ISBN: 978-3-95450-238-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2023-THPDP101

CREATING OF HDF5 FILES AS DATA SOURCE FOR ANALYSES USING THE EXAMPLE OF ALPS IIc AND DOOCS CONTROL SYSTEM

S. Karstensen[1] , P. Gonzalez-Caminal[3], G. Günther[2], A. Lindner[1], O. Mannix[2], I. Oceano[1], V. Rybnikov[1], K. Schwarz[1], G. Sedov[1]
[1]DESY Hamburg, Germany
[2]Helmholtz-Zentrum Berlin für Materialien und Energie GmbH (HZB), Germany
[3]Fusion for Energy (ATG Science & Technology, S.L.), Barcelona, Spain

Abstract

A versatile and graphical HDF5 file generation project. In the realm of physical experiments, data is typically gathered through the utilization of measurement devices intricately integrated into control systems. These control systems employ diverse methods for archiving historical data. Nevertheless, users consistently grapple with a recurring challenge: how to access the data, decipher the intricacies of the control system's structure, and understand the formatting. The prevailing approach involves the creation of scripts or programs tasked with extracting data from these control systems, and if necessary, followed by the requisite data pre-processing for subsequent analysis. Typically, this pre-processing is carried out in formats known exclusively to the individual user. Moreover, the longevity of data utility often teeters on the brink when users depart or deviate from established conventions.

To address these challenges, we are actively developing software that serves a dual purpose: firstly, to provide an API for a control system (in this instance, DOOCS), and secondly, to transform the extracted data into a universally recognized format, such as HDF5, complete with all relevant metadata.

The ALPS IIc experiment serves as an ideal testbench for this software development task due to its compatibility in terms of state and timing.

ABOUT HDF5

The Hierarchical Data Format Version 5 (HDF5) 0 is a unique high-performance technology suite that consists of an abstract data model, library, and file format for storing and managing extremely large and/or complex data collections. The technology is used worldwide by government, industry, and academia in a wide range of science, engineering, and business disciplines.

Advantages of HDF5

- Versatile data model that can represent very complex, heterogeneous data objects and a wide variety of metadata through an unlimited variety of datatypes.
- Ready for high-speed raw data acquisition
- Portable and extensible with no limits on file size, allowing applications to evolve in their use of HDF5.
- Self-describing, requiring no outside information for applications to interpret the structure and contents of a file.
- Robust software ecosystem of open-source tools and applications for managing, manipulating, viewing, and analyzing data.
- Architecturally independent software library that runs on a wide range of computational platforms (from laptops to massively parallel systems) and programming languages (including C, C++, Fortran 90, and Java interfaces)
- Advanced performance features that allow for access time and storage space optimizations through customizable product packaging, compression, and encryption
- Long-term data archiving solution
- Self-explaining data structure

METADATA

Metadata is structured data that contains overarching information about a resource. Metadata is used to describe the measured data (measurement data) with additional information, enabling their machine and automated processing, as well as providing insight into their origin and improving understanding. Taking metadata into account in the analysis can lead to better results.

Metadata could include:

- Process-specific information
- Images, Log Files, Links
- Hazards
- Responsible Users
- Storage Location
- Information about locations
- Permissions
- Publications
- Derived Objects
- Comments
- Hardware information
- eLogbook entries
- And many other information

ISBN: 978-3-95450-238-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2023-THPDP101

HDF5 GENERATOR

The HDF5 Generator is an innovative software product currently under development by DESY in Hamburg. It serves as a bridge between experiments, control systems, and databases, seamlessly transforming collected data into a new HDF5 file.

Figure 1 provides a comprehensive overview of the HDF5 data generation process.

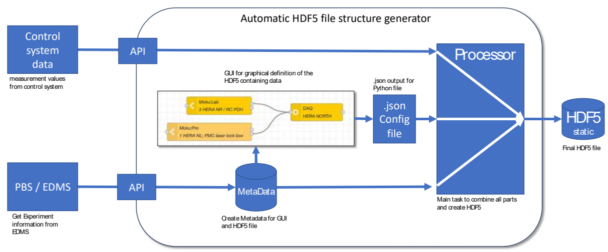

Figure 1: Principle of HDF5 generator.

To the left, control system data is received and meticulously processed via an API. Metadata containing essential information about the control system data and setup of the whole experiment for potential source reconstruction, follows a parallel procedure. Additionally, certain metadata serves the crucial role of furnishing users with a tailored graphical interface, simplifying the process of data selection. These elements are seamlessly amalgamated by the processor and subsequently archived as an HDF5 file.

Actually, we are working on a first version of the HDF generator. First tests are made and they are promising.

GUI CAPABILITIES

The graphical interface (Fig. 2) encompasses all experiment-related devices and channels, providing the user with the capability to choose which data to record in the HDF5 file. Each data value is supplemented with metadata, facilitating precise one-to-one associations. The graphical structure is documented and stored in a JSON file, ensuring a well-defined structure for ease of reading. As a proof of concept, we initially utilized Node-RED to conduct a feasibility analysis, with the final product slated to feature a distinct GUI.

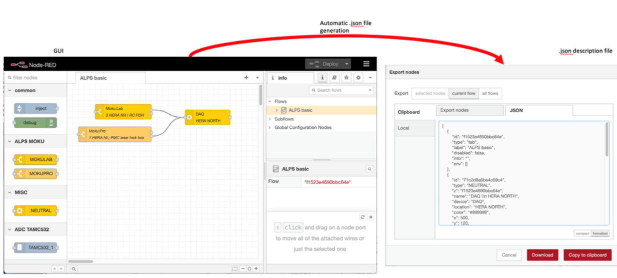

Figure 2: GUI interface and json output.

