2nd FAIRmat workshop on data exchange and storage in photoemission spectroscopy

Survey (1/2)

0 1 9

What are your expectations for the workshop?
(1/2)

- Understand NXmpes standard
- Information exchange understand what exists and future prospects
- Ho to smoothly prepare data in nexus format
- NEXUS mpes format defined. I would also like to know more about NOMAD and electronic lab notebooks but that might be another occasion. Learn to know people working with NEXUS.
- Learn more about MPES
- Learn more how far standardization of NXmpes is
- Learn more about nexus format
- I would like to know the plans on how data files from different instruments will be implemented into the Nexus format and NOMAD
- learn about the latest developments in MPES for using in NOMAD
- Very high
- Update about fairmat status
- Collect input from users and instrument manufacturers

Survey (1/2)

0 1 9

What are your expectations for the workshop?
(2/2)

- Developing rhe community and collaboration between academia and tech partners.
- Learn about recent developments and RDM possibilities.
- To learn how I can use Nexus for my data to make it FAIR
- Learn more about NeXus format and FAIRmat
- Finish a fair data standard for pes.
- What can instrument manufacturers provide and how can we assist.
- Learn about interoperability

Ranking poll

Survey (2/2)

0 2 1

What's the closest description of your intent?

1. Supporting scientists in providing data — 1.95
2. Creating data in NeXus format — 1.86
3. Learning about FAIR data and NeXus — 1.76
4. Publishing data in NOMAD — 0.91

Research data management, FAIRmat and NOMAD

Heiko B. Weber,
Friedrich-Alexander-Universität Erlangen-Nürnberg

Research data

Traditional data lifetime

New data lifecycle

Open data

Open access (~2001) to peer reviewed literature, being enforced by funding agencies

FAIR principles[1] (2016); statement at G20 summit in China[2]

Data as resource[3]: A European strategy for data (2020)

Expectations: DFG and other funding agencies expect FAIR data from us researchers

(funding is connected to FAIR data pledges)

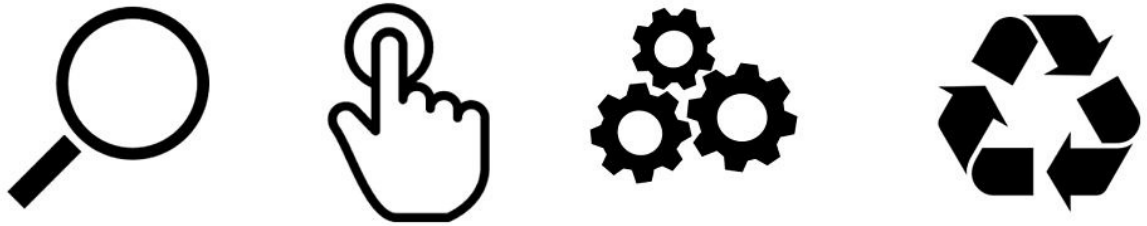
[1] Scientific Data 3, 160018 (2016)

[2] https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_16_2967

[3] https://ec.europa.eu/info/sites/default/files/communication-european-strategy-data-19feb2020_en.pdf

Underlying principles for open data

Findable Accessible Interoperable Reusable



Important: community agreement!

F indable and A rtificial I ntelligence R eady

FAIR data: benefit or burden?

Rich opportunities

- Big data opportunities in science
- Data analysis as novel discipline (besides experiment and theory)
- Accelerated materials discovery
- Enhanced data competence of scientists
- Reduction of misleading interpretations and fraud

Significant investment

- The transition to FAIR data demands for reorganizing scientists' workflow entirely

Nationale Forschungsdaten-Infrastruktur (NFDI)

(polycentered bottom-up network)

1st round (10/2020)

- DataPLANT: Plant research data
- GHGA: German Human Genome–Phenome Archive
- KonsortSWD: Consortium for the Social, Educational, Behavioural and Economic Sciences
- NFDI4Biodiversity: Biodiversity, Ecology and Environmental Data
- NFDI4Cat: NFDI for sciences related to catalysis
- NFDI4Chem: Chemistry consortium for the NFDI
- NFDI4Culture: Consortium for Research Data on Material and Immaterial Cultural Heritage
- NFDI4Health: NFDI personal health data
- NFDI4Ing: NFDI for Engineering Sciences

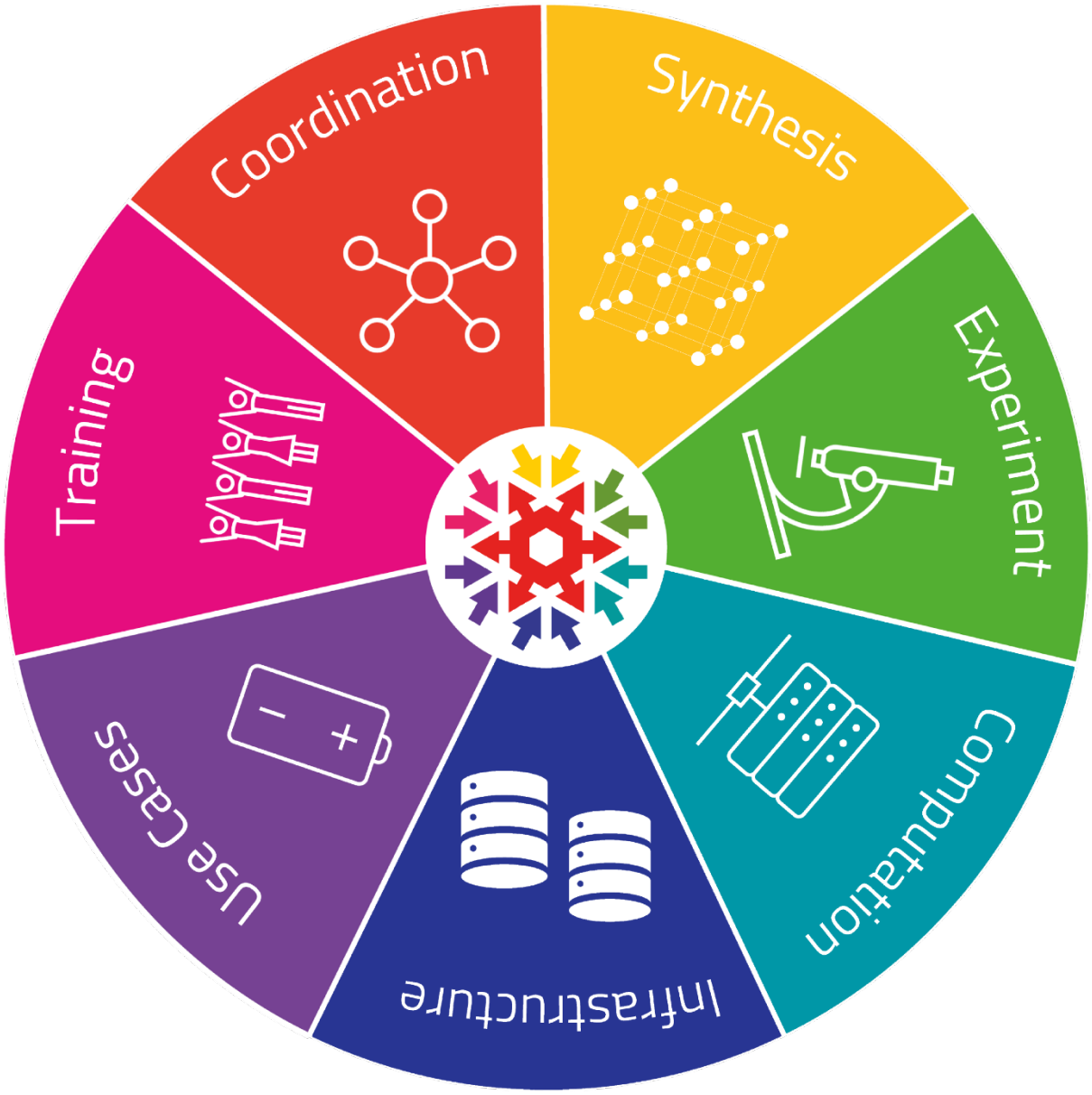
2nd round (10/2021)

- BERD@NFDI: NFDI for Business, Economic and Related Data
- **DAPHNE4NFDI: Data from PHoton and Neutron Experiments for NFDI**
- **FAIRmat: FAIR Data Infrastructure for Condensed-Matter Physics and the Chemical Physics of Solids**
- MaRDI: Mathematical Research Data Initiative
- NFDI4DataScience: NFDI for Data Science and Artificial Intelligence
- NFDI4Earth: NFDI Consortium Earth System Sciences
- NFDI4Microbiota: NFDI for Microbiota Research
- **NFDI-MatWerk: NFDI for Materials Science and Materials Engineering**
- PUNCH4NFDI: Particles, Universe, NuClei and Hadrons for the NFDI
- Text+: Language and text-based research data infrastructure

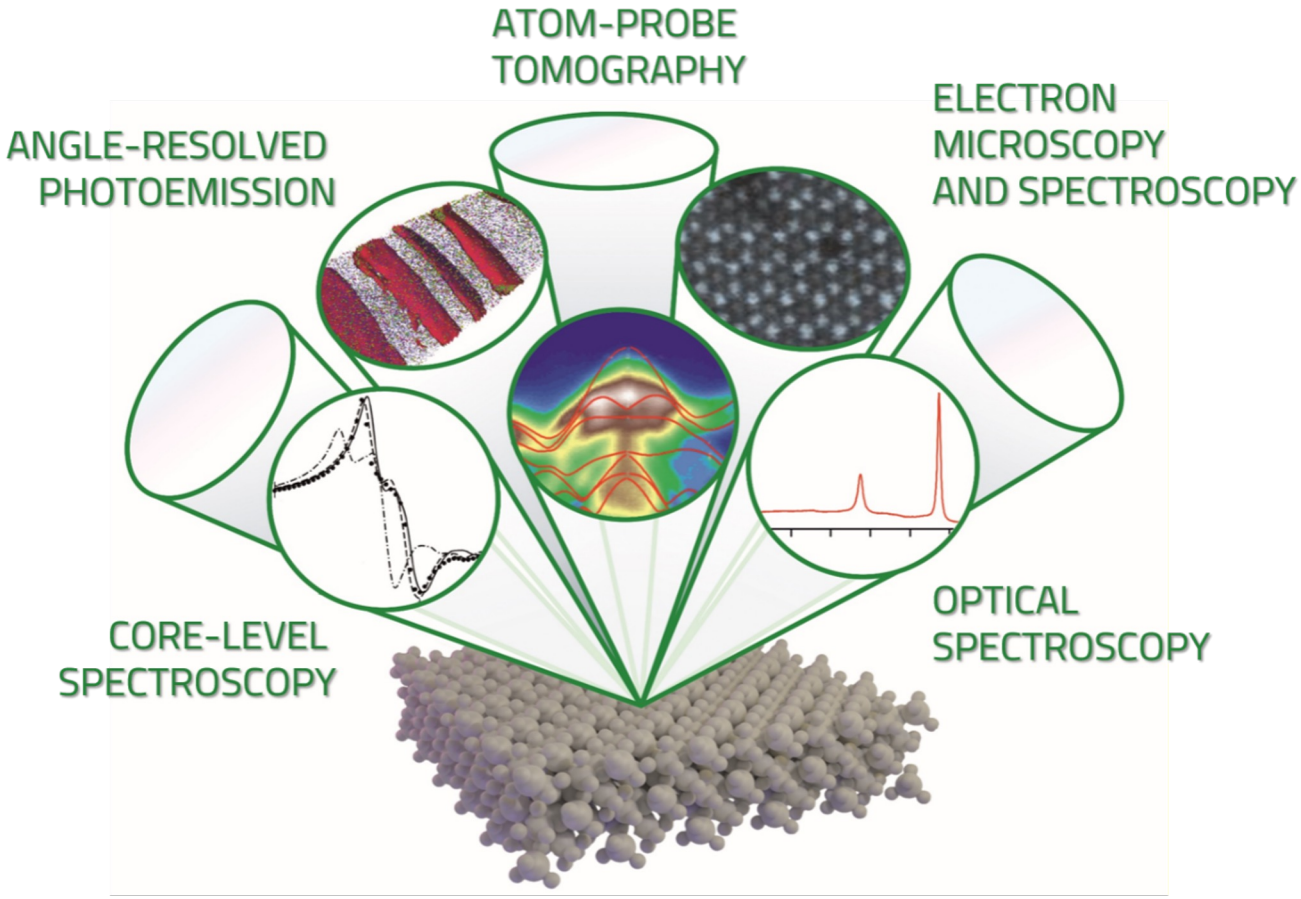
3rd round (3/2023)

- FAIRagro: FAIR Data Infrastructure for Agrosystems
- NFDI4BIOIMAGE: National research data infrastructure for microscopy and bioimage analysis
- NFDI4Energy: National Research Data Infrastructure for Interdisciplinary Energy System Research
- NFDI4Immuno: National Research Data Infrastructure for Immunology
- NFDI4Memory: The Consortium for the Historically Oriented Humanities
- NFDI4Objects: Research Data Infrastructure for the Material Remains of Human History
- NFDIxCS: National Research Data Infrastructure for and with Computer Science

