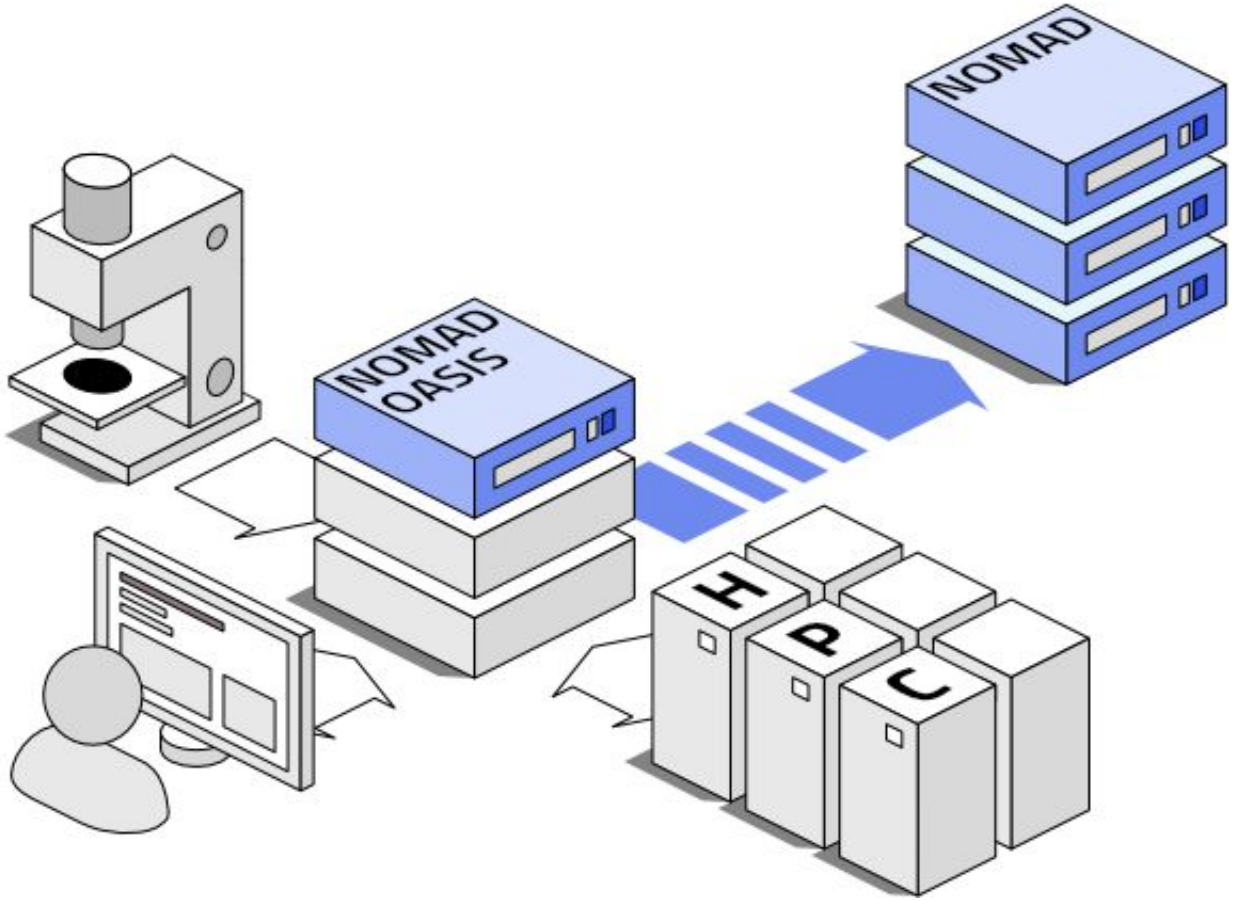


FAIRmat Tutorial 8:

Using NOMAD as an Electronic lab notebook (ELN) for FAIR data

Organized by FAIRmat Area A Synthesis

FAIRmat Area A: Synthesis

Why Structured Data?

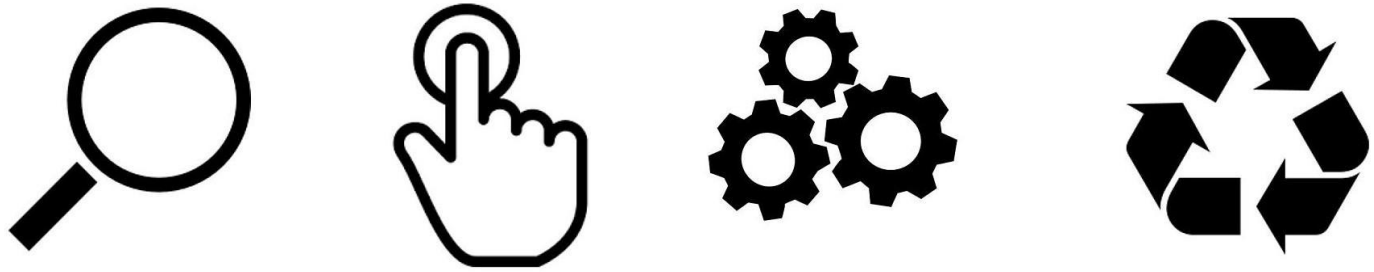
Structured Data

FAIRmat

Structured Data

Why Structured Data?

- the structure highlights relationships between data points (semantic **interoperability**)
- **machine-readable**
- ideal for classification, regression, and clustering with **AI**
- **searchable**
- **shareable** (using a community agreed structure)
- leads to data-driven decisions



FAIRmat

Research Data Management

Raw Data

ELN

Data Analysis

Data Modelling

FAIRmat

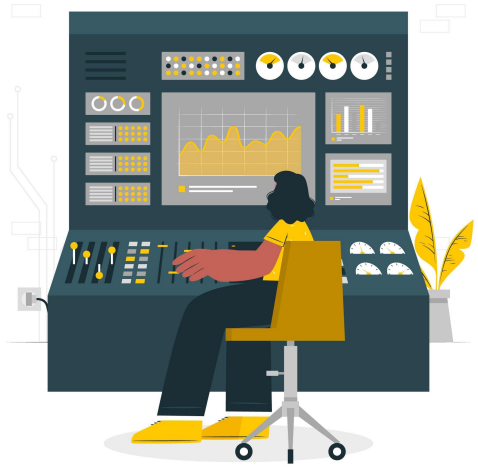
Research Data Management

Raw Data

ELN

Data Analysis

- Log files from instruments
- Recipe files from process software
- Spreadsheet files



https://de.freepik.com/vektoren-kostenlos/illustration-des-bedienfeldkonzepts_13662974.htm

Data Modelling

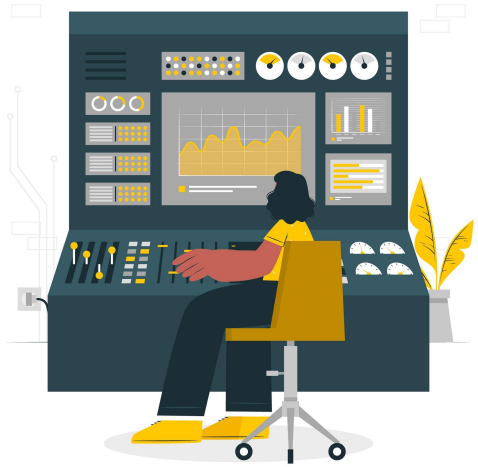
FAIRmat

Research Data Management

Data Modelling

Raw Data

- Log files from instruments
- Recipe files from process software
- Spreadsheet files



https://de.freepik.com/vektoren-kostenlos/illustration-des-bedienfeldkonzepts_13662974.htm

ELN

- Manually entered data and metadata

https://de.freepik.com/vektoren-kostenlos/laptop-mit-bildungssymbol-isoliert_11691038.htm

Data Analysis

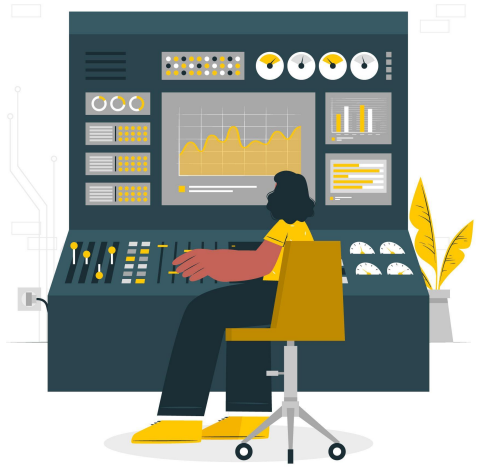
FAIRmat

Research Data Management

Data Modelling

Raw Data

- Log files from instruments
- Recipe files from process software
- Spreadsheet files



https://de.freepik.com/vektoren-kostenlos/illustration-des-bedienfeldkonzepts_13662974.htm

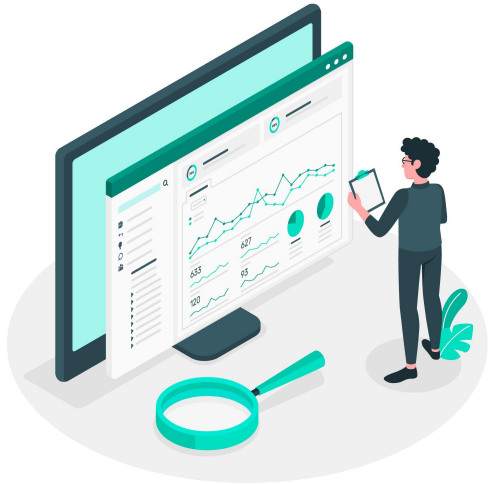
ELN

- Manually entered data and metadata

https://de.freepik.com/vektoren-kostenlos/laptop-mit-bildungssymbol-isoliert_11691038.htm

