HMC HELMHOLTZ Metadata Collaboration

HIM HELMHOLTZ Helmholtz-Institut Mainz

BONN

PUNCH 4NFDI

DESY.

GSI HZB Helmholtz Zentrum Berlin

IR of FAIR - Principles at the Instrument Level

Gerrit Günther[1] // Sebastian Baunack[2] // Luigi Capozza[2] // Oliver Freyermuth[3] //
Pau Gonzalez-Caminal[4] // Boxing Gou[5] // Johann Isaak[6] // Sven Karstensen[4] // Axel Lindner4 // Frank Maas[2] //
Oonagh Mannix[1] // Andrew Mistry[7] // Isabella Oceano[4] // Christiane Schneide[4] // Thomas Schörner-Sadenius[4] //
Kilian Schwarz[4] // Vivien Serve[1] // Lisa-Marie Stein[4] // Stefan Typel[6] // Malte C. Wilfert[2]

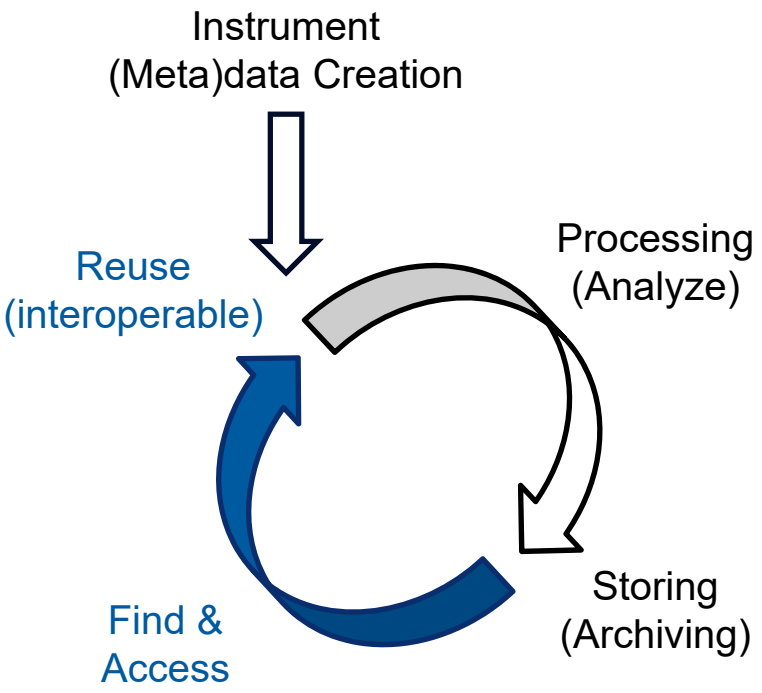
[1] Helmholtz-Zentrum Berlin für Materialien und Energie // [2] Helmholtz-Institut Mainz //
[3] Rheinische Friedrich-Wilhelms-Universität Bonn // [4] Deutsches Elektronen-Synchrotron //
[5] Institute of Modern Physics, Chinese Academy of Sciences // [6] Technische Universität Darmstadt //
[7] GSI Helmholtzzentrum für Schwerionenforschung

www.helmholtz-metadaten.de

FAIR Data Life Cycle

Closing the current (meta)data workflow by ensuring:

