



NOMAD for experimental materials science data

FAIRmat Users Meeting

16.11.2023

Andrea Albino, Hampus Näsström, Rubel Mozumder, Markus
Kühbach, Sandor Brockhauser & Sebastian Brückner
FAIRmat Area A – Synthesis & B – Experiment Team

NOMAD for experimental materials science data



Schedule

Thursday 16th Nov., 13:00 - 16:00h:

- Introduction: Sebastian and Sandor (30 + 10 min)
- Demo & hands-on part 1: built-in ELNs and customization – Hampus & Andrea (~40 min):
- Coffee break (~20mins)
- Demo & hands-on part 2: XRD – Hampus and Rubel (~30 min)
- Demo of Area A and B examples: (~30min):
 - Pulsed Laser Deposition - Hampus
 - MOVPE Experiment – Andrea
 - EBSD – Markus

Outline



- Research data management
- What is a Schema?
- Schemas in NOMAD
 - Custom Schemas and Schema/Parser Plugins
 - Base Sections
 - Levels of Schema
- Implementation Strategy
- Standardization → Sandor
- Demo/Hands-On

Why do we need good research data management?



1. By policy



2. Knowledge retention

Stonenhöhe: 34 mm

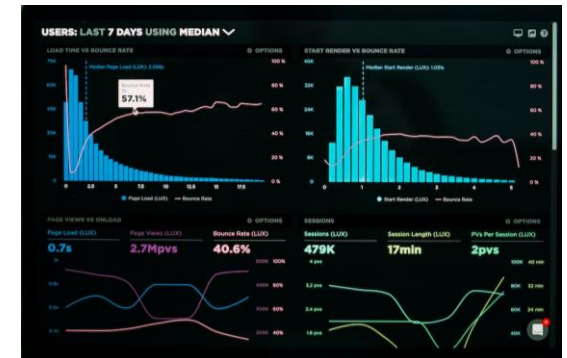
Bearbeiter/-in	Sus. P. P.	Versuch Nr.	984
Datum	8.10.19	Anlage	1505/1 1505/2 1510 (1520)
Material: Wacker	Wacker	Dotierung	1
AusgangsØ (mm)	Soll Ø (mm)	sonstige	Ph ₂
103	103	rundgeschliffen	BtH ₂
Induktor	HV 5.8.10 ⁻⁸ [mbar]	LR (gut) (schlecht)	P 1.200 [mbar]
Versuchsziel	SG Ar 15 [l/min] + N ₂ 0.35 [%] +	Zündposition	
Induktortest <100> vfr (PTP)			



