



# Towards structured data in electron spectroscopy

Heiko B. Weber and the FAIRmat team

Friedrich-Alexander-Universität Erlangen-  
Nürnberg

# Outline

- The FAIRmat approach for making experimental data FAIR
- Application definitions and NEXUS
- MPES example



# Open data

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



**FAIR principles**<sup>1</sup> (2016); statement at G20 summit in China<sup>2</sup>

**Data as resource**<sup>3</sup>: 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)

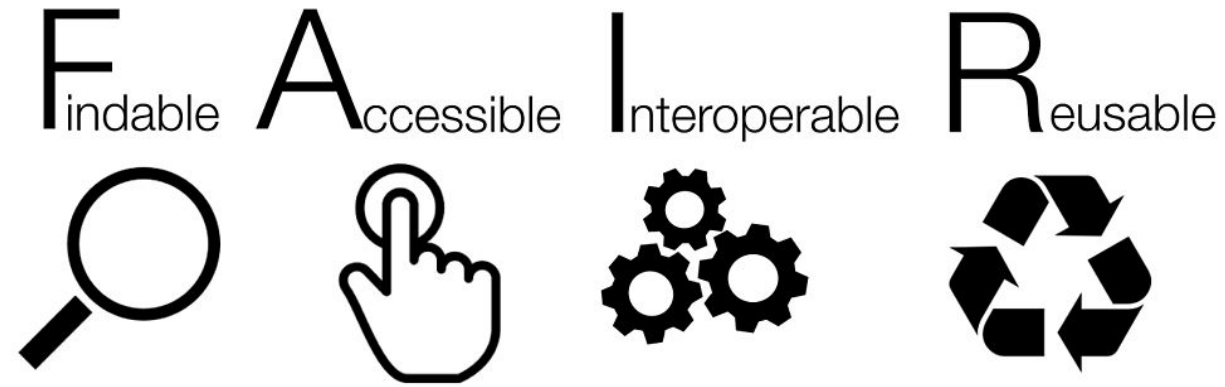
<sup>1</sup>[Scientific Data 3, 160018 \(2016\)](#)

<sup>2</sup>[https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT\\_16\\_2967](https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_16_2967)

<sup>3</sup>[https://ec.europa.eu/info/sites/default/files/communication-european-strategy-data-19feb2020\\_en.pdf](https://ec.europa.eu/info/sites/default/files/communication-european-strategy-data-19feb2020_en.pdf)



# Underlying principles for open research data



**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)

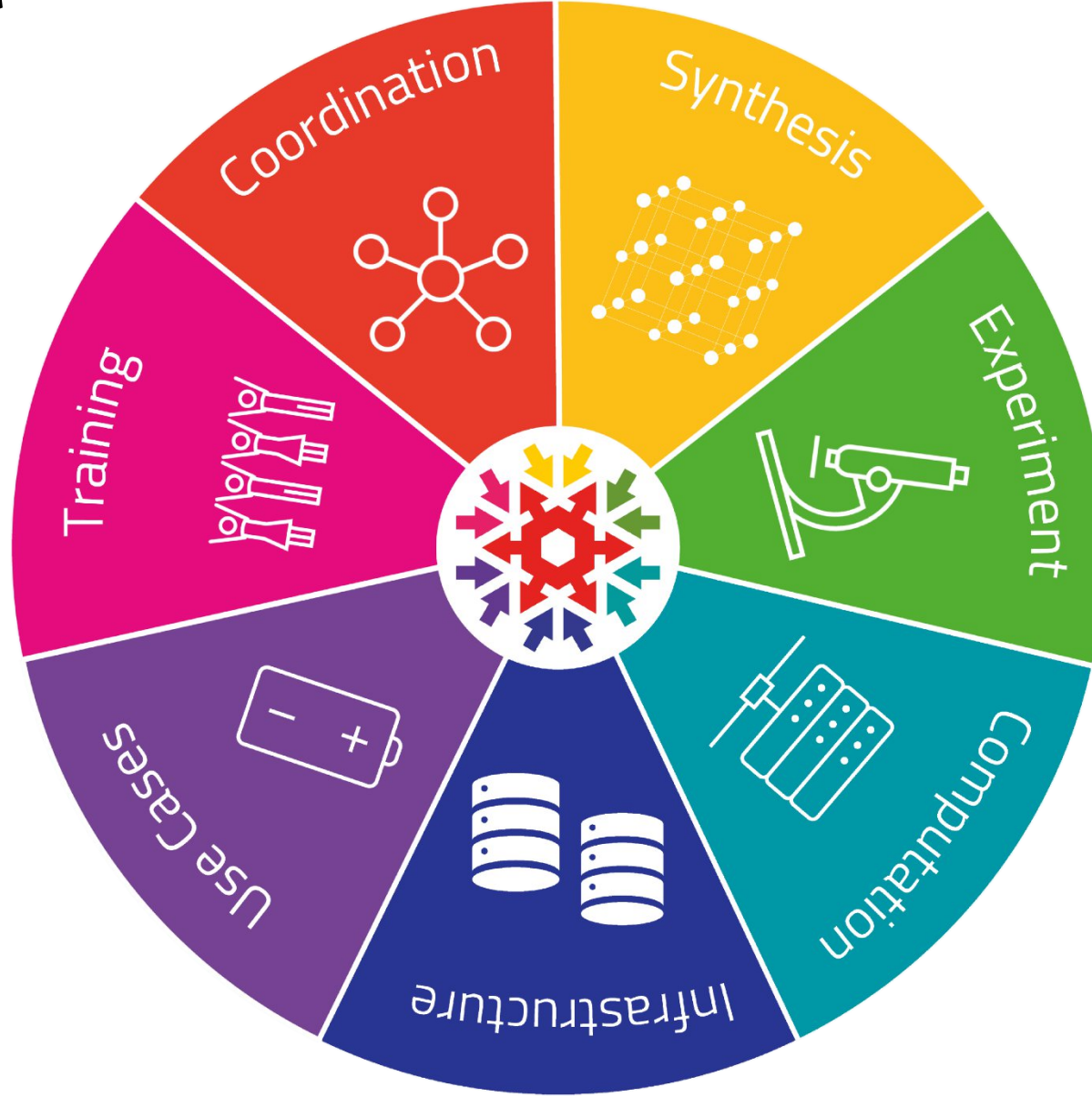
- [BERD@NFDI](#): NFDI for Business, Economic and Related Data
- [DAPHNE4NFDI](#): Data from PHoton and Neutron Experiments for NFDI
- [FAIRmat](#): FAIR Data Infrastructure for Condensed-Matter Physics and the Chemical Physics of Solids
- [MaRDI](#): Mathematical Research Data Initiative
- [NFDI4DataScience](#): 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)

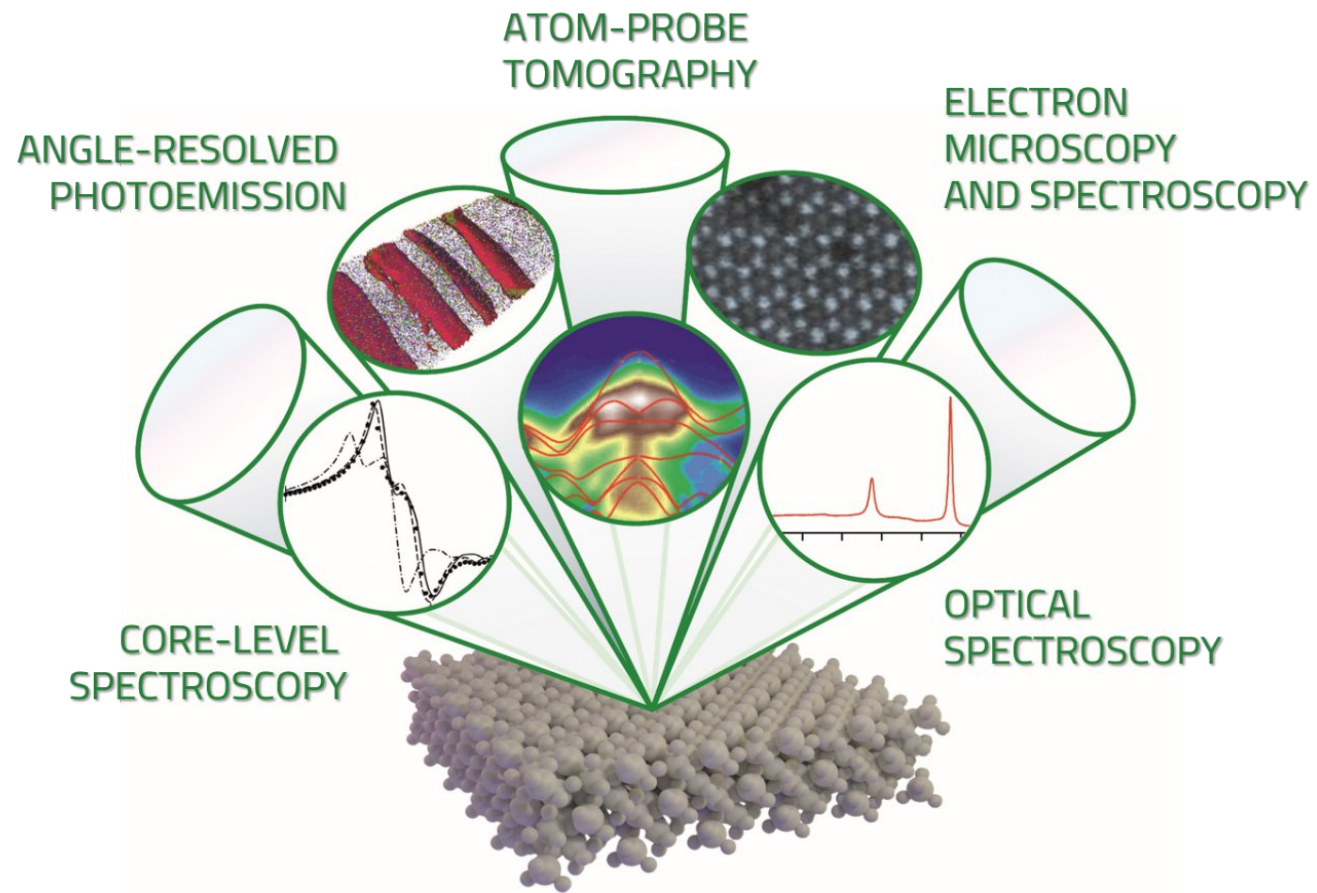
- [FAIRagro](#): FAIR Data Infrastructure for Agrosystems
- [NFDI4BIOIMAGE](#): National research data infrastructure for microscopy and bioimage analysis
- [NFDI4Energy](#): National Research Data Infrastructure for Interdisciplinary Energy System Research
- [NFDI4Immuno](#): National Research Data 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