- **F**indability
- **A**ccessibility
- **I**nteroperability
- **R**eusability



FAIR Data Life Cycle

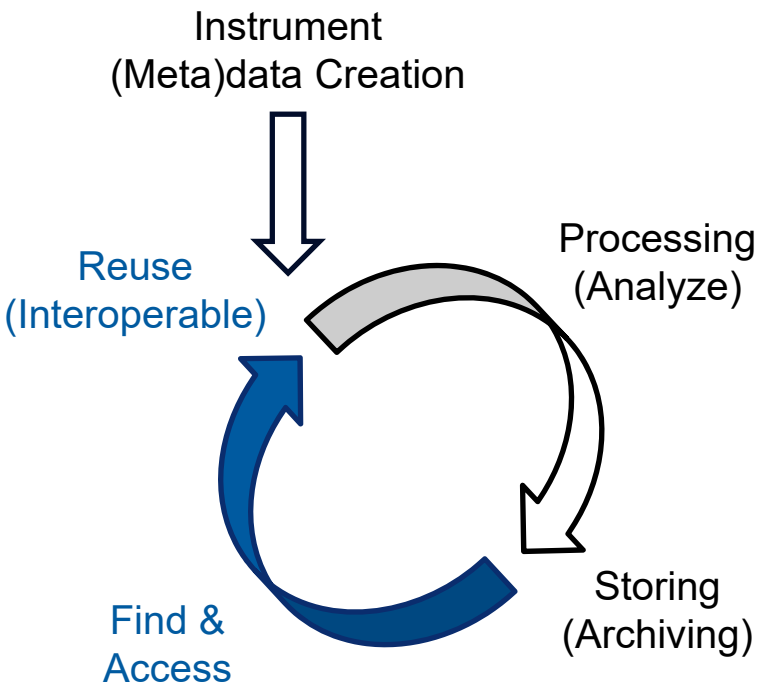
Closing the current (meta)data workflow by ensuring:

- **F**indability
- **A**ccessibility

addressed by repository

- Interoperability
- **R**eusability

basis for reuse at instrument level

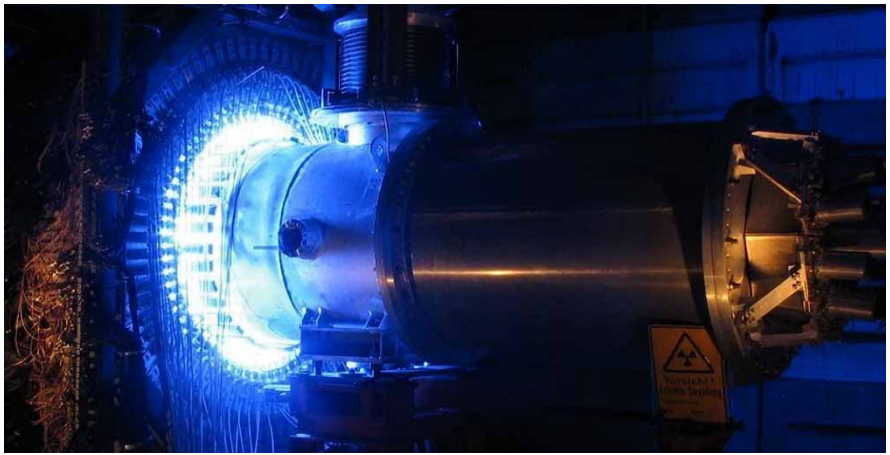


Interoperable and reusable raw (meta)data:

1. **Physical representation (file format)**
2. **Knowledge representation (of file content)**
3. **Context**

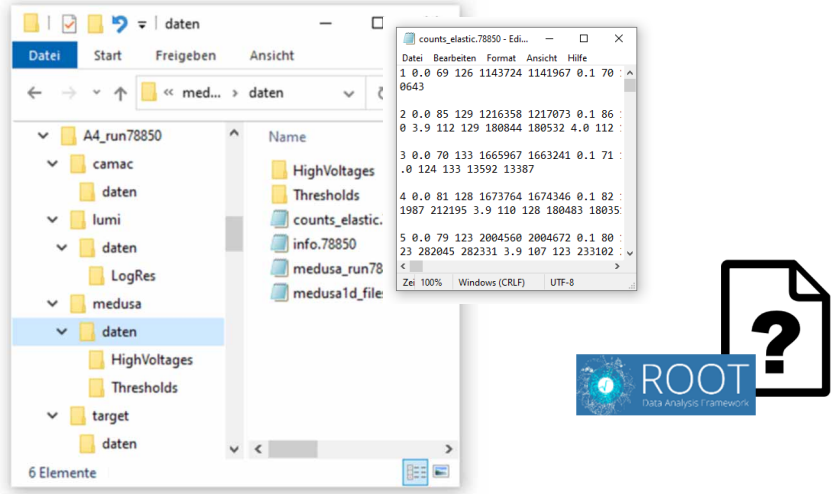
Use Cases

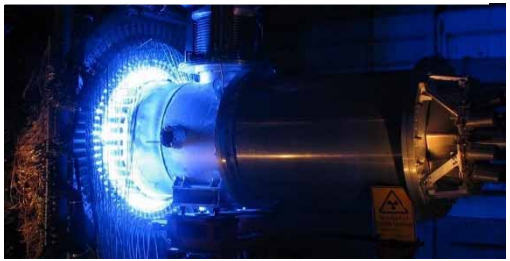
- A4 experiment @ MAMI
- Already dismounted (still producing results)