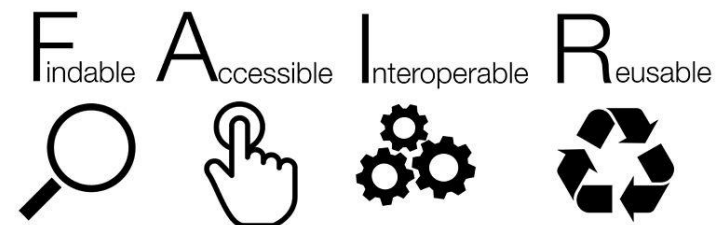
3. Reproducibility



4. Automatization, access to ML/AI tools

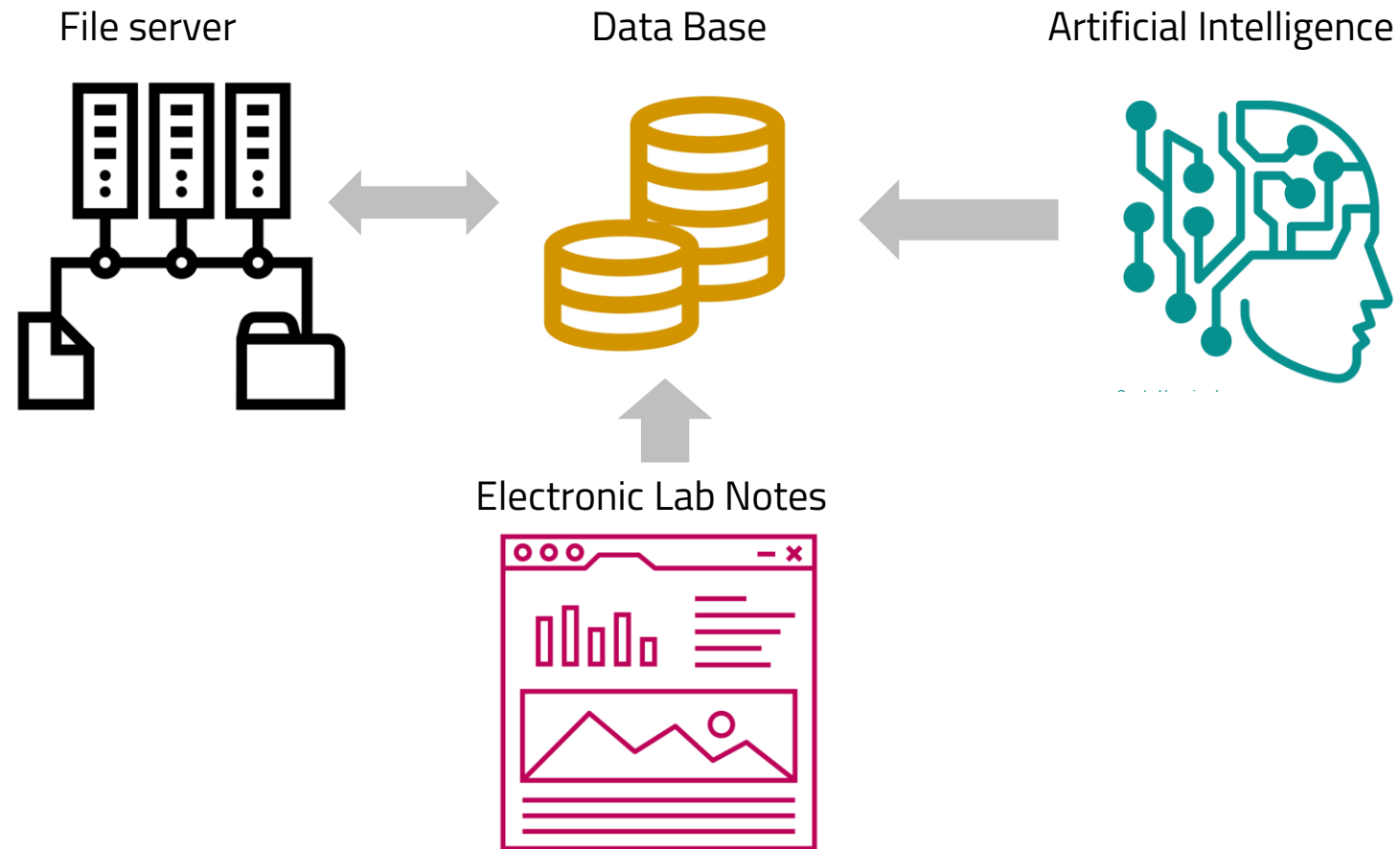


5. Prepare for the future:





Research data management (RDM) with structured data



- Data base with **structured data** → **machine readable and shareable** → **FAIR**
- ELN as an interface to the database
 - more than the simple electronic substitution of written lab notes!
 - ELN entries require a **structured** approach of collecting data and a metadata structure



Drag & Drop

Upload file-by-file or zip
and upload whole directory structures.



Get a DOI

NOMAD allows you to publish and archive
your data for free. Assign a DOI to uploaded
datasets and reference your papers.



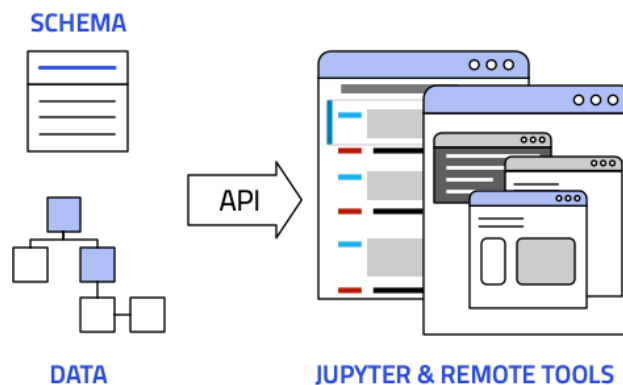
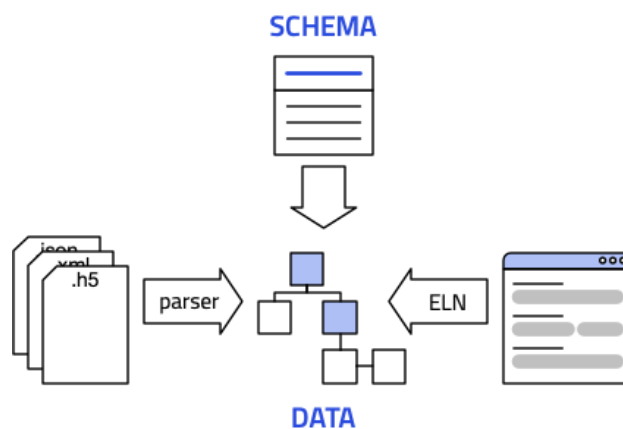
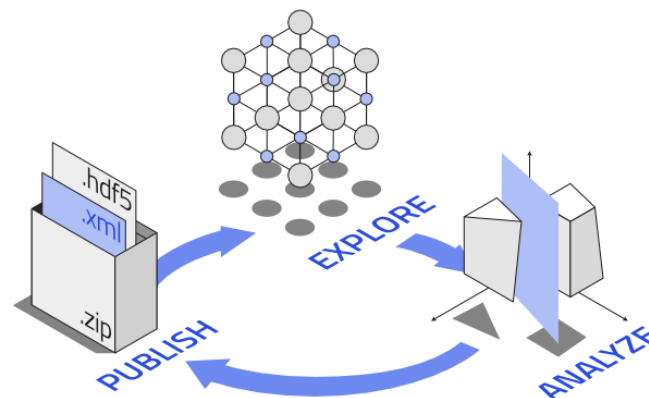
You Control Your Data

Not ready to publish? Organize data and
collaborate in private first.



Ready to use Tools

Run Jupyter and other tools directly on NOMAD.





Drag & Drop

Upload file-by-file or zip
and upload whole directory structures.



Get a DOI

NOMAD allows you to publish and archive
your data for free. Assign a DOI to uploaded
datasets and reference your papers.



You Control Your Data

Not ready to publish? Organize data and
collaborate in private first.



Ready to use Tools

Run Jupyter and other tools directly on NOMAD.



Runs on your premises

Runs behind your firewall and inside your
VPN. Use your own resources for processing
and running analysis tools.



Custom ELNs

Extend and customize NOMAD's schema to
create specialized editors to document your work.



Your own parsers

Support your own file formats. Add parsers
and normalization routines.



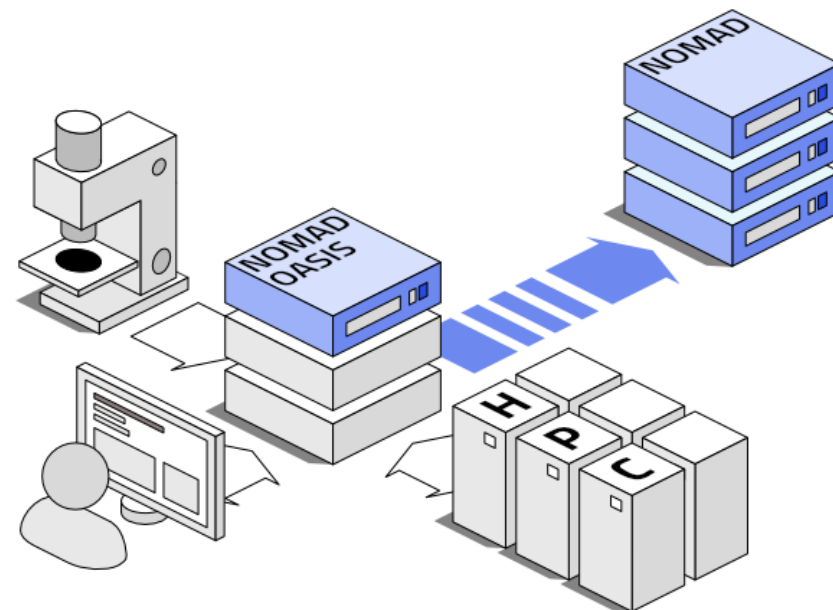
Use your own tools

Add tools and notebooks that directly access and
manipulate your data in NOMAD.