# Area B and its pilot experiments





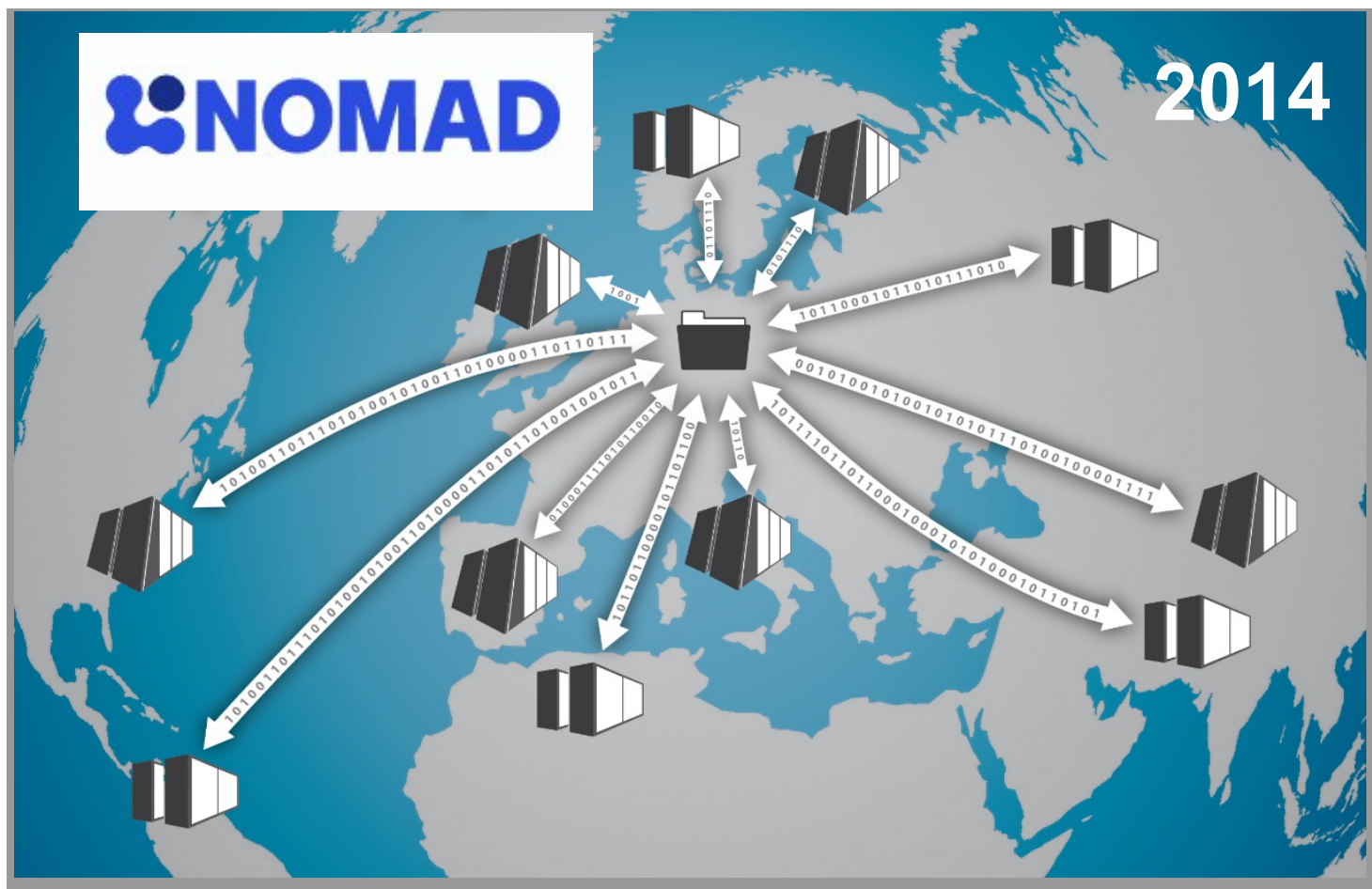
# Nomad repository

World's largest collection of computational materials science data

Make data repurposable



Recycle the waste!



The **N**ovel **M**aterials **D**iscovery Laboratory





# Computational Solid-state science

40 community software packages supported

Thus 40 parsers serve as *converters*

Normalized data in the *NOMAD Archive*

Same units, formats

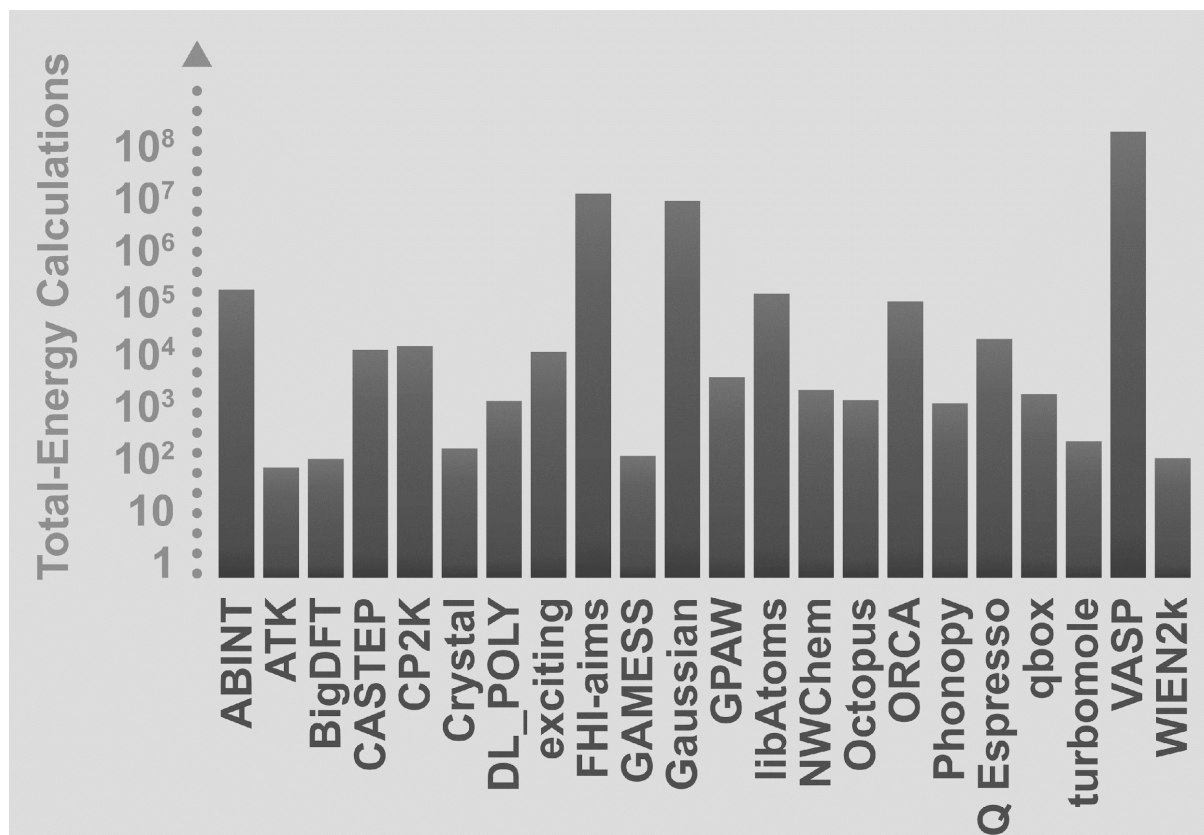
>100 mio. calculations

billions of CPU core hours  
at HPC centers, worldwide

Metadata items

several 1000

to uniquely describe all  
the computed data



# Added value – the NOMAD Encyclopedia

Forerunners



FAIRmat

AgFeO<sub>3</sub> - space group 221

### Structure

Legend: Ag (grey), Fe (orange), O (red)