FAIRmat


Coordination
Synthesis
Experiment
Computation
Infrastructure
Use cases
Training

Area B and its pilot experiments

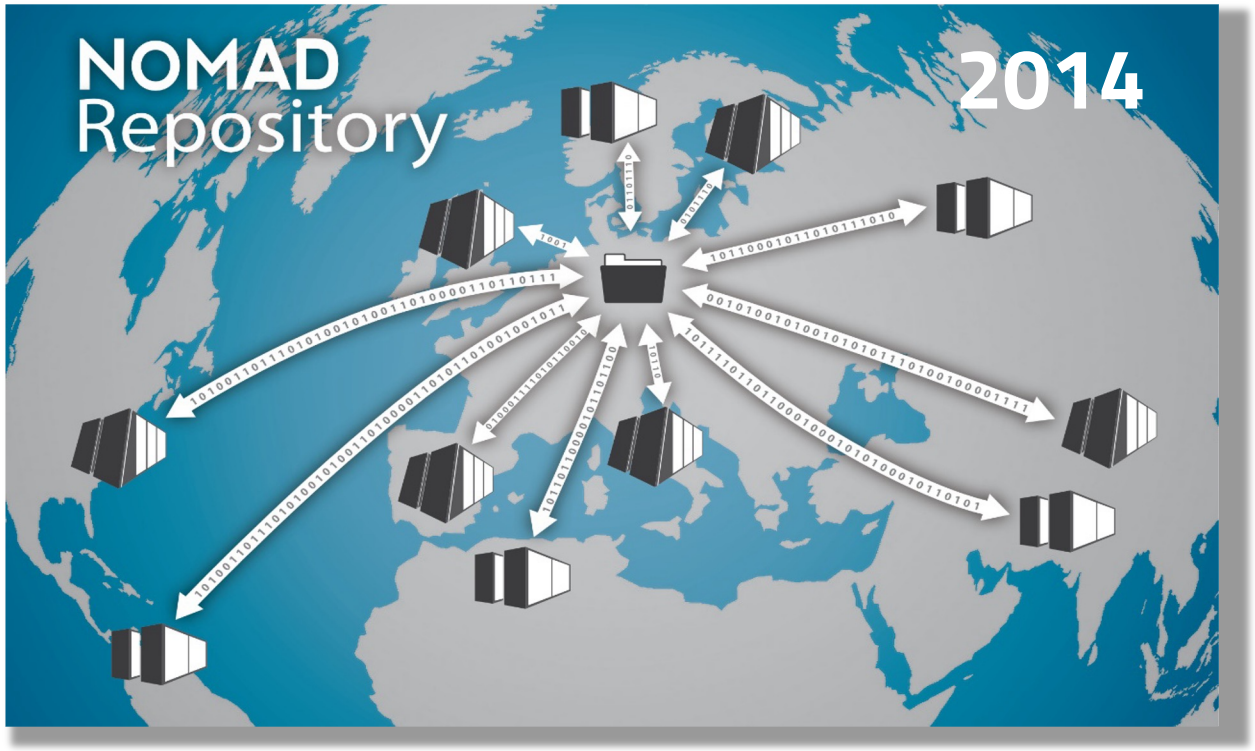


The NOMAD central hub

NOMAD

Materials data managed and shared

The Novel Materials Discovery Laboratory



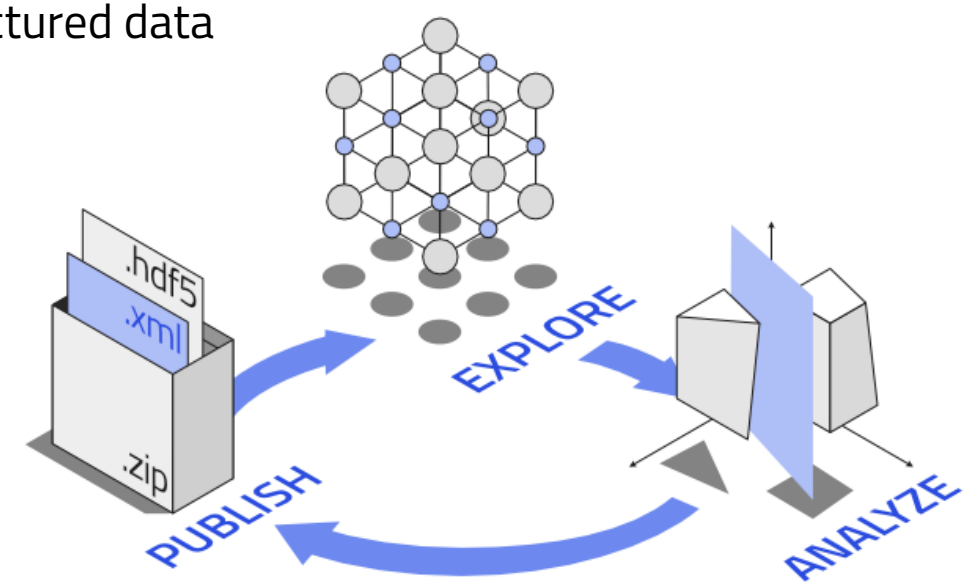
UPLOADED ENTRIES	REPRESENTED MATERIALS	UPLOADED FILES
13,009,223	3,220,189	119.4 TB

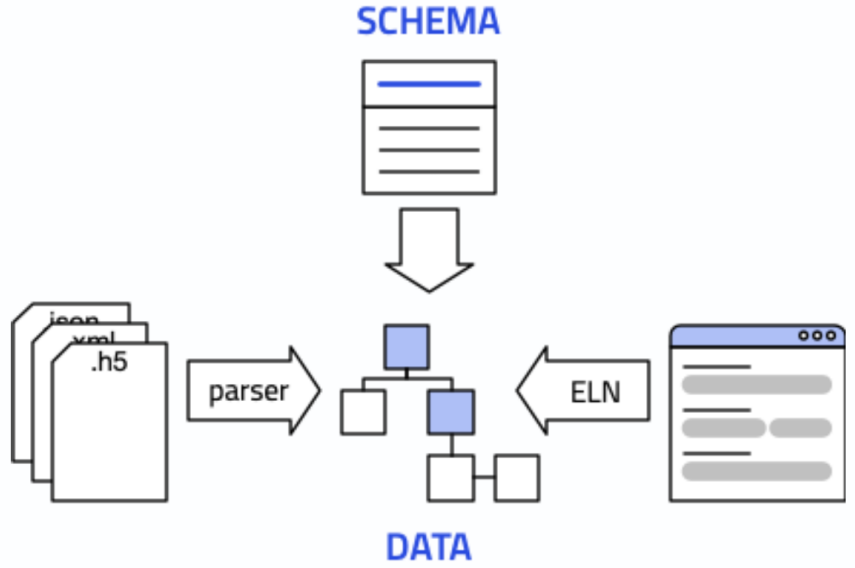
as of September 10, 202

The NOMAD central hub

NOMAD

FAIRmat hub for **harmonizing** and **reusing** structured data

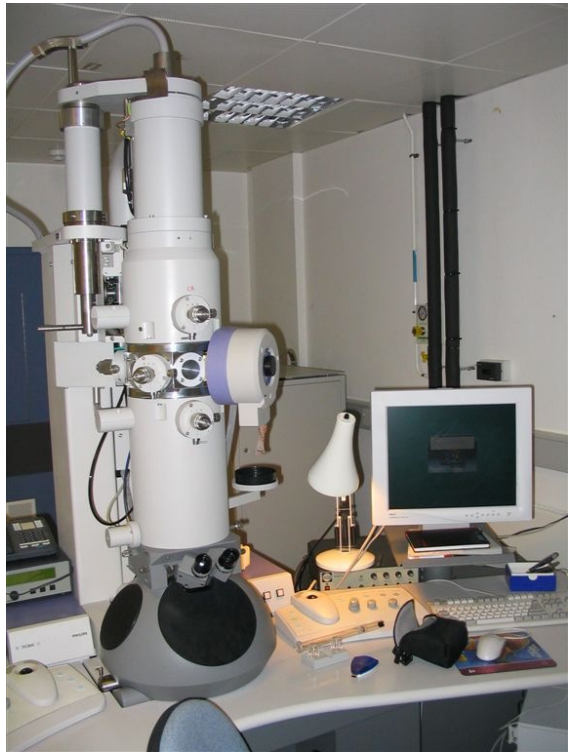


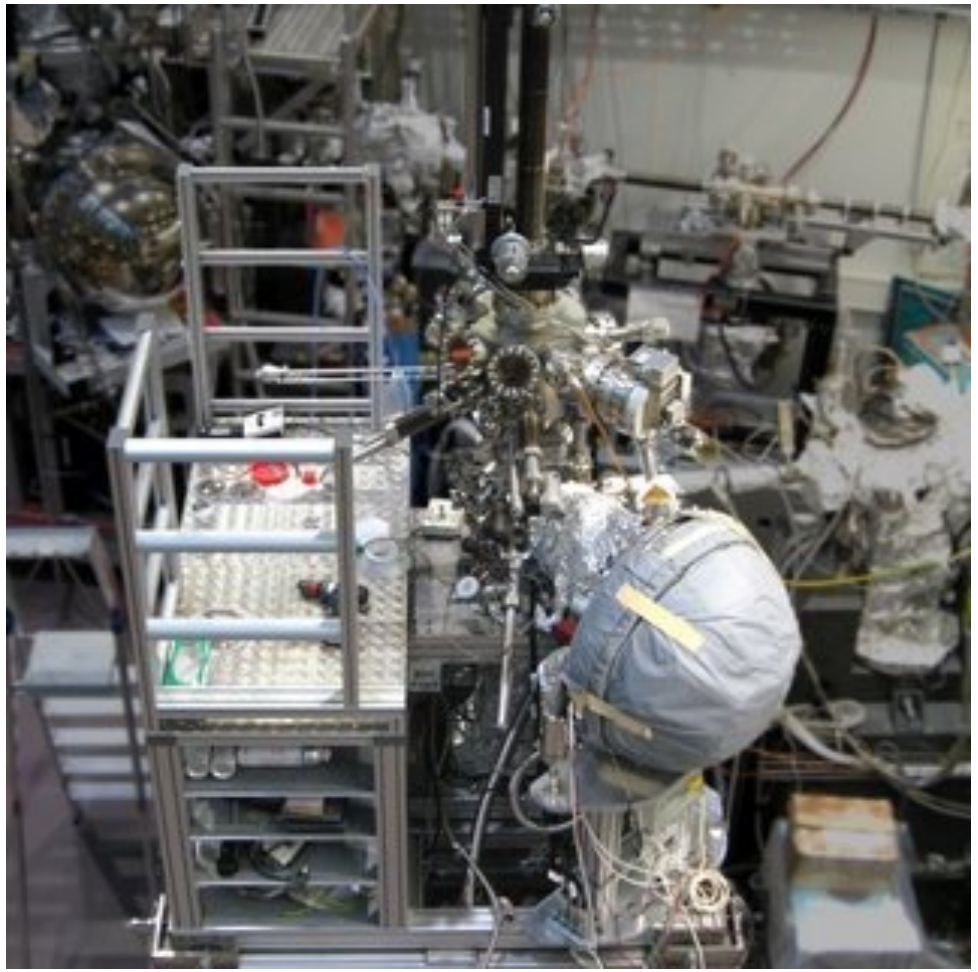


All data is organized in well-defined structures that are described by a formal schema.

