From the Past To the Future: Legacy Data in Small and Medium-Scale “PUNCH“ Experience – a Blueprint for PUNCH and Other Disciplines

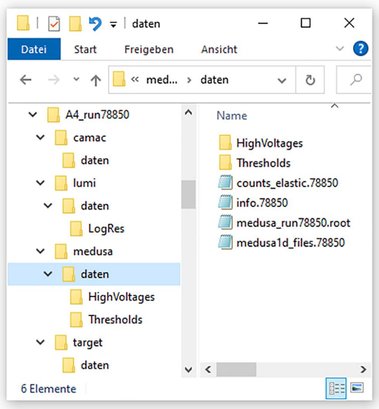
PATOF

Ding-Ze Hu[1] 0009-0007-8191-6035 \\ **Lisa-Marie Stein**[1] 0000-0001-7905-0462 \\ **Dr. Martin Köhler**[1] 0000-0003-0617-3319 \\ **Dr. Thomas Schoerner**[1] 0000-0002-7213-0352

[1] **Deutsches Elektronen-Synchrotron DESY**

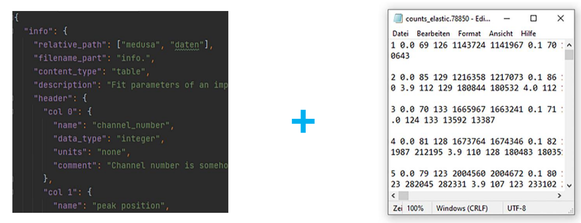
Data from the “PUNCH” (Particles, Universe, NuClei & Hadrons) disciplines particle / astroparticle / hadron & nuclear physics, and astronomy are valuable and often allow for new scientific insights not expected during an experiment’s lifetime. The PUNCH disciplines are very experienced in data management, largely due to early digital data acquisition systems, high data rates & volumes to be managed, and globally distributed user communities.
The experience and strategy of metadata management from PUNCH will be utilized in the PATOF project and we will go beyond. As the DESY library, we provide a “cookbook” capturing the conceptual 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 last version of the DataCite metadata schema without discarding the original individual metadata concepts.

Motivation: The particle physics experiment A4 at MAMI 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. Here, we report from our approach to make the data hoard of a dismounted project sustainable and publishable according to the FAIR principles which, finally, would make the data intelligible to, both, humans and machines.



BUT

- Hierarchical folder structure creates vague context
- Heterogenous file formats
- Minimal metadata provided
- Keeping structure and naming for convenience



Enrichment *ASCII table*

```
<column12>
  <name rel="rdfs:seeAlso" resource="https://doi.org/10.1351/goldbook.P04712">y_0 for polarization 0</name>
  <units rel="owl:sameAs" property="qudt:Number" resource="https://doi.org/10.1351/goldbook.C01370">count</units>
  <meanValue>78005.2908628204</meanValue>
  <rangeBottom>216.166</rangeBottom>
  <rangeTop>146147.084</rangeTop>
  <dataset type="xsd:float">
    <value>34889.059</value>
    <value>57867.741</value>
    <value>48403.713</value>
    <value>76495.574</value>
    <value>97206.345</value>
    <value>120959.259</value>
    <value>104223.584</value>
    <value>42705.002</value>
    <value>62608.645</value>
    <value>57638.797</value>
    <value>75986.432</value>
```