Show axis    Show bonds

System type: bulk  
Space group: 221  
Structure type: CaO3Ti (0)

### Electronic structure

#### Band structure

#### DOS

From calculation 383297 (GGA - VASP)

Search: Fe & O

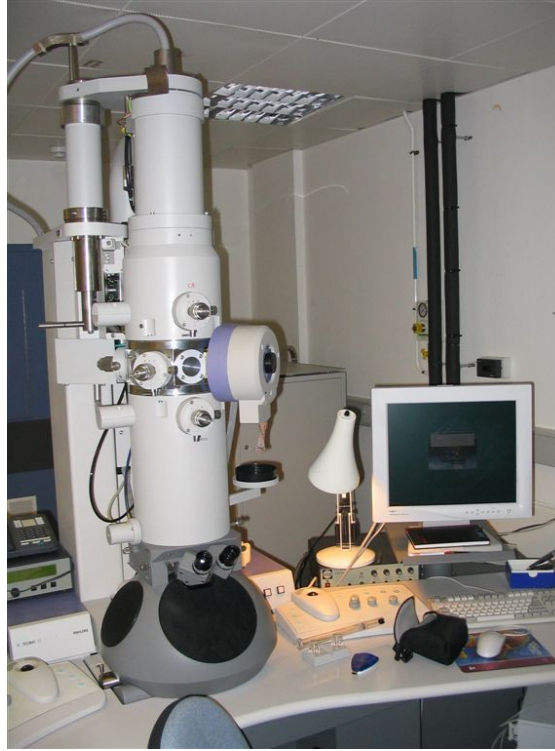
Exclusive search

Element   Formula/Material   Properties   AND   OR   NOT   ( )

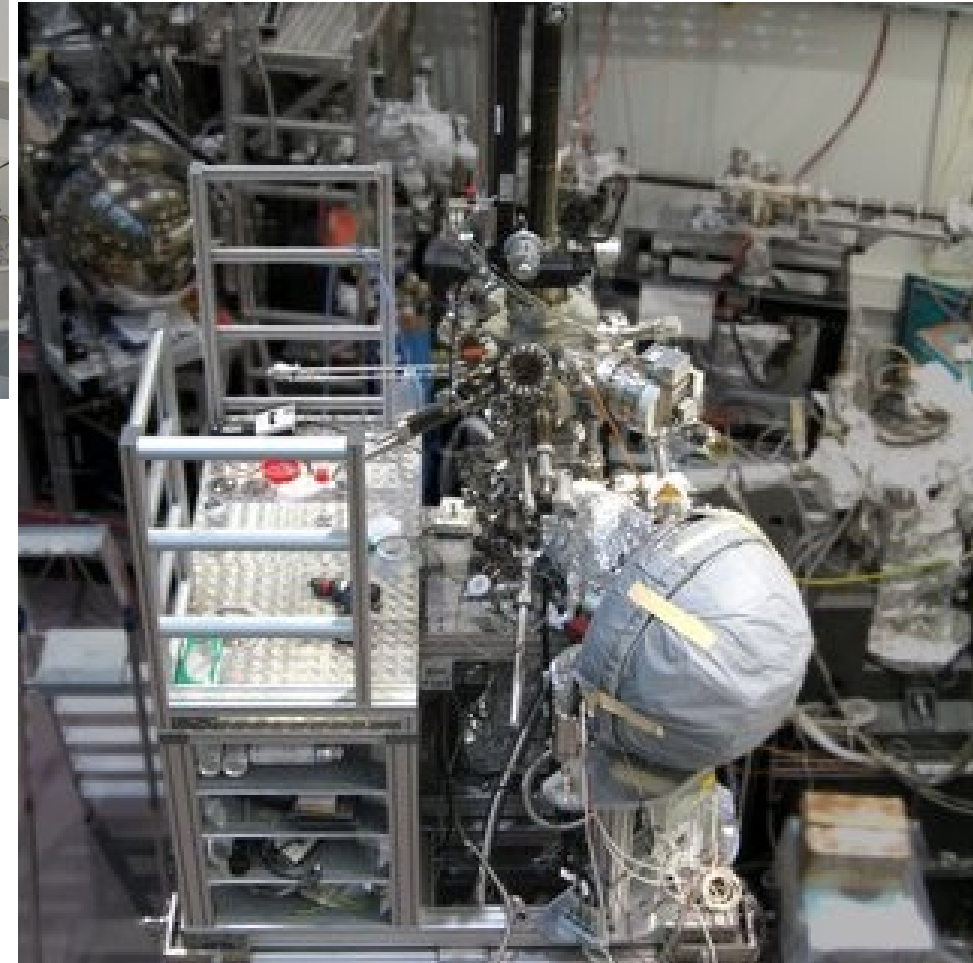
H	He																	He						
1	2																	2						
Li	Be							B	C	N	O	F	Ne							Ne				
3	4							5	6	7	8	9	10							10				
Na	Mg							Al	Si	P	S	Cl	Ar							Ar				
11	12							13	14	15	16	17	18							18				
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr							Kr
19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36							36
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe							Xe
37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54							54
Cs	Ba	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn							Rn	
55	56	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86							86	

## Experiments

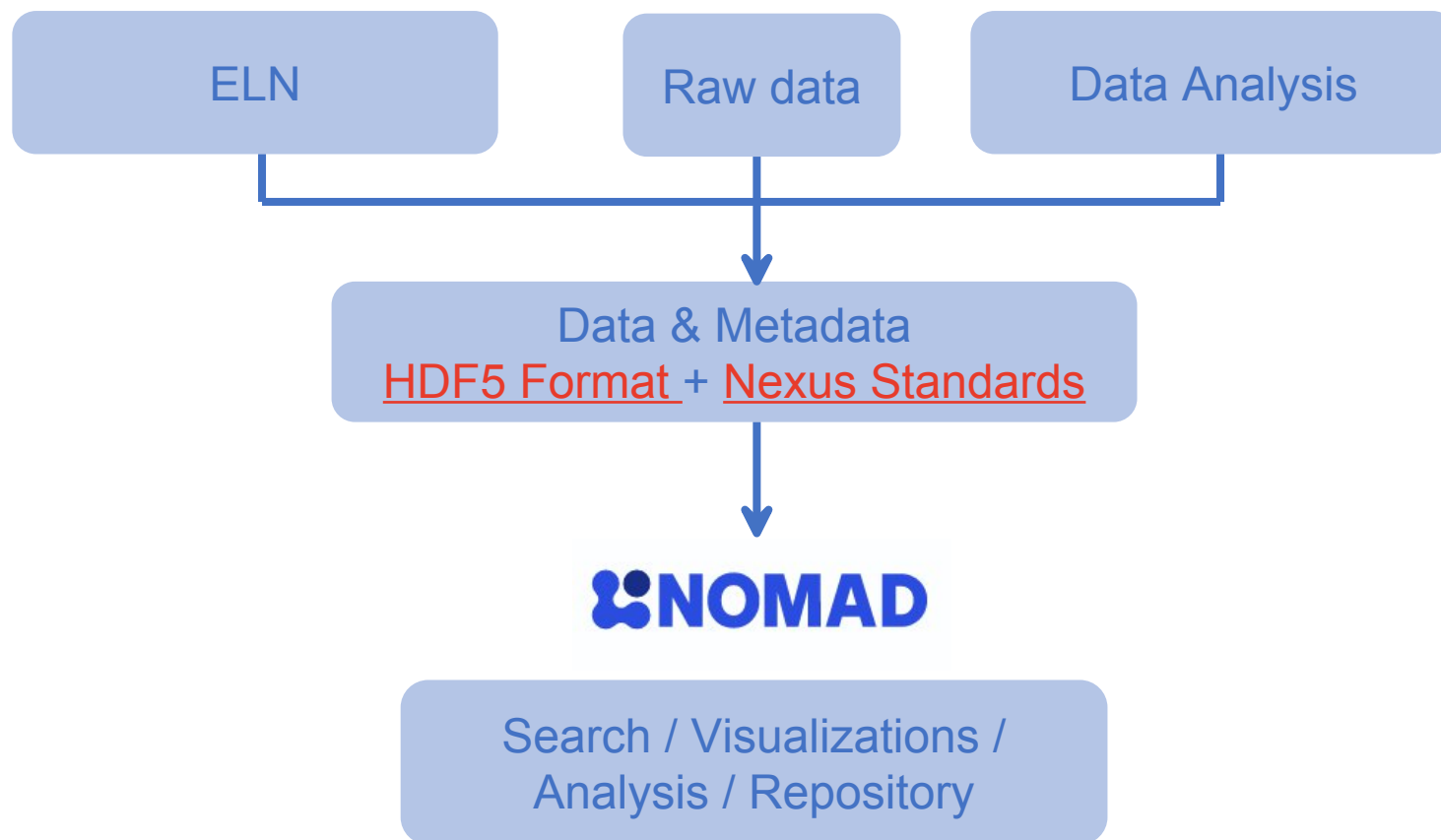
- Instrumentation
- Sample Preparation
- Sample Environment
- Monitors and Detectors
- Data Processing
- Notes



In comparison to large-scale experiments in physics,  
solid-state physics is extremely heterogeneous.