Publish selected data

Oasis will soon be connected to the public NOMAD service.

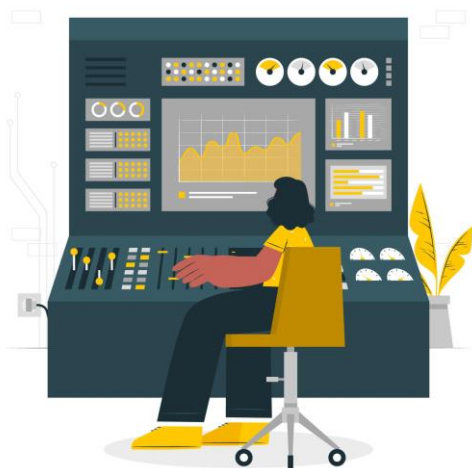


Research Data Management



Raw Data

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



https://de.freepik.com/vektoren-kostenlos/illustration-des-bdienfeldkonzepts_13662974.htm

ELN

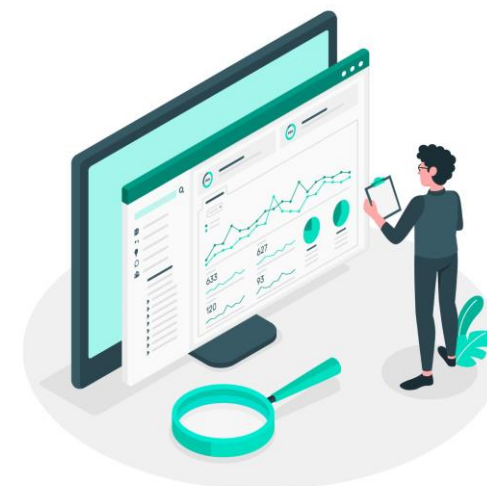
- Manually entered data and metadata



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

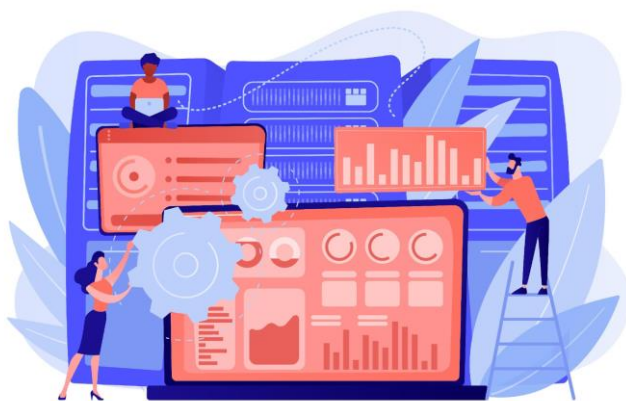
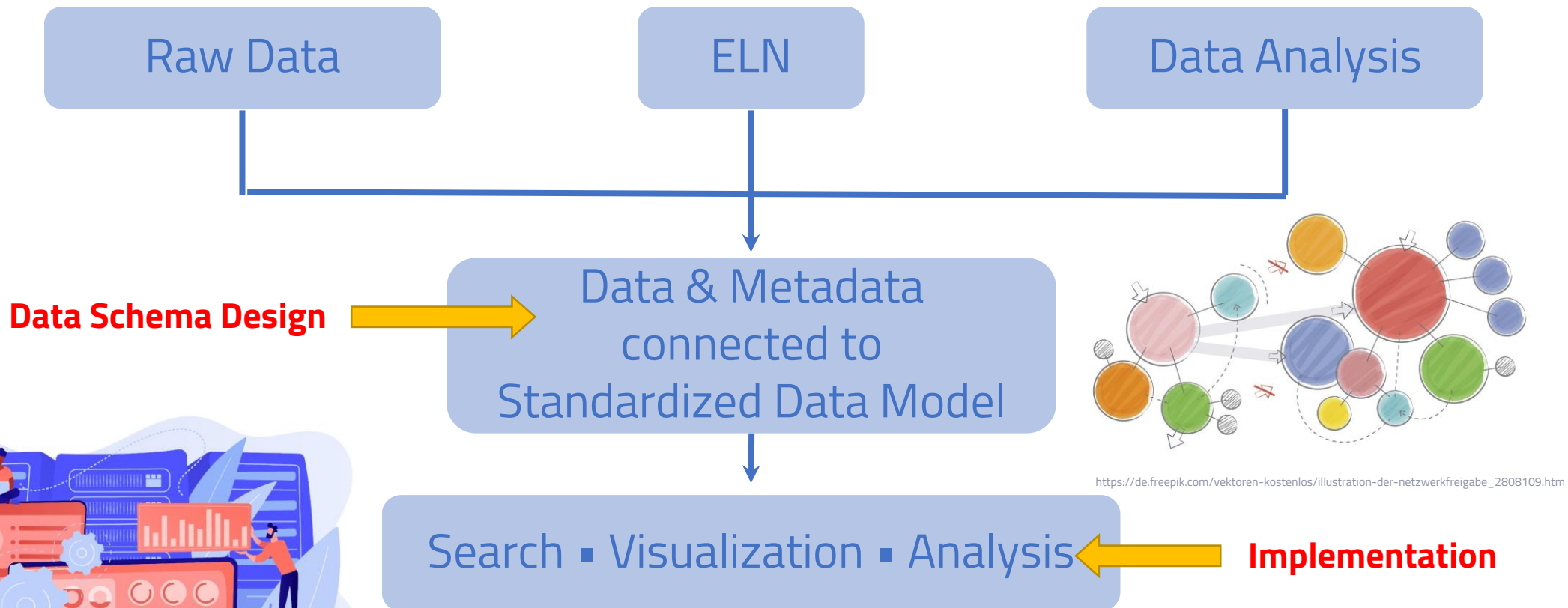
Data Analysis