Data Analysis

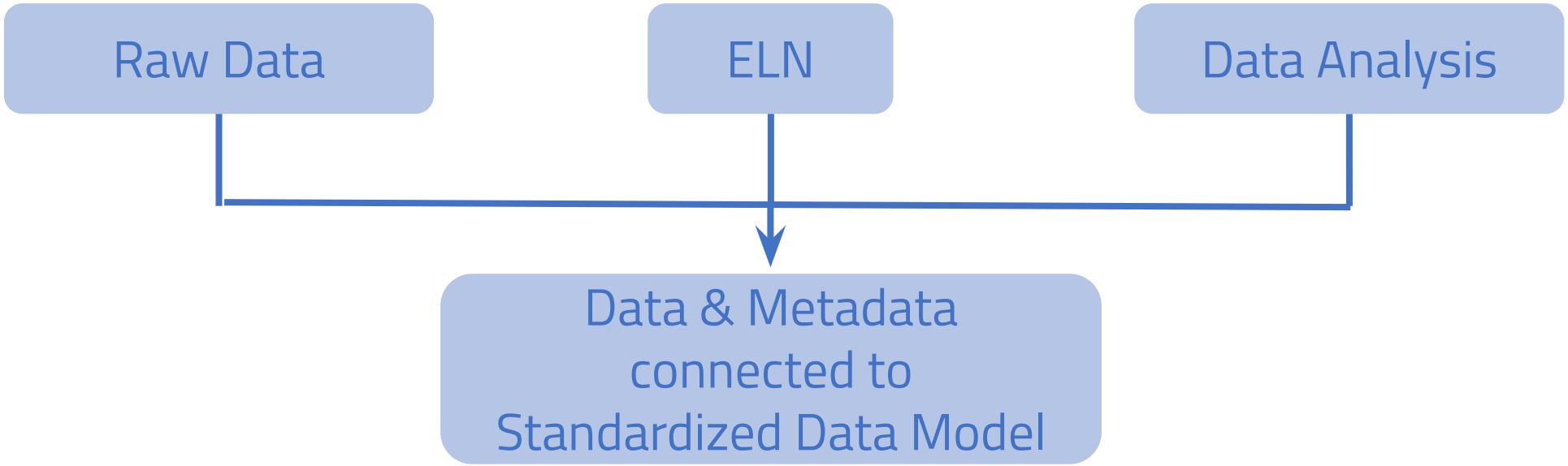
- Post processing software
- User-tailored scripts



https://de.freepik.com/vektoren-kostenlos/site-statistik-konzeptillustration_7140739.htm

FAIRmat

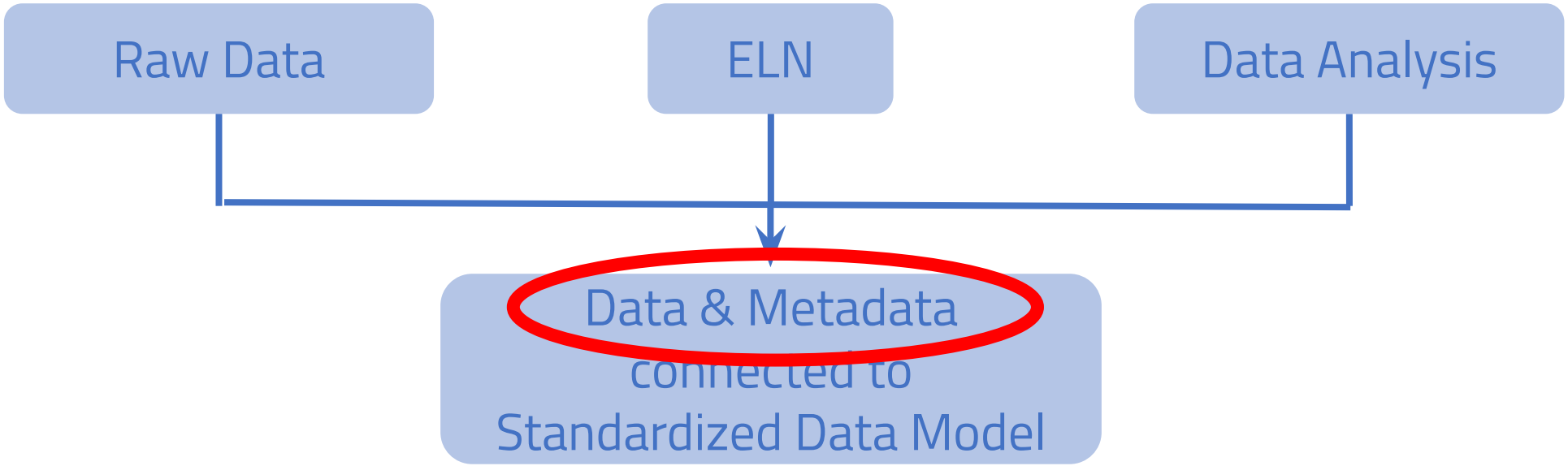
Research Data Management



Data Modelling

FAIRmat

Research Data Management

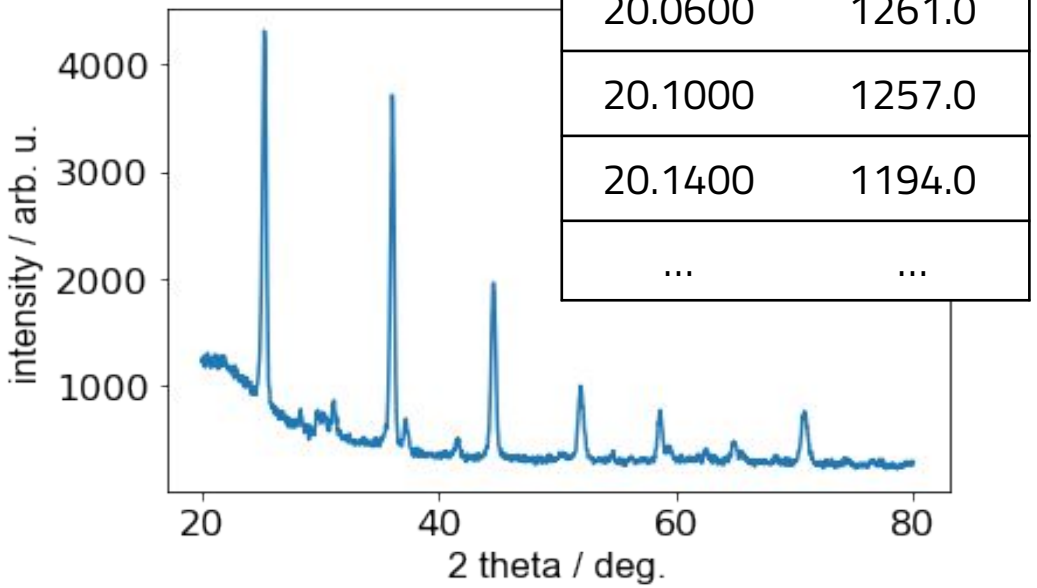


Data Modelling

FAIRmat

Data & Metadata

Data:
actual content of information



2 theta	intensity
20.0200	1224.0
20.0600	1261.0
20.1000	1257.0
20.1400	1194.0
...	...

Metadata

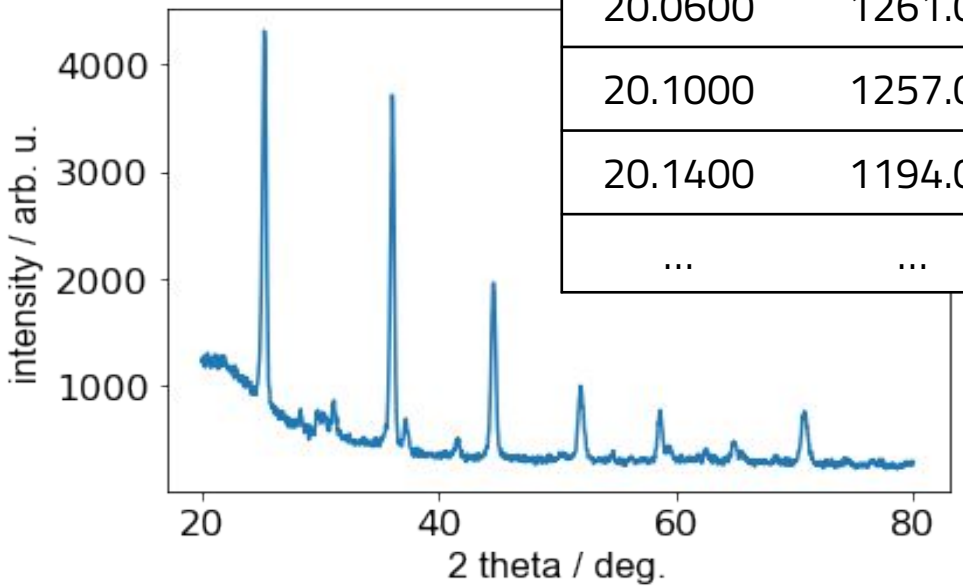
FAIRmat

Data & Metadata

Data:
actual content of information

Metadata:
data that provides information about other data

2 theta	intensity
20.0200	1224.0
20.0600	1261.0
20.1000	1257.0
20.1400	1194.0
...	...



Metadata

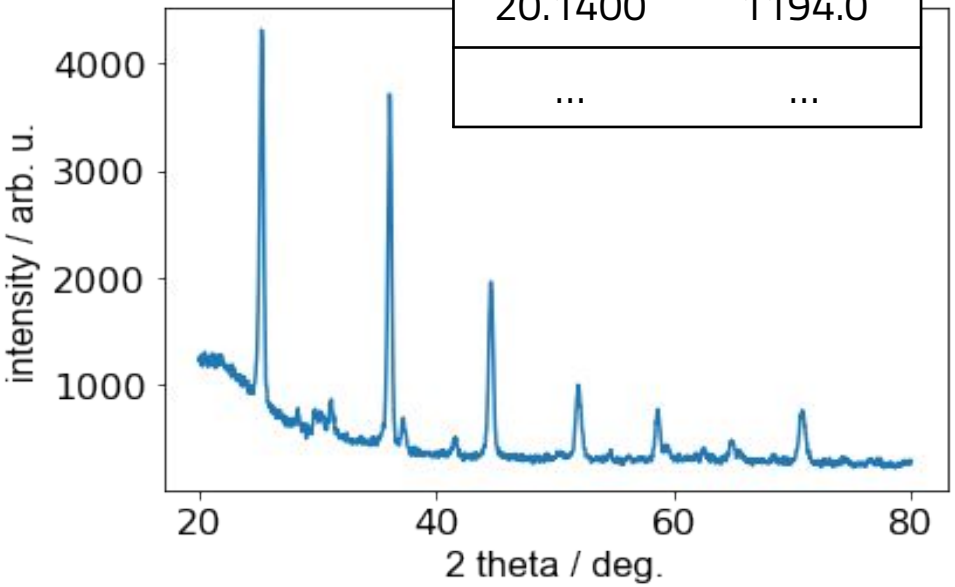
FAIRmat

Data & Metadata

Metadata

Data:
actual content of information

2 theta	intensity
20.0200	1224.0
20.0600	1261.0
20.1000	1257.0
20.1400	1194.0
...	...



