

IR of FAIR - Principles at the Instrument Level

Gerrit Günther¹ // Sebastian Baunack² // Luigi Capozza² // Oliver Freyermuth³ //

Pau Gonzalez-Caminal⁴ // Boxing Gou⁵ // Johann Isaak⁶ // Sven Karstensen⁴ // Axel Lindner⁴ // Frank Maas² //

Oonagh Mannix¹ // Andrew Mistry⁷ // Isabella Oceano⁴ // Christiane Schneide⁴ // Thomas Schörner-Sadenius⁴ //

Kilian Schwarz⁴ // Vivien Serve¹ // Lisa-Marie Stein⁴ // Stefan Typel⁶ // Malte C. Wilfert²

¹ Helmholtz-Zentrum Berlin für Materialien und Energie // ² Helmholtz-Institut Mainz //

³ Rheinische Friedrich-Wilhelms-Universität Bonn // ⁴ Deutsches Elektronen-Synchrotron //

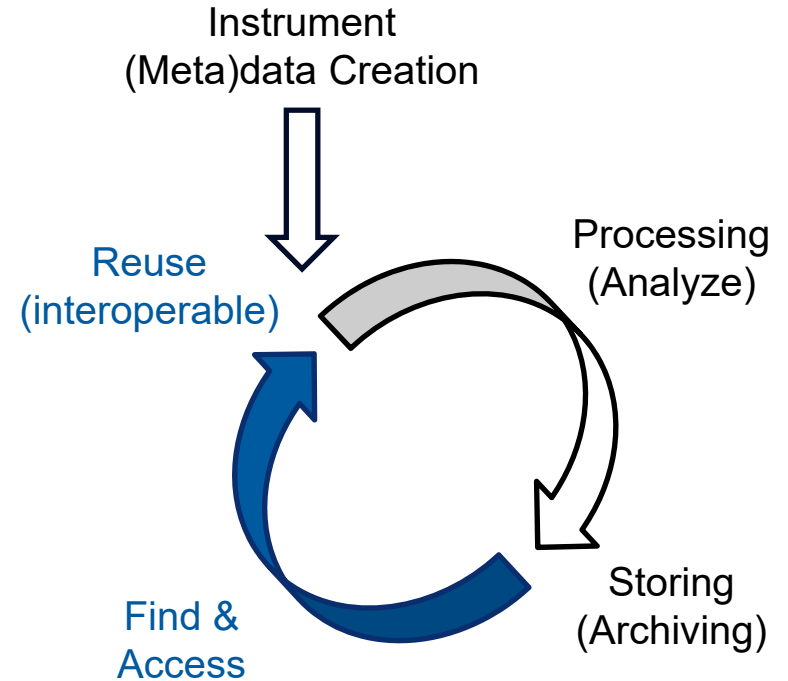
⁵ Institute of Modern Physics, Chinese Academy of Sciences // ⁶ Technische Universität Darmstadt //

⁷ GSI Helmholtzzentrum für Schwerionenforschung

FAIR Data Life Cycle

Closing the current (meta)data workflow by ensuring:

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



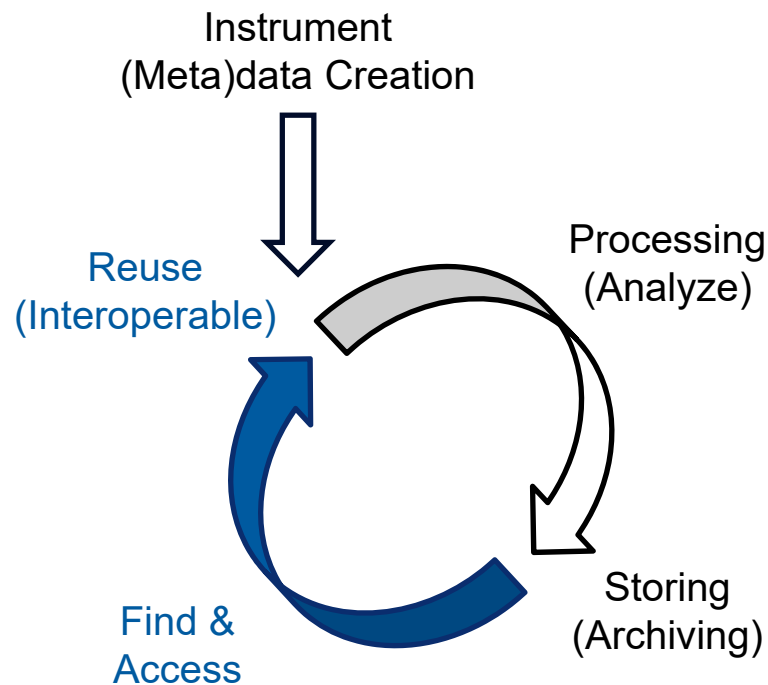
FAIR Data Life Cycle

Closing the current (meta)data workflow by ensuring:

- **Findability**
 - **Accessibility**
 - **Interoperability**
 - **Reusability**
- } addressed by repository
- } basis for reuse at instrument level

Interoperable and reusable raw (meta)data:

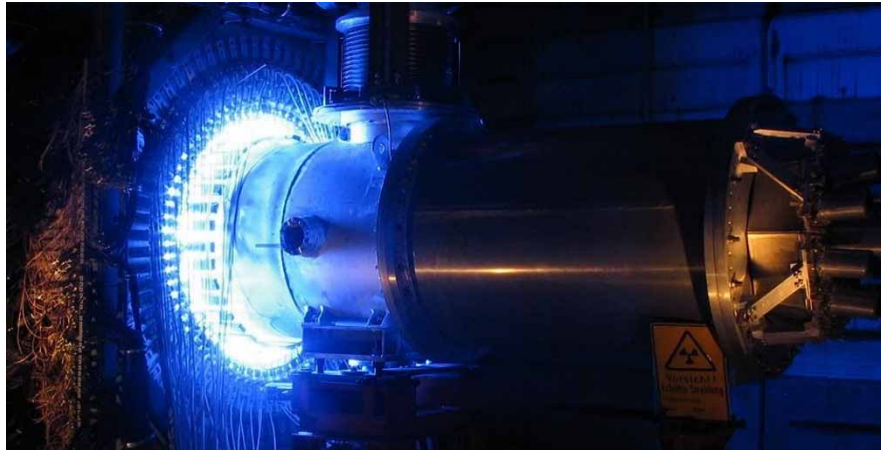
- 1. Physical representation (file format)**
- 2. Knowledge representation (of file content)**
- 3. Context**



Use Cases



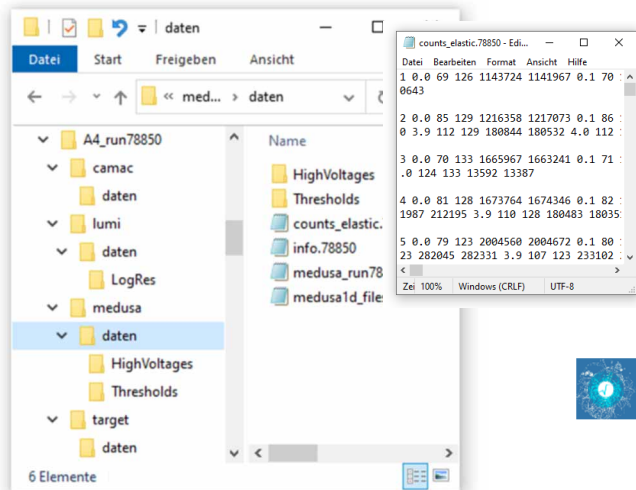
- A4 experiment @ MAMI
- Already dismantled (still producing results)