Experiments

- Instrumentation
- Sample Preparation
- Sample Environment
- Monitors and Detectors
- Data Processing
- Notes

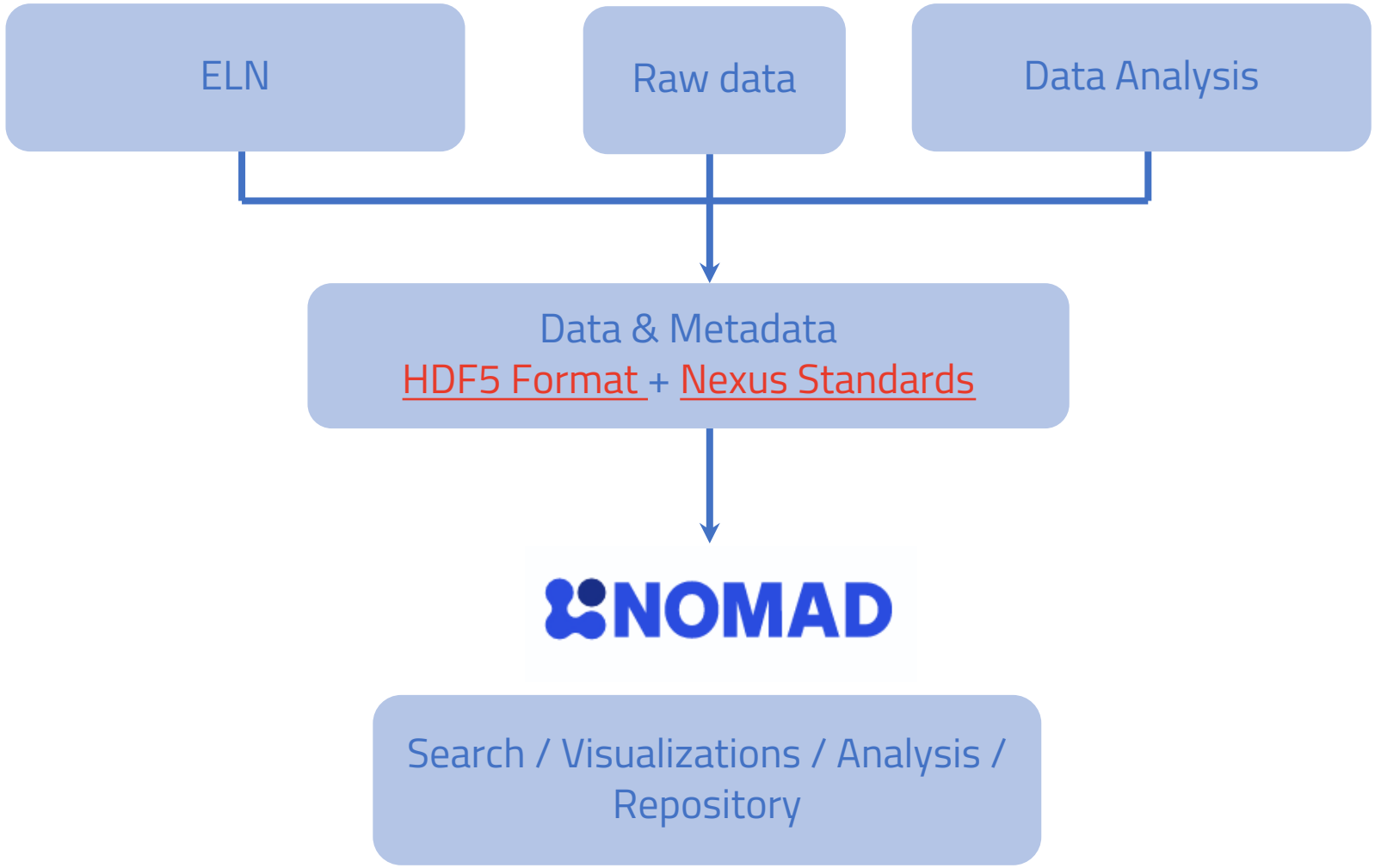




In comparison to large-scale experiments in physics,

solid-state physics is extremely heterogeneous.

Workflow for experimental data



Community standards: the NEXUS proposal

NeXus: ≈30 year-old description schema for experiments in the electron/neutron/photon beam community (aligned with Daphne4NFDI).

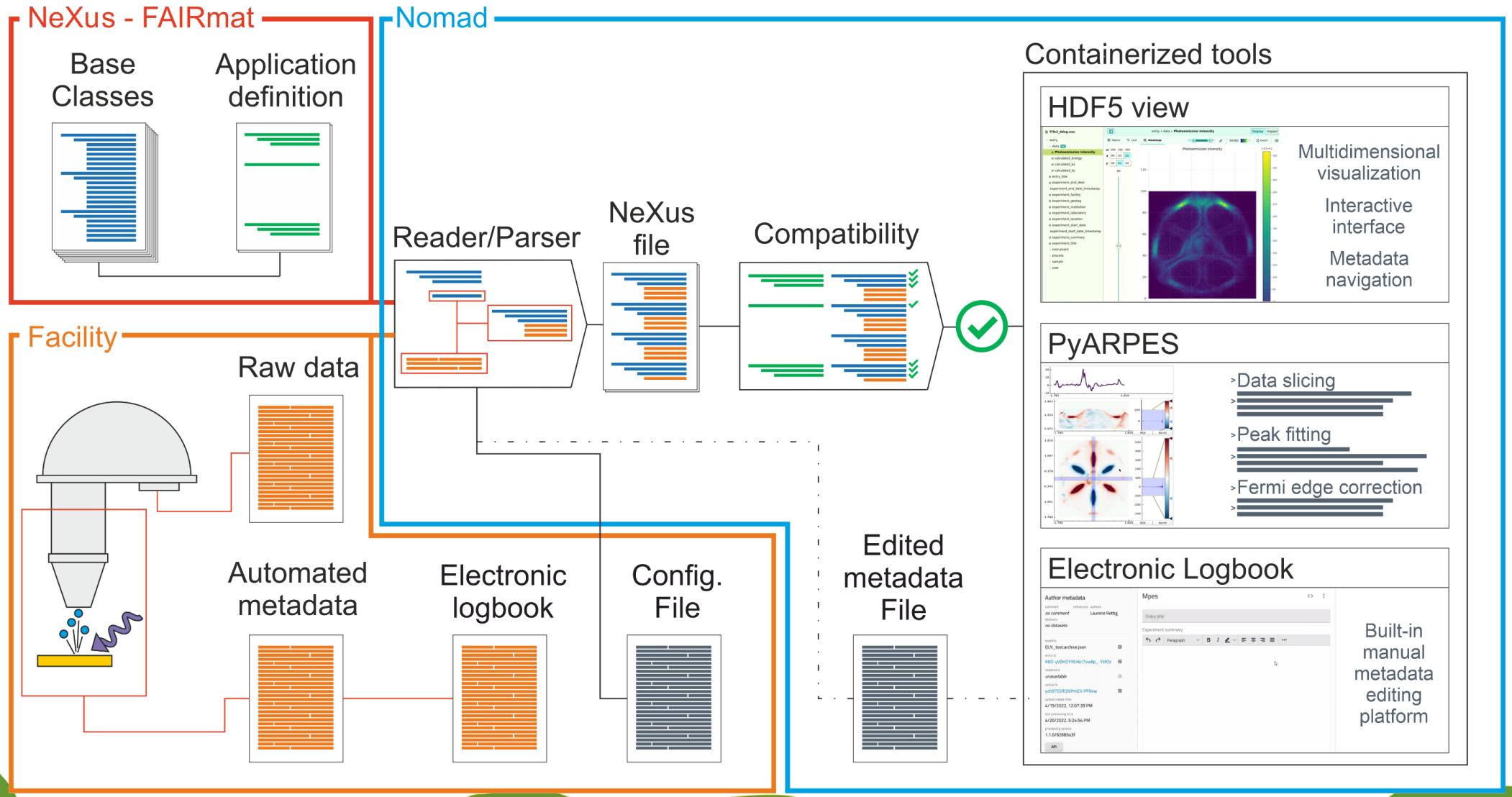
FAIRmat contributes most actively to new standards.

Pre-harmonization of experimental data

https://fairmat-nfdi.github.io/nexus_definitions/mpes-workshop-jan24

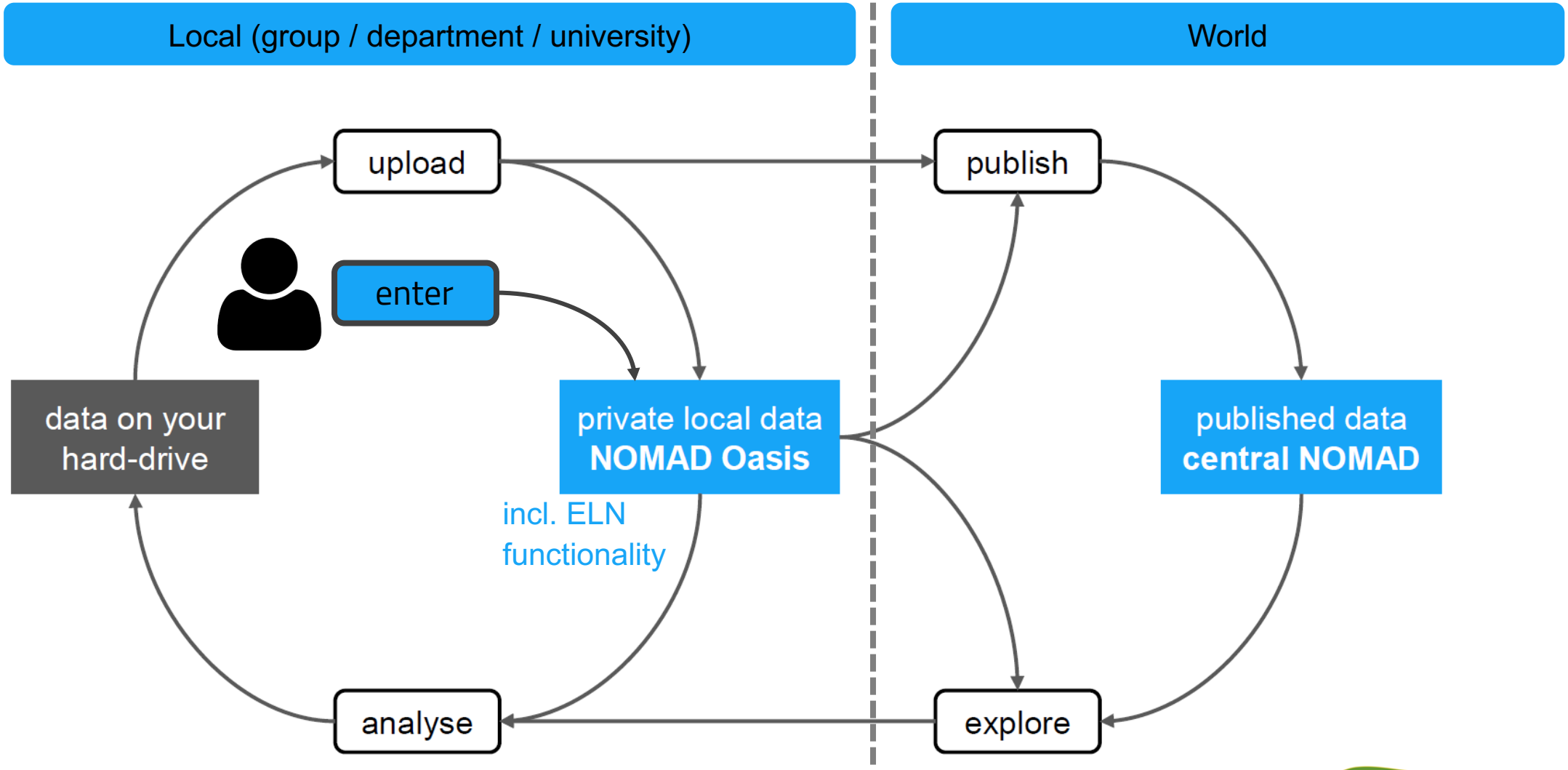
PaNET ontology as taxonomy for experiment types.

Photoemission workflow



NOMAD OASIS

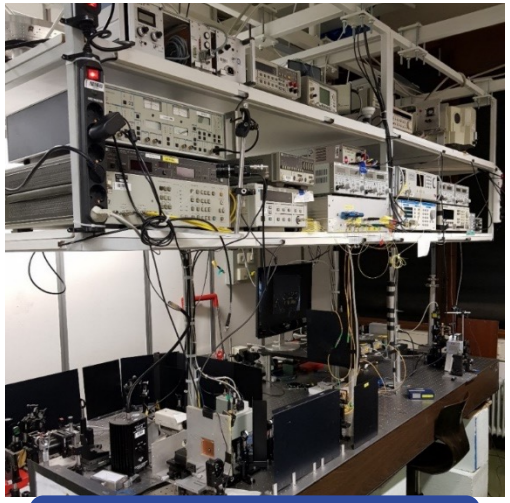
NOMAD



Further fields of action

- Configurable Lab Control Software (CAMELS)
- Electronic Laboratory Notebooks (ELNs)
- Organizing community standards
- Driving forward concepts for data literacy

NOMAD CAMELS concept



Experimental setup

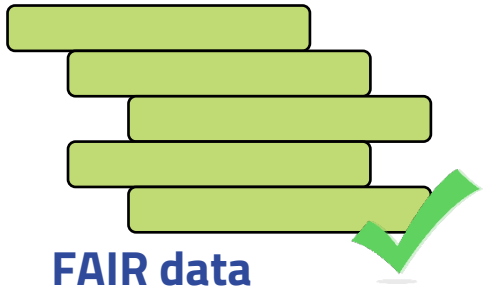
Communication:
- VISA
- EPICS
- etc …

NOMAD CAMELS

GUI

User

- No programming skills required
- Fast configuration of measurement protocol



FAIR data
- Raw data
- Metadata incl. devices, settings, protocol, …
- Structured data format (HDF5)
- Fed into NOMAD OASIS

Electronic lab notebooks (ELN)

Requirements:

- Easy input
- Easy output

Data structure first !

More important than the tools used is the **data structure** !

Data literacy in the physics curriculum

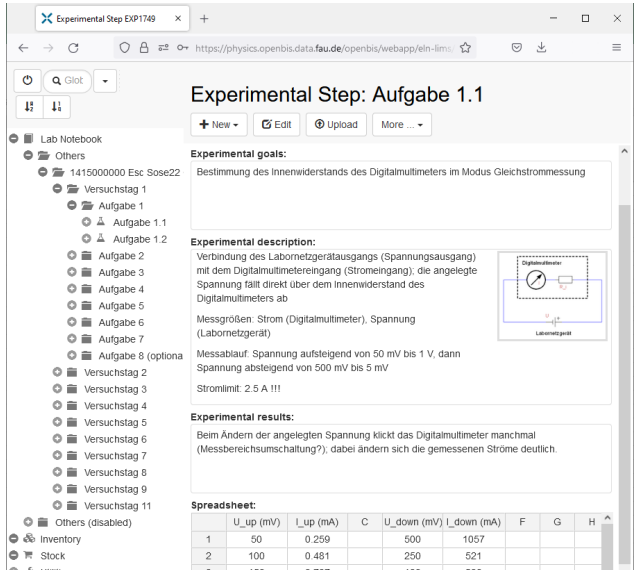
FAIRmat can not cover everything!