Structural Metadata:
provides information about containers of data

```
XRD Measurement:
  2 theta:
    description: The 2-theta angle of the diffractogram.
    unit: °
    type: float
  intensity:
    description: The count at each 2-theta value.
    unit: dimensionless
    type: float
```

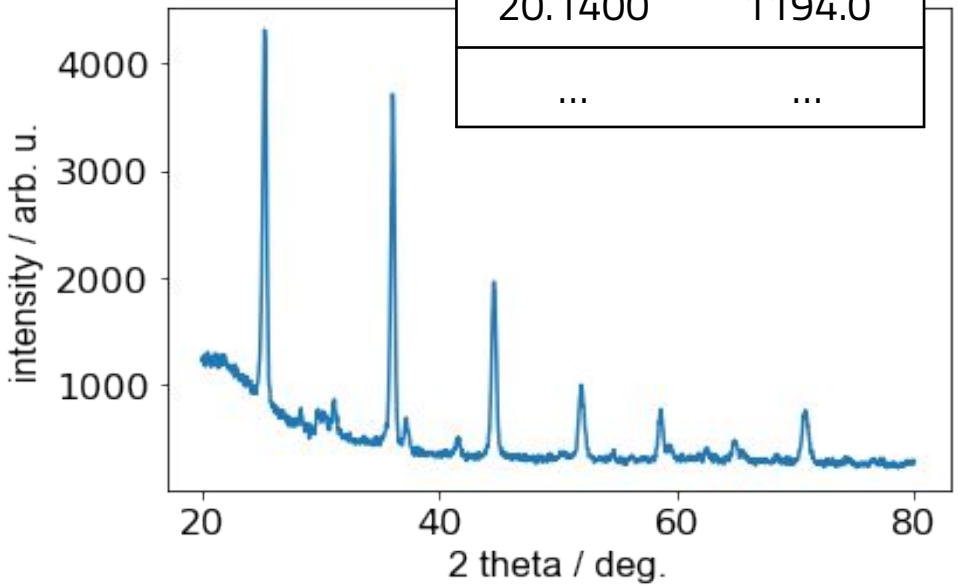
FAIRmat

Data & Metadata

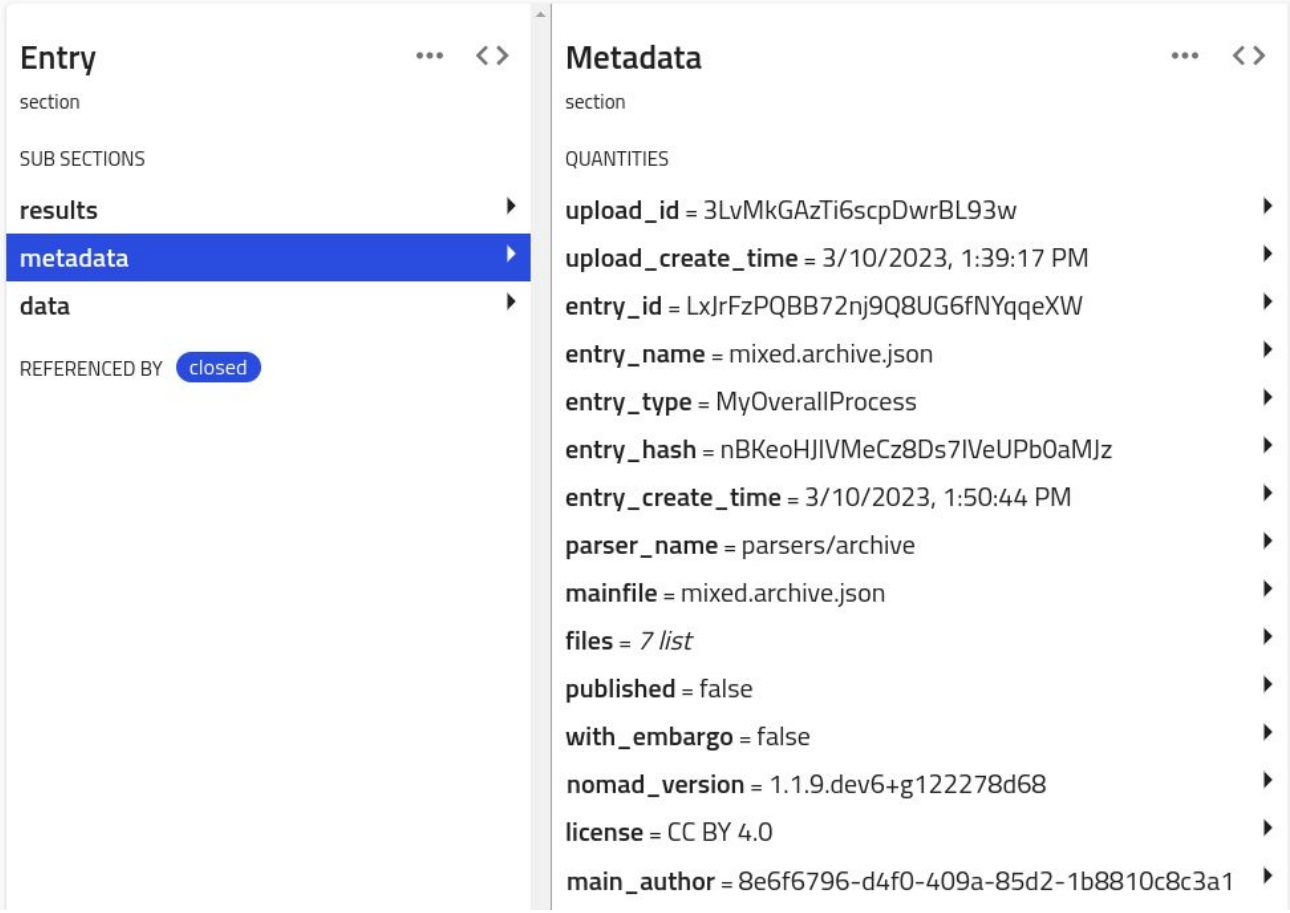
Metadata

Data:
actual content of information

2 theta	intensity
20.0200	1224.0
20.0600	1261.0
20.1000	1257.0
20.1400	1194.0
...	...



Structural Metadata:
provides information about containers of data



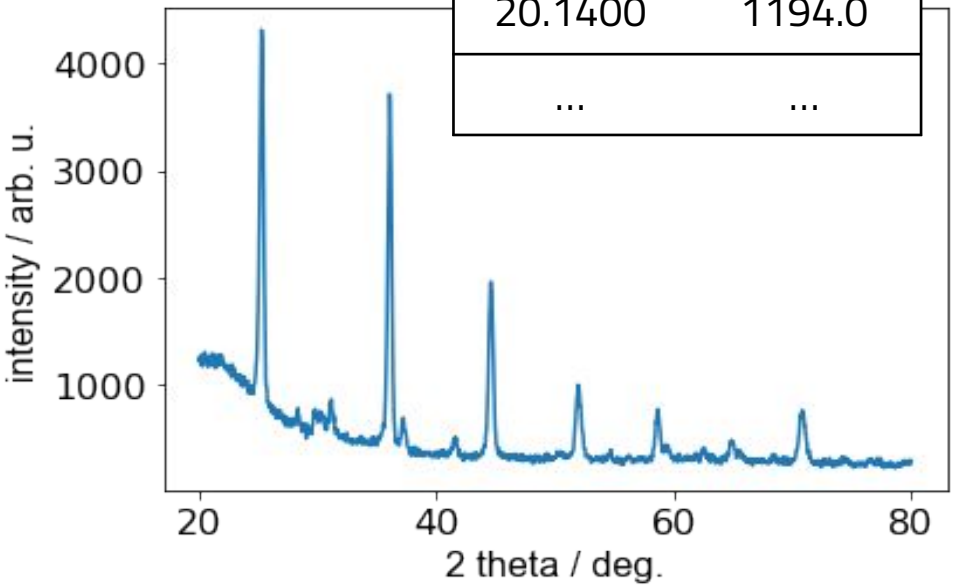
FAIRmat

Data & Metadata

Metadata

Data:
actual content of information

2 theta	intensity
20.0200	1224.0
20.0600	1261.0
20.1000	1257.0
20.1400	1194.0
...	...



Descriptive Metadata:
provides important context about data

Date of the experiment?

Temperature?

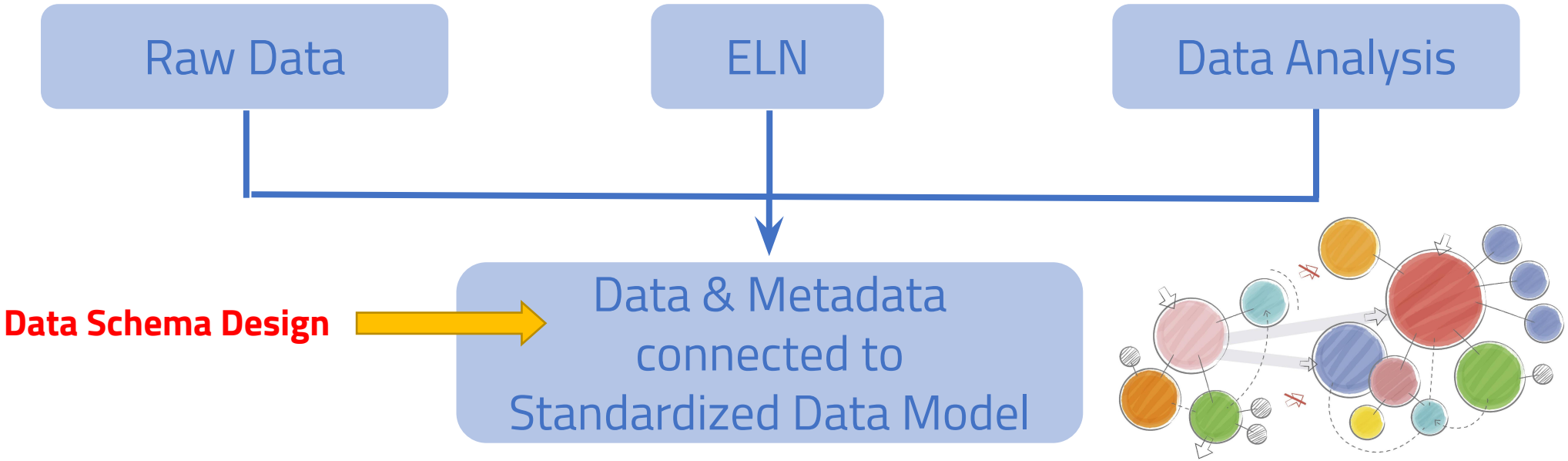
Instrument specifications?