- Any Light Particle Search @ DESY
- In commissioning phase

Physical Representation: File Format – A4



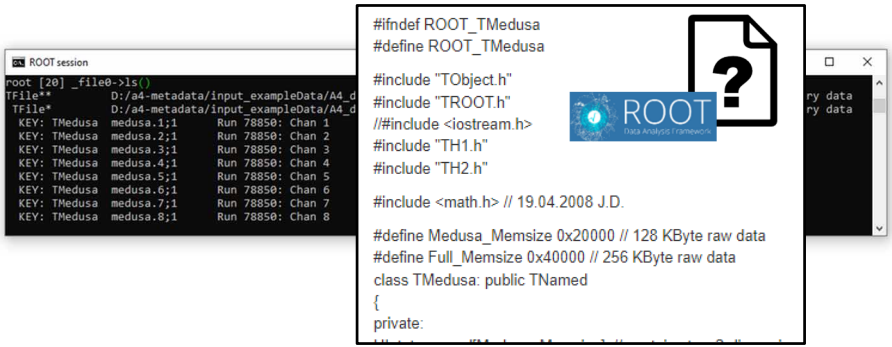


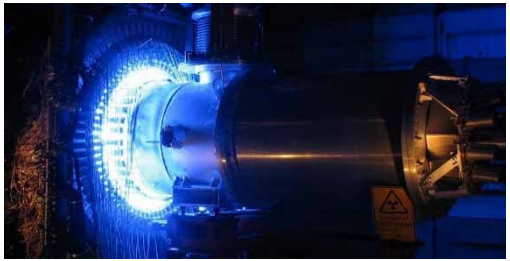
A4 raw instrument output:

- Nested folder structure (vague context)
- ASCII Files: Unformatted, minimum metadata, contain device output and preliminary results
  ⇨ XML (machine-readable)
- ROOT Files: community-specific standard, contain mainly detector data

Physical Representation: File Format – A4

Customized C++ class required to read ROOT files:



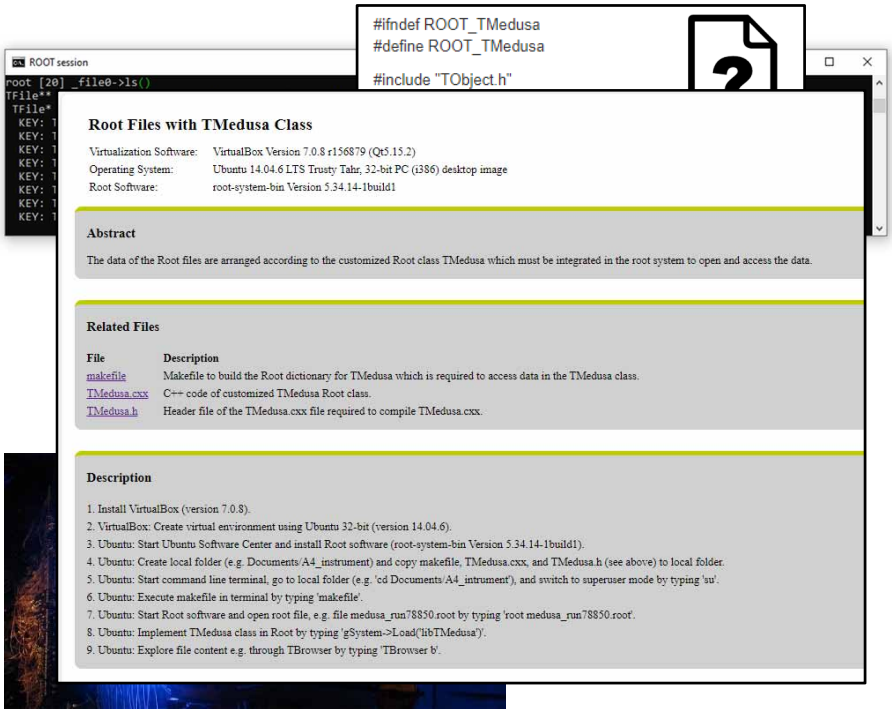


A4 raw instrument output:

- Nested folder structure (vague context)
- ASCII Files: Unformatted, minimum metadata, contain device output and preliminary results
  ⇨ XML (machine-readable)
- ROOT Files: community-specific standard, contain mainly detector data

Physical Representation: File Format – A4

Customized C++ class required to read ROOT files:



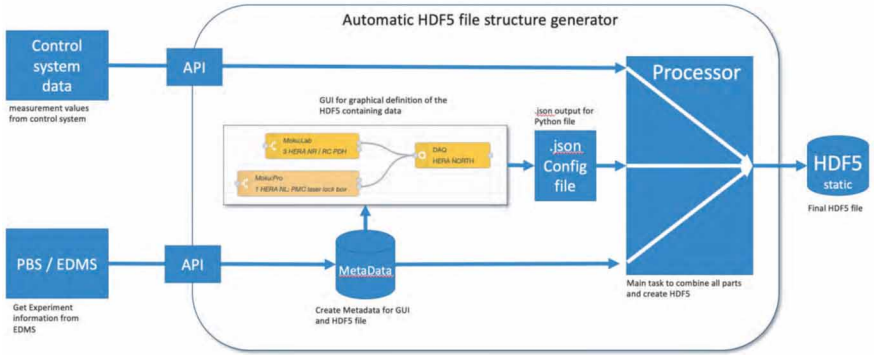
A4 raw instrument output:

- Nested folder structure (vague context)
- ASCII Files: Unformatted, minimum metadata, contain device output and preliminary results
  ⇨ XML (machine-readable)
- ROOT Files: community-specific standard, contain mainly detector data
  ⇨ manual HTML page to install customized class

Physical Representation: File Format – ALPS II

ALPS II raw instrument output:

- Stores (meta)data in local database
- Converted to HDF5 in certain intervals and ingested to globally accessible repository
  - ⇨ HDF5 is de-facto standard in photon and neutron science
  - ⇨ Various software to read HDF5 files



Knowledge Representation – A4

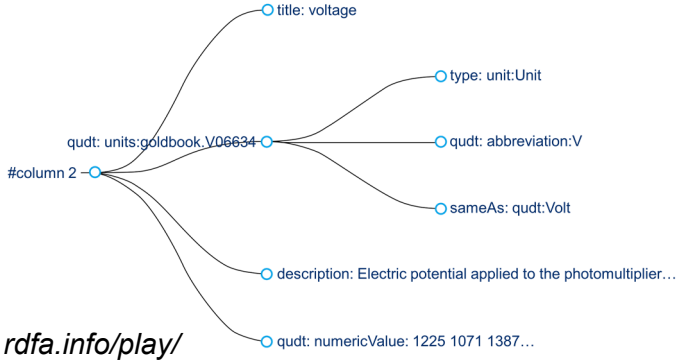
```
1 1225
2 1071
3 1387
4 1111
5 1251
6 1223
7 1250
8 1158
9 1093
10 1278
11 1111
12 1500
13 1304
14 1359
15 1028
16 1061
```

+

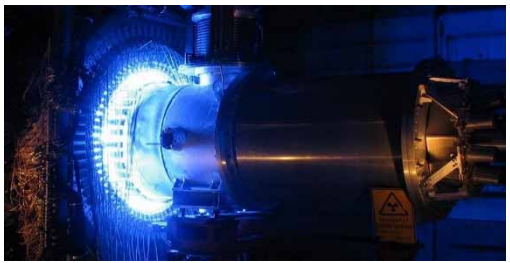
```
"col 0": {
  "name": "channel number",
  "data_type": "integer",
  "units": "dimensionless",
  "comment": "Numbering of de
},
"col 1": {
  "name": "voltage",
  "data_type": "integer",
  "units": "Volt",
  "comment": "Electric potent
}
```