- Any Light Particle Search @ DESY
- In commissioning phase



Physical Representation: File Format – A4



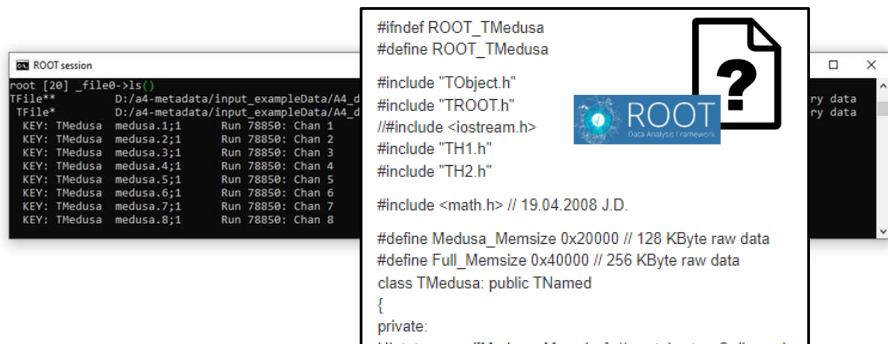
A4 raw instrument output:

- 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:



```
ROOT session
root [20] _file0->ls()
TFile*      D:/a4-metadata/input_exampleData/A4_d
TFile*      D:/a4-metadata/input_exampleData/A4_d
KEY: TMedusa  medusa.1;1  Run 78850: Chan 1
KEY: TMedusa  medusa.2;1  Run 78850: Chan 2
KEY: TMedusa  medusa.3;1  Run 78850: Chan 3
KEY: TMedusa  medusa.4;1  Run 78850: Chan 4
KEY: TMedusa  medusa.5;1  Run 78850: Chan 5
KEY: TMedusa  medusa.6;1  Run 78850: Chan 6
KEY: TMedusa  medusa.7;1  Run 78850: Chan 7
KEY: TMedusa  medusa.8;1  Run 78850: Chan 8
```

```
#ifndef ROOT_TMedusa
#define ROOT_TMedusa

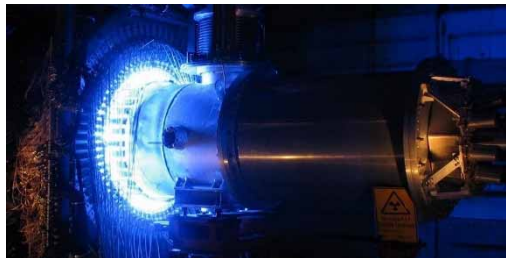
#include "TObject.h"
#include "TROOT.h"
#include <iostream.h>
#include "TH1.h"
#include "TH2.h"

#include <math.h> // 19.04.2008 J.D.

#define Medusa_Memsize 0x20000 // 128 KByte raw data
#define Full_Memsize 0x40000 // 256 KByte raw data
class TMedusa: public TObject
{
private:
```

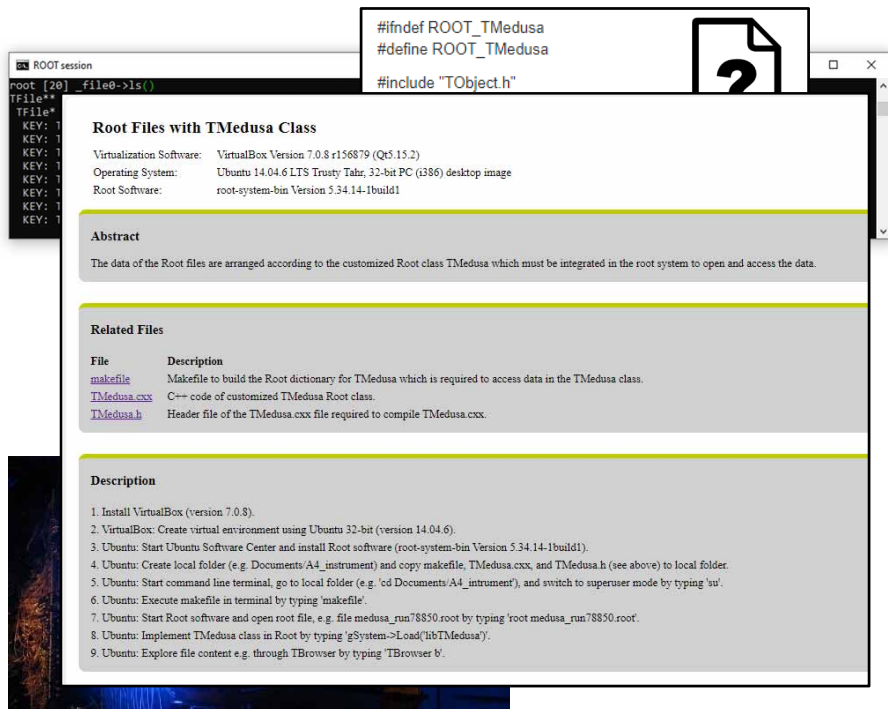
A4 raw instrument output:

- 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:



The image shows a ROOT session terminal window on the left and a web browser window on the right. The terminal window displays the command `file0->ls()` and the output `TFFile*`. The web browser window displays the documentation for the TMedusa class, which is a customized C++ class required to read ROOT files. The documentation includes a code snippet for the class definition, an abstract description, related files, and a detailed description of the installation and usage steps.

```
#ifndef ROOT_TMedusa
#define ROOT_TMedusa

#include "TObject.h"
```

Root Files with TMedusa Class

Virtualization Software: VirtualBox Version 7.0.8 r156879 (Qs5.15.2)
Operating System: Ubuntu 14.04.6 LTS Trusty Tahr, 32-bit PC (i386) desktop image
Root Software: root-system-bin Version 5.34.14-1build1

Abstract

The data of the Root files are arranged according to the customized Root class TMedusa which must be integrated in the root system to open and access the data.

Related Files

File	Description
makefile	Makefile to build the Root dictionary for TMedusa which is required to access data in the TMedusa class.
TMedusa.cxx	C++ code of customized TMedusa Root class.
TMedusa.h	Header file of the TMedusa.cxx file required to compile TMedusa.cxx.

Description

1. Install VirtualBox (version 7.0.8).
2. VirtualBox: Create virtual environment using Ubuntu 32-bit (version 14.04.6).
3. Ubuntu: Start Ubuntu Software Center and install root software (root-system-bin Version 5.34.14-1build1).
4. Ubuntu: Create local folder (e.g. Documents/A4_instrument) and copy makefile, TMedusa.cxx, and TMedusa.h (see above) to local folder.
5. Ubuntu: Start command line terminal, go to local folder (e.g. 'cd Documents/A4_instrument'), and switch to superuser mode by typing 'su'.
6. Ubuntu: Execute makefile in terminal by typing 'makefile'.
7. Ubuntu: Start Root software and open root file, e.g. file medusa_run78850.root by typing 'root medusa_run78850.root'.
8. Ubuntu: Implement TMedusa class in Root by typing 'gSystem->Load("libTMedusa")'.
9. Ubuntu: Explore file content e.g. through TBrower by typing 'TBrower b'.

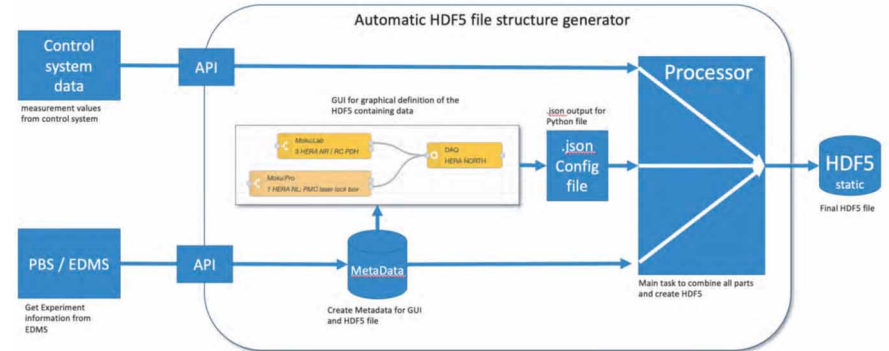
A4 raw instrument output:

- 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
}
    
```

ASCII

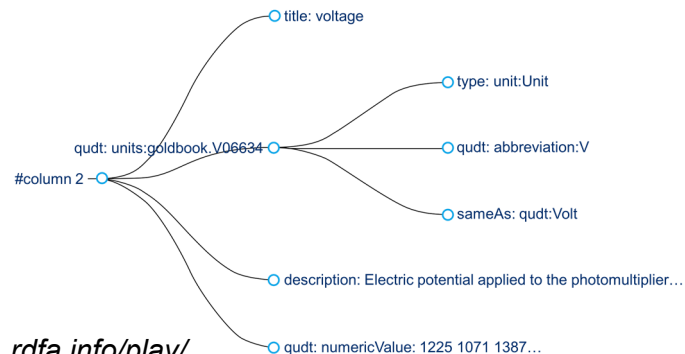
Metadata enrichment

=

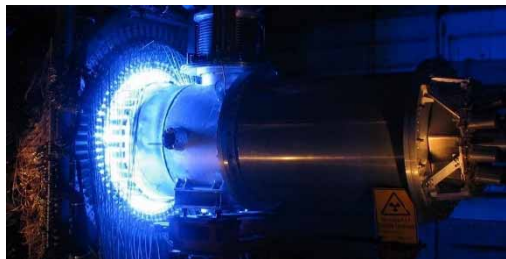
```

<-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>
  </dataset>
  </column2>
    