- Post processing software
- User-tailored scripts



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

Research Data Management



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

NOMAD

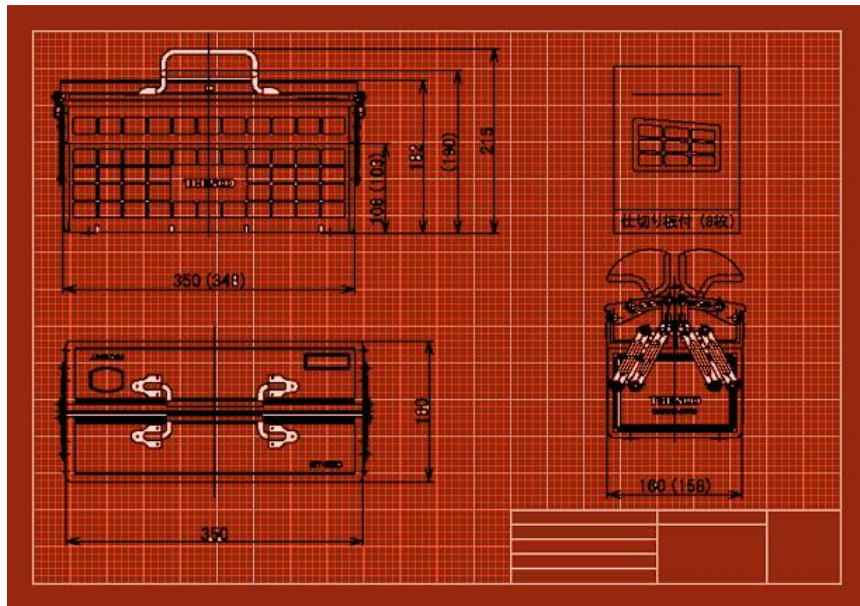
FAIRmat follows a **bottom-up approach** that is driven by the needs of scientists and already enjoys strong support from the community



Schema and Template Concepts

Schema:

formal description of data, data types, and data file structure, such as XML files



Blueprint of a toolbox

Template:

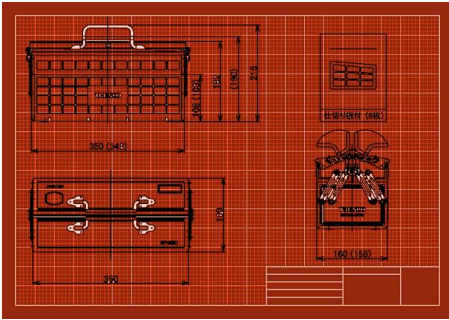
physical object, or instance, generated from a schema



Toolbox tailored on specific set of tools



Schema and Template Concepts



Schema



Template



www.freepik.com

Data



Structured Archive File



YAML files to describe data structures

It is a Markup Language defining hierarchy with indentation

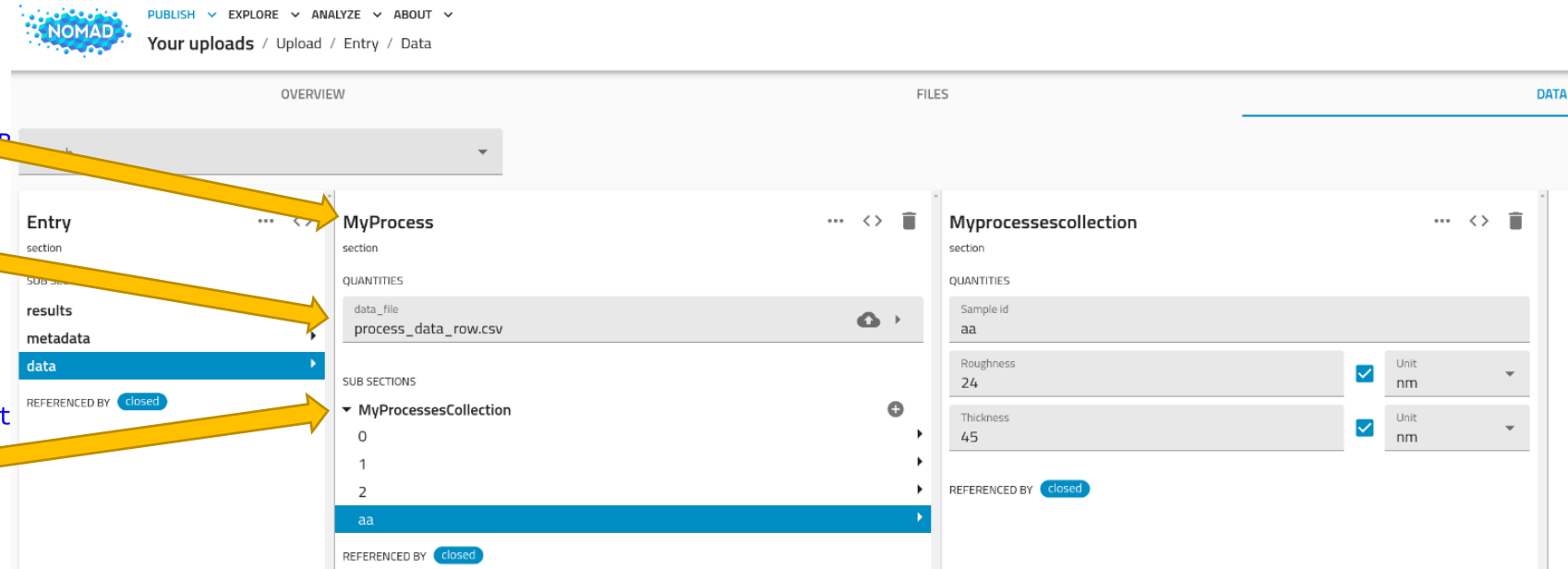
```
definitions:
  name: My Custom Schemas
  sections:
    MyProcess:
      base_sections:
        - nomad.datamodel.metainfo.eln.BasicEln
      quantities:
        data_file:
          type: str
        m_annotations:
          browser:
            adaptor: RawFileAdaptor
          eln:
           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

- extensible
- structured
- plain text
- human readable



Custom Yaml Schemas in NOMAD

```
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)

→ Public examples: github.com/FAIRmat-NFDI/AreaA-data_modeling_and_schemas



“Base sections” approach

Boil down the data structure to elemental building blocks:



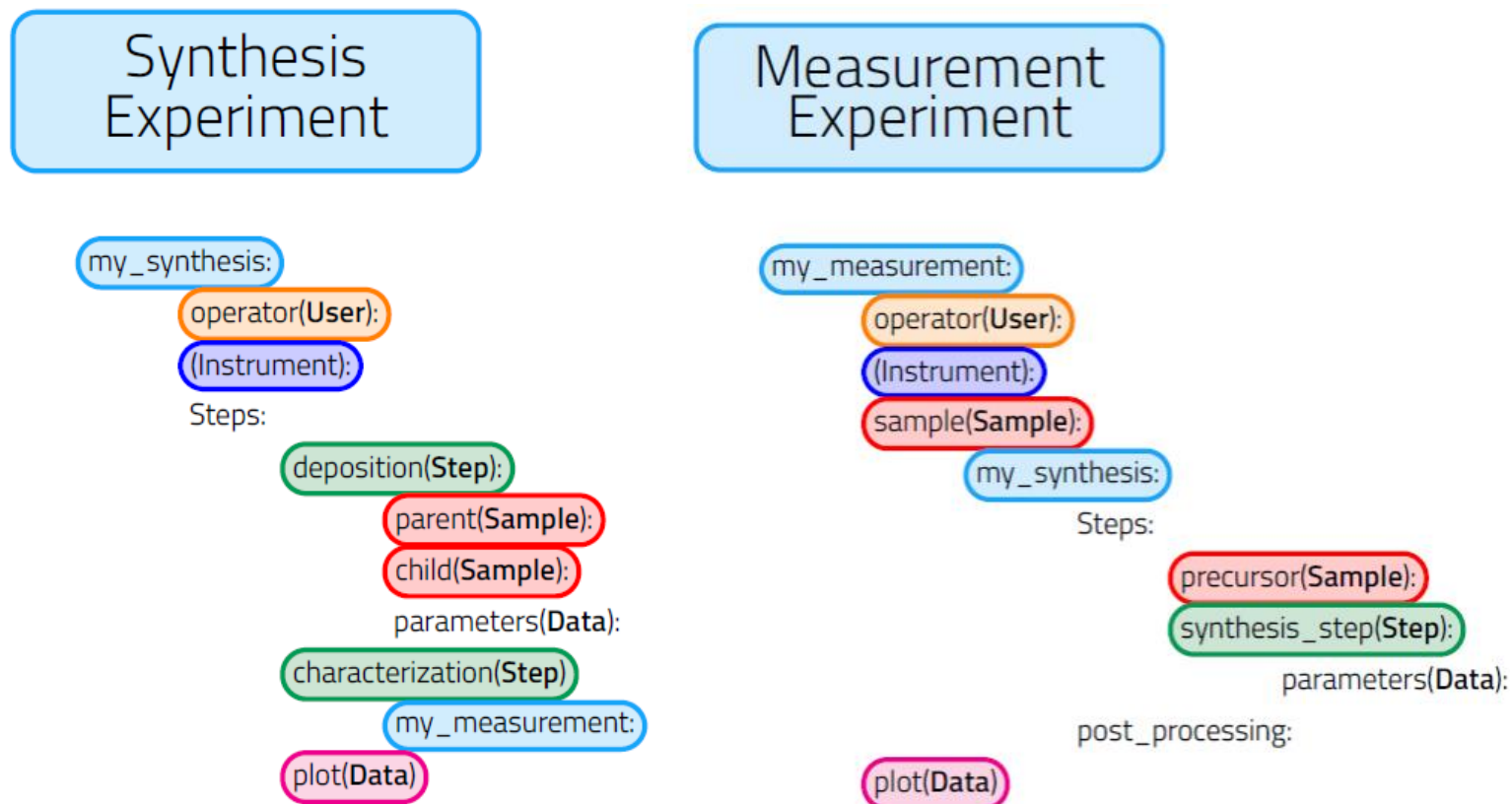
Each Base Section contains a set of properties (Quantities)

Allows for additional searchability and processing capabilities in Nomad!



Modularity and Flexibility

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



Representing Hierarchy: Coding a Schema



Inheritance

My Gaussmeter



"is a"

Instrument

"My Gaussmeter"
inherits
the properties of
"Instrument"

&

Composition

Experiment



"has a"

My Gaussmeter

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

How it looks like in NOMAD YAML files:



Schema types in NOMAD

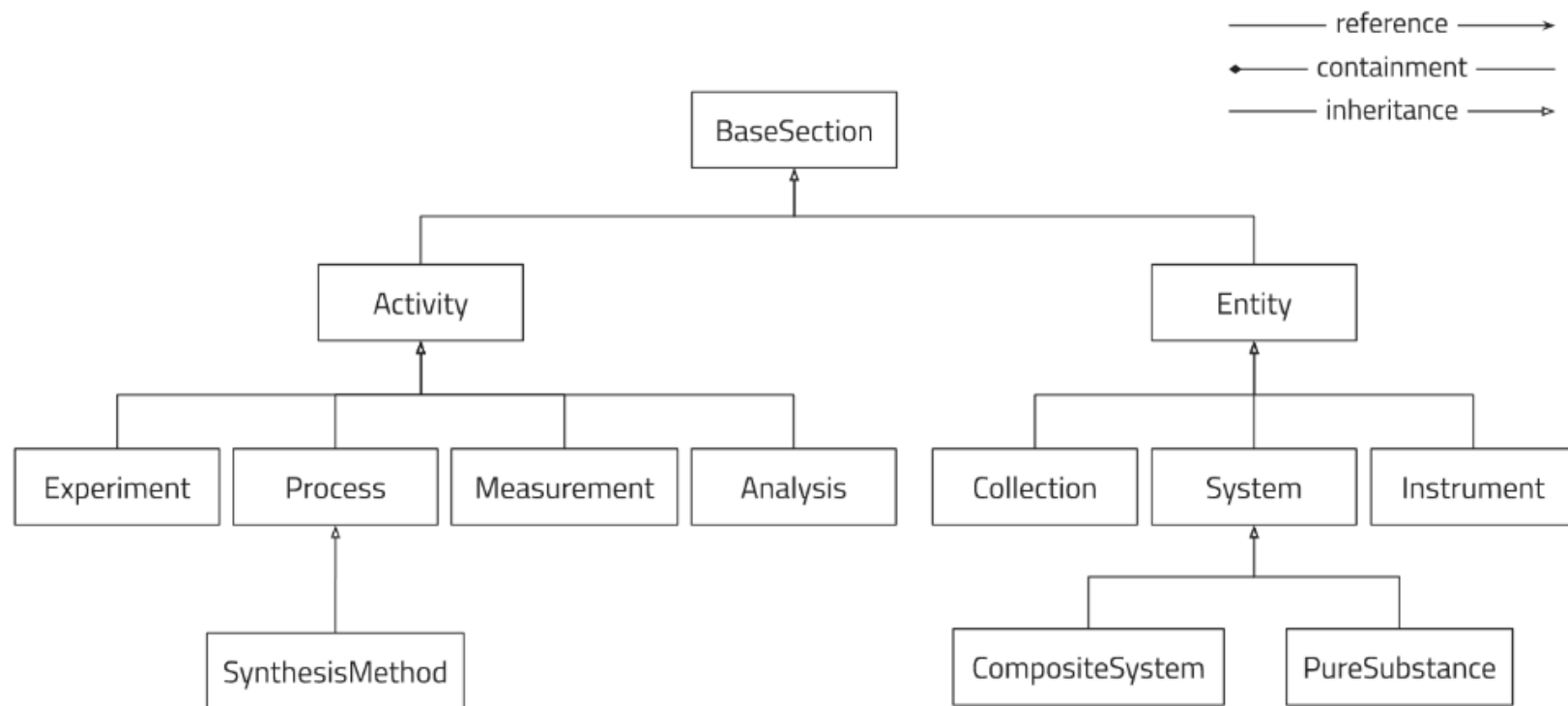


- Custom Yaml schemas
 - Yaml text file
 - Can be uploaded in any NOMAD instance
- Python schema/parser plugins
 - Python package
 - More powerful → file reading, normalizer functions
 - Plugin needs to be mounted in a NOMAD (Oasis) instance
- NeXus Application definitions
 - Yaml text file
 - Community standards
 - Works also outside NOMAD ecosystem

NOMAD's Base sections



→ general data model in NOMAD to ensure interoperability

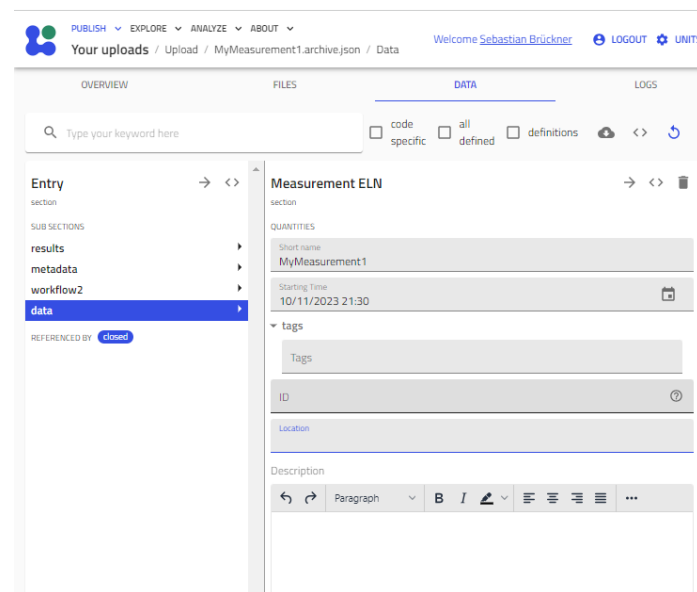
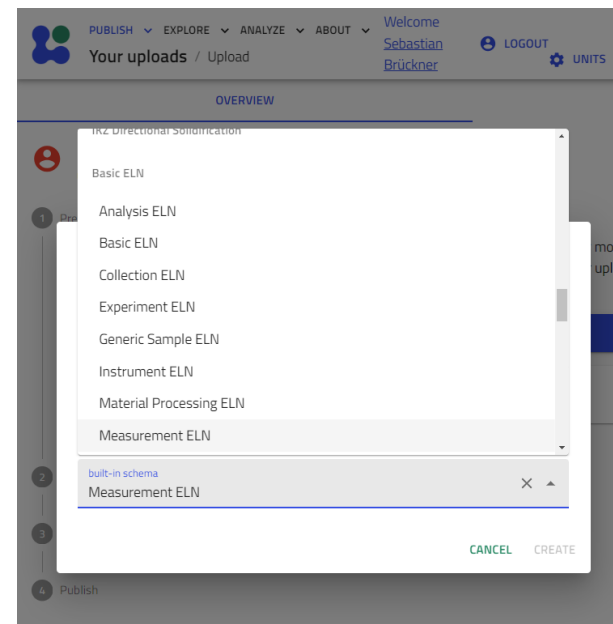


- entity-activity model → relation between different data (sample, instrument, process, measurement, analysis, experiment) and standardized structures
- Foundation to build general tools in NOMAD (search app, workflow visualizer, etc.)

Base Sections: Built-in ELNs

Use NOMAD without creating schemas and make use of data model - Customize later

- “built-in ELNs” in NOMAD based on the base sections: generic ELNs for entries like sample, instrument, measurement, process etc. → built on the base section.
- partly structured data which exhibits the functionalities and relationships between different entries given by the base sections.
- can be used in NOMAD without the need to write a schema
→ familiarize with NOMAD’s data model, start customize from there by building individual user schema on top.





allow for useful
explore apps and
tools for the **Oasis**,
like "**Explore
Measurements**" or
"**Explore Samples**"

PUBLISH EXPLORE ANALYZE ABOUT

Measurements | Processes

Welcome [Pepe Marquez](#) LOGOUT UNITS

FILTERS

- Material
- Elements / Formula
- Electronic Lab Notebook
- User Defined Quantities
- Author / Origin / Dataset
- Visibility / IDs / Schemas
- Optimade