=

```
−<column2 about="#column2">
   <name type="xsd:string" property="dc:title">voltage</name>
  −<units typeOf="unit:Unit" resource="https://doi.org/10.1351/goldbook.V0
     Volt
     <abbrev property="qudt:abbreviation" type="xsd:string">V</abbrev>
     <rel rel="owl:sameAs" resource="qudt:Volt"/>
   </units>
  −<comment type="xsd:string" property="dc:description">
     Electric potential applied to the photomultiplier (PMT) coupled to the cen
   </comment>
   <meanValue>1250.3649706457925</meanValue>
   <rangeBottom>987.0</rangeBottom>
   <rangeTop>1500.0</rangeTop>
  −<dataset property="qudt:numericValue" type="xsd:integer">
     <value>1225</value>
     <value>1071</value>
     <value>1387</value>
```



ASCII *Metadata enrichment* *Enriched, machine-readable XML* *rdfa.info/play/*



- ROOT files are to some extend self-describing
- ASCII files are neither human nor machine readable:
  - ⇨ Convert to human/machine-readable xml format
  - ⇨ Add metadata through json configuration file (e. g. units)
  - ⇨ RDF provide machine-readable definitions (e.g. using ontology qudt)

Knowledge Representation – ALPS II

NeXus Definition Language (NXDL):

- Standard in photon and neutron science to store experimental data
- Hierarchical, defined structure (definition)
- Persistent web description (definition, URL)

⇨ Human- and machine-readable

3.3.1.17. NXentry

Status:

base class, extends NXobject

Description:

(**required**) NXentry describes the measurement.

The top-level NeXus group which contains all the data and associated information tha
surement. It is mandatory that there is at least one group of this type in the NeXus file

Symbols:

No symbol table

Groups cited:

NXcollection, NXdata, NXinstrument, NXmonitor, NXnote, NXparameters, NXprocess, NX
NXuser

Structure:

@default: (optional) NX_CHAR

Declares which NXdata group contains the data to be shown by default. It is used to resolve ambiguity when one NXdata group exists. The value names a child group. If that group itself has a `default` attribute, continue this chain until an NXdata group is reached.

For more information about how NeXus identifies the de
section.