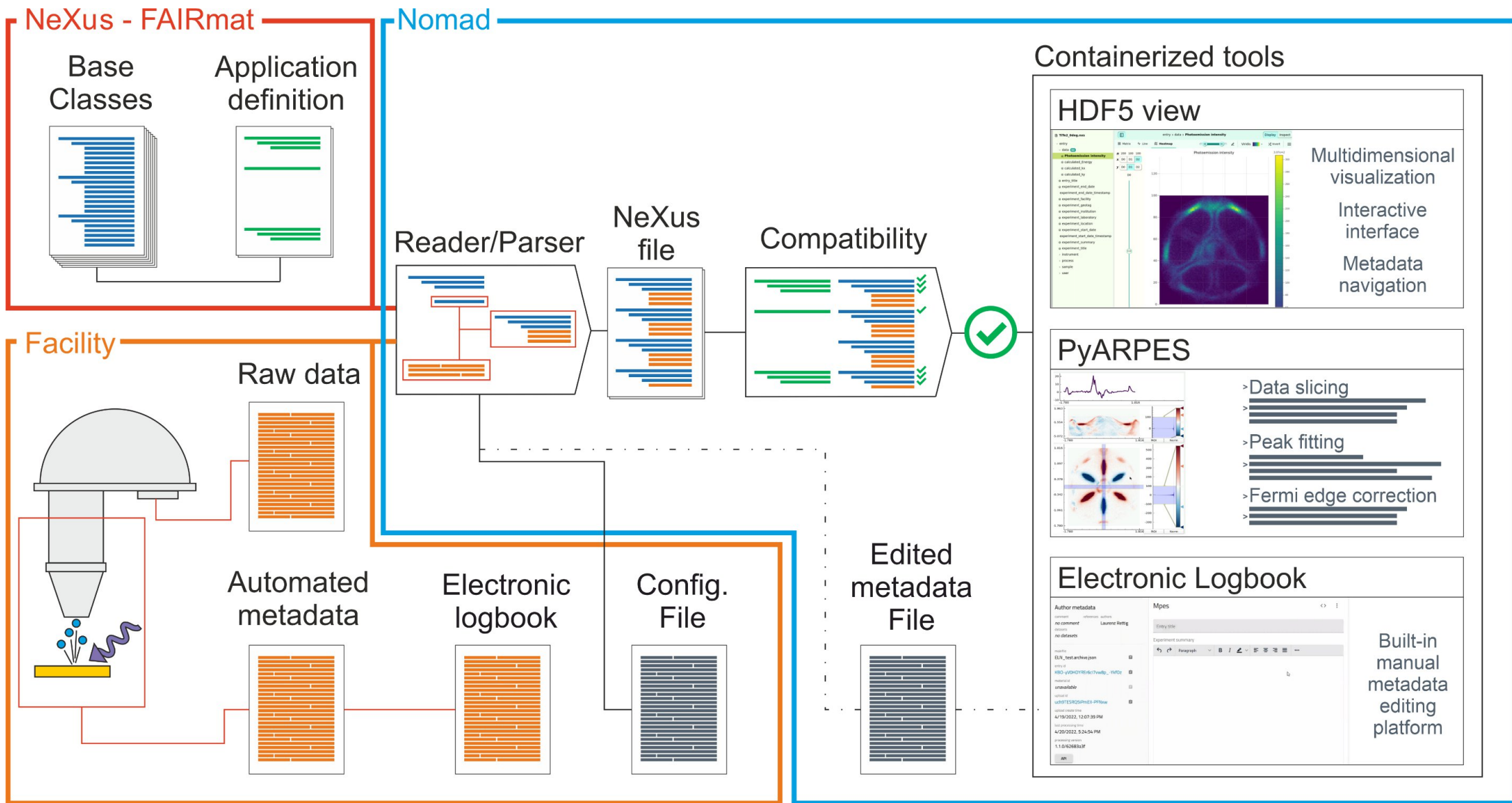


# FAIRmat: experimental research data



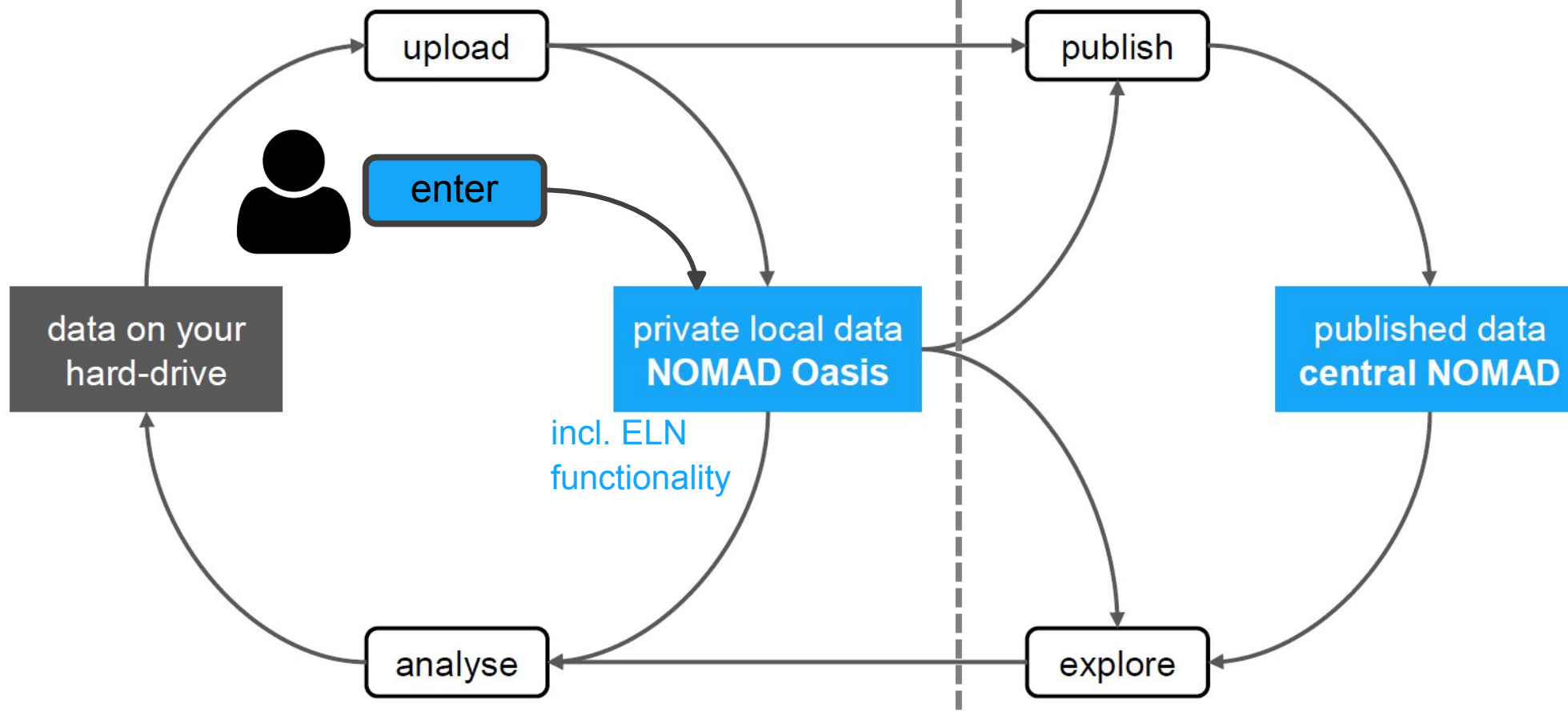
# Photoelectron spectroscopy

What to do?



Local (group / department / university)

World



Local/global



FAIRmat

## FAIRmat's fields of action for experimental physics

- Configurable Lab Control Software
- Electronic Laboratory Notebooks (ELNs)
- FAIR-ready data management
- Community standards
- Involving technology partners
- Workflow in the NOMAD environment
- Broad data expertise







# Application Definitions and the NeXus Standard

Sandor Brockhauser and the FAIRmat  
team

Center for Materials Science Data,  
Humboldt-Universität zu Berlin, Germany

## What to Model

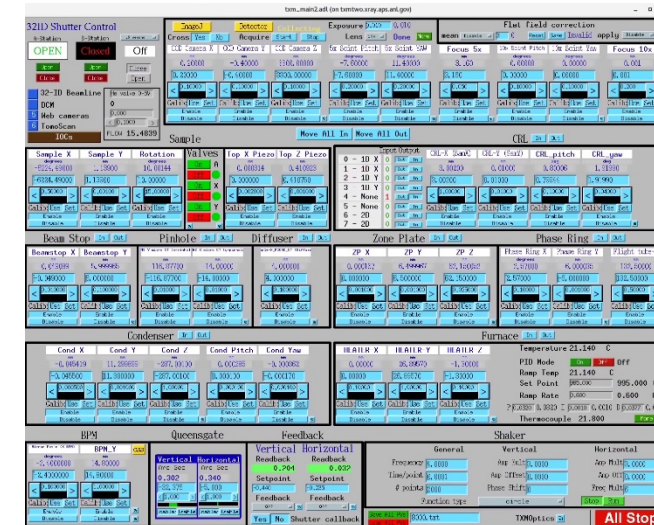
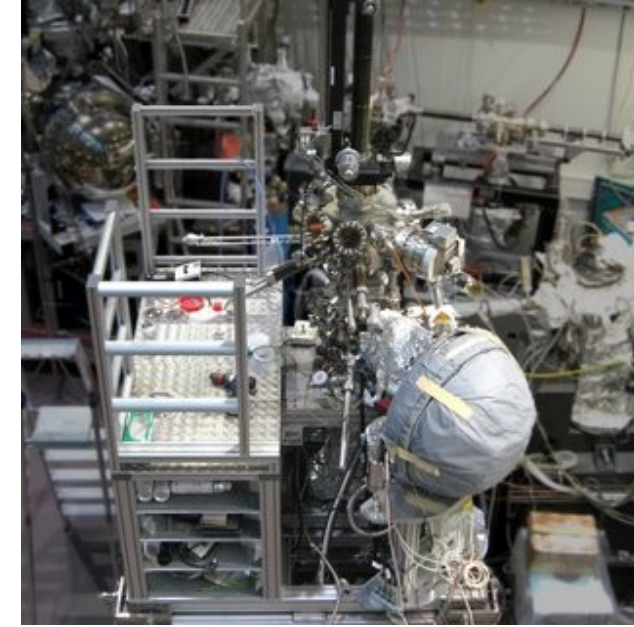
- **Experiment**