EXPERIMENT

All experiments
Search your experimental data

[Measurements | Processes](#)
Search your measurements and processes

Samples
Search your samples

USE CASES

Solar Cells
Search solar cells

ALL

Entries
Search entries across all domains

Type your query or keyword here

+ HISTOGRAM + SCATTER PLOT + PERIODIC TABLE ↺ <> ^

Name	linear
Pepe Marquez	38

SHOWING THE ONLY ITEM

Methods linear

Type here

- X-Ray Diffraction (XRD) 18
- Evaporation 4
- Spin Coating 4
- Annealing 2
- Hot plate annealing 2
- HySprint_108_HyVap_EQMeasurement 2
- X-ray diffraction (XRD) 2

SHOWING TOP 7 ITEMS

Upload Create Time ☐ autorange linear

20/38 search results

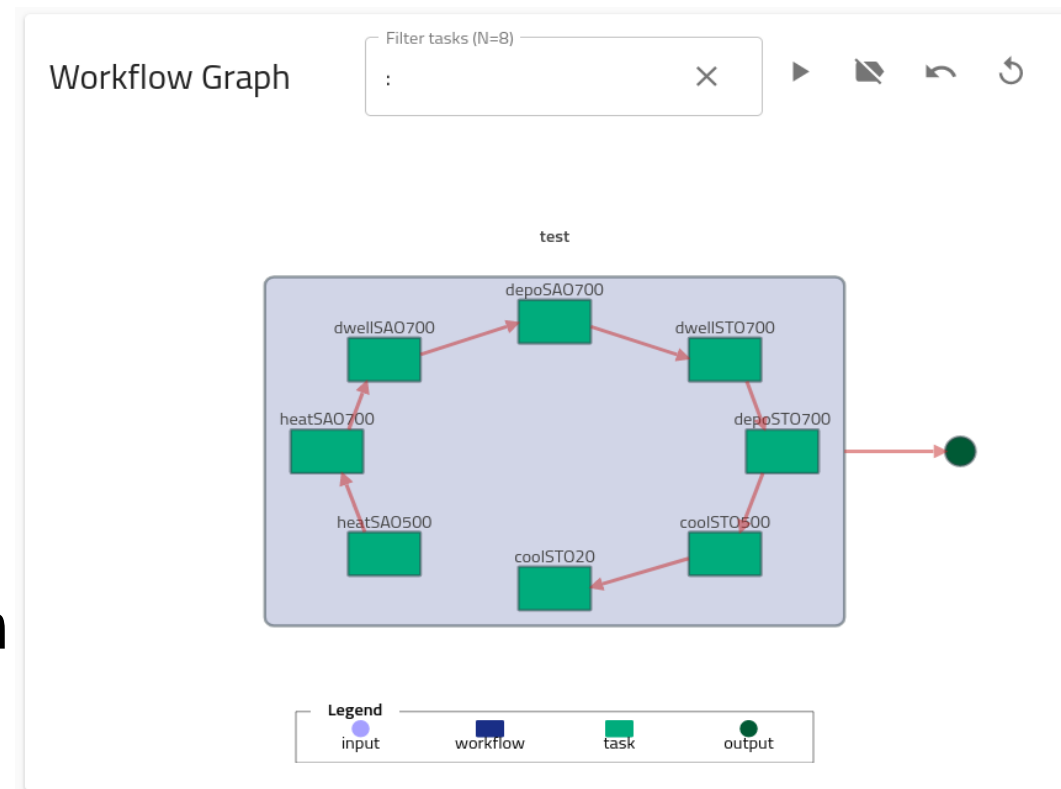
Methods	Entry time ↓	Entry type	Authors
HySprint_108_HyVap_EQMeasurement	2023-05-23T12:29:00.329000+00:00	HySprint_108_HyVap_EQMeasurement	Pepe Marquez →
HySprint_108_HyVap_EQMeasurement	2023-05-23T12:09:12.077000+00:00	HySprint_108_HyVap_EQMeasurement	Pepe Marquez →
JV Measurement	2023-05-23T12:09:11.912000+00:00	HySprint_108_HyVap_JVmeasurement	Pepe Marquez →
Evaporation	2023-05-23T11:53:48.392000+00:00	HySprint_108_HyVap_Evaporation	Pepe Marquez →
Evaporation	2023-05-23T11:53:48.023000+00:00	HySprint_108_HyVap_Evaporation	Pepe Marquez →
Annealing	2023-05-23T11:53:47.689000+00:00	HySprint_108_HySpin_ThermalAnnealing	Pepe Marquez →
Spin Coating	2023-05-23T11:53:47.309000+00:00	HySprint_108_HySpin_SpinCoating	Pepe Marquez →

Base Sections: Workflow



Use workflow section for experimental data:

- data schema automatically populate the workflow section in NOMAD if built on the **activity base sections** (measurement, process, analysis, experiment).
- different tasks, inputs and outputs of an activity are visualized in the overview section of a NOMAD entry.



Base Sections: Entity History Card



New “Card” in overview section of NOMAD upload

Entries based on the “**Entity**” **base section** exhibit a *History Card* listing all “Activities” they have been used in → e.g. sample or instrument history

History			
4/4 activities			☰
Name	Entry type	Entry creation time ↑	
STO-SAO-STO-Alev	IKZPulsedLaserDeposition	11/8/2023, 5:25:40 PM	→
Annealing	ELNProcess	11/13/2023, 2:35:12 PM	→
XRD	ELNXRayDiffraction	11/13/2023, 2:39:13 PM	→
Phase Identification	ELNAnalysis	11/13/2023, 2:41:27 PM	→