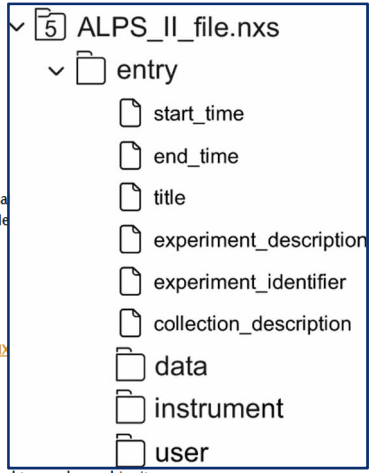
@IDF_Version: (optional) NX_CHAR

ISIS Muon IDF_Version

title: (optional) NX_CHAR

Extended title for entry

experiment_identifier: (optional) NX_CHAR



Context – A4

A4 dataset

Boxing Gou, Frank Maas

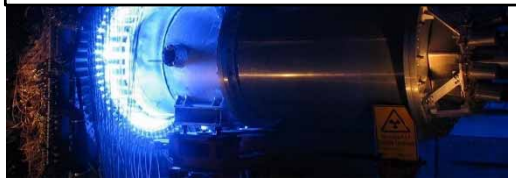
Abstract

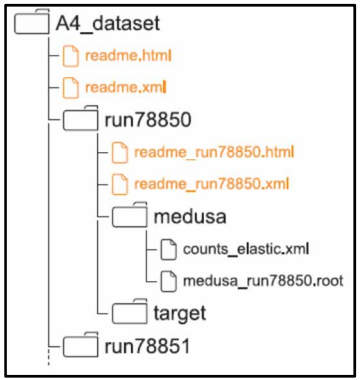
We report on a new measurement of the beam transverse single spin asymmetry in electron-proton elastic scattering at five beam energies from 315.1 to 1508.4 MeV and at a scattering angle of 30 < I < 40. The covered Q2 values are 0.032, 0.057, 0.082, 0.218, 0.613 (GeV/c)2. The measurement clearly indicates significant inelastic contributions to the two-photon-exchange (TPE) amplitude in the low-Q2 kinematic region. No theoretical calculation is able to reproduce our result. Comparison with a calculation based on unitarity, which only takes into account elastic and lN inelastic intermediate states, suggests that there are other inelastic intermediate states such as N, K, and N. Covering a wide energy range, our new high-precision data provide a benchmark to study those intermediate states.

Included Data Sets

File	Description
logbook\elog_beam191010.html	Electronic Lab-Notebook containing automatic and manual entries.
run78850\README_run78850.html	Measurement with Foerster: 18015, run time: 300s
run78871\README_run78871.html	Measurement with Foerster: 18017, run time: 500s
A4DataMeaning.pdf	Description of the files of a measurement and their content.

Note: Files of the data set were processed with the data converter producing the log file: _conversion_report\report_data_converter.html.