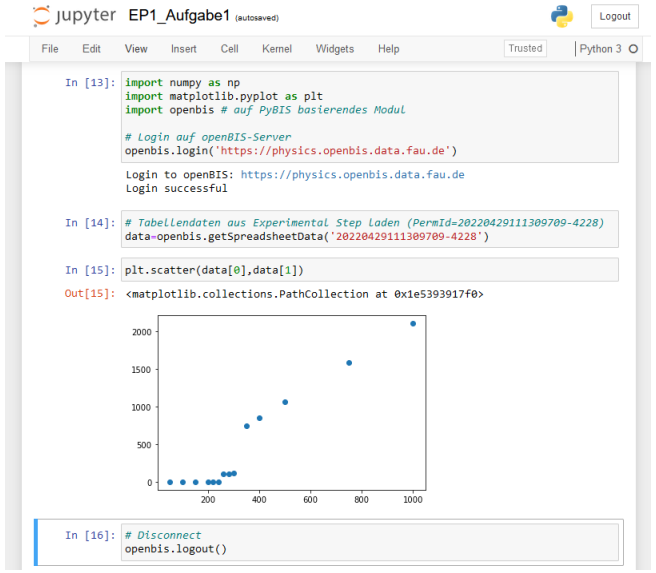
Erlangen:

- ➢ Obligatory python™ courses in the first semester
- ➢ ELNs in the lab courses

See: *Früh zur Datenkompetenz*
Physik Journal 21 (2022)

Sandbox for trying out ELN concepts

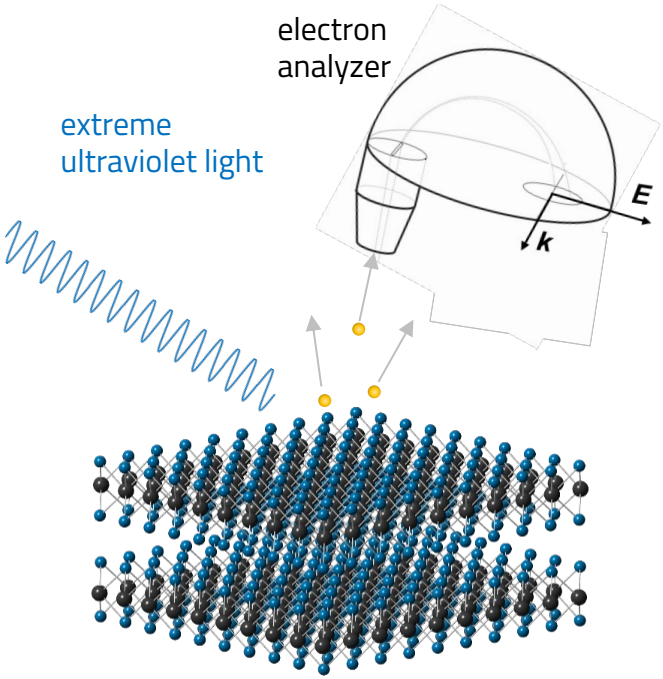




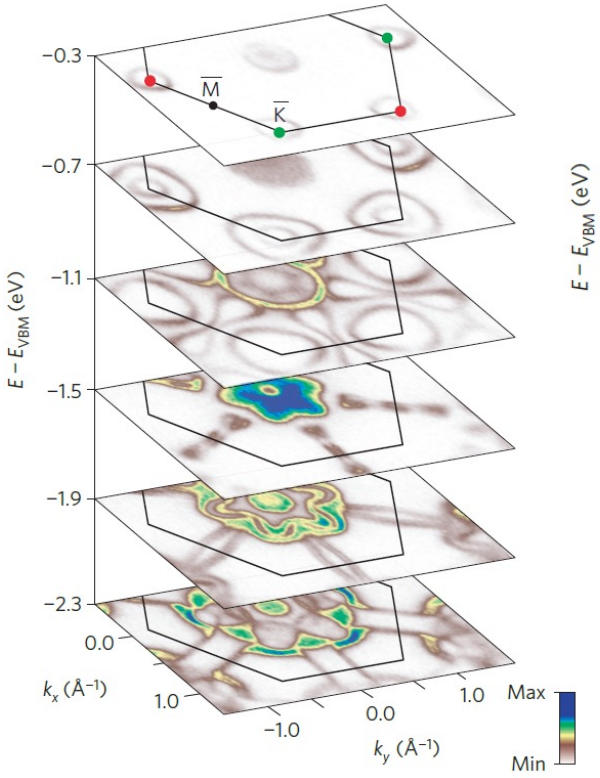
Introduction to NeXus and PES schema

Multidimensional photoemission spectroscopy

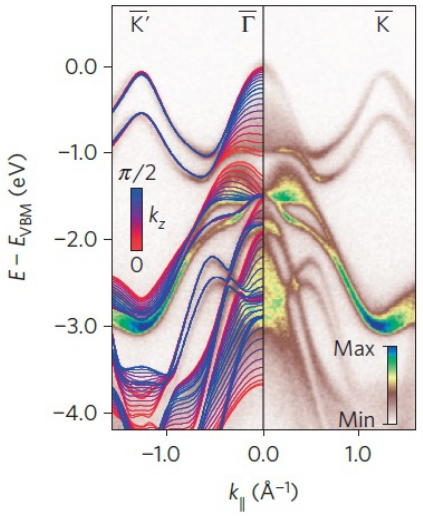
experiment



data: band mapping

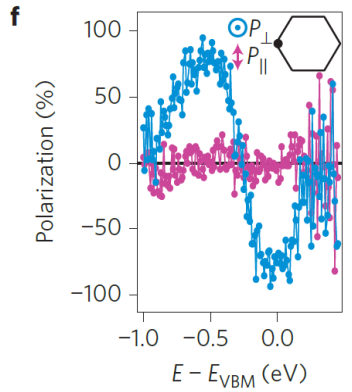


energy band dispersion



Also possible:

Spin polarization



Example: valence band structure of WSe_2
Riley *et al*, Nature Phys. 10, 835 (2014)

- momentum ***k*** *(k_x, k_y)*
- energy *E*
- spin ***S***

} "multidimensional PES" (more than 3D)

Schönhense et al., New J. Phys. 20, 033004 (2018)

Extending the parameter space

$$I \rightarrow I(\mu_{mat}, E, k_x, k_y, S, t, \omega_p, \sigma, x, y, T, \dots)$$

→ Dependence on materials and preparation procedures (μ_{mat})

→ Spin-resolving detectors: S

→ Photon energy and polarization ω_p

→ Electronic structure of **non-equilibrium states** (time-resolved ARPES):
time t, pump/probe photon energy ω_p, polarization, …

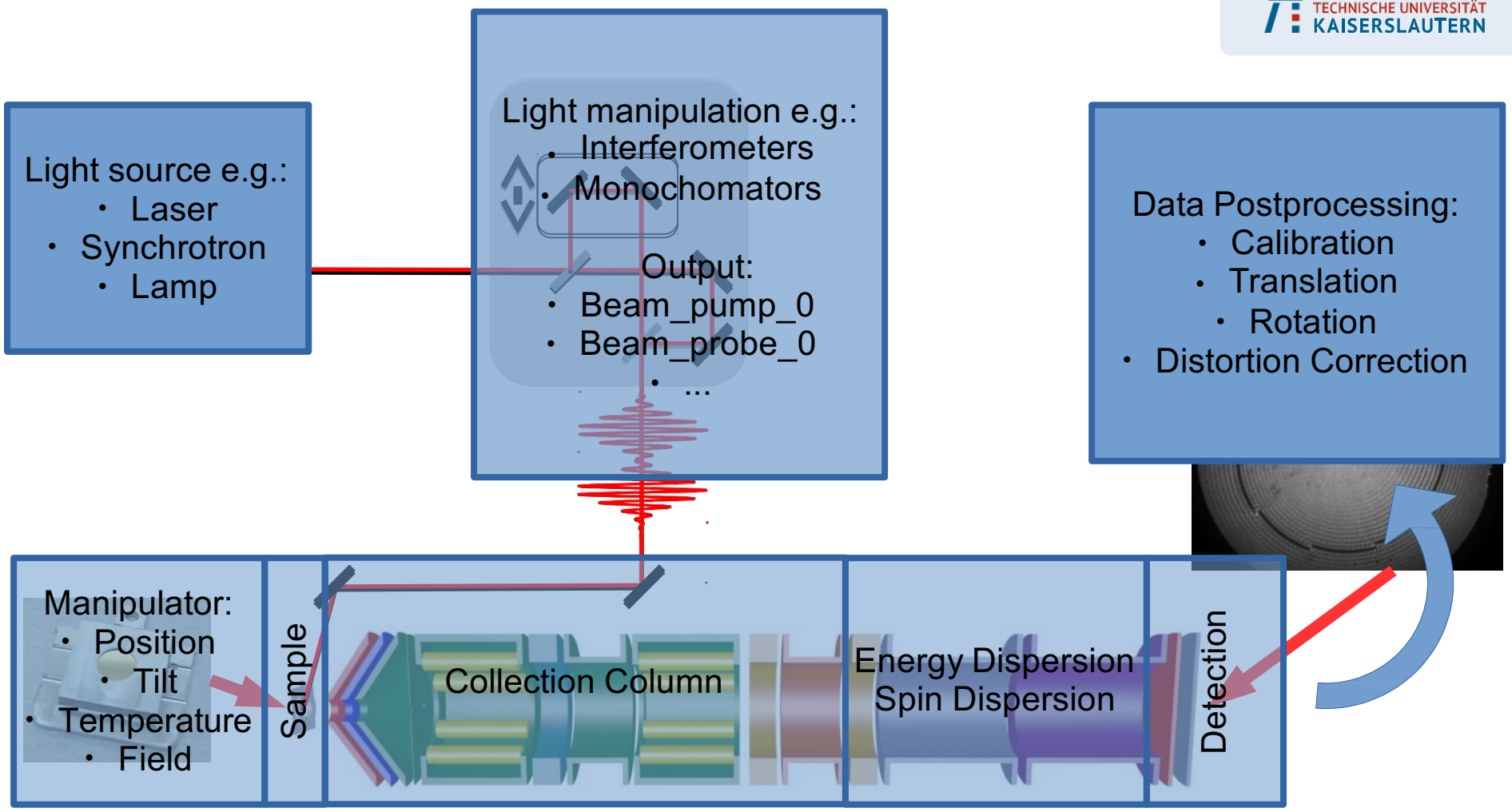
→ Dependence on sample parameters:
strain σ, sample position x, y, temperature T

→ …

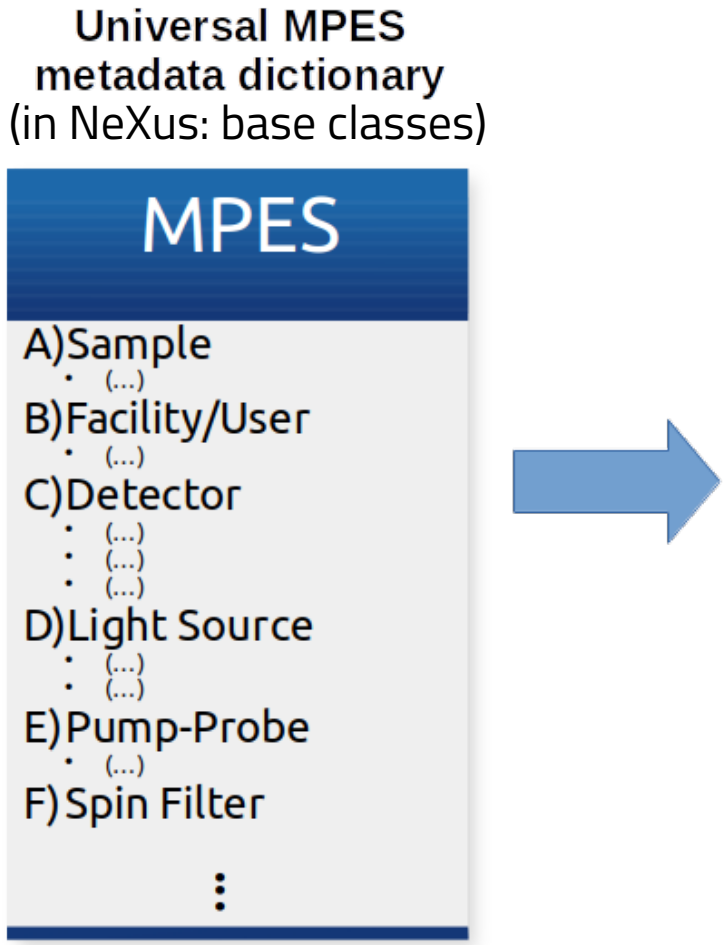
→ huge parameter space
→ Mostly only subspace both experiment-tally accessible and interesting

