Contribution ID: **17**   Type: **Poster**

PATOF: From the Past To the Future: Legacy Data in Small and Medium-Scale "PUNCH"Experiments - a Blueprint for PUNCH and Other Disciplines

Tuesday 10 October 2023 14:30 (15 minutes)

The PATOF project builds on work at MAMI particle physics experiment A4. A4 produced a stream of valuable data for many years which already released scientific output of high quality and still provides a solid basis for future publications. The A4 data set consists of 100 TB and 300 million files of different types (hierarchical folder structure and file format with minimal metadata provided create vague context). Recent work with consulting support from the HMC hub "Matter"helped to identify problems and potential solutions for a FAIRification of A4 data. We would like to go beyond and build a FAIR Metadata Factory that can be used across research fields. The first focus will be on creating machine-readable XML files containing metadata from the logbook and other sources and to further enrich them, other challenges will be an automatised treatment of personalised logbook information.

In this project, we intend to conclude the work on A4 data, to extract the lessons learned there in the form of a cookbook, and to apply them to four other experiments: The ALPS II axion and dark matter search experiment at DESY is expected to collect 1 TB of data per week. The PRIMA experiment at MAMI in Mainz for measuring the pion transition form factor is taking data of 3 TB per week in 2023. The upcoming nuclear physics experiment P2 at MESA in Mainz is expected to collect 3 TB of data per week. These are real data mixed with calibration data and polarimetry data. Finally, the LUXE experiment at DESY planned to start in 2026 and will collect 1.5 PB of data per year.

The focus of PATOF is on making the data of A4 (and ALPS II, PRIMA, P2, and LUXE) fully publicly available. We refer to these four future experiments jointly as "APPLe". In order to achieve this, a "metadata factory" will be implemented, the concept as follows:
- DESY library, provide a "cookbook"capturing the methodology for making individual experiment-specific metadata schemas FAIR and describing a "FAIR Metadata Factory", i.e. a process to create a naturally evolved metadata schema by extending the DataCite schema without discarding the original metadata concepts.
We first consult the domain experts from the concrete experiments (e.g., what data must be in the metadata) and design the metadata schema which partially follows the DataCite metadata schema as the core of it, plus experiment-specific metadata fields. Based on the consultation and experience that we have, we cross-reference the metadata of different experiments to find out the best strategies for automatically developing metadata schemas that can be used for different experiments, and even newly developing experiments.

The objectives of the project are i) a FAIR Metadata Factory (i.e. a cookbook of (meta)data management recommendations), and ii) the FAIRification of data from concrete experiments. Both aspects are inherently open in nature so that everybody can profit from PATOF results. The cookbook is expected to be further enhanced with contributions from other experiments even after PATOF ("living cookbook").

Please assign your contribution to one of the following topics

Metadata annotation and management close to the research process

Please specify "other" (stakeholder)

In addition please add keywords.

Metadata, Scientific Data Management, FAIR

Please assign yourself (presenting author) to one of the stakeholders.

Data professionals and stewards

Primary author: HU, Ding-Ze (Deutsches Elektronen-Synchrotron DESY)

Co-authors: Dr ENKE, Harry (Leibniz-Institut für Astrophysik Potsdam (AIP)); STEIN, Lisa-Marie (DESY); Dr KÖHLER, Martin (Deutsches Elektronen-Synchrotron DESY); Dr SCHOERNER, Thomas (Deutsches Elektronen-Synchrotron DESY)

Presenter: HU, Ding-Ze (Deutsches Elektronen-Synchrotron DESY)

Session Classification: Poster session

Track Classification: Poster session

Primary author: HU, Ding-Ze (Deutsches Elektronen-Synchrotron DESY)

Co-authors: Dr ENKE, Harry (Leibniz-Institut für Astrophysik Potsdam (AIP)); STEIN, Lisa-Marie (DESY); Dr KÖHLER, Martin (Deutsches Elektronen-Synchrotron DESY); Dr SCHOERNER, Thomas (Deutsches Elektronen-Synchrotron DESY)

Presenter: HU, Ding-Ze (Deutsches Elektronen-Synchrotron DESY)

Session Classification: Poster session

Track Classification: Poster session