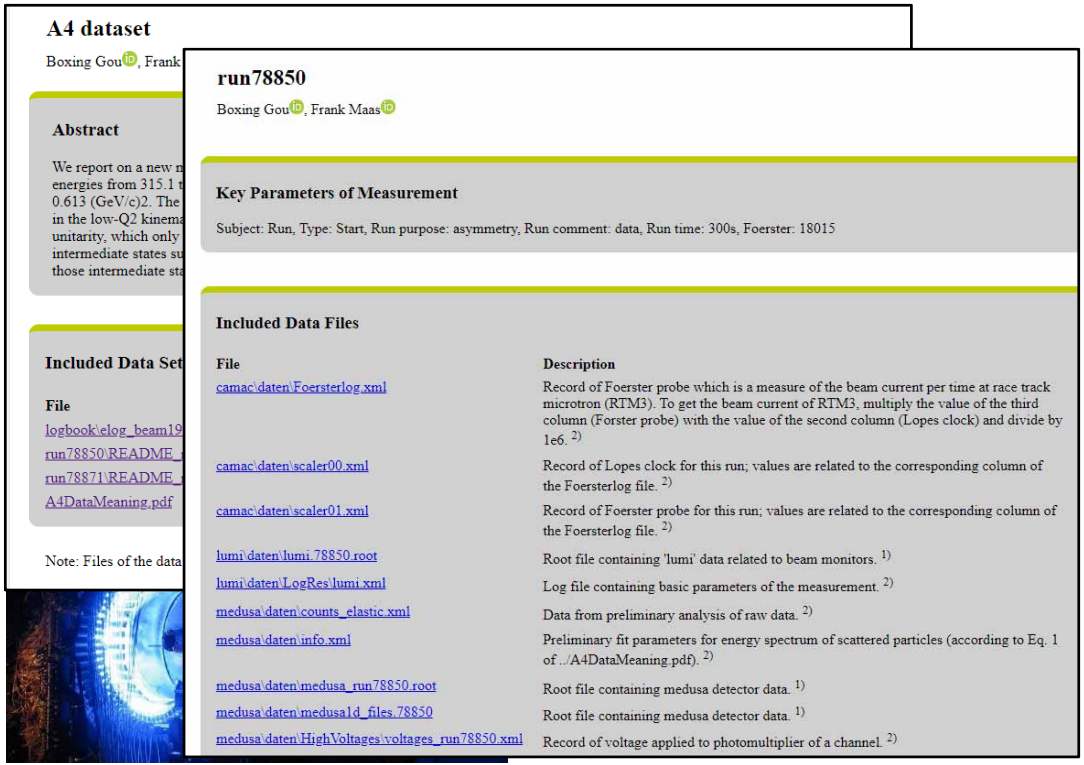


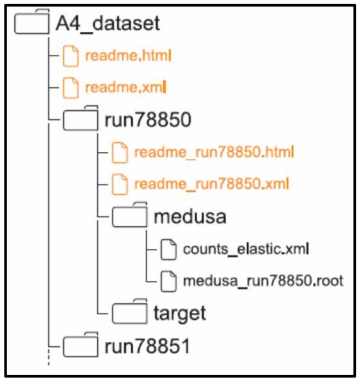
Create Context

- Additional Readme HTML file

  ⇨ Human-friendly, browsable

Context – A4



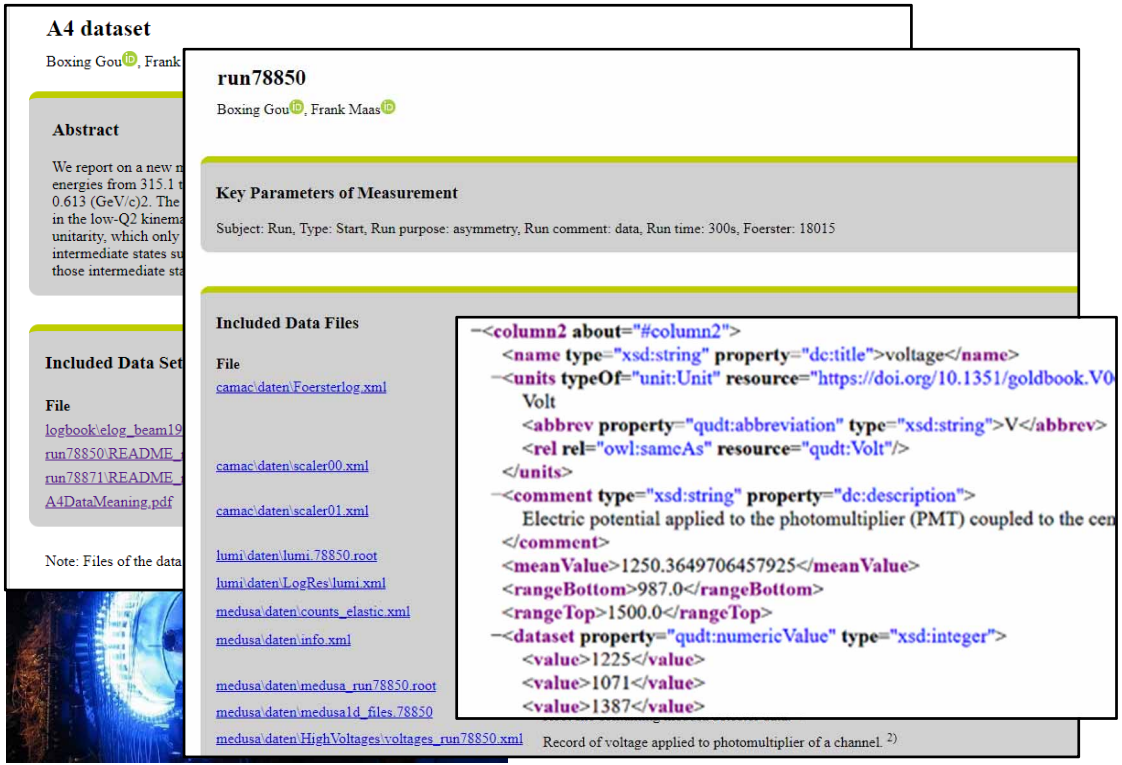


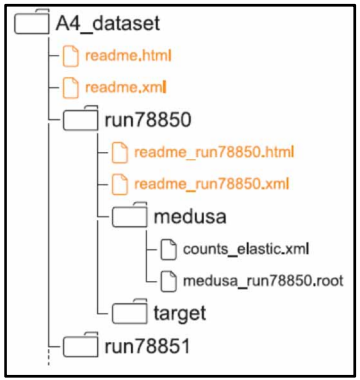
Create Context

- Additional Readme HTML file

  ⇨ Human-friendly, browsable

Context – A4



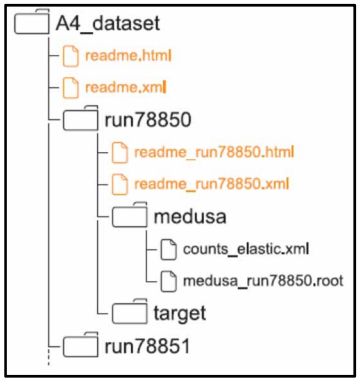


Create Context

- Additional Readme HTML file

  ⇨ Human-friendly, browsable

Context – A4



Create Context

- Additional Readme HTML file

  ⇨ Human-friendly, browsable

- Separate Readme XML file

  ⇨ Using DataCite (machine-readable)