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

- **Observation**

in

- **Well-controlled environment**



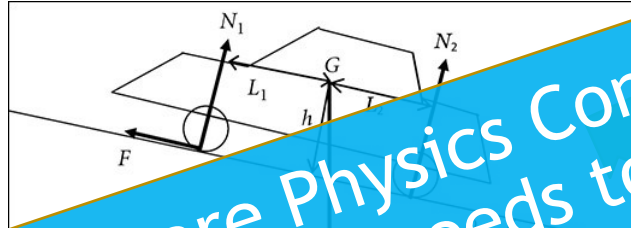
## What to Model

### • Experiment

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

## Experiment Data Models

### Physics Model



Experiment Parameters are Physics Concepts to describe what and when needs to be

- Set or Maintained
- Measured or Calculated

Experiment	
• start time, $t_0$	1
• duration, $t$	1
• distance, $d$	1
Slope	
• angle, $\theta$	1
• length, $l$	1

Car	
• mass, $m$	
• center of mass, $[G, h]$	
• wheel positions, $[L_1, L_2]$	
• force on wheels, $[N_1, N_2]$	
• resistive force, $F$	

## Experiment Parameters

## What to Model

- **Experiment**

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

## Conceptual Design Model

- *Implementation independent Physics Model*
- *Note: Does not tell how exactly it is performed*

### Samples:

- Composition
- Geometry
- History
- ...

### Sample Environment.:

- Temperature
- Pressure
- Magnetic field
- ...

### Instrument:

- ARPES experiment
- XRD
- XRF measurement
- I-V measurement
- ...

### Data Processing:

- Data correction
- Calibration
- Data reduction
- Scientific analysis
- ...



## Data-modelling for Reproducibility

- **Experiment**

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

### Technical Design Model

- *Implementation specific*
- *Not only Physics Model,  
But also Technical Details*

*E.g.*

*what was the pressure in a chamber*



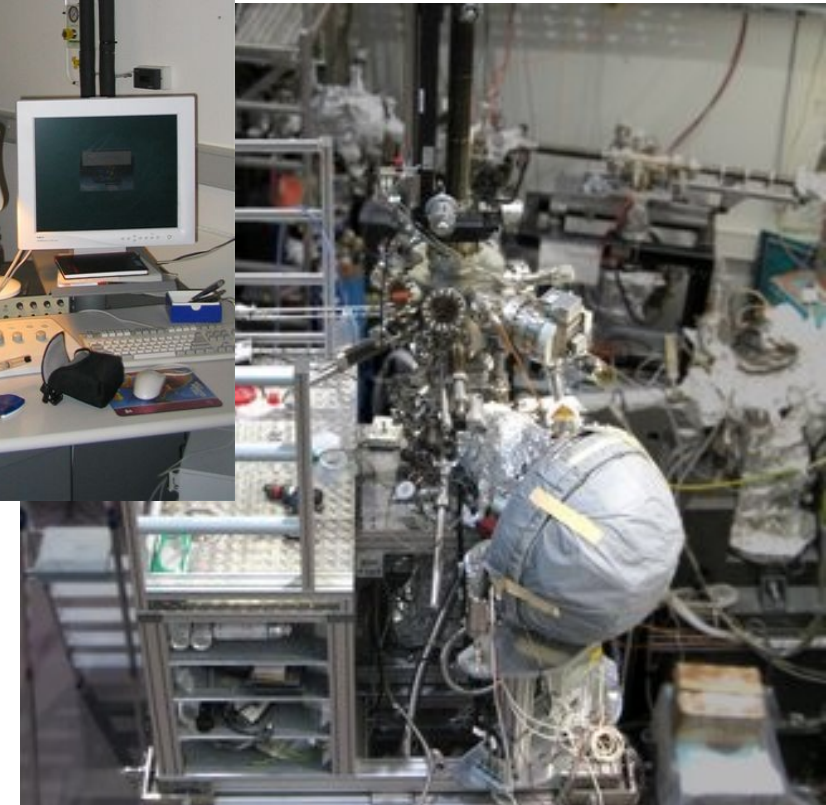
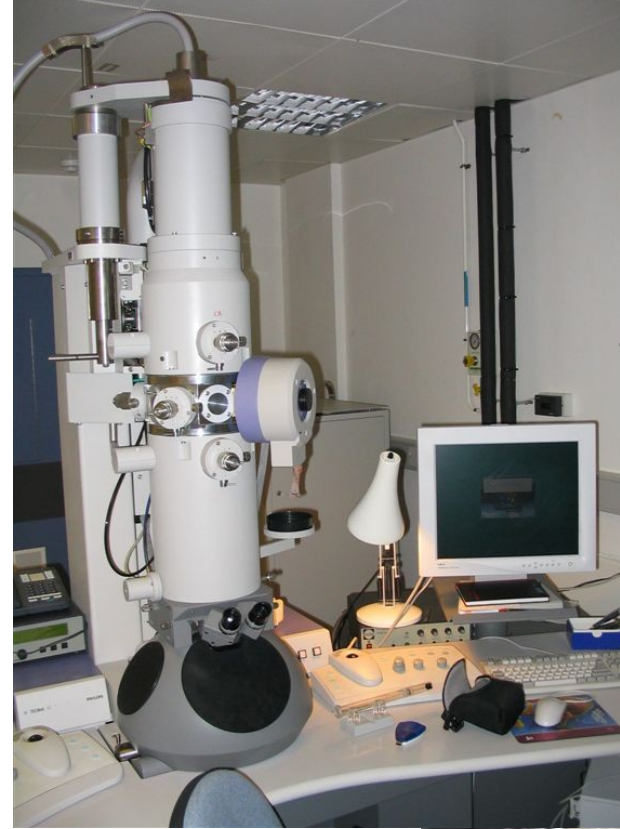
*how it has been produced and maintained*



## Data-modelling for Reproducibility

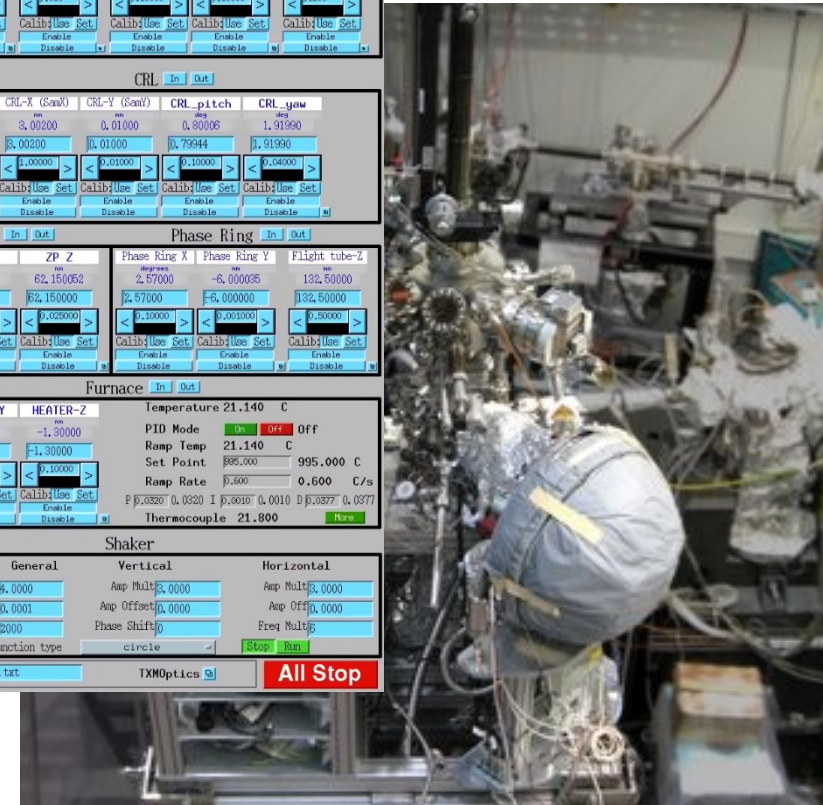
- **Experiment**

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



# Data-modelling for Reproducibility

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

# Data-modelling for Reproducibility

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



The screenshot displays the 'motorx\_all.adl' control window. It features several sections for configuring motor parameters:

- Drive Section:** Shows 'motor 1' (rs:m1) as an 'asynMotor' in 'EGU:degrees' mode. Parameters include Hi limit (0.00000), Readback (1600.00000), MoveAbs (1600.00000), Lo limit (0.00000), MoveRel (0.00000), and Tweak (1.00000). Buttons for Stop, Pause, Move, Enable, and Disable are present.
- Calibration Section:** Includes 'Cal Use Set' (Off), 'Variable' (0.00000), and 'Dir' (Pos, Neg).
- Dynamics Section:** Configures 'Normal', 'Backlash', and 'Jog' modes. Parameters include Maximum Spd (0.00000), Speed (1.00000), Base Speed (0.10000), Accel. sec (0.20000), Backlash distance (0.00000), and Move Fraction (1.00000).
- Servo Section:** Configures 'Proportional', 'Integral', and 'Derivative' gains, all set to 0.00000.
- Resolution Section:** Sets Motor resolution (0.01000), Encoder res. (0.01000), Readback res. (0.00000), and Retry deadband (0.01000).
- STATUS Section:** Shows 'NO\_ALARM' and various state indicators like State (0x0x90a), CurrDir (0), Moving (0), At Home (1), MotorPos (160000), Encoder (159998), MIP (0x0x0), Err (0.00000), Version (6.90), VME Card# (0), Precision (5), Torque (Disable/Enable), and FWD LINK (0).

A red circle highlights the 'Focus' parameter in the '32ID Shutter Control' section, which is set to 3.160.





# Data-modelling for Reproducibility

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



The screenshot displays a motor control interface with several panels:

- motor 1 (rls:m1) asynMotor EGU: degrees**
- Drive** table:
 

Parameter	User	Dial	Raw
Hi limit	0.00000	0.00000	
Readback	1600.00000	1600.00000	160000
MoveAbs	1600.00000	1600.00000	160000
Lo limit	0.00000	0.00000	
MoveRel	0.00000	JogR	JogF
Tweak	< 1.0000 >	HomR	HomF
- Calibration** table:
 

Variable	Value	Dir
Cal Use	Set	Off
Variable	0.00000	Pos
- Dynamics** table:
 

Parameter	Normal	Backlash	Jog
Proportional	0.00000	1.00000	1.00000
Integral	0.00000	0.00000	0.00000
Derivative	0.00000	0.00000	0.00000
- Status** section:
 

State	0x0x90a
CurrDir	0
Moving	0
At Home	1
MotorPos	160000
Encoder	159998
MIP	0x0x0
Err	0.00000
Version	6.90
VME Card#	0
Precision	5
Torque	Disable Enable

**Flow Diagram:**

```

    graph TD
        EGU[EGU's] --> RVAL
        RVAL --> X((X))
        MRES --> X
        X --> DVAL
        DVAL --> Y((Y))
        DIR[DIR: +/- 1] --> Y
        Y --> Z((Z))
        OFF[OFF] --> Z
        Z --> VAL
        VAL --> User[User]
        User --> Dial
        Dial --> Raw
    
```



# Data-modelling for Reproducibility

## • Experiment

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

How deep a Technical Design Model should go for reproducibility, it must be defined by the community.  
 Note that local requirements in LIMS/ELN may defer.

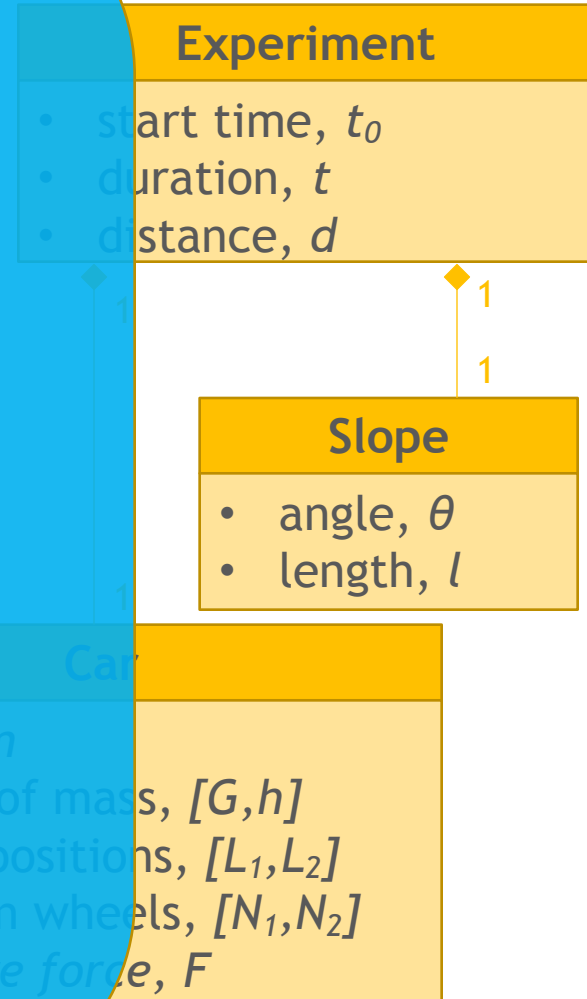


motor 1	(rls:m1)	EGU: degrees
Drive	User	Dial
Hi limit	0.00000	0.00000
Readback	1600.00000	1600.00000
MoveAbs	1600.00000	1600.00000
Lo limit	0.00000	0.00000

The diagram illustrates a control logic for a motor system. It includes nodes for REP (Repeat), RMP (Repeat Move Parameter), RRBV (Repeat Readback Value), RDBL Link (Repeat Drive Back Link), RRES (Repeat Readback Error Signal), DRBV (Drive Readback Value), and RBV (Readback Value). The logic involves multiplication (X) and division (X/+, X/-) operations, along with conditional logic like 'URIP == NO' and 'DIR: +/- 1'.

## Data-modelling for a Community

- Controlled Vocabulary of Concepts
  - Clear relationships between concepts
  - Cardinality and Optionality
  - Standard units
  - Community agreement
  - Agreed data format or API
  - Common tools and applications
- NeXus <https://www.nexusformat.org/>
- Extendable Data Modelling (NXDL)
  - Automated documentation:
    - HTML vocabulary - Humans*
    - OWL ontology OWL - Machines*
  - Community STANDARDIZATION process (NIAC)
  - Tools
    - Verification and Validation
    - $NXDL (schema) \leftrightarrow NXS (data)$
    - Visualisation



# Data-modelling in NeXus

- Base Classes

57

+106

- NXinstrument, NXsample, NXprocess, NXuser  
NXnote
- NXdetector, NXsensor, NXmonitor, NXbeam,  
NXmirror, NXgeometry,...