Goal to develop flexible, community-driven data + metadata format for such MPES data

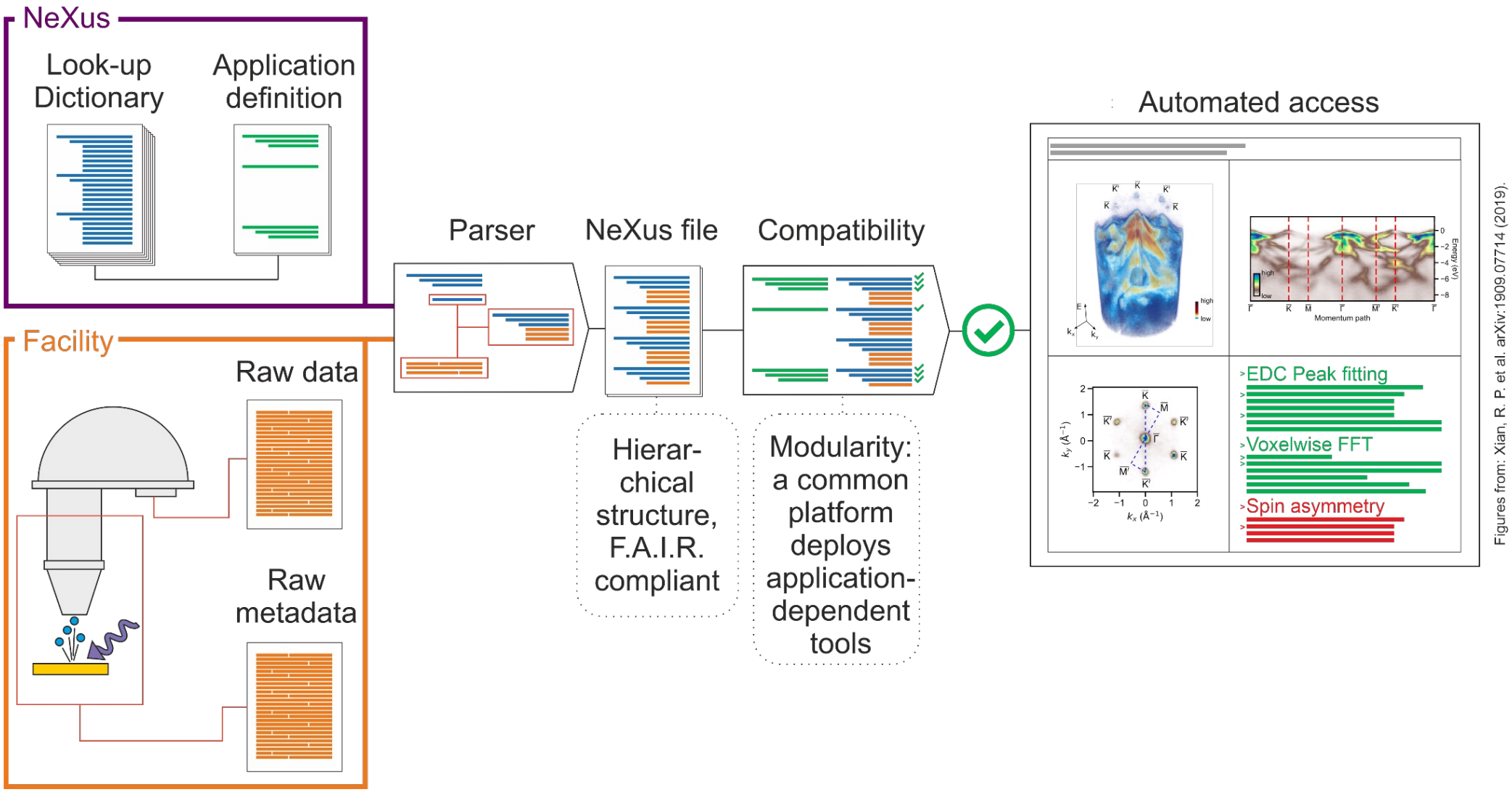
Ingredients of a photoemission experiment



Hierarchy of application definitions



Simpify analysis



Figures from: Xian, R. P. et al. arXiv:1909.07714 (2019).

Application definition: NXmpes

Application Definitions

We created two new application definitions:

NXmpes:
A general appdef with minimalistic metadata requirements, apt to describe all photemission experiments.

Base Classes

We developed entirely new base classes:

NXelectronanalyser:
A base class to describe electron kinetic energy analizers. Contains the collective characteristics of the instrument such as energy resolution, and includes the following subclasses:

NXcollectioncolumn:
Base class to describe the set of electronic lenses in the electron collection column (standard, PEEM, momentum-microscope, etc.).

NXenergydispersion:
Base class to describe the energy dispersion sytem (hemispherical, time-of-flight, etc.).

NXspindispersion:
Base class to describe the set of electronic lenses in the electron collection column.

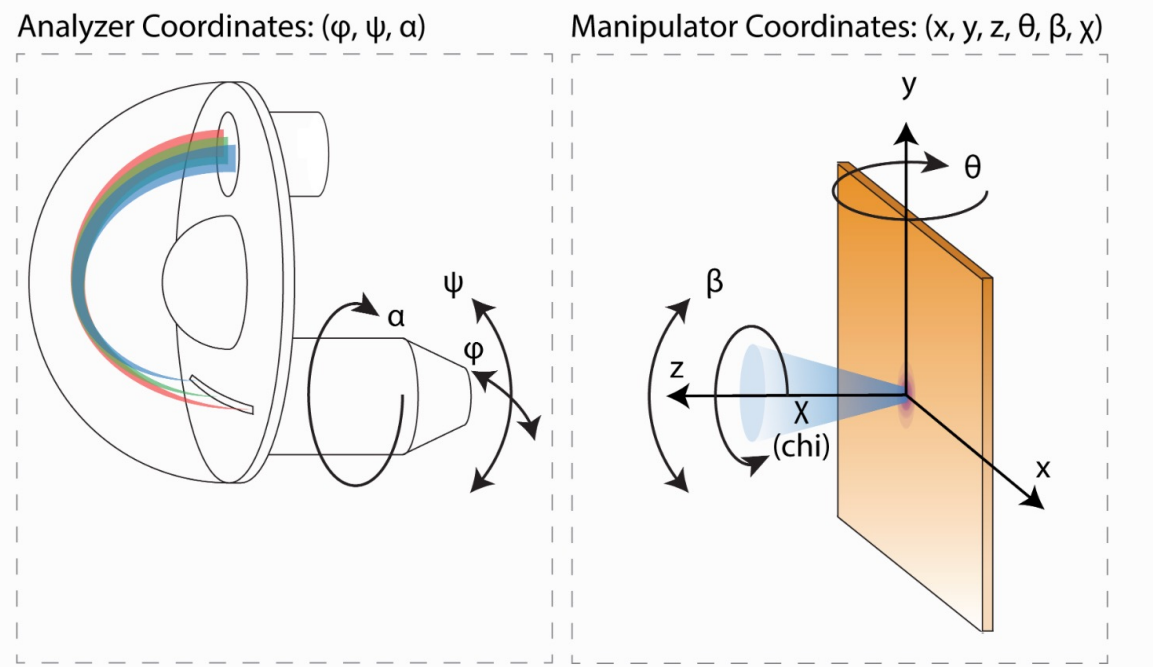
NXmanipulator:
A base class to describe the complex manipulators used in photoemission experiments, often with > 4 degrees of freedom, cryogenic cooling and other advanced features.

Nxmpes_ARPES: application definition for ARPES data

https://fairmat-nfdi.github.io/nexus_definitions/mpes-workshop-jan24

Import to pyARPES

- We drafted a mechanism to automatically interpret collected data
- Test implementation with the open source software pyArpes