Context – ALPS II

NeXus Definition Language (NXDL) contains pre-defined terms providing high-level information, such as

- Start/end time of measurement, duration
- Title, experiment description
- Citations to related publications
- Contacts to experimental team (user)

⇨ Human- and machine-readable

3.3.1.17. NXentry

Status:

base class, extends NXobject

Description:

(required) NXentry describes the measurement.

The top-level NeXus group which contains all the data and associated information
surement. It is mandatory that there is at least one group of this type in the NeXu

Symbols:

No symbol table

Groups cited:

NXcollection, NXdata, NXinstrument, NXmonitor, NXnote, NXparameters, NXproces
NXuser

Structure:

@default: (optional) NX_CHAR

Declares which NXdata group contains the data to be shown by default. It is used to resolve ambiguity when one NXdata group exists. The value names a child group. If that group itself has a `default` attribute, continue this chain until an NXdata group is reached.

For more information about how NeXus identifies the de
section.

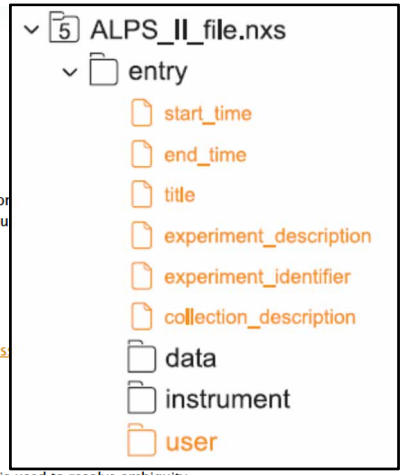
@IDF_Version: (optional) NX_CHAR

ISIS Muon IDF_Version

title: (optional) NX_CHAR

Extended title for entry

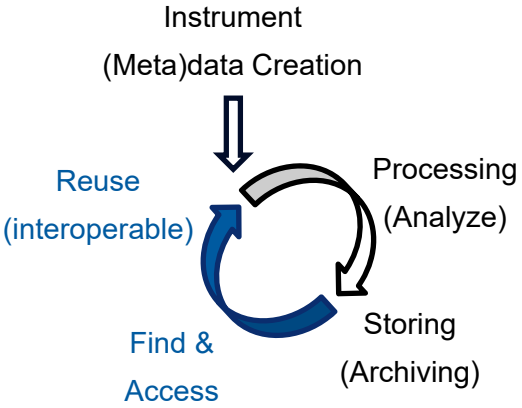
experiment_identifier: (optional) NX_CHAR



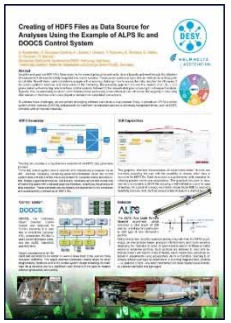
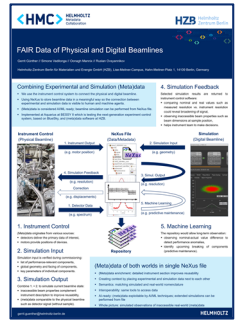
Summary

FAIRness of raw (meta)data can be improved on the instrument control level:

- Physical representation (file format):
  - open, readable
  - provide manual for reading if required (e. g. C++ files for Root classes)
- Knowledge representation
  - Use existing Standard (e.g. DataCite)
  - Adopting related standard (NeXus for ALPS II)
  - Create your own logic (e. g. through RDF)
- Context:
  - Using schema (e. g. NeXus terms)
  - Additional Readme file



Poster ALPS II



Poster FAIR Data