The Distributed Object-Oriented Control System - DOOCS - provides a versatile software framework for creating accelerator-based control system applications 0. These can range from monitoring simple temperature sensors up to high-level controls and feedback of beam parameters as required for complex accelerator operations (Fig. 3).

CONTROL SYSTEM

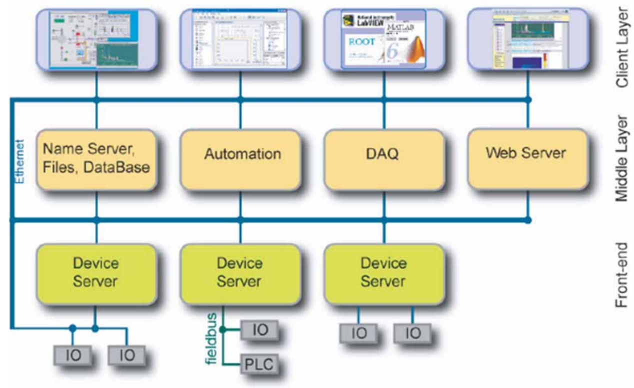

Figure 3: DOOCS control system schema.

DOOCS is based on a distributed client-server architecture combined with a device-oriented view. The devices are the basic entities and can be virtual integrated software instruments or real ones with hardware attached. Each control system parameter is made accessible via network calls through a device application.

Its transportation layer is based on the standardized, industrial RPC protocol and allows for a robust and efficient data transfer. Support for integrating a timing system providing clock, trigger and other time-based accelerator information is built-in into the core software and comfortably accessible within the framework.

DETECTOR AND DATA SOURCE

The Any Light Particle Search II (ALPS II) 0, pictured in Fig. 4 , is a light-shining-through-a-wall (LSW) experiment based at DESY in Hamburg, Germany, that will search for axions and axion-like particles down to the coupling of the axion to two photons of $g_{a\gamma\gamma} > 2 \times 10^{-11}$ GeV^{-1} for masses below 0.1 meV. ALPS II will use two strings of superconducting dipole magnets that are over one hundred meters in length, as well as optical cavities before and after the wall to boost the effective signal rate of the regenerated photons by more than 12 orders of magnitude when compared to previous generations of LSW experiments. Data taking with a simplified optical system has started in May 2023.

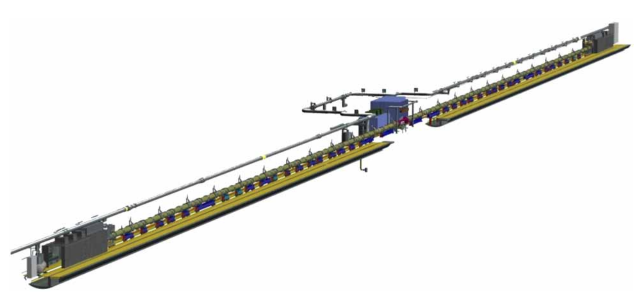

Figure 4: ALPS IIc at DESY, length: 280 m.

ISBN: 978-3-95450-238-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2023-THPDP101

The ALPS experiment, which stretches a total length of 250 metres, is looking for a particularly light type of new elementary particle.

Figure 5 illustrates the fundamental concept of photon-axion and axion-photon conversion within ALPS II, employing Feynman diagrams.

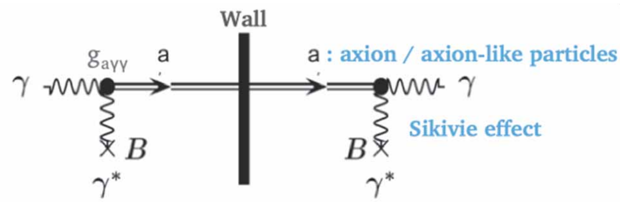

Figure 5: Feynman diagrams of the photon-axion and axion-photon coupling in ALPS II.

On the left-hand side, we observe a photon (γ) traversing a robust magnetic field (B) and undergoing conversion into an axion (a), which can subsequently pass through the light-tight barrier. The right portion of the schematic depicts the inverse process: the axion (a) encountering a strong magnetic field and undergoing reconversion into a photon.

Using twenty-four recycled superconducting magnets from the HERA 0 accelerator (Figure 6), an intense laser beam, precision interferometry and highly sensitive detectors are the heart of ALPS.

Figure 6: Inside the HERA tunnel.

CONCLUSION

The goal is to make data easily accessible for analysis, with a well-defined HDF5 format structure to select all needed data by using a graphical tool for selecting dedicated data. The present idea and partial implementation indicate that the focus will no longer be on cumbersome programming to extract data. Instead, the primary focus will be on data analysis. This way, valuable time and manpower can be saved or used more effectively.

We expect to have a first functional product by the end of 2023.

ACKNOWLEGEMENTS

This publication was supported within Hub Matter at the Helmholtz-Zentrum Berlin by the Helmholtz Metadata Collaboration (HMC), an incubator-platform of the Helmholtz Association within the framework of the Information and Data Science strategic initiative.

REFERENCES

[1] The HDF Group 2006, https://www.hdfgroup.org/solutions/hdf5

[2] DOOCS Homepage, https://doocs.desy.de

[3] Aaron D. Spector for the ALPS Collaboration, “Approaching the first any light particle search II science run”, Deutsches Elektronen Synchrotron DESY, Hamburg, Germany, Apr. 2023. doi:10.21468/SciPostPhysProc.12.039, https://inspirehep.net/literature/2674982

[4] HERA, https://en.wikipedia.org/wiki/HERA_(particle_accelerator)