Hemispherical Analyzer Angular Conventions

ELSEVIER

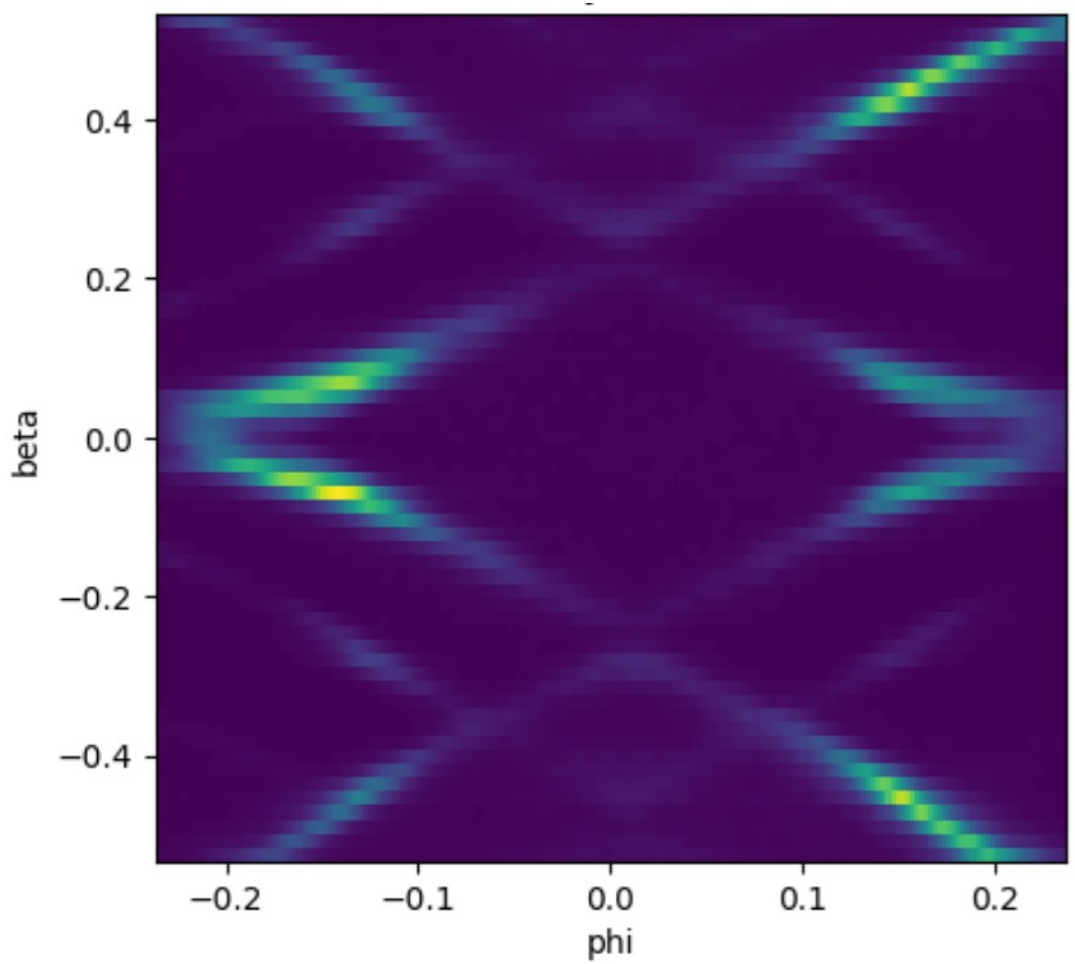
SoftwareX

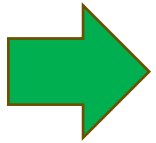
Volume 11, January–June 2020, 100472

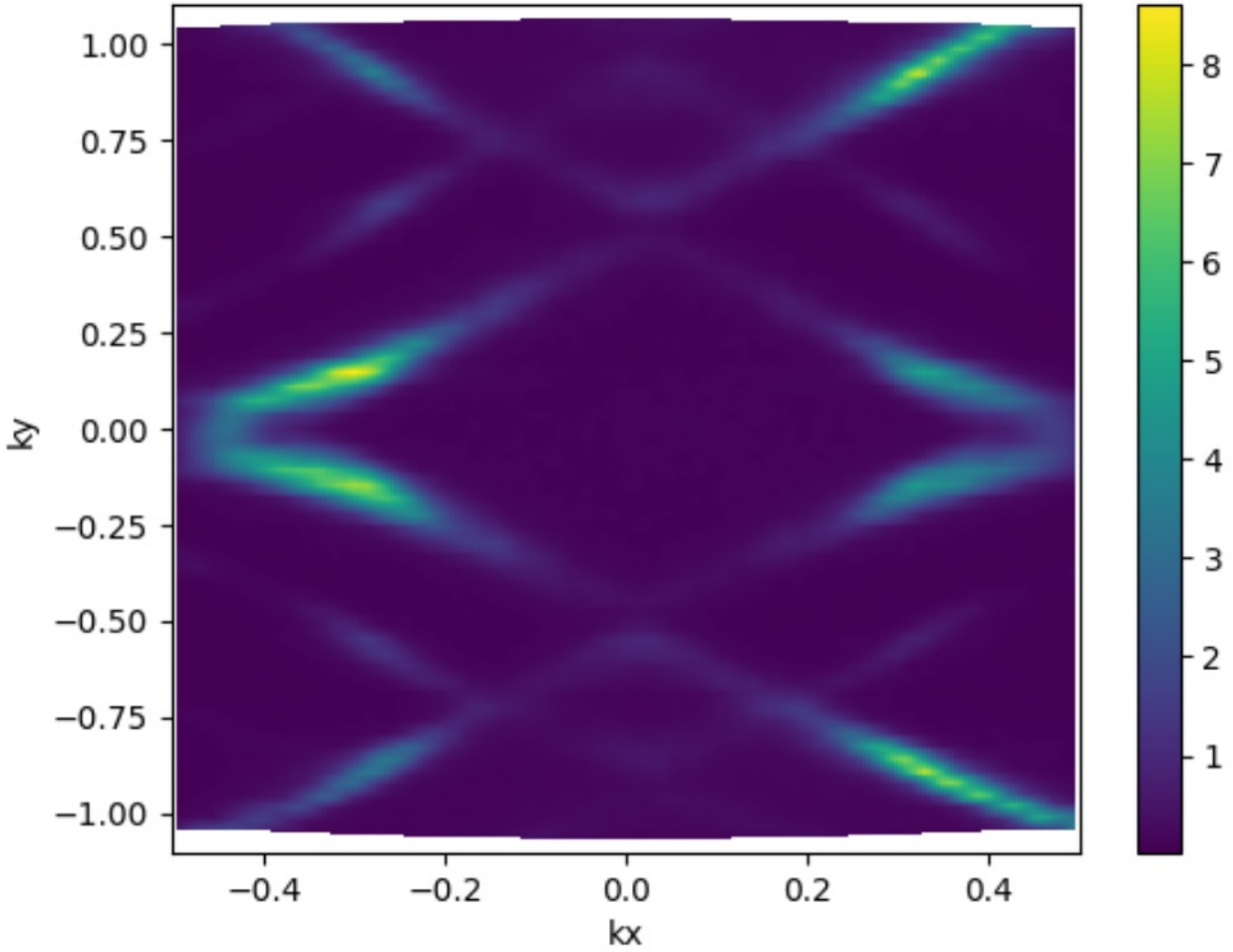
Original software publication

PyARPES: An analysis framework for

Conversion between angle and k-space





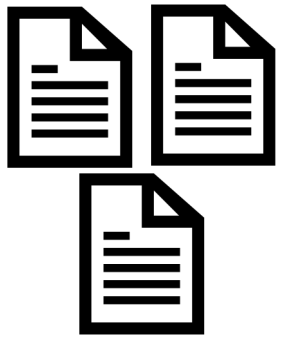


Loading of Nxmpes_ARPES files
Automatically detect and assign angles and coordinates

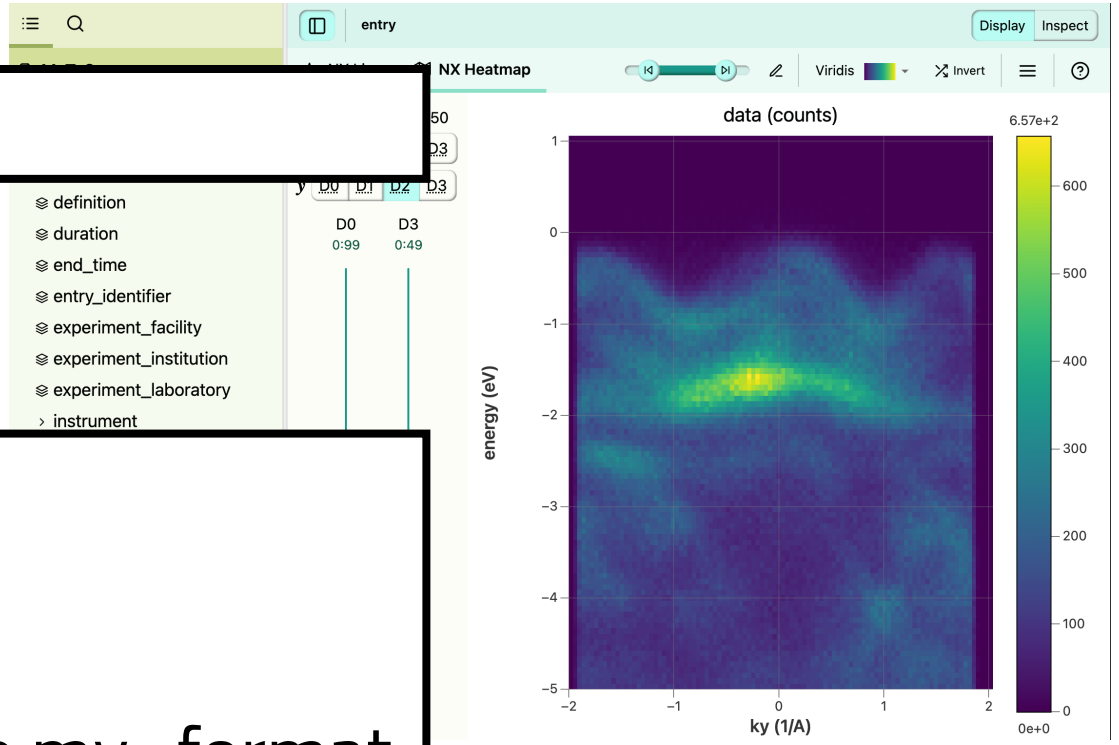
Conversion from angles to kspace for various geometries

Overview of tools

Pynxtools converts data into NeXus



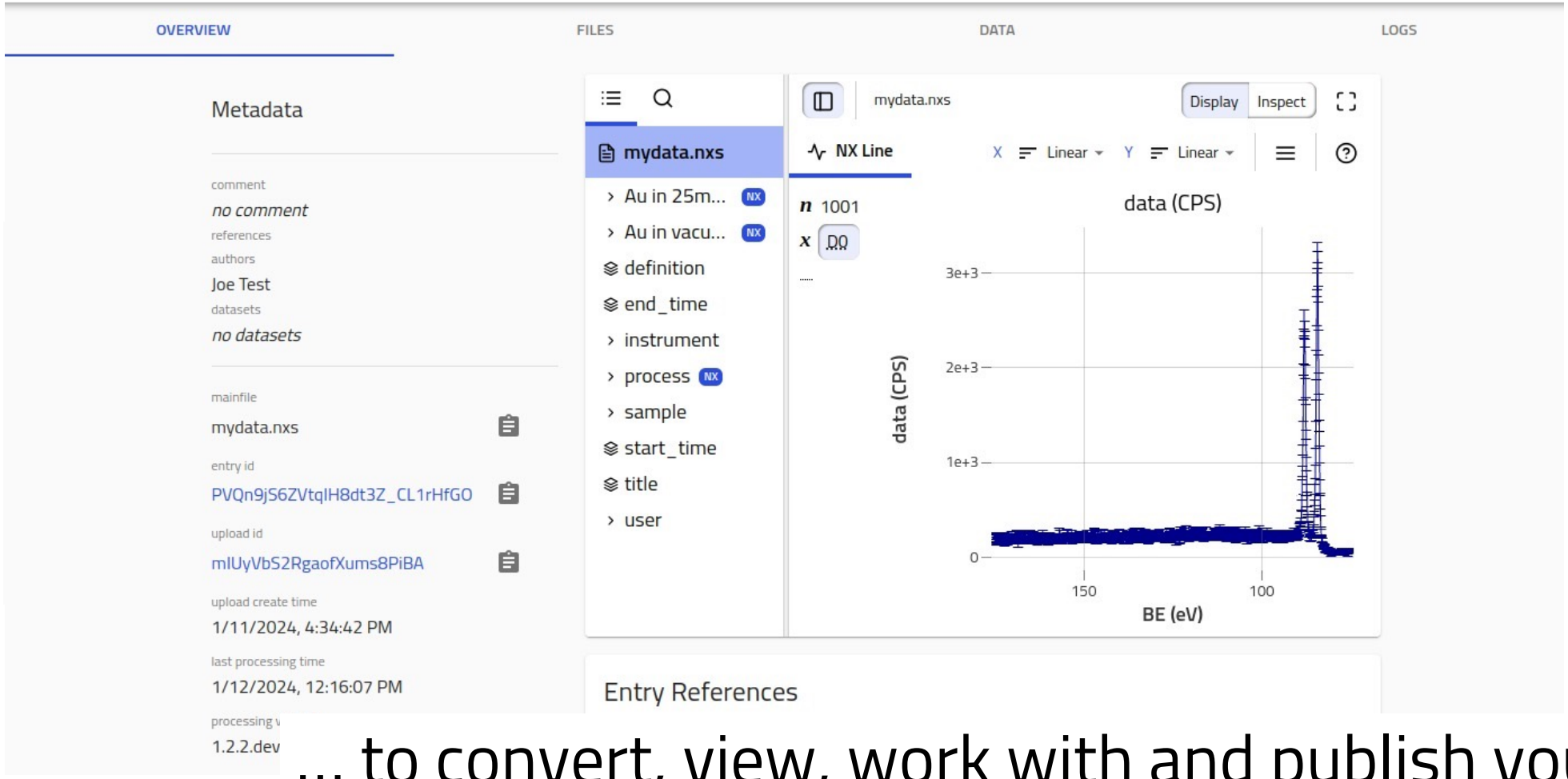
```
> pip install pynxtools
```




```
> dataconverter
        --reader mpes
        --nxdl NXmpes
        --input-file input_file.my_format
        --input-file config_file.json
```

https://fairmat-nfdi.github.io/pynxtools/tutorial/converting-data-to-nexus.html

NeXus is integrated with NOMAD



… to convert, view, work with and publish your data

https://fairmat-nfdi.github.io/pynxtools/tutorial/nexus-to-nomad.html

A short introduction to NeXus notation

INSTRUMENT: (required) NXinstrument ⇦

Freely choosale name

beam_TYPE: (required) NXbeam ⇦

All names of the form beam_*, e.g., beam_probe, beam_pump

energy_calibration: (optional) NXcalibration ⇦

Fixed name

Application defs vs. Base classes

Application definitions

- Collects required and/or additional fields
- Fields are required if not stated otherwise
- Re-uses base classes

Base classes

- A collection of terms
- All fields are optional

The contents of the NXmpes appdef

User	| Experiment user(s)
Instrument	| Instrument metadata
Sample	| Sample metadata
Process	| Post-processing information
Data	| Numerical data

FAIRmat

Addressed issues from the last workshop

Addressed issues

- Connection to ISO standards
- Proper referencing of calibrations
- Calibration classes for atom-level energy calibration
- General representation of resolutions
- Improvements for sample metadata (e.g., storage of preparation history)
- Constructs for setting and reading monitoring sensor values
- Generally less required fields
- Support for writing partial appdefs (e.g., only a detector for NXmpes)
- Drafts for tailored PEEM, ARPES, XPS, Liquidjet appdefs
- Draft for connection to evaluation software (Nxmpes_arpes + pyarpes)

NXmpes refers to ISO standards

- Implemented a standard way of referencing other standards in NeXus
- Nxmpes references ISO 18115-1:2023

```
energy_resolution(NXresolution):
  exists: recommended
  doc:
  - |
    Overall energy resolution of the MPES instrument
  - |
    xref:
      spec: ISO 18115-1:2023
      term: 10.7 ff.
      url: https://www.iso.org/obp/ui/en/#iso:std:iso:18115:-1:ed-3:v1:en:term:10.7
  - |
    xref:
      spec: ISO 18115-1:2023
      term: 10.24
      url: https://www.iso.org/obp/ui/en/#iso:std:iso:18115:-1:ed-3:v1:en:term:10.24
```

10.7

relative resolution of a spectrometer

<angle, energy, frequency, mass, time, velocity, wavelength, wavenumber> ratio of the *resolution of a spectrometer* (10.6) at a specified *measurand* (3.3) value to that measurand value

10.24

energy resolution

quantity which describes the minimum distinguishable energy separation between peaks or regions in an energy spectrum

NXcalibration collects calibration data

Defined calibration processes (based on ISO standards)

energy_calibration: (optional) NXcalibration ⇆

▼ Calibration event on the energy axis. ...

Calibration event on the energy axis.

For XPS, the calibration should ideally be performed according to ISO 15472:2010 specification.

calibrated_axis: (recommended) NX_FLOAT ⇆

This is the calibrated energy axis to be used for data plotting.

angular_calibration: (optional) NXcalibration ⇆

calibrated_axis: (recommended) NX_FLOAT ⇆

This is the calibrated angular axis to be used for data plotting.

spatial_calibration: (optional) NXcalibration ⇆

calibrated_axis: (recommended) NX_FLOAT ⇆

This is the calibrated spatial axis to be used for data plotting.

momentum_calibration: (optional) NXcalibration ⇆

calibrated_axis: (recommended) NX_FLOAT ⇆

This is the momentum axis to be used for data plotting.

energy_referencing: (optional) NXcalibration ⇆

▶ For energy referencing, the measured energies are corrected for the chargi ...

reference_peak: (required) NX_CHAR

▼ Reference peak that was used for the calibration. ...

Reference peak that was used for the calibration.

For example: adventitious carbon | C-C | metallic Au | elemental Si | Fermi edge | vacuum level

binding_energy: (recommended) NX_FLOAT

▼ The binding energy (in units of eV) that the specified emission line app ...