Three Levels of Schema in NOMAD

Base sections

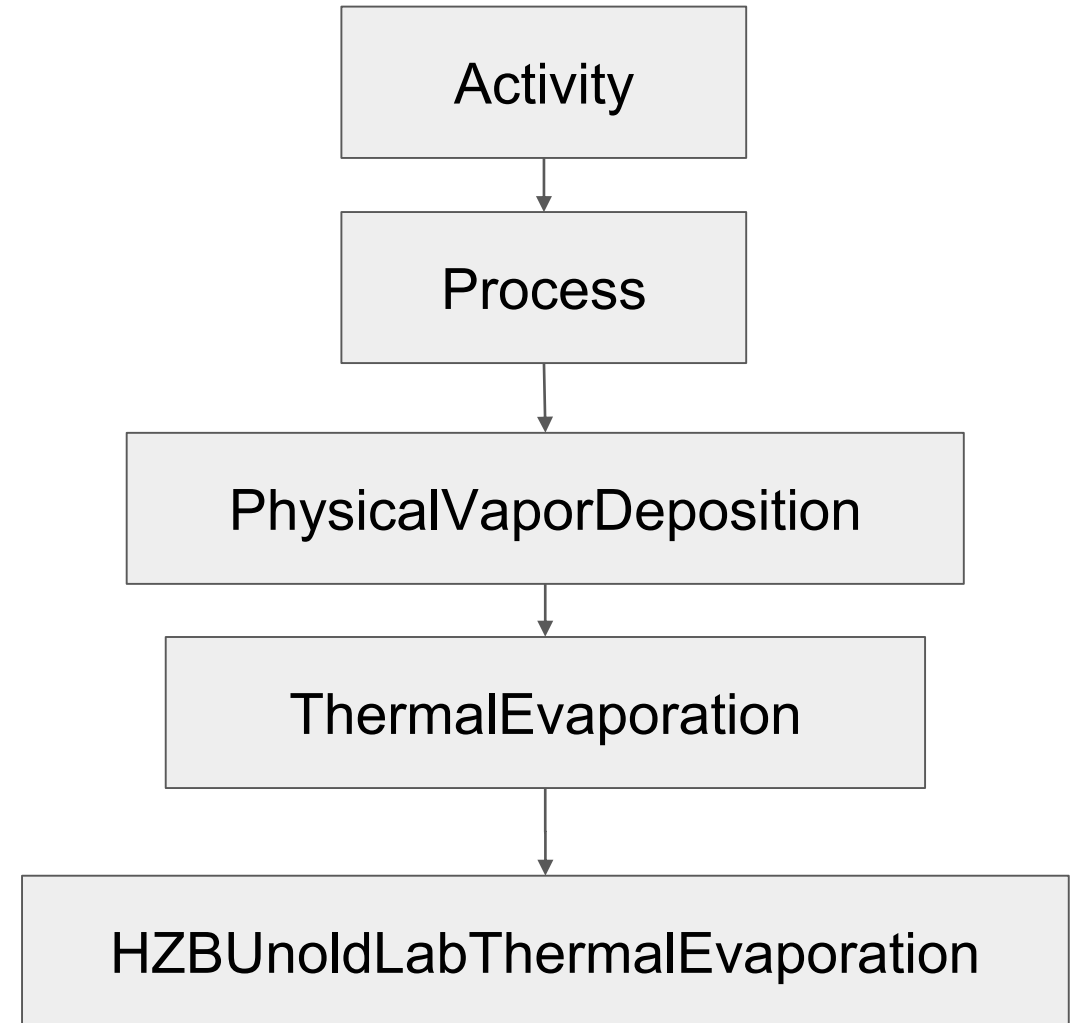
- General definitions
- Not for specific field
- Maintained by FAIRmat
- No EntryData

Standard plugins

- Not for specific lab/instrument
- nomad-material-processing
- Community (CHMO, PVDO, AutoPeroSol, etc.)

User schema

- Applies one or more base sections
- Specific to lab/instrument
- Developed by users
- Creates instances by inheriting EntryData





Three Levels of Schema in NOMAD

Base sections

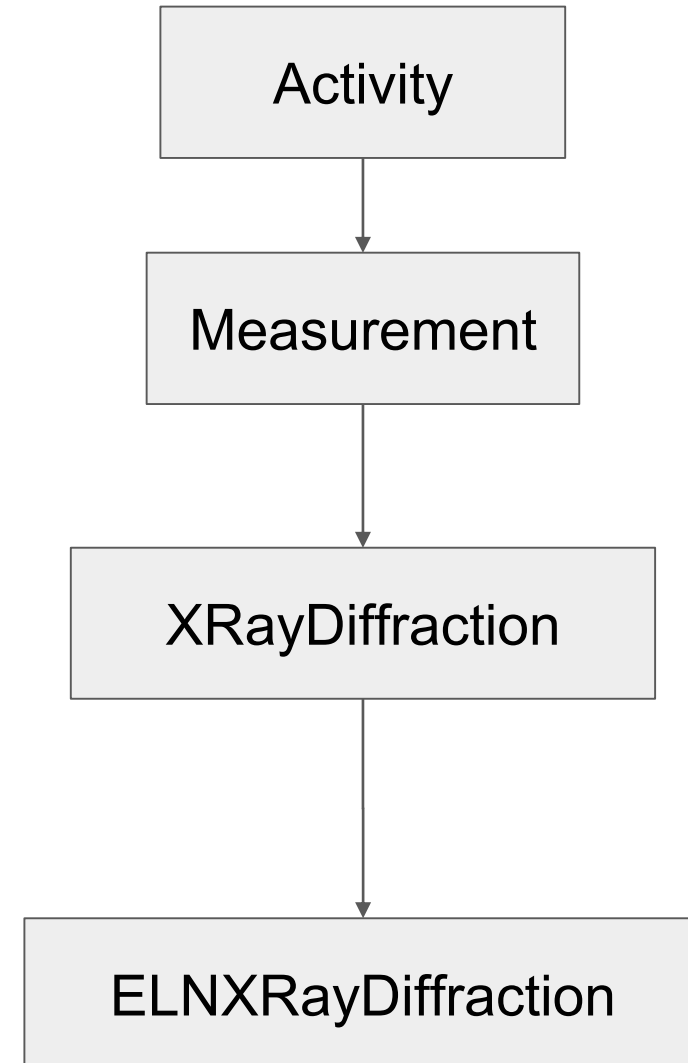
- General definitions
- Not for specific field
- Maintained by FAIRmat
- No EntryData

Standard plugins

- Not for specific lab/instrument
- nomad-measurements
- Community (CHMO, AutoPeroSol, NeXus, etc.)

User schema

- Applies one or more base sections
- Specific to lab/instrument
- Developed by users
- Creates instances by inheriting EntryData



Schema implementation strategy in NOMAD



New Measurement in NOMAD:

1. Built-in ELN Measurement
2. Custom YAML Schema
3. Plugin with Normalizer
 - a. File reader separated out
 - b. Schema as NOMAD “parser”
4. Standardized NeXus Reader

Sandors Part...

