.

The CERN technical infrastructure is a representative example of a critical infrastructure composed of several subsystems, each one responsible for a given critical functionality (e.g. distribute electrical power, cooling fluids, provide essential communication systems, etc.). Each subsystem is composed of thousands of components and is managed by a different monitoring system. A monitoring system collects all data related to the subsystem: physical signals of the equipment, reports, and logs related to failures, expert evaluations, maintenance interventions, installation dates of equipment and more.

The engine of CASO is a machine-learning core, composed of algorithms developed for the component-level and high-level system dependencies analysis.

Furthermore, CASO can be dynamically linked to an enterprise asset management system and the AFT to automatically report findings in order to select the most appropriate and efficient maintenance program and to guide and implement the required consolidation strategies to tackle aging problems of components, whilst maintaining the highest level of performance at a reasonable cost.

Failure prediction in PSB Power supplies

The LASER database is used at CERN to store system alarms. It is used by machine operators to actively monitor the status of accelerator systems. If an alarm occurs, operators have to react with appropriate recovery procedures. Alarms have different severities (from “0 = negligible effects”, to “3 = direct effect on operation”). In the CERN injectors, and particularly in the PSB, LASER is the main diagnostic tool used in the Control Room to identify the root cause of problems potentially affecting accelerator operation. An initiative was launched to verify the possibility to predict level 3 alarms by a combination of alarms occurring in the past. The focus of the study was on PSB power converters, as this is a key system for the availability of the accelerator. The study made use of state-of-the-art machine learning methods for timeseries classification. The problem formulation allowed scanning a number of hyperparameters, amongst which the prediction time, i.e. the time that the operations team and experts would have to react before the level 3 alarm is actually observed. This aspect is very relevant as it could yield significant improvements in scheduling interventions in the accelerator at convenient times (e.g. in the shadow of longer interventions or during technical stops).

With this goal in mind, the adopted machine learning approach was complemented with explainable-AI techniques, to potentially assist experts in the analysis of non-trivial dependencies identified by the model. An example of the use of the algorithm is reported in Fig. 4. The results from the model have been validated by power converter experts, who confirmed the validity of the findings and identified system dependencies.

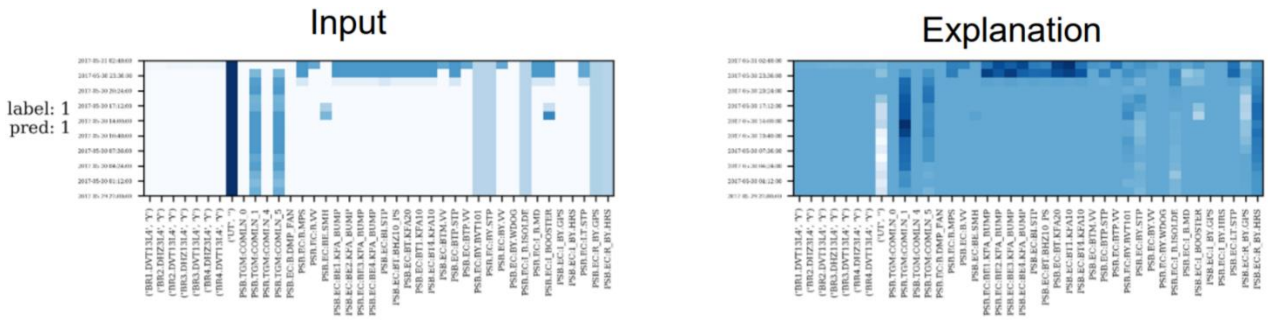

Figure 4 – Failure prediction for PSB power converters, complemented with explainable-AI techniques.

The developed approach has general validity and is based on state-of-the-art methods for timeseries classification. Going forward, one could envisage an extension of this approach to other systems and machines and, in the mid-to-long term adopting it as a tool to assist accelerator operators in the Control Room for failure prognostics.

Breakdown prediction in CLIC RF structures

RF (Radio Frequency) cavities are one of the essential building blocks of particle accelerators. They provide the necessary beam acceleration to reach the required energy for different accelerator applications. The development of high-gradient accelerating structures is one of the most challenging and interesting topics in the accelerator community and is of strategic interest for next-generation particle colliders. When designing high-gradient structures, one of the known issues is the occurrence of so-called RF breakdowns. These phenomena consist in the local formation of arcs and plasma, with possible damage of the cavity surface, due to small surface deformations yielding to local field enhancement.

The occurrence of breakdowns limits the performance of particle accelerators and can lead to damage of RF equipment. In this context, an initiative was launched at CERN to study the possibility to predict the occurrence of breakdowns in CLIC (Compact LInear Collider) accelerating structures with the use of machine learning methods. The data set used in this study is provided by the CLIC team, covering a period of 6 months for the so-called Xbox 2 test bench (see Fig. 5).