Specimen specifications?

Humidity in the room?

FAIRmat

Research Data Management



https://de.freepik.com/vektoren-kostenlos/illustration-der-netzwerkfreigabe_2808109.htm

Research Data Management

Data Modelling

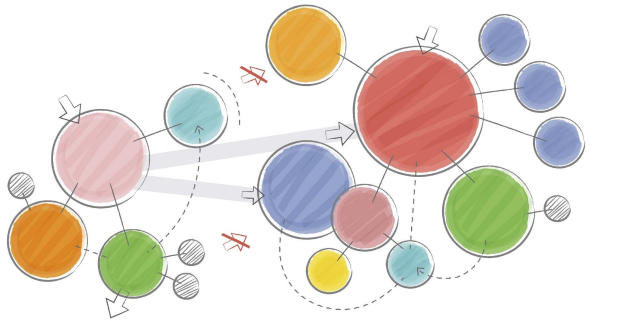
Raw Data

ELN

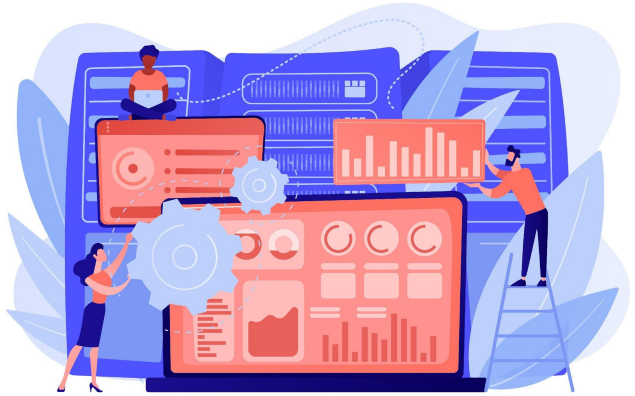
Data Analysis

Data Schema Design

Data & Metadata connected to Standardized Data Model



https://de.freepik.com/vektoren-kostenlos/illustration-der-netzwerkfreigabe_2808109.htm



https://de.freepik.com/vektoren-kostenlos/laptop-mit-datenvisualisierungssoftware-und-entwicklern-arbeiten-big-data-visualisierung-big-data-analyse-konzept-der-visualisierungssoftware_11667704.htm

Search ▪ Visualization ▪ Analysis

Implementation

FAIRmat

Research Data Management

Data Modelling

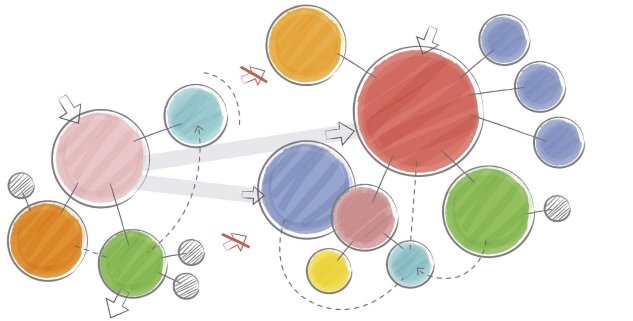
Raw Data

ELN

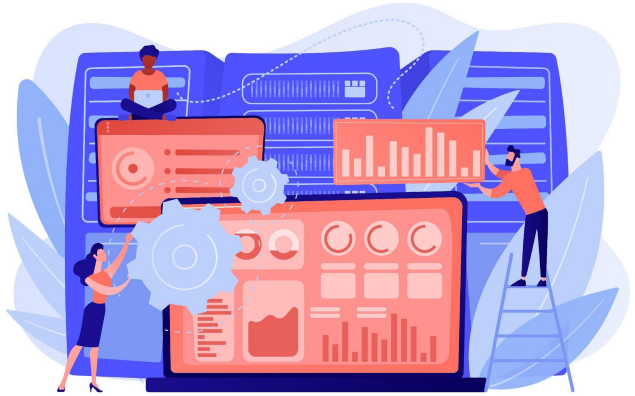
Data Analysis

Data Schema Design

Data & Metadata connected to Standardized Data Model



https://de.freepik.com/vektoren-kostenlos/illustration-der-netzwerkfreigabe_2808109.htm



https://de.freepik.com/vektoren-kostenlos/laptop-mit-datenvisualisierungssoftware-und-entwicklern-arbeiten-big-data-visualisierung-big-data-analyse-konzept-der-visualisierungssoftware_11667704.htm

Search ▪ Visualization ▪ Analysis

Implementation

NOMAD

FAIRmat follows a **bottom-up approach** that is driven by the needs of scientists and already enjoys strong support from the community

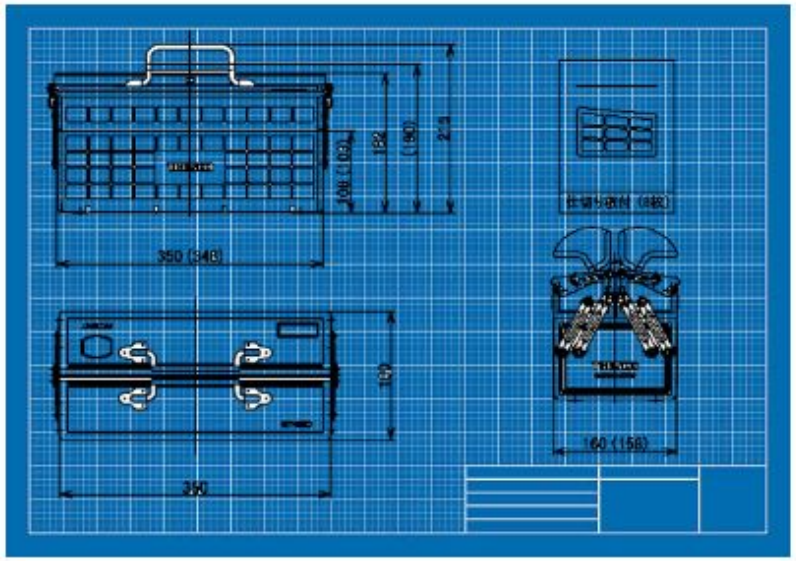
FAIRmat

Schema and Template Concepts

Data Modelling

Schema:

A formal description of data, data types, and data file structures, such as XML files.



The blueprint of a toolbox

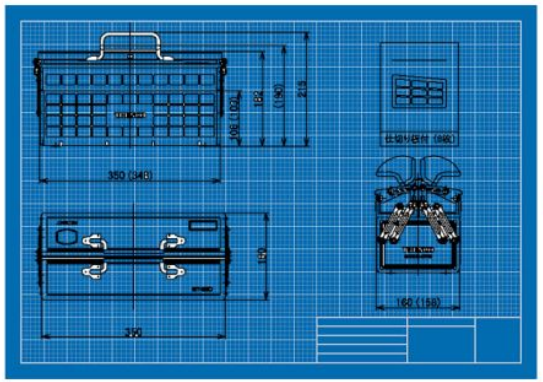
Template:

A physical object from which other objects are based or derived.