```

Enriched, machine-readable XML



rdfa.info/play/



- ROOT files are to some extent 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 that describes the measurement. 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](#), [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 default group, see the [NeXus User Manual](#).

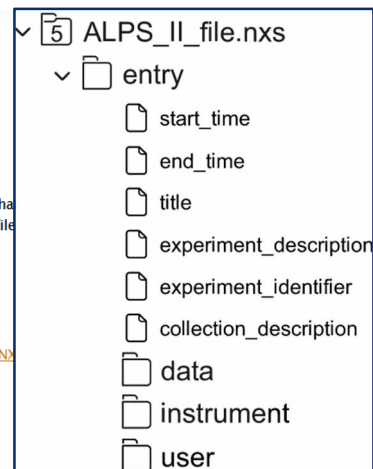
@IDF_Version: (optional) [NX_CHAR](#)

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 

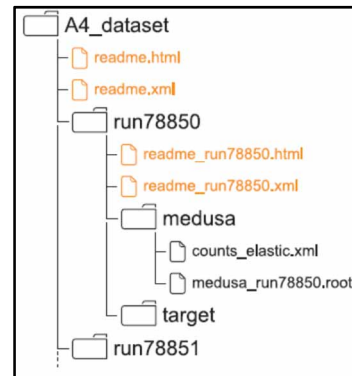
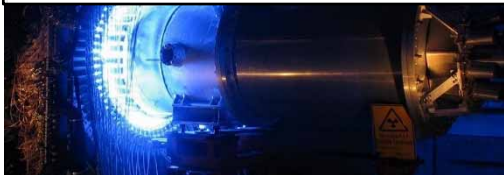
Abstract

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 Q^2 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- Q^2 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 $\mathbb{P}N$ inelastic intermediate states, suggests that there are other inelastic intermediate states such as N , K , and Λ . 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

A4 dataset

Boxing Gou¹, Frank

Abstract

We report on a new measurement of the beam energies from 315.1 to 0.613 (GeV/c)². The measurement is done in the low-Q² kinematic region, which only includes intermediate states such as those intermediate states

Included Data Set

File

[logbook/elog_beam19](#)
[run78850_README.html](#)
[run78871_README.html](#)
[A4DataMeaning.pdf](#)

Note: Files of the data



run78850

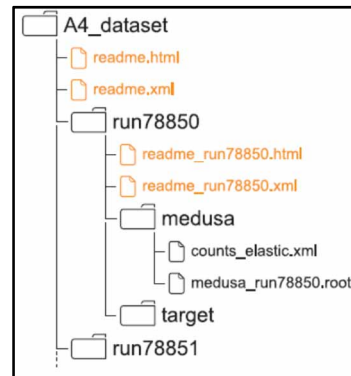
Boxing Gou¹, Frank Maas²

Key Parameters of Measurement

Subject: Run, Type: Start, Run purpose: asymmetry, Run comment: data, Run time: 300s, Foerster: 18015

Included Data Files

File	Description
camac/daten/Foersterlog.xml	Record of Foerster probe which is a measure of the beam current per time at race track microtron (RTM3). To get the beam current of RTM3, multiply the value of the third column (Foerster probe) with the value of the second column (Lopes clock) and divide by $1e6$. ²⁾
camac/daten/scaler00.xml	Record of Lopes clock for this run; values are related to the corresponding column of the Foersterlog file. ²⁾
camac/daten/scaler01.xml	Record of Foerster probe for this run; values are related to the corresponding column of the Foersterlog file. ²⁾
lumi/daten/lumi_78850.root	Root file containing 'lumi' data related to beam monitors. ¹⁾
lumi/daten/LogRes/lumi.xml	Log file containing basic parameters of the measurement. ²⁾
medusa/daten/counts_elastic.xml	Data from preliminary analysis of raw data. ²⁾
medusa/daten/info.xml	Preliminary fit parameters for energy spectrum of scattered particles (according to Eq. 1 of ../A4DataMeaning.pdf). ²⁾
medusa/daten/medusa_run78850.root	Root file containing medusa detector data. ¹⁾
medusa/daten/medusaId_files_78850	Root file containing medusa detector data. ¹⁾
medusa/daten/HighVoltages/voltages_run78850.xml	Record of voltage applied to photomultiplier of a channel. ²⁾




Create Context

- Additional Readme HTML file
⇒ Human-friendly, browsable

Context – A4

A4 dataset

Boxing Gou , Frank

Abstract

We report on a new n...
energies from 315.1 t...
0.613 (GeV/c)². The...
in the low-Q² kinema...
unitarity, which only...
intermediate states su...
those intermediate sta...


Key Parameters of Measurement

Subject: Run, Type: Start, Run purpose: asymmetry, Run comment: data, Run time: 300s, Foerster: 18015

Included Data Files

File
camaci/daten/Foersterlog.xml
camaci/daten/scaler00.xml
camaci/daten/scaler01.xml
lumi/daten/lumi_78850.root
lumi/daten/LogRes/lumi.xml
medusa/daten/counts_elastic.xml
medusa/daten/info.xml
medusa/daten/medusa_run78850.root
medusa/daten/medusaId_files_78850
medusa/daten/HighVoltages/voltages_run78850.xml

Note: Files of the data

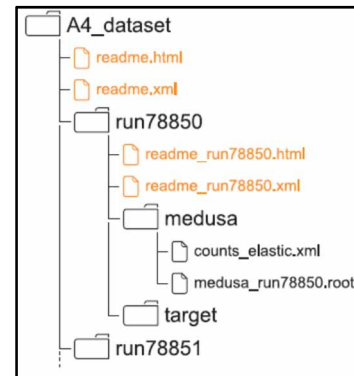