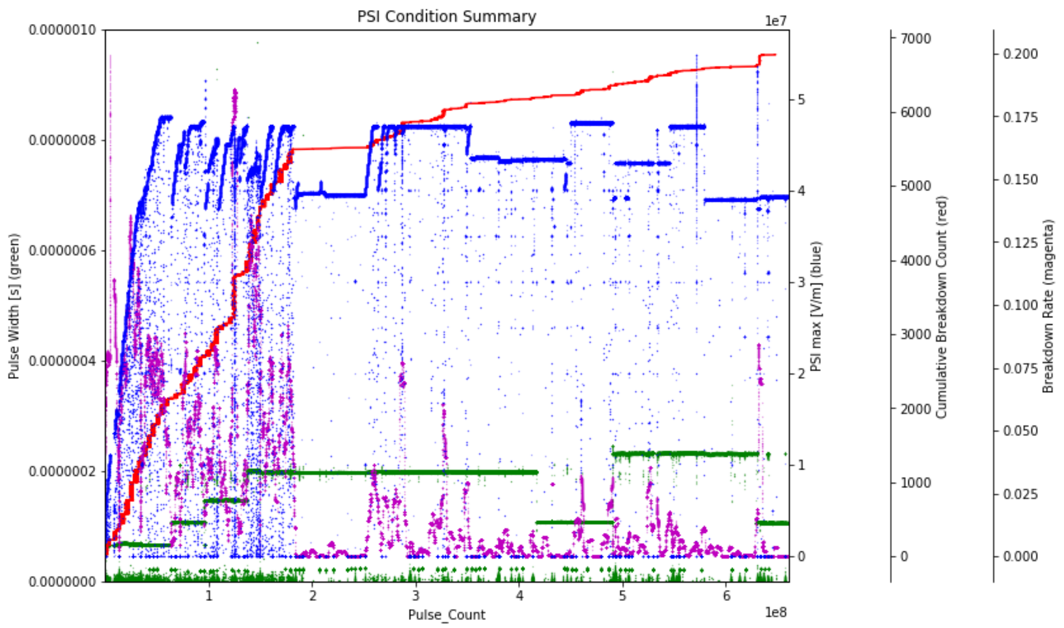

Figure 5 – Overview of Xbox2 CLIC RF structure tests.

Both supervised and unsupervised machine learning methods were employed for this use case. The initial data exploration phase was aimed at visualizing the high dimensional data, using unsupervised methods

(e.g. 2D-tSNE), to obtain insights about the data properties and to verify different labeling and data cleaning approaches.
This was followed by a breakdown prediction study with supervised methods. In particular, LSTM (Long Short-Term Memory) networks and convolutional neural networks for time series classification were used. The interpretation of the model parameters after training further enabled the reverse engineering of physical properties in the test bench, thanks to the applied 'explainable-AI' methods.
The results are currently being validated by system experts, as the developed models have potentially highlighted the presence of unforeseen breakdown precursors. Similar models could be applied to other sensitive RF cavities at CERN like the Linac4 RFQ, and are of general interest for the RF community.

Future Developments

Based on what was illustrated in the previous paragraphs, the present knowledge, tools and infrastructures available at CERN represent a very good basis towards the idea of having an AI-assisted system for operation and maintenance of particle accelerators. Additional developments and studies on various levels would nevertheless be required to achieve this ambitious target and generalize the system to accommodate requirements from any type of accelerator (including medical Linacs) and potential industrial partners. See Fig. 6 for a conceptual sketch allowing for AI-assisted operation and maintenance.
The following steps and activities are identified as needed going forward:

1. Development of a common framework to develop, test and validate ML-based tools on different components and systems to ensure their robustness and continuous improvement. Emphasis should be put on explainable-AI for enhanced failure diagnostics and maintenance.
2. Adaptation of CERN systems and technologies like AFT and NXCALS for use in different domains (e.g. by using open source technologies and abstracting from the CERN context)
3. Demonstrate the ability to perform transfer learning (e.g. Linac4 to medical Linacs). Ideally, models developed at CERN for specific systems should be validated and tested on similar systems from different machines (e.g. power supplies or vacuum systems).

The SMARTlinac could be an excellent test bed to implement a proof of concept for such a tool and systematically address the points above. CERN would highly profit from these activities, ensuring the improvement and further development of presently used machine learning techniques and information technologies. Furthermore, investing in these activities would allow CERN to develop expertise in a trending topic in industry, which is expected to become even more important in the context of next-generation machines, for which AI-assisted operation and maintenance will likely be a necessity.

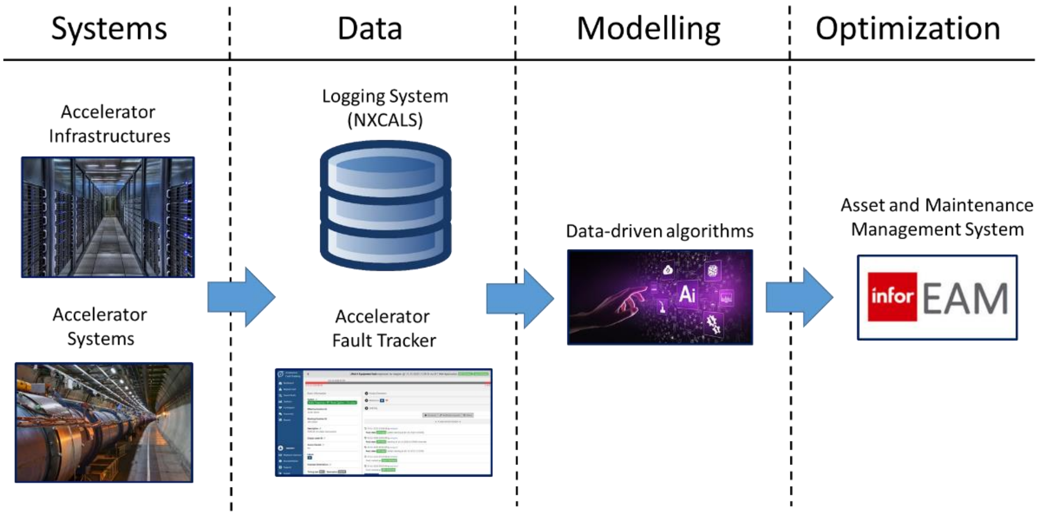

Figure 6 – Possible data-flow and architecture for AI-assisted operation and maintenance of particle accelerators.