A toolbox made for a specific set of tools

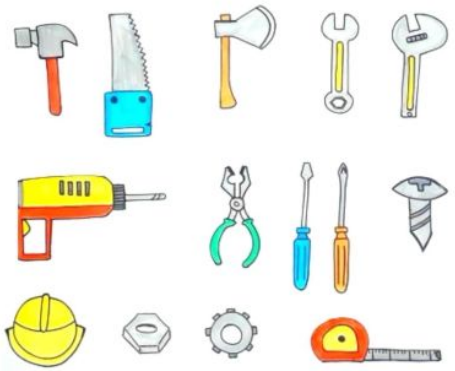
FAIRmat

Schema and Template Concepts



Schema

Template



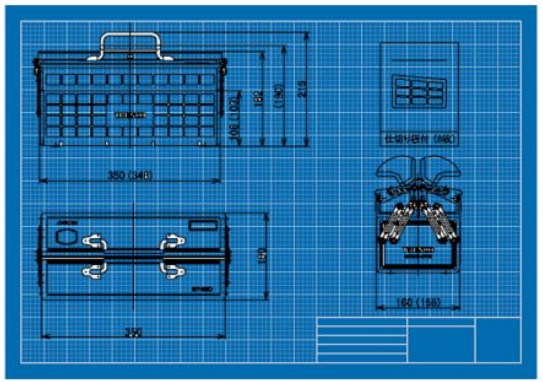
Data

Structured Archive File

Data Modelling

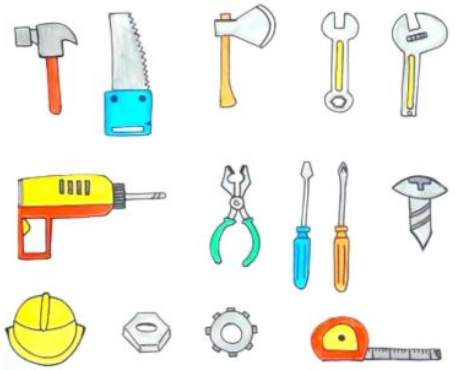
FAIRmat

Schema and Template Concepts



Schema

Template



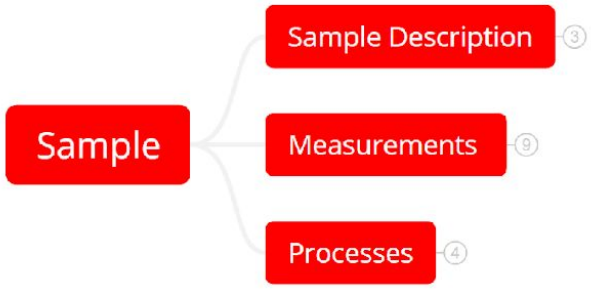
Data

Structured Archive File

Data Modelling

FAIRmat

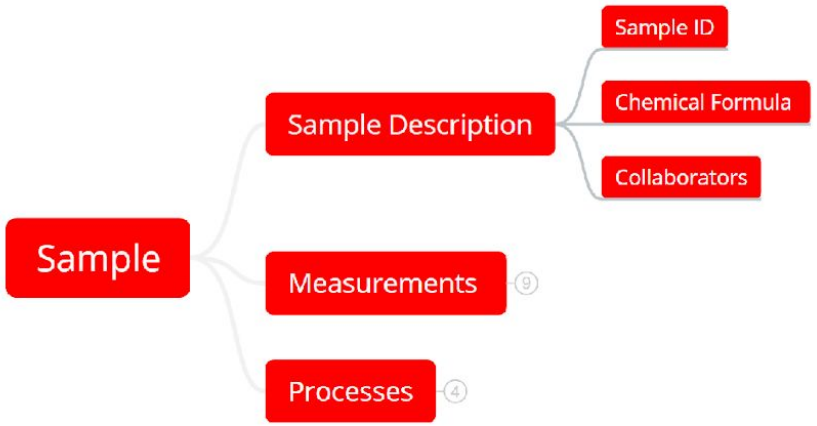
Hierarchical organization



Example of file formats: **XML, JSON, YAML, or HDF5**

Shared Metadata for Data-Centric Materials Science, Ghiringhelli et al. arXiv:2205.14774

Hierarchical organization



Example of file formats: **XML, JSON, YAML, or HDF5**

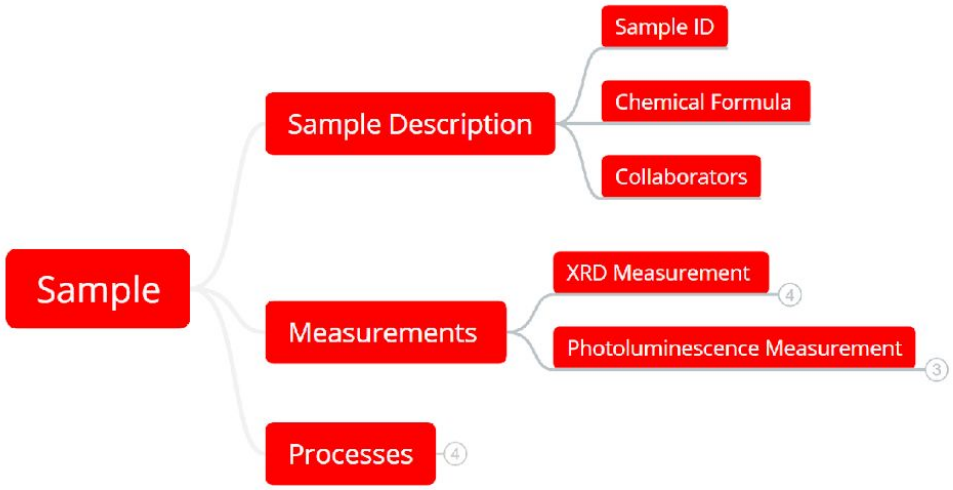
Shared Metadata for Data-Centric Materials Science, Ghiringhelli et al. arXiv:2205.14774

Structured Data

FAIRmat

Structured Data

Hierarchical organization



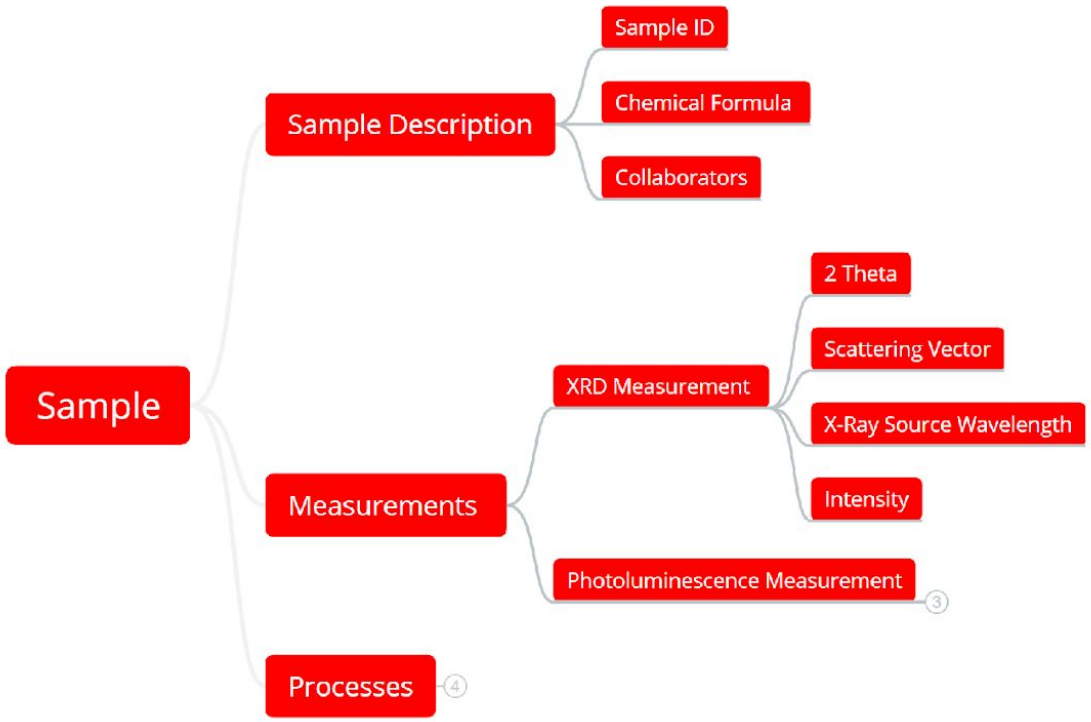
Example of file formats: **XML, JSON, YAML, or HDF5**

Shared Metadata for Data-Centric Materials Science, Ghiringhelli et al. arXiv:2205.14774

FAIRmat

Structured Data

Hierarchical organization



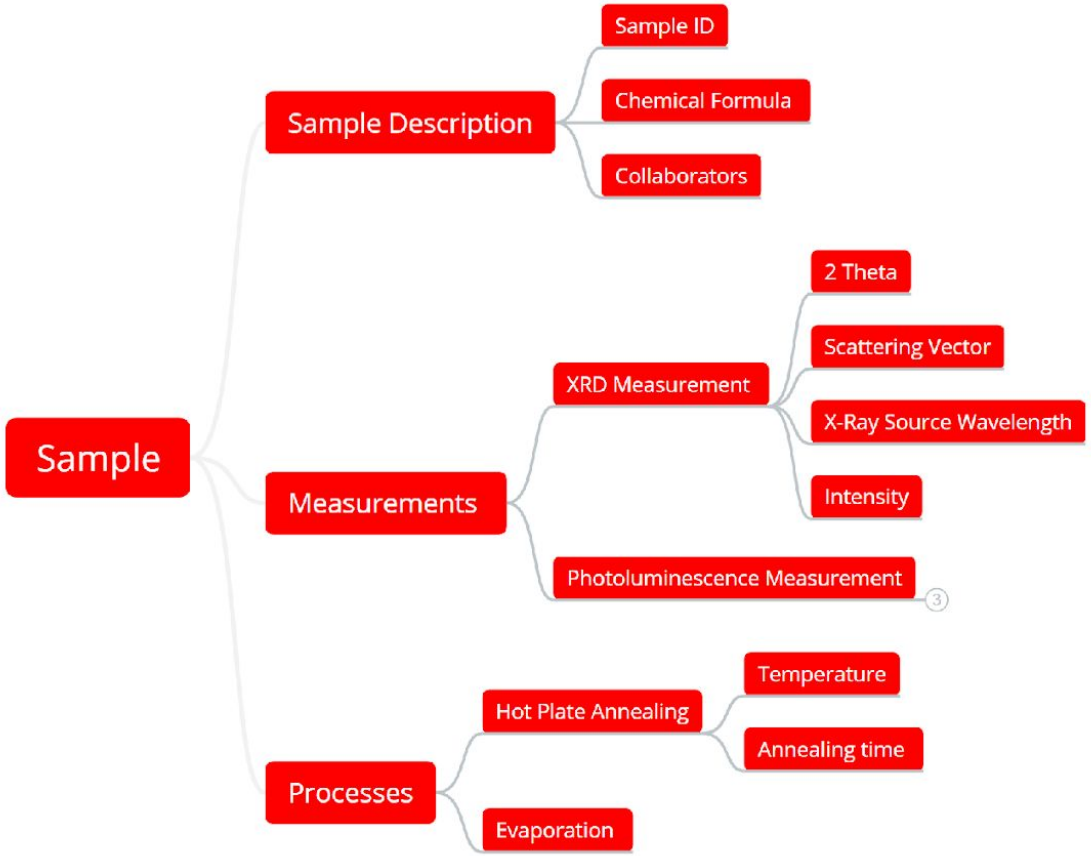
Example of file formats: **XML, JSON, YAML, or HDF5**

Shared Metadata for Data-Centric Materials Science, Ghiringhelli et al. arXiv:2205.14774

FAIRmat

Structured Data

Hierarchical organization



Example of file formats: **XML, JSON, YAML, or HDF5**

Shared Metadata for Data-Centric Materials Science, Ghiringhelli et al. arXiv:2205.14774

FAIRmat

Structured Data

YAML files to describe data structures

It is a Markup Language defining hierarchy with indentation

```yaml
definitions:
  name: My Custom Schemas
  sections:
    MyProcess:
      base_sections:
        - nomad.datamodel.metainfo.eln.BasicEln
      quantities:
        data_file:
          type: str
          m_annotations:
            browser:
              adaptor: RawFileAdaptor
            eln:
              component: FileEditQuantity
      sub_sections:
        MyProcessesCollection:
          section:
            quantities:
              sample_id:
                type: str
                m_annotations:
                  eln:
                    component: StringEditQuantity
```