- Application Definitions

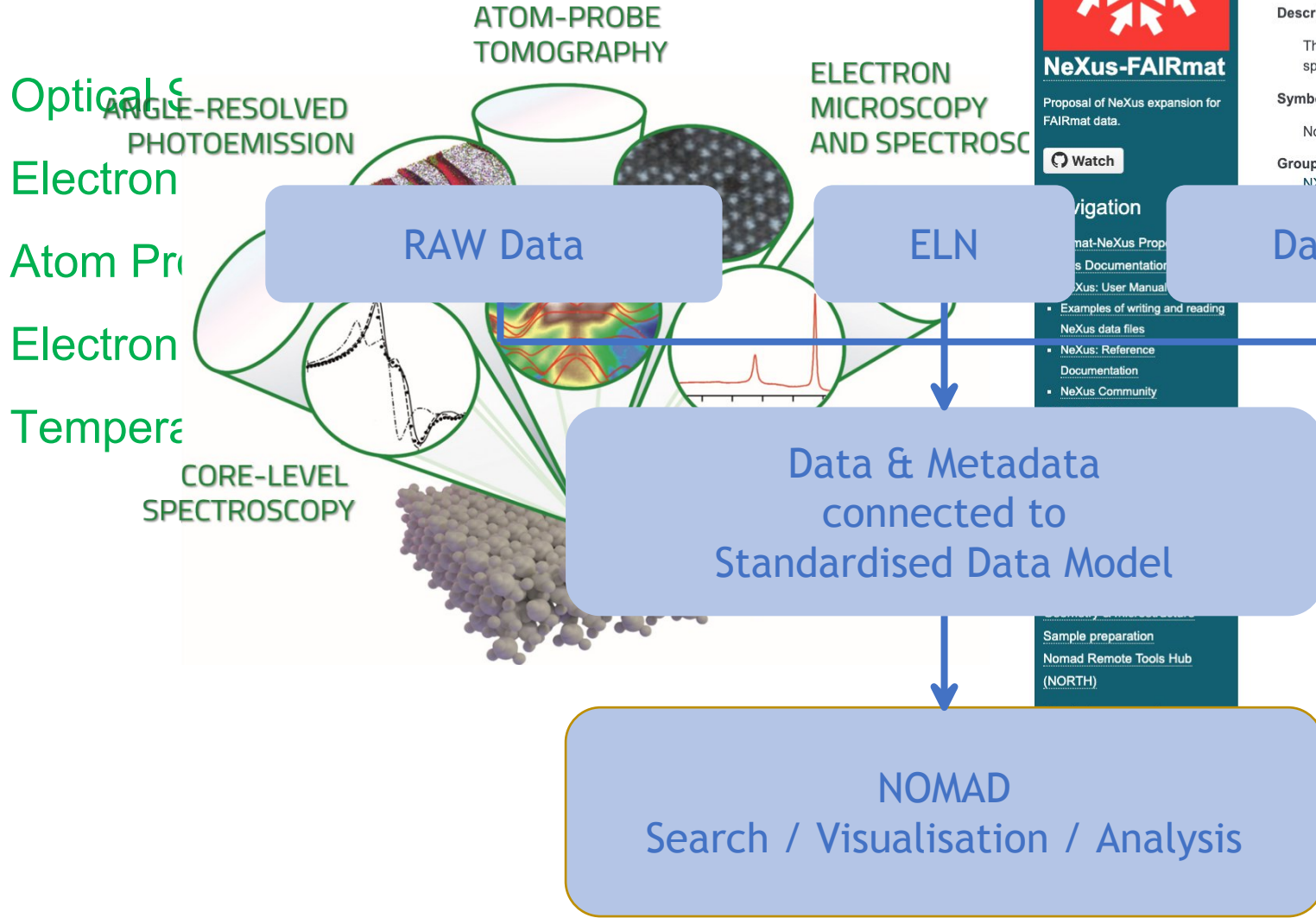
34

+38

- NXarpes, NXxas, NXem, NXapm, NXmpes  
NXellipsometry, NXsensor\_scan  
...



# Data Models for Experiment



## NXmpes

Status: application definition, extends [NXObject](#)

Description: This is the most general application definition for multidimensional photoelectron spectroscopy.

Symbols: No symbol table

Groups cited: [NXaperture](#), [NXbeam](#), [NXcalibration](#), [NXcollectioncolumn](#), [NXdata](#), [NXdetector](#), [NXinstrument](#), [NXmanipulator](#), [NXnote](#), [NXobject](#), [NXoutput](#), [NXparameter](#), [NXprocess](#), [NXsample](#), [NXsource](#), [NXtarget](#)

title: (required) [NX\\_CHAR](#)

start\_time: (required) [NX\\_DATE\\_TIME](#)  
Datetime of the start of the measurement.

definition: (required) [NX\\_CHAR](#)  
Obligatory value: [NXmpes](#)

@version: (required) [NX\\_CHAR](#)

USER: (required) [NXUser](#)  
Contact information of at least the user of the instrument or the investigator who performed this experiment. Adding multiple users if relevant is recommended.

name: (required) [NX\\_CHAR](#)  
Name of the user.

affiliation: (recommended) [NX\\_CHAR](#)  
Name of the affiliation of the user at the point in time when the experiment was performed.

address: (recommended) [NX\\_CHAR](#)  
Full address (street, street number, ZIP, city, country) of the user's affiliation.

email: (required) [NX\\_CHAR](#)  
Email address of the user.

orcid: (recommended) [NX\\_CHAR](#)  
Author ID defined by <https://orcid.org/>.

INSTRUMENT: (required) [NXInstrument](#)

energy\_resolution: (required) [NX\\_FLOAT](#) {units=[NX\\_ENERGY](#)}

SOURCE: (required) [NXsource](#)  
The source used to generate the primary photons. Properties refer strictly to parameters of the source, not of the output beam. For example, the energy of the photons is the energy of the source, not the energy of the output beam.



<https://fairmat-nfdi.github.io/nexus-fairmat-proposal/>

## Data-modelling for a Community

- Initial Proposal with Examples
- Community Feedbacks
- Common Proposal with Technology Partners
- Standardisation



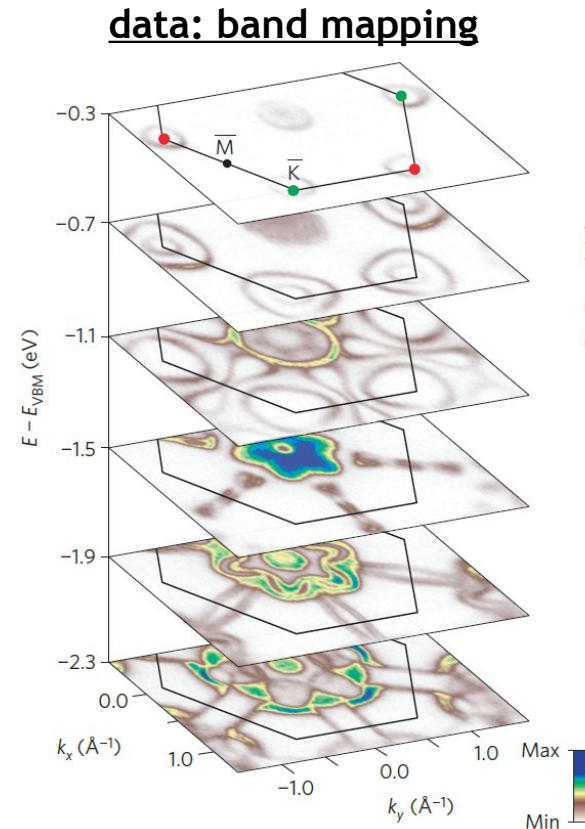
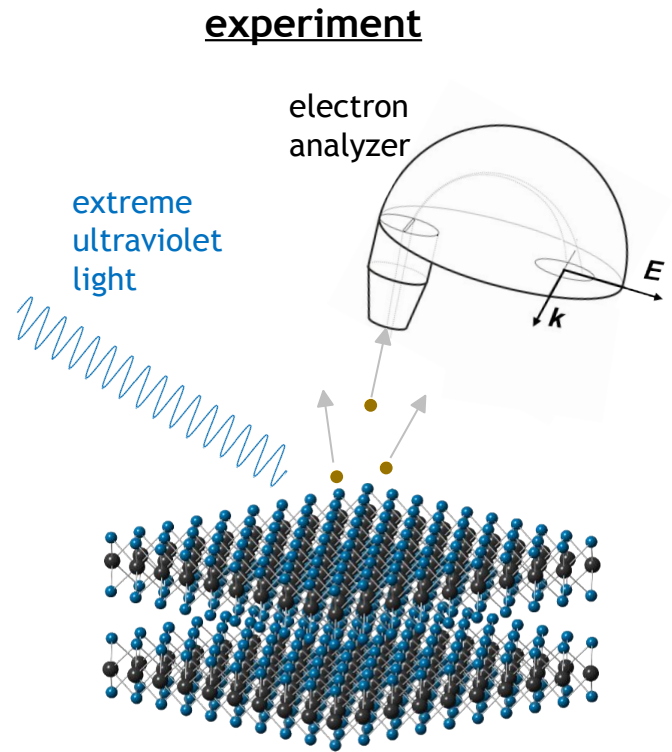


# NXmpes: structure and example in NOMAD

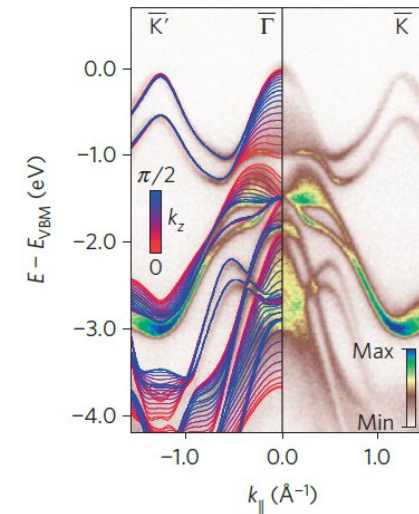
Laurenz Rettig and the FAIRmat team

Fritz Haber Institute of the Max Planck  
Society, Berlin

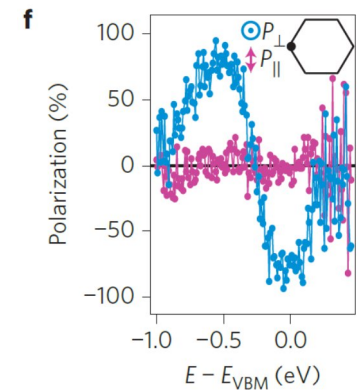
# Multidimensional photoemission spectroscopy



**energy band dispersion**



**Also possible:  
Spin polarization**



Example: valence band structure of  $\text{WSe}_2$   
Riley *et al.*, *Nature Phys.* 10, 835 (2014)

- momentum  $\mathbf{k}$  ( $k_x, k_y$ )
- energy  $E$
- spin  $\mathbf{S}$

}

“multidimensional PES” (more than 3D)