```

--<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>

```

Record of voltage applied to photomultiplier of a channel. ²⁾



Create Context

- Additional Readme HTML file
⇒ Human-friendly, browsable

Context – A4

A4 dataset

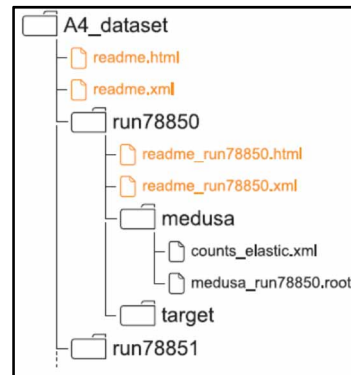
```
Boxing Gou</titles>
<title type="xsd:string" xml:lang="en-US">A4 dataset</title>
</titles>
<creators>
<creator>
<creatorName type="xsd:string" nameType="Personal">Gou, Boxing</creatorName>
<givenName type="xsd:string">Boxing</givenName>
<familyName type="xsd:string">Gou</familyName>
<nameIdentifier type="xsd:string" nameIdentifierScheme="ORCID">0000-0002-8918-3514</nameIdentifier>
<affiliation type="xsd:string">
Chinese Academy of Sciences, Institute of Modern Physics
</affiliation>
<affiliationIdentifier type="xsd:string" affiliationIdentifierScheme="ROR">https://ror.org/03x8rhq63</affiliationIdentifier>
</creator>
<creator>
<creatorName type="xsd:string" nameType="Personal">Maas, Frank</creatorName>
<givenName type="xsd:string">Frank</givenName>
<familyName type="xsd:string">Maas</familyName>
<nameIdentifier type="xsd:string" nameIdentifierScheme="ORCID">0000-0002-9271-1883</nameIdentifier>
<affiliation type="xsd:string">Helmholtz Institute Mainz</affiliation>
<affiliationIdentifier type="xsd:string" affiliationIdentifierScheme="ROR">https://ror.org/024thra40</affiliationIdentifier>
</creator>
</creators>
<contributors>
<contributor contributorType="DataCurator">
<creatorName type="xsd:string" nameType="Personal">Mannix, Oonagh</creatorName>
<givenName type="xsd:string">Oonagh</givenName>
<familyName type="xsd:string">Mannix</familyName>
<nameIdentifier type="xsd:string" nameIdentifierScheme="ORCID">0000-0003-0575-2853</nameIdentifier>
<affiliation type="xsd:string">
Helmholtz-Zentrum Berlin fuer Materialien und Energie
</affiliation>
<affiliationIdentifier type="xsd:string" affiliationIdentifierScheme="ROR">https://ror.org/02aj13c28</affiliationIdentifier>
</contributor>
</contributors>
<publisher type="xsd:string">GSI Helmholtzzentrum fuer Schwerionenforschung</publisher>
<publicationYear type="xsd:integer">2023</publicationYear>
<resourceTypeGeneral type="xsd:string">Dataset/Processed Data</resourceTypeGeneral>
<resourceType type="xsd:string">Dataset</resourceType>
```

Included

File

logbook/els
run78850.R
run78871.UR
A4DataMeta

Note: Files



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. 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](#), [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 default group, see the [NeXus User Guide](#).

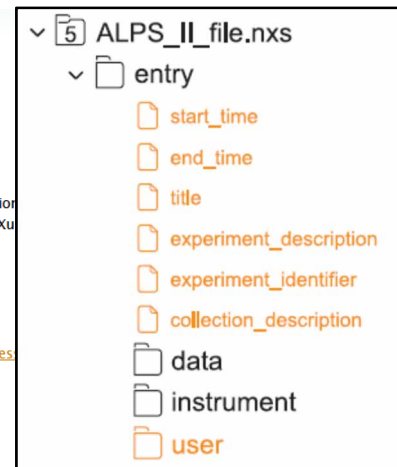
@IDF_Version: (optional) [NX_CHAR](#)

ISIS Muon IDF_Version

title: (optional) [NX_CHAR](#)

Extended title for entry

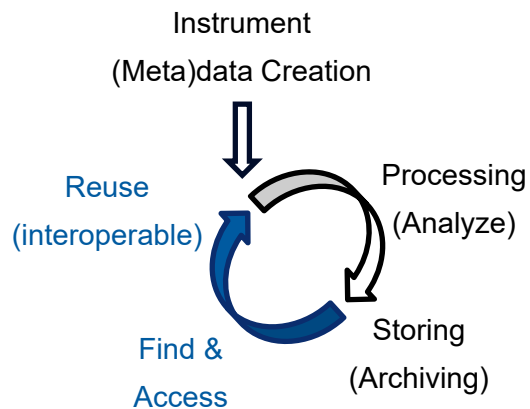
experiment_identifier: (optional) [NX_CHAR](#)



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