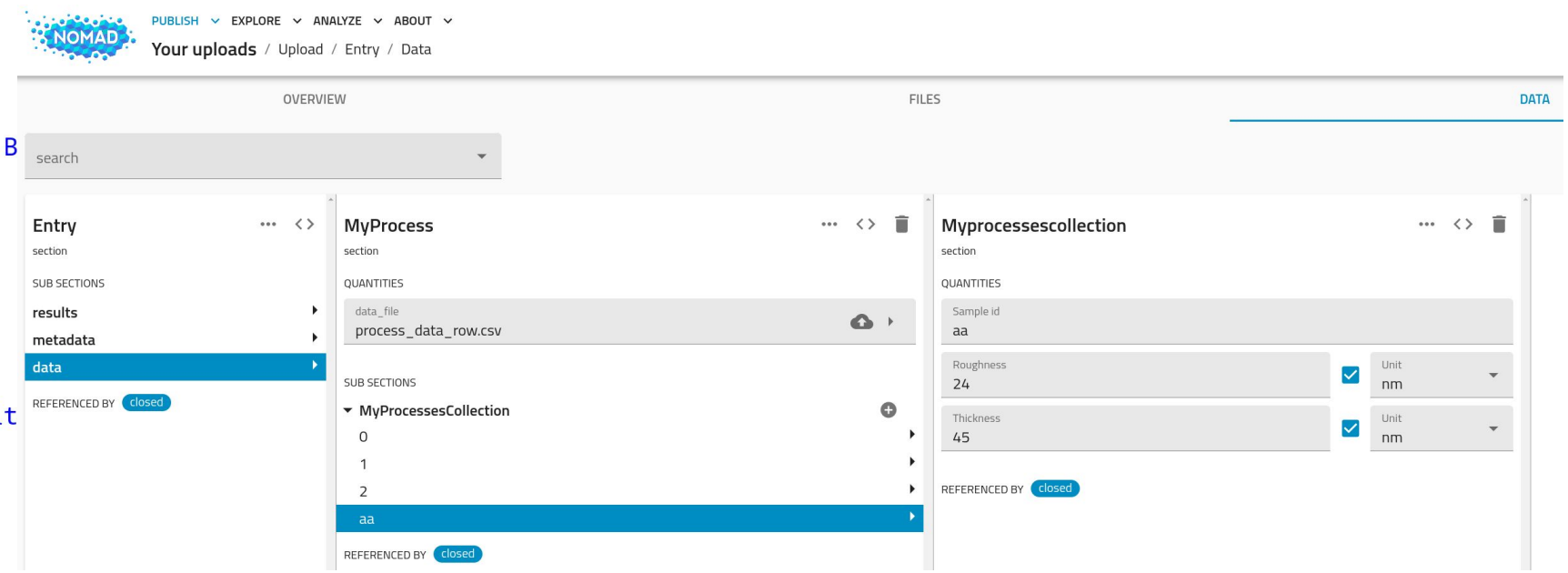
- extensible
- structured
- plain text
- human readable

FAIRmat

Using Schemas in Nomad

Structured Data

```yaml
definitions:
  name: My Custom Schemas
  sections:
    MyProcess:
      base_sections:
        - nomad.datamodel.metainfo.eln.B
      quantities:
        data_file:
          type: str
          m_annotations:
            browser:
              adaptor: RawFileAdaptor
            eln:
              component: FileEditQuantit
      sub_sections:
        MyProcessesCollection:
          section:
            quantities:
              sample_id:
                type: str
                m_annotations:
                  eln:
                    component: StringEditQuantity
              roughness:
                type: np.float64
                unit: nm
                m_annotations:
                  eln:
                    component: NumberEditQuantity
                    defaultDisplayUnit: nm
```

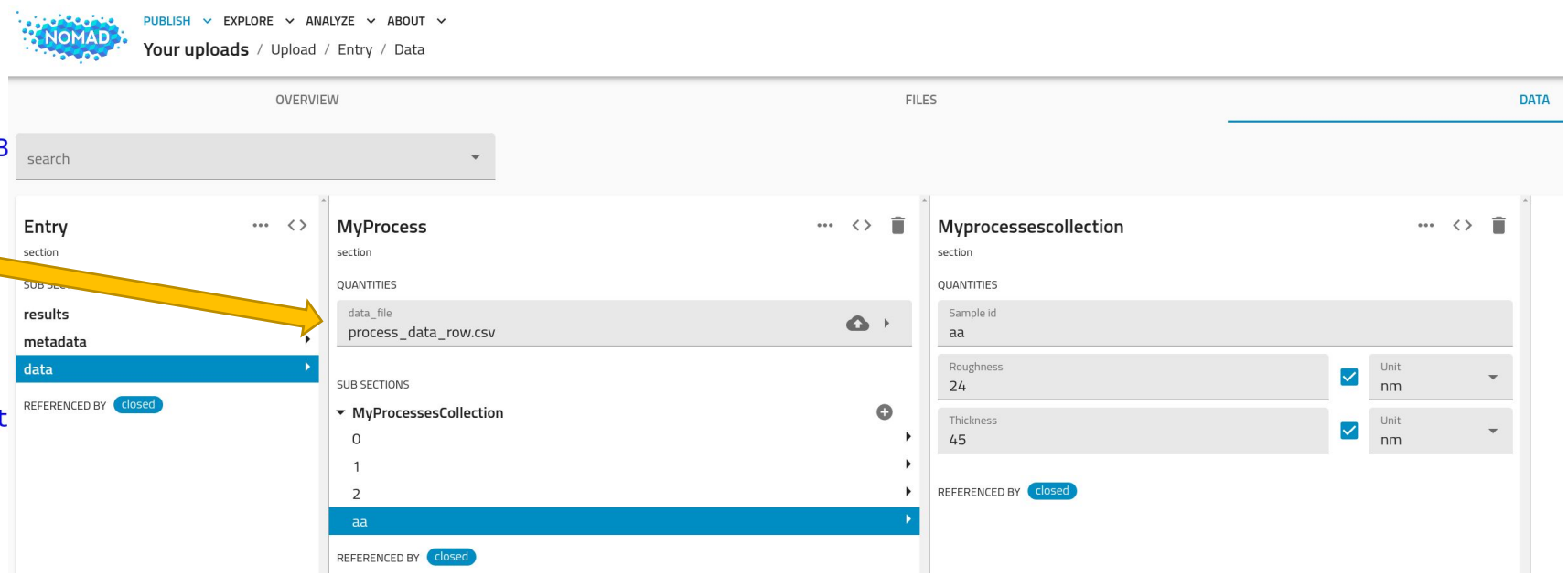


FAIRmat

Using Schemas in Nomad

Structured Data

```yaml
definitions:
  name: My Custom Schemas
  sections:
    MyProcess:
      base_sections:
        - nomad.datamodel.metainfo.eln.B
      quantities:
        data_file:
          type: str
          m_annotations:
            browser:
              adaptor: RawFileAdaptor
            eln:
              component: FileEditQuantit
      sub_sections:
        MyProcessesCollection:
          section:
            quantities:
              sample_id:
                type: str
                m_annotations:
                  eln:
                    component: StringEditQuantity
              roughness:
                type: np.float64
                unit: nm
                m_annotations:
                  eln:
                    component: NumberEditQuantity
                    defaultDisplayUnit: nm
```



Main elements:

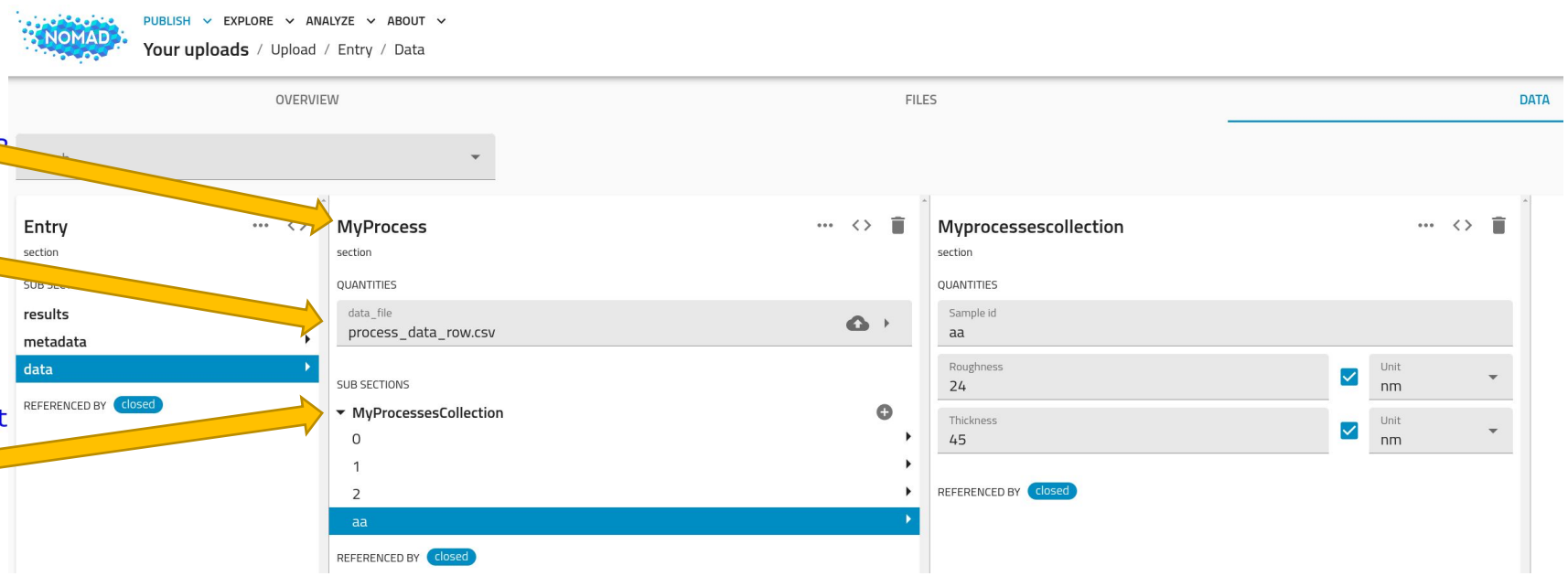
- Quantities (data fields int, float, str, datetime)
- Attributes (type, shape, unit, annotations)