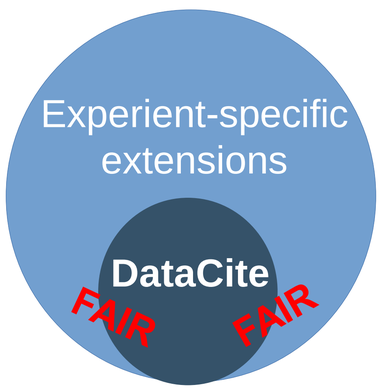
Enriched, machine-readable (xml) file

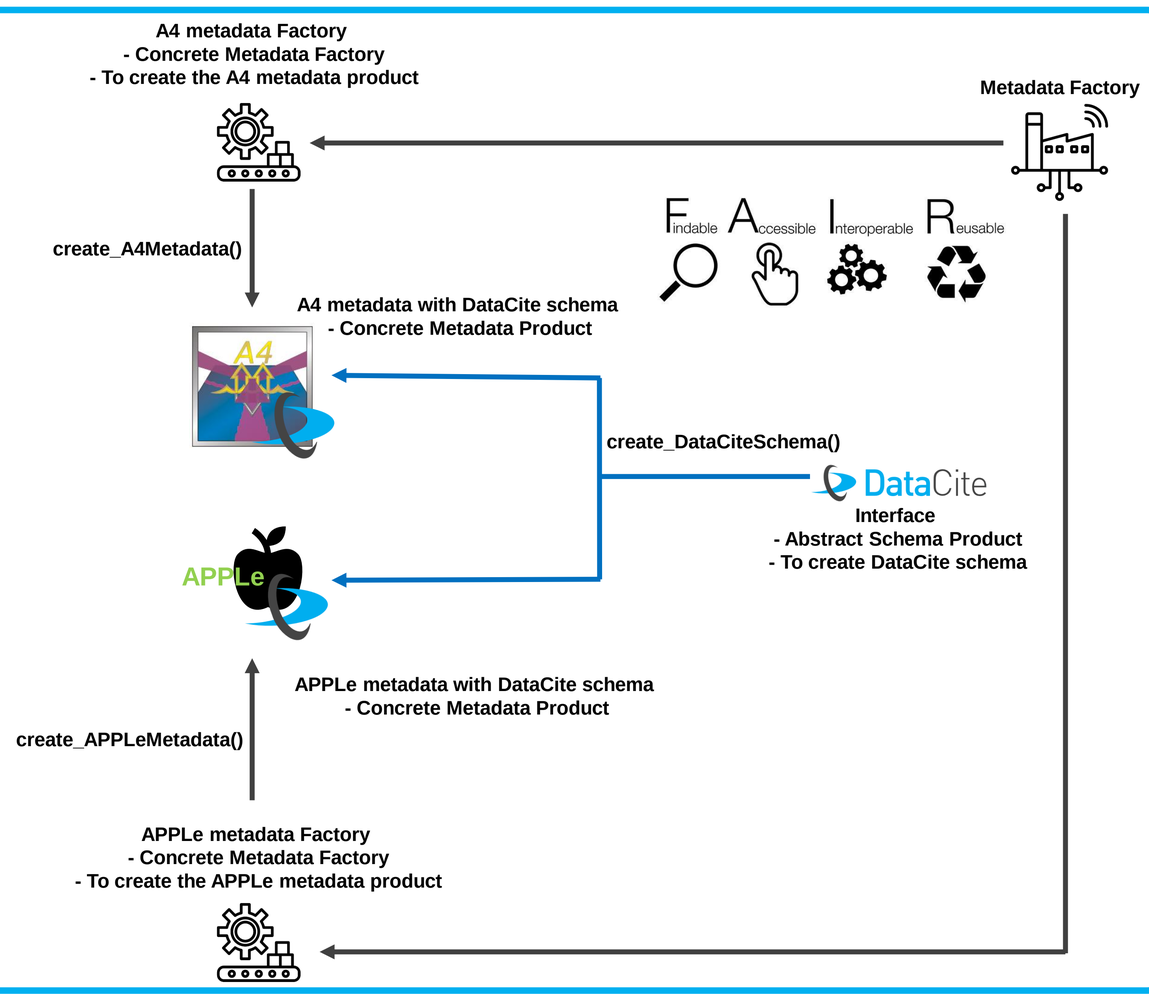
Add Metadata

leads to the concept of **Metadata Factory …**

Through the Metadata Factory Pattern, we define a framework that allows us to create related or dependent products that follow a general pattern.

The focus of PATOF is on making the data of A4 and APPLe (four future experiments ALPS II, PRIMA, P2, and LUXE) fully publicly available. Consider the experience from A4 and the recommendations from the DESY library; define metadata schemas and ways to implement the FAIR principles at an early stage in their planning, so that the FAIR Metadata Factory is ready to be used at an early state of APPLe experiments.





More concretely, the following deliverables are foreseen:

- D1 [30 Jun 2023]: Finalisation of the data collection and data availability, as well as metadata polishing for the A4 experiment (HI Mainz).
- D2 [31 Dec 2023]: Report on the A4 experiences and intermediate recommendations for PUNCH experiments (HI Mainz, DESY).
- D3 [29 Feb 2024]: Alpha version of the description of the “FAIR Metadata Factory” based on D2 (HI Mainz, DESY).
- D4 [30 Jun 2024]: Report on the application of D3 on APPLe (HI Mainz, DESY).
- D5 [31 Dec 2024]: “Cookbook” describing a general “FAIR Metadata Factory” based on D3 and reflecting the experience from D4 with the perspective of establishing a “living” cookbook (DESY).