The binding energy (in units of eV) that the specified emission line appeared at, after adjusting the binding energy scale.

This concept is related to term 12.16__ff_ of the ISO 18115-1:2023 standard.

offset: (recommended) NX_FLOAT ⇆

▼ Offset between measured binding energy and calibrated binding energy of ...

Offset between measured binding energy and calibrated binding energy of the emission line.

calibrated_axis: (recommended) NX_FLOAT ⇆

▶ This is the calibrated energy axis to be used for data plotting. ...

level: (recommended) NXelectron_level

Electronic core or valence level that was used for the calibration.

transmission_correction: (optional) NXcalibration ⇆

▶ In the transmission correction, each intensity measurement for electrons o ...

transmission_function: (recommended) NXdata ⇆

▼ Transmission function of the electron analyser. ...

Transmission function of the electron analyser.

The transmission function (TF) specifies the detection efficiency for electrons of different kinetic energy passing through the electron analyser.

@signal: (required) NX_CHAR ⇆

Obligatory value: `relative_intensity`

@axes: (required) NX_CHAR ⇆

Obligatory value: `kinetic_energy`

kinetic_energy: (required) NX_FLOAT (Rank: 1, Dimensions: [n_transmission_function]) {units=NX_ENERGY}

Kinetic energy values

relative_intensity: (required) NX_FLOAT (Rank: 1, Dimensions: [n_transmission_function]) {units=NX_UNITLESS}

Relative transmission efficiency for the given kinetic energies

Improvements for sample metadata

Nxsample_history keeps track of sample-related activities

- Preparation history
- *In situ* treatments
- Destructive measuremens

sample_history: (recommended) NXsample_history

▼ A set of activities that occurred to the sample prior to/during photoemiss ...

A set of activities that occurred to the sample prior to/during photoemission experiment.

sample_preparation: (recommended) NXphysical_process

▼ Details about the sample preparation for the MPES experiment (e.g. UHV c ...

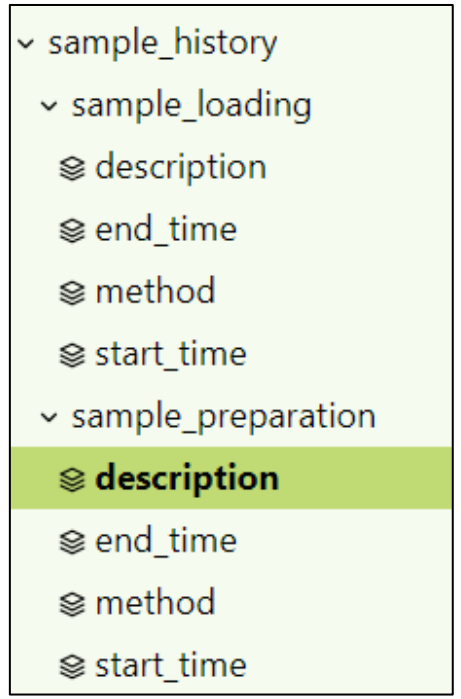
Details about the sample preparation for the MPES experiment (e.g. UHV cleaving, in-situ growth,

start_time: (required) NX_DATE_TIME ⇦

end_time: (recommended) NX_DATE_TIME ⇦

method: (recommended) NX_CHAR ⇦

Details about the method of sample preparation before the MPES experiment.



HDF5
NeXus file

NXenvironment: sensor/actuators settings and readout

Previously: all environmental conditions stored on NXsample

Now: NXsample contains *environments* linking to sensors and actuators on NXinstrument

NXmpes/ NXsample

temperature: (recommended) NXenvironment

Sample temperature (either controlled or just measured).

temperature_sensor: (required) NXsensor ⇆

▼ Temperature sensor measuring the sample temperature. ...

Temperature sensor measuring the sample temperature. This should be a link to /entry/instrument/manipulator/temperature_sensor.

sample_heater: (optional) NXactuator ⇆

▶ Device to heat the sample. ...

cryostat: (optional) NXactuator ⇆

▶ Cryostat for cooling the sample. ...

gas_pressure: (recommended) NXenvironment ⇆

Gas pressure surrounding the sample.

pressure_gauge: (required) NXsensor ⇆

▶ Gauge measuring the gas pressure. ...

NXmpes/ NXinstrument

MANIPULATOR: (optional) NXmanipulator

Manipulator for positioning of the sample.

temperature_sensor: (recommended) NXsensor ⇆

name: (recommended) NX_CHAR ⇆

measurement: (required) NX_CHAR ⇆

Obligatory value: `temperature`

type: (optional) NX_CHAR ⇆

value: (required) NX_FLOAT ⇆

sample_heater: (optional) NXactuator ⇆

name: (recommended) NX_CHAR ⇆

physical_quantity: (required) NX_CHAR ⇆

Obligatory value: `temperature`

type: (optional) NX_CHAR ⇆

heater_power: (required) NX_FLOAT ⇆

PID: (recommended) NXpid ⇆

setpoint: (recommended) NX_FLOAT ⇆

Less required fields

We decreased the total ratio of required to total fields and groups from **45%** to **15%**

Required 24 (32) Recommended 74 (29) Optional 60 (10)

Total fields and attributes: 158 (71)

NeXus supports partial appdefs

- It's as simple as adding a list of concepts to the root level **partial** attribute.
- pynxtools is able to merge multiple partial appdefs into a fully supported data file.

How does it look like?

/@partial = [NXmpes/NXinstrument/NXdetector, ...]

Proper referencing of calibrations

- Formula fits
- Linear fits
- Mapping
- External references:
  Standard procedure,
  calibration file online
  or integrated into file
- May have multiple input references

original_axis: (optional) NX_FLOAT (Rank: 1, Dimensions: [ncal]) {units=NX_ANY}

Vector containing the data coordinates in the original uncalibrated axis

@symbol: (optional) NX_CHAR

▶ The symbol of the axis to be used in the fit_function, e.g., `energy`, `E`. ...

@input_path: (optional) NX_CHAR

▶ The path from which this data is derived, e.g., raw detector axis. ...

input_SYMBOL: (optional) NX_FLOAT (Rank: 1, Dimensions: [ncal]) {units=NX_ANY}

▶ Additional input axis to be used in the formula. ...

@input_path: (optional) NX_CHAR

▶ The path from which this data is derived, e.g., raw detector axis. ...

coefficients: (optional) NX_FLOAT (Rank: 1, Dimensions: [ncoeff]) {units=NX_ANY}

▶ For non-linear energy calibrations, e.g. in a TOF, a polynomial function is fi ...

fit_function: (optional) NX_CHAR

▶ For non-linear energy calibrations. Here we can store the formula of the ...

scaling: (optional) NX_FLOAT {units=NX_ANY}

▶ For linear calibration. Scaling parameter. ...

offset: (optional) NX_FLOAT {units=NX_ANY}

▶ For linear calibration. Offset parameter. ...

mapping_MAPPING: (optional) NX_FLOAT

General representation of resolutions

- Allows to collect a
  single resolution
  or an array for each data point
- Can have a type of:
  estimated, derived, calibrated
  or other
- Support for
  response functions, formulas and calibration

▶ The physical quantity of the resolution, e.g., ...

type: (optional) NX_CHAR

▶ The process by which the resolution was determined. ...

resolution: (optional) NX_FLOAT {units=NX_ANY}

The resolution of the physical quantity.

resolution_errors: (optional) NX_FLOAT {units=NX_ANY}

Standard deviation of the resolution of the physical quantity.

formula_SYMBOL: (optional) NX_CHAR

▶ A symbol linking to another path in this appdef to be referred to from the ...

resolution_formula: (optional) NX_CHAR

▶ A resolution formula to determine the resolution from a set of symbols as ...

note: (optional) NXnote

Additional details of the estimate or description of the calibration procedure

response_function: (optional) NXdata

▶ The response of the instrument or part to a infinitesimally sharp input signal ...

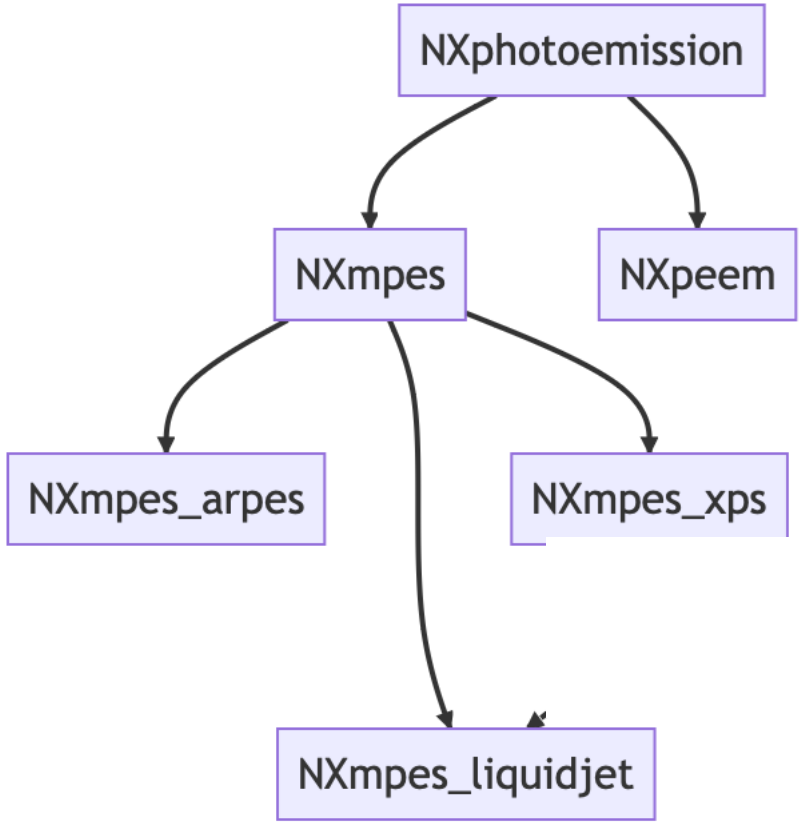
input: (optional) NX_FLOAT {units=NX_ANY}

▶ The input axis or grid of the response function. ...

magnitude: (optional) NX_FLOAT

▶ The magnitude of the response function corresponding to the points ...

The community built a hierarchy of techniques



Future plans

- Base class inheritance
- Umbrella definition NXphotoemission
- Metainformation for each entry to connect to other sources (e.g., connection to NOMAD or other NeXus entries)

www.fairmat-nfdi.eu

@FAIRmat_NFDI

fairmat@physik.hu-berlin.de

company/fairmat-nfdi

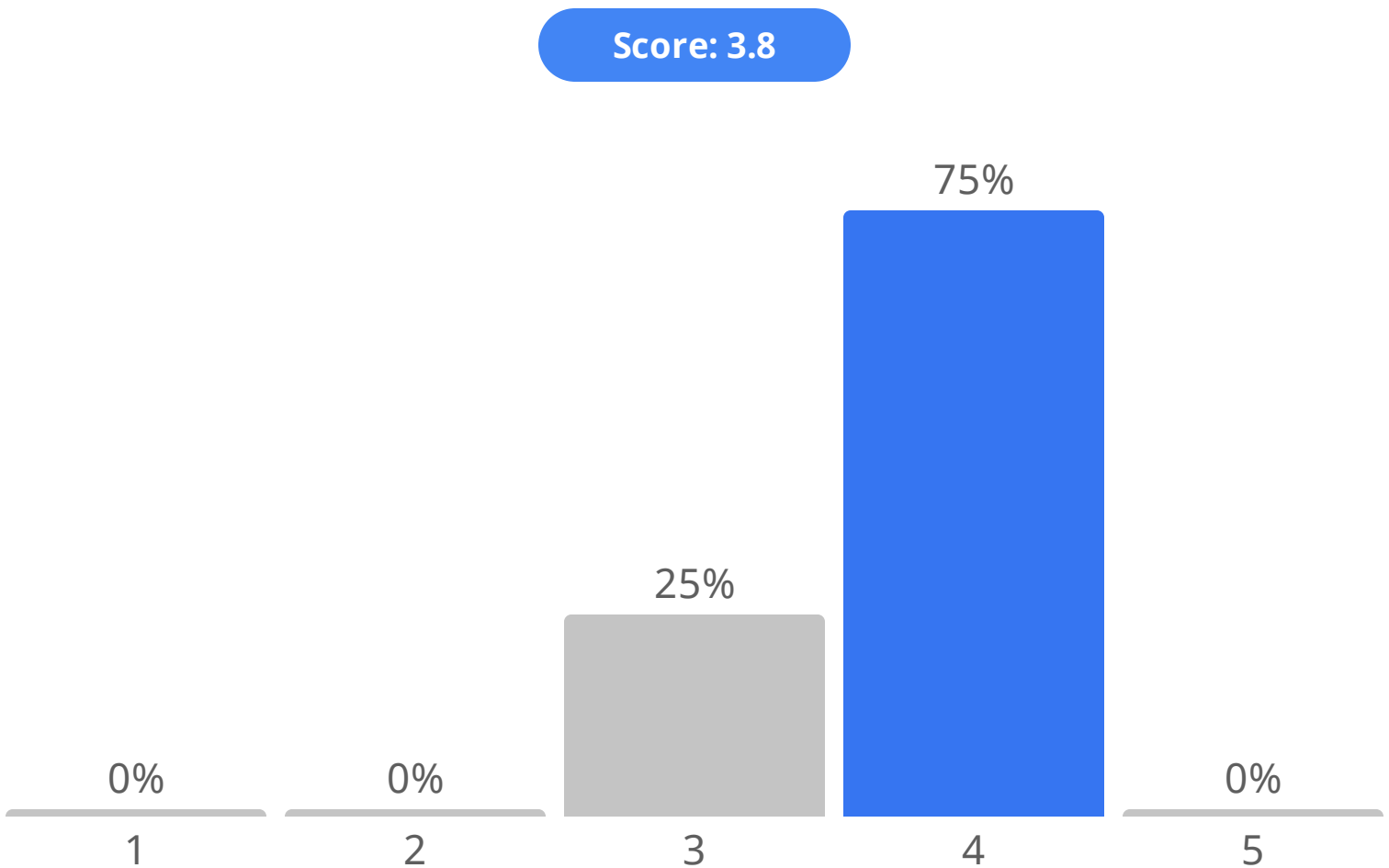
www.nomad-lab.eu

Rating poll

Survey (1/4)

Are the definitions complete and ready to use?