FAIRmat

Using Schemas in Nomad

Structured Data

```yaml
definitions:
  name: My Custom Schemas
  sections:
    MyProcess:
      base_sections:
        - nomad.datamodel.metainfo.
      quantities:
        data_file:
          type: str
          m_annotations:
            browser:
              adaptor: RawFileAdaptor
            eln:
              component: FileEditQuantit
      sub_sections:
        MyProcessesCollection:
          section:
            quantities:
              sample_id:
                type: str
                m_annotations:
                  eln:
                    component: StringEditQuantity
              roughness:
                type: np.float64
                unit: nm
                m_annotations:
                  eln:
                    component: NumberEditQuantity
                    defaultDisplayUnit: nm
```



Main elements:

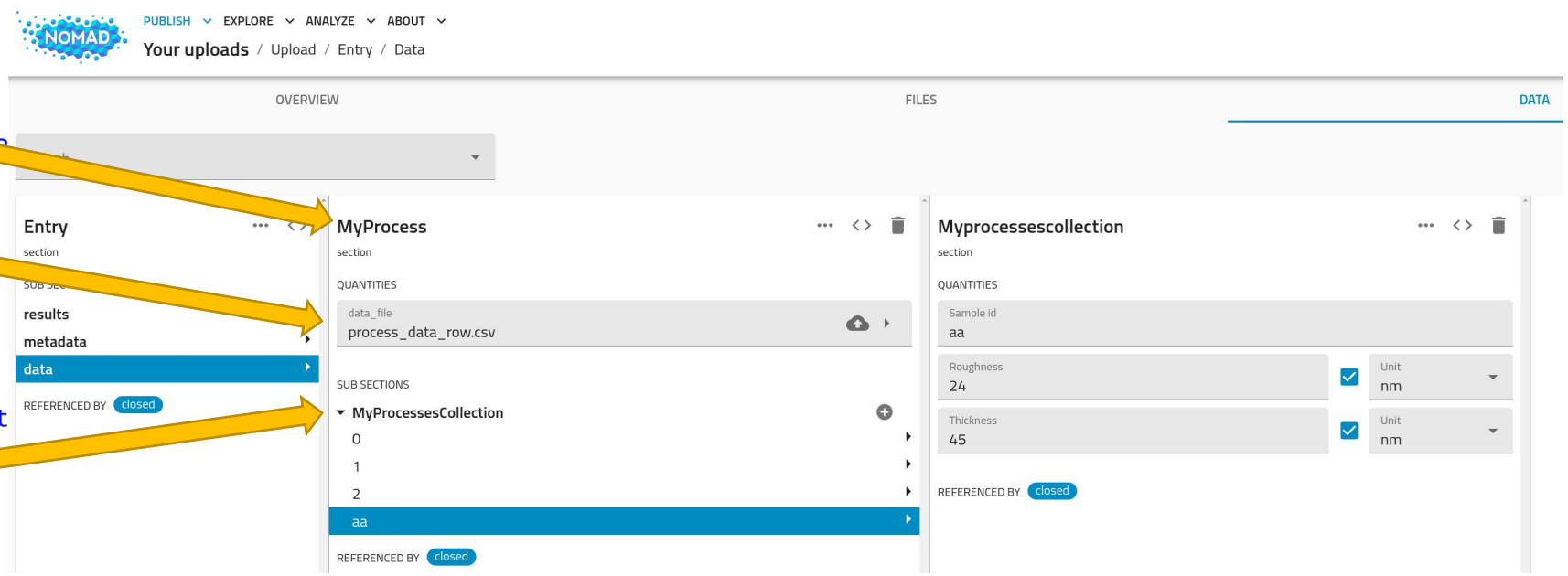
- Quantities (data fields int, float, str, datetime)
- Attributes (type, shape, unit, annotations)
- Sections (or Classes, collections of Quantities)

FAIRmat

Using Schemas in Nomad

Structured Data

```yaml
definitions:
  name: My Custom Schemas
  sections:
    MyProcess:
      base_sections:
        - nomad.datamodel.metainfo.c
      quantities:
        data_file:
          type: str
          m_annotations:
            browser:
              adaptor: RawFileAdaptor
            eln:
              component: FileEditQuantit
      sub_sections:
        MyProcessesCollection:
          section:
            quantities:
              sample_id:
                type: str
                m_annotations:
                  eln:
                    component: StringEditQuantity
              roughness:
                type: np.float64
                unit: nm
                m_annotations:
                  eln:
                    component: NumberEditQuantity
                    defaultDisplayUnit: nm
```

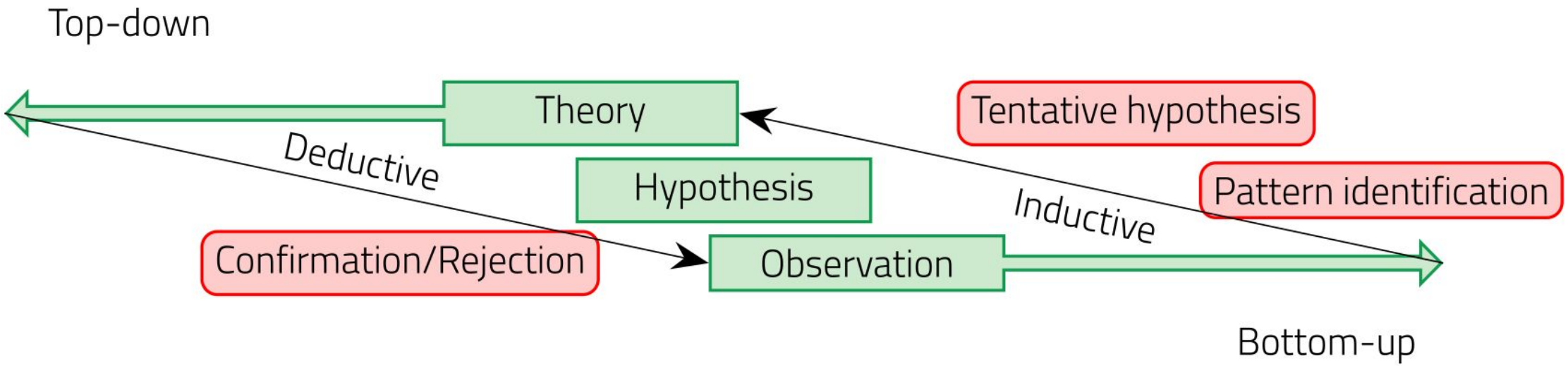


Several examples are already public:

github.com/FAIRmat-NFDI/AreaA-data_modeling_and_schemas

FAIRmat

Towards a Standard / an Ontology



An iterative exchange among scientific communities

is necessary to acquire generality, consistency, usefulness

“Base Classes” approach

Boil down the data structure to elemental building blocks:

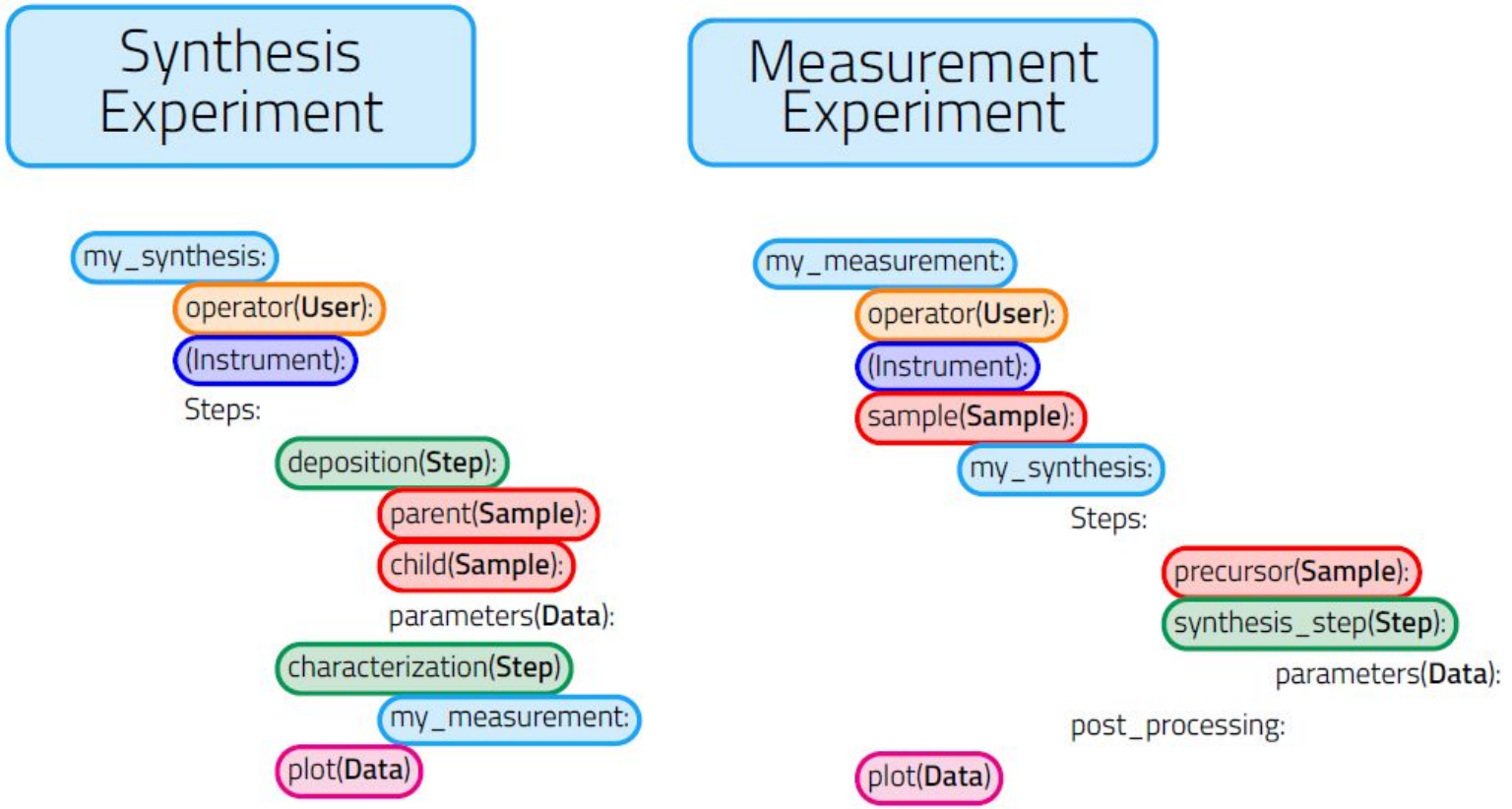
Each Base Class contains a set of properties (Quantities)

Allows for additional searchability and processing capabilities in Nomad!