0 0 8

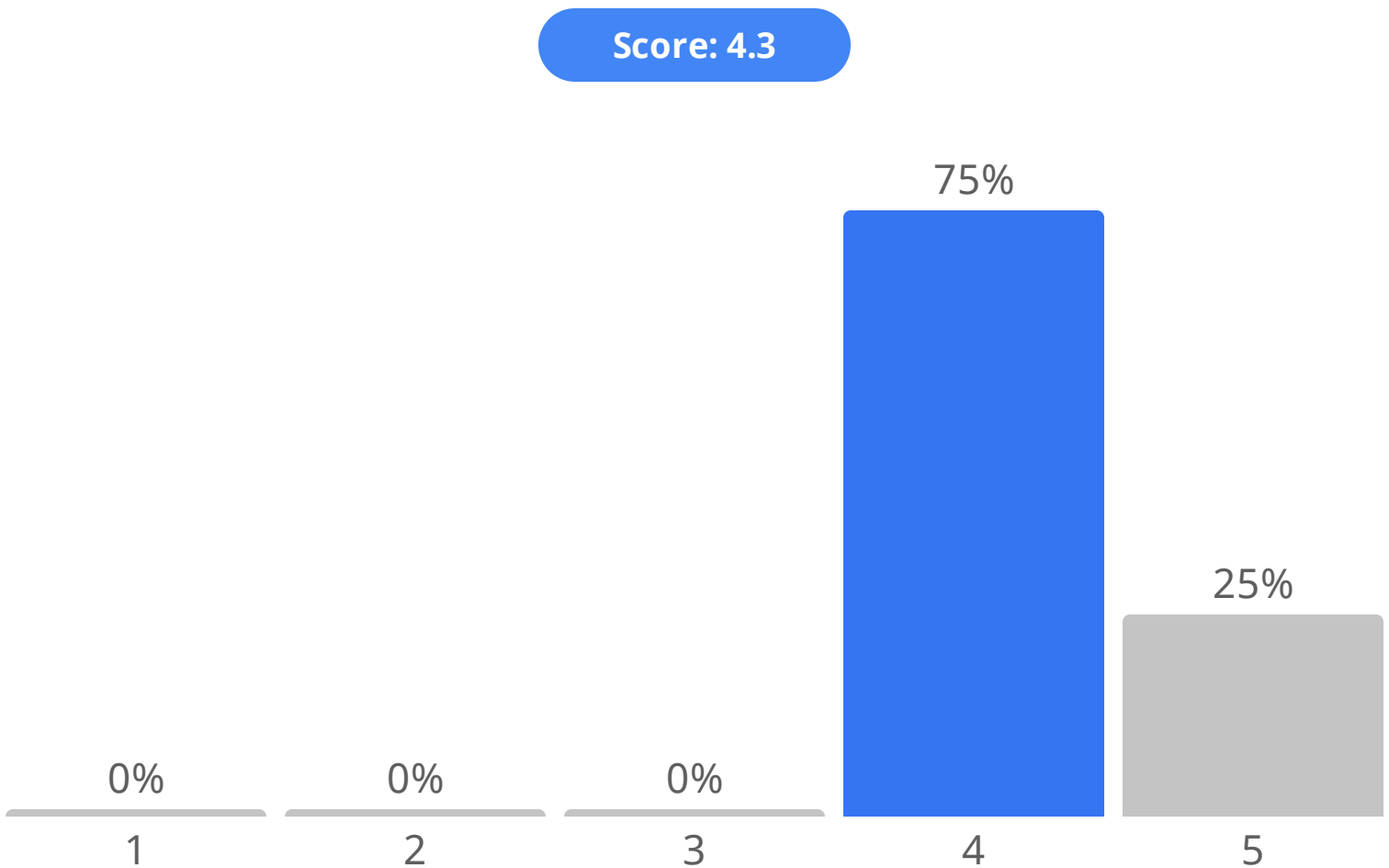


Rating poll

Survey (2/4)

Do you see major design flaws in the NXmpes?

0 0 8

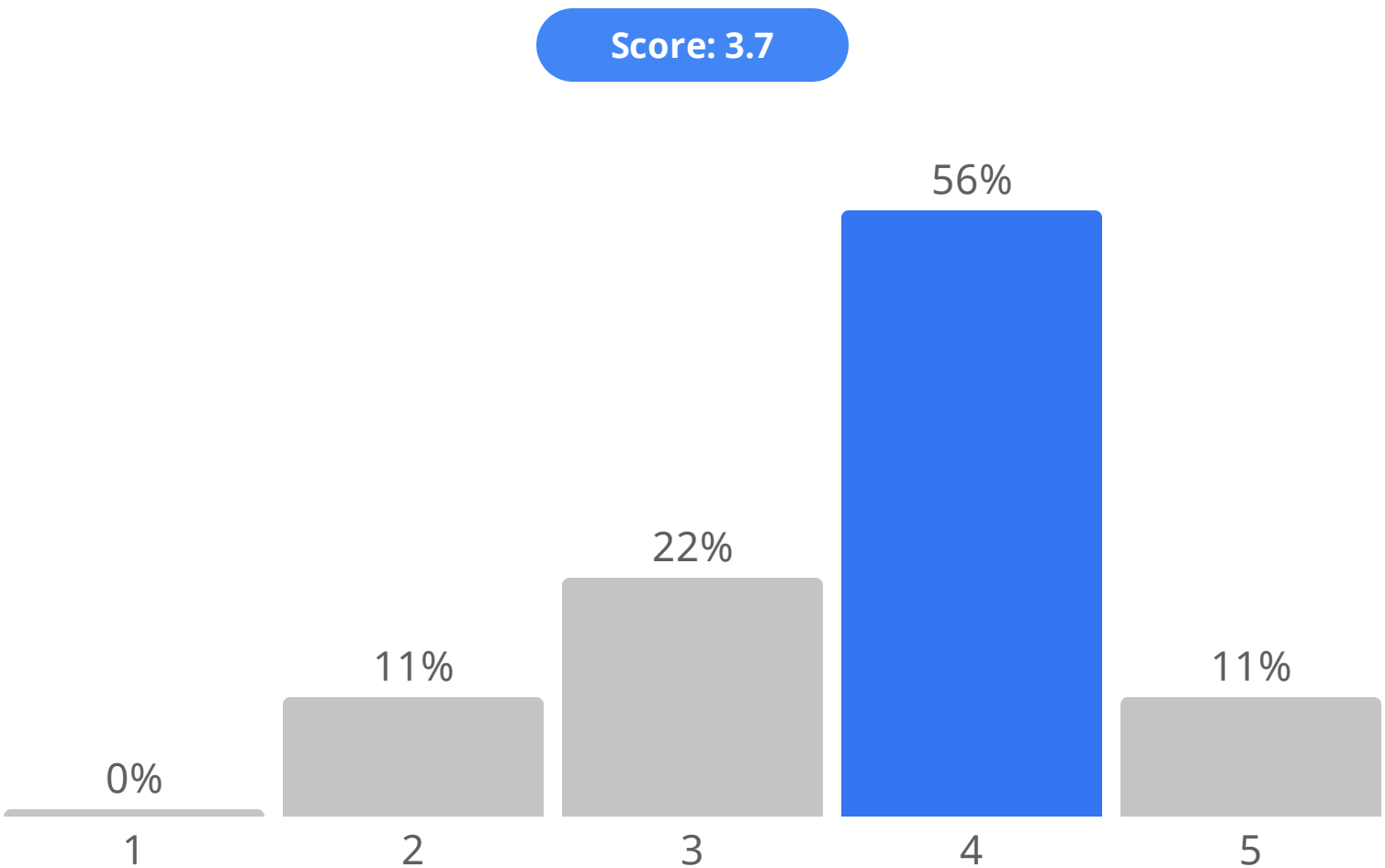


Rating poll

Survey (3/4)

NeXus is...

0 0 9



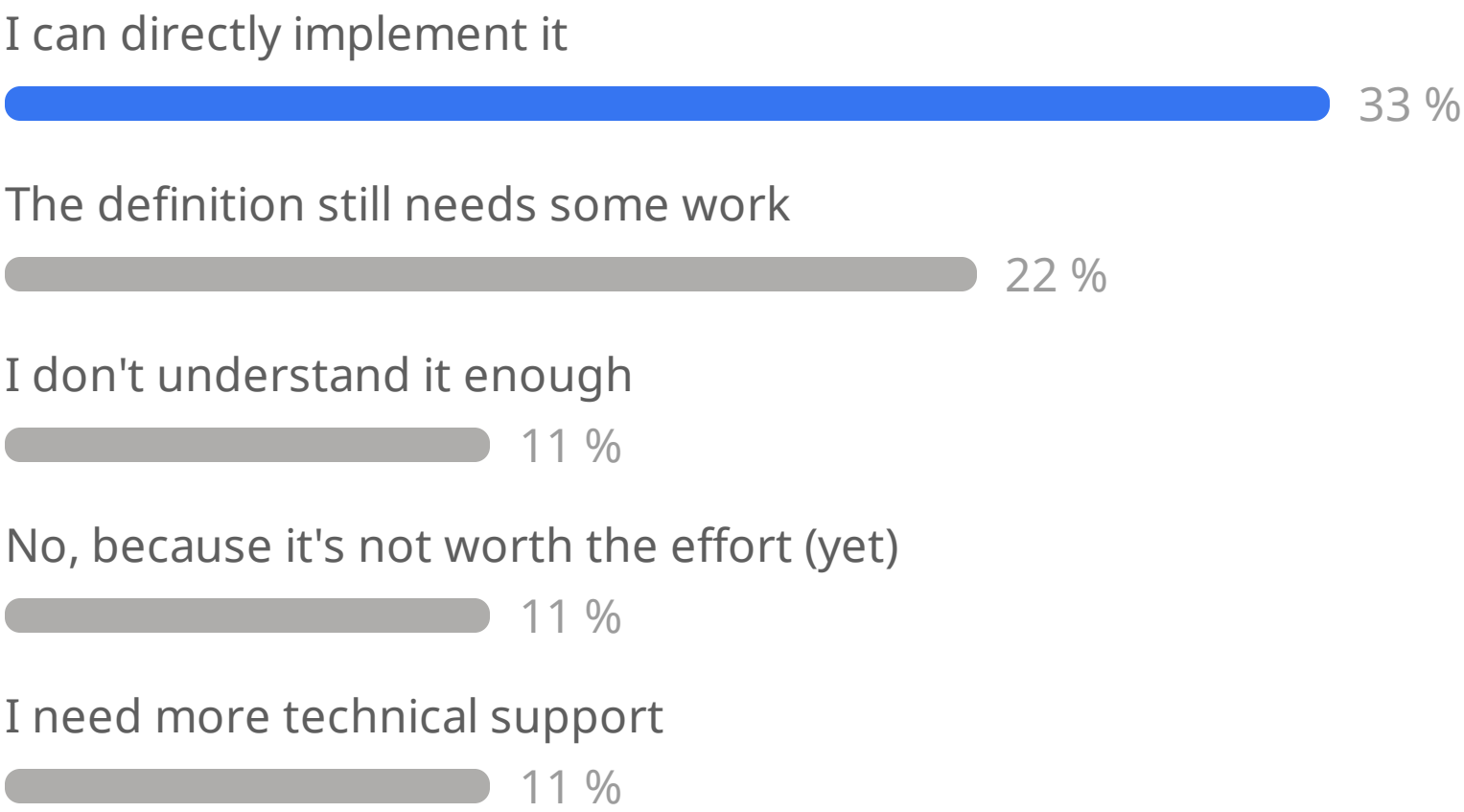
Multiple-choice poll

Survey (4/4)

0 0 9

Could you currently export to NXmpes?

(1/2)



Survey (4/4)

0 0 9

Could you currently export to NXmpes?

(2/2)

No, but for another reason

11 %