Towards a Standard

Modularity and Flexibility

Combine base classes into complex structures, depending on single user needs

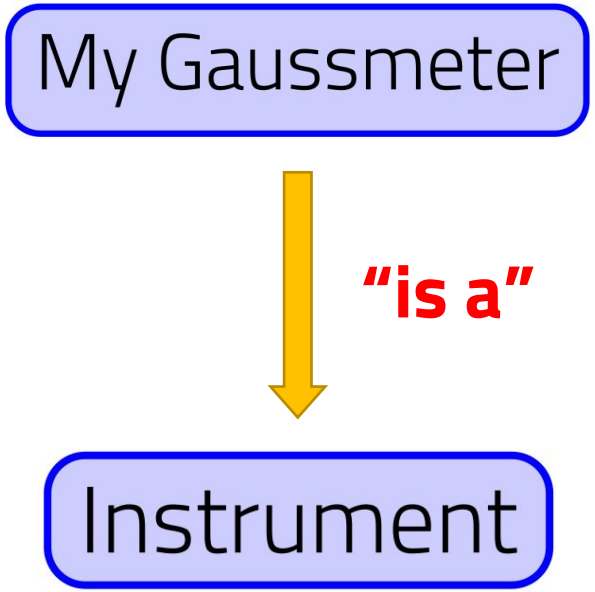


FAIRmat

Representing Hierarchy: Coding a Schema

Representing Hierarchy: Coding a Schema

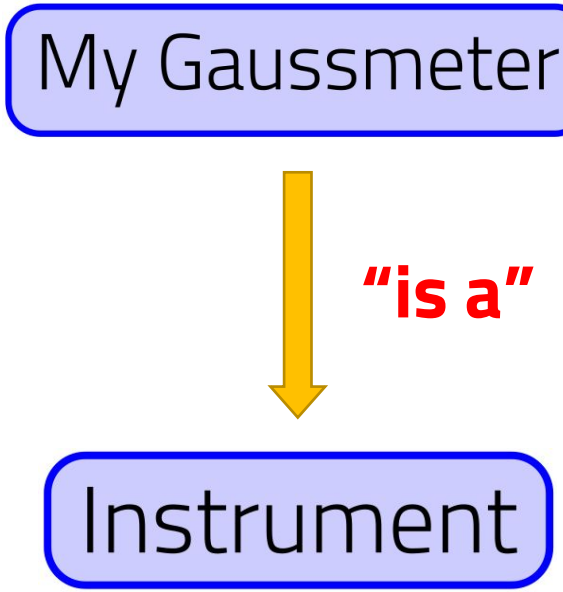
Inheritance



"My Gaussmeter"
inherits
the properties of
"Instrument"

Representing Hierarchy: Coding a Schema

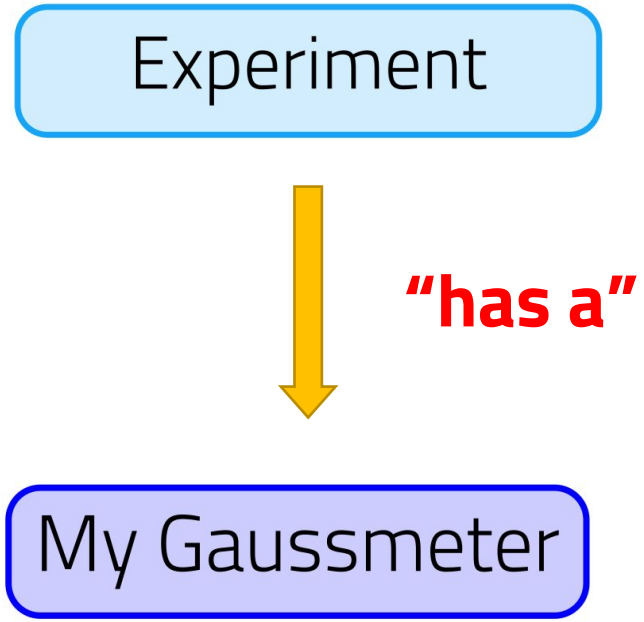
Inheritance



"My Gaussmeter"
inherits
the properties of
"Instrument"

&

Composition



"Experiment"
is composed by
an "Instrument"
(a "User", a "Sample", etc.)

How it looks like in Nomad YAML files:

```yaml
definitions:
  name: My Custom Schemas
  sections:
    MyProcess:
      quantities:
        data_file:
          type: str
        duration:
          type: np.float64
          unit: s
    MySpecifiedProcess:
      base_sections:
        - '#/MyProcess'
      quantities:
        carrier_gas:
          type: str
      sub_sections:
        MyProcessesCollection:
          section:
            quantities:
              sample_id:
                type: str
                m_annotations:
                  eln:
                    component: StringEditQuantity
```

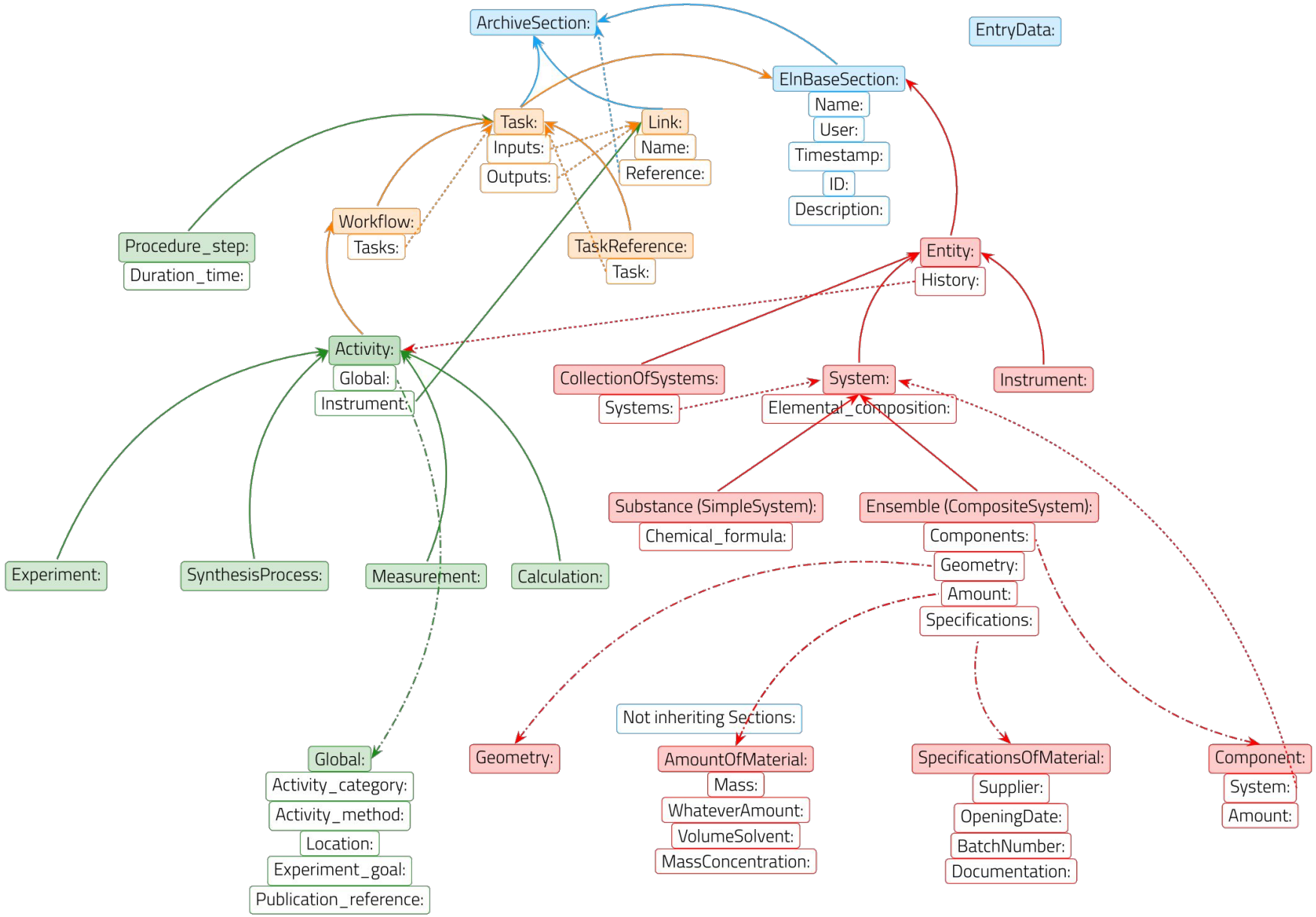
Inheritance

Composition

"MySpecifiedProcess"
is a
"MyProcess"

"MySpecifiedProcess"
has
quantities
and
sub_sections

Experimental data model is under development



Our state-of-the-art

FAIRmat

Get Involved

sebastian.brueckner@physik.hu-berlin.de

github.com/FAIRmat-NFDI/AreaA-data_modeling_and_schemas

github.com/nomad-coe/nomad

Contacts

FAIRmat

Contacts

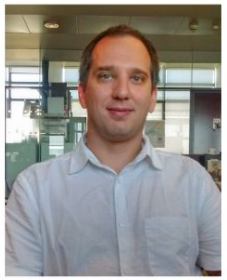
Acknowledgments

Area D Team: development of the infrastructure

Markus Scheidgen

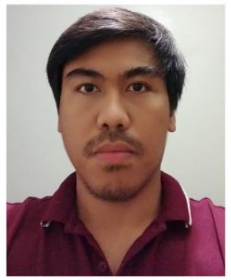
Theodore Chang

Adam Fekete



Mohammad Nakhaee

Alvin Ladines



Lauri Himanen

Amir Golparvar

FAIRmat

Thank you!