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





# Extending the parameter space

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

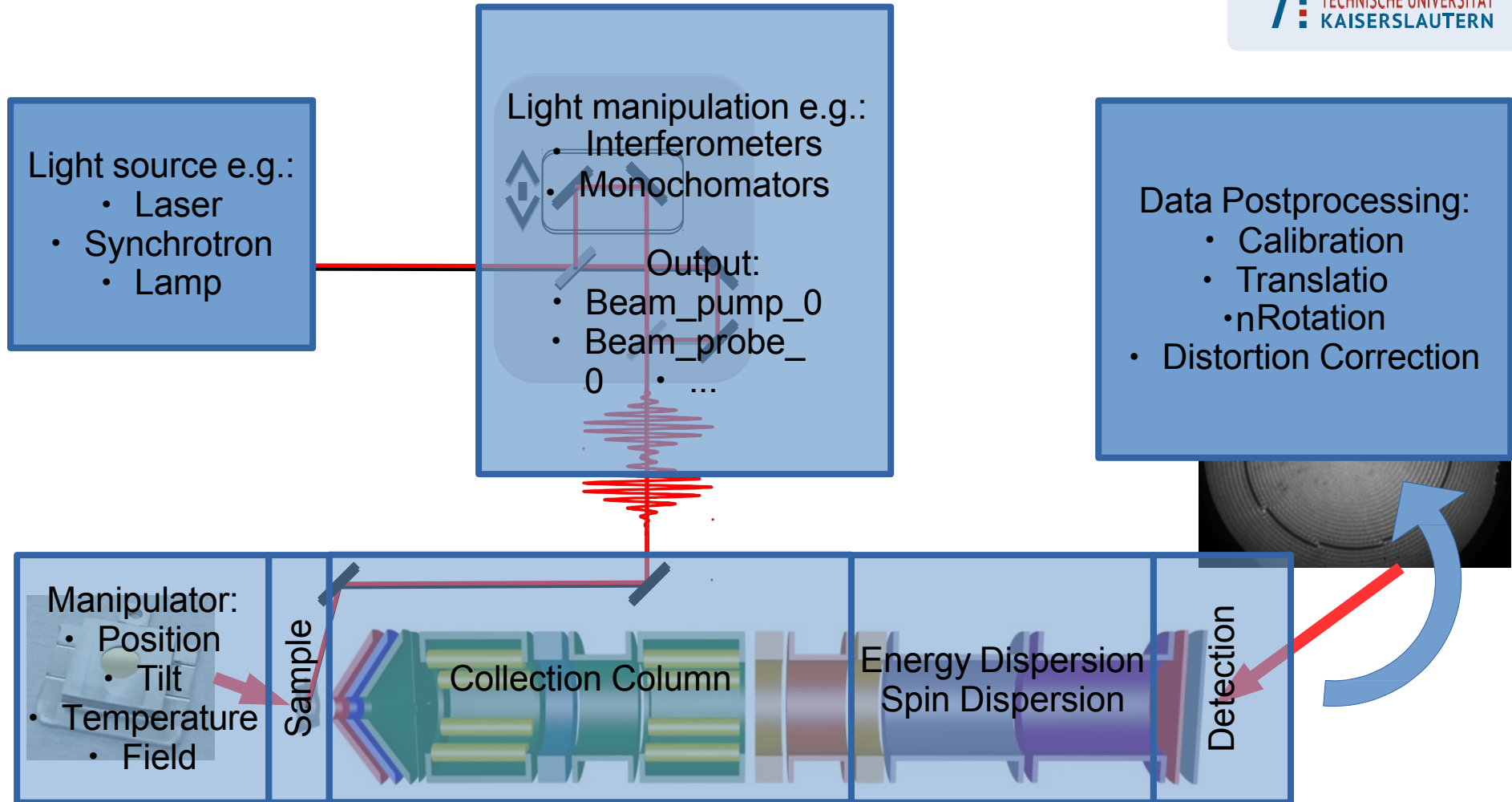
- € Dependence on materials and preparation procedures ( $\mu_{mat}$ )
- € Spin-resolving detectors:  $S$
- € Photon energy and polarization  $\omega_p$
- € Electronic structure of **non-equilibrium states** (time-resolved ARPES):  
time  $t$ , pump/probe photon energy  $\omega_p$ , polarization, ...
- à Dependence on sample parameters:  
strain  $\sigma$ , sample position  $x, y$ , temperature  $T$
- € ...

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

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

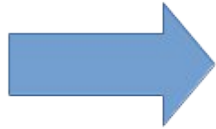
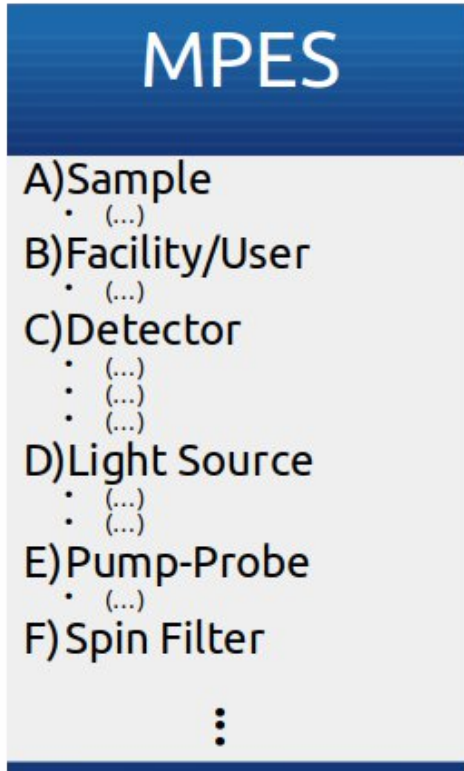


# Ingredients of a photoemission experiment

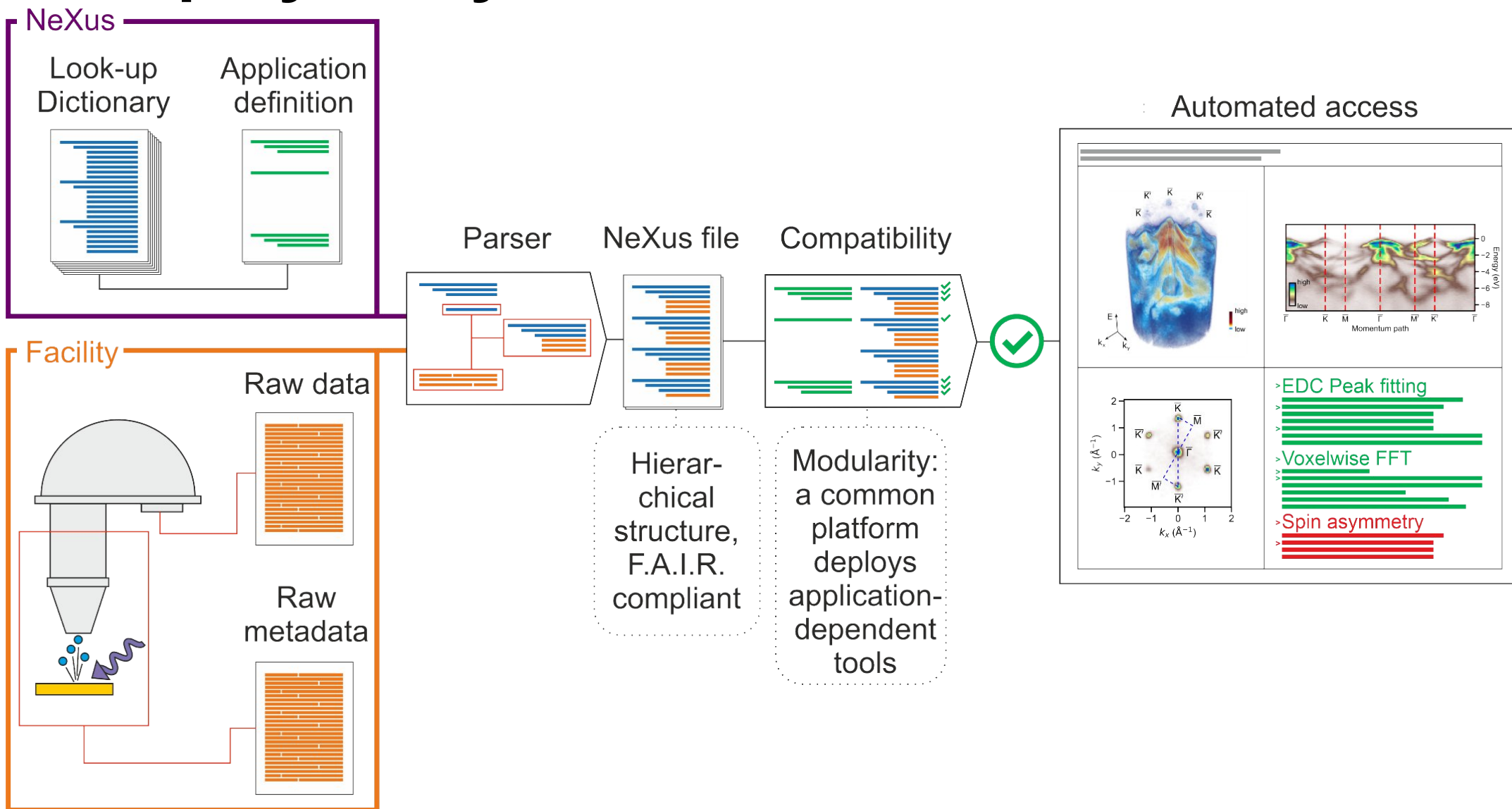


# Hierarchy of application definitions

Universal MPES  
metadata dictionary  
(in NeXus: base classes)



## Simplify analysis



# Application definition: NXmpes

## Application Definitions

We created two new application definitions:

### NXmpes:

A general appdef with minimalistic metadata requirements, apt to describe all photoemission experiments.

## Base Classes

We developed entirely new base classes:

### NXelectronanalyser:

A base class to describe electron kinetic energy analyzers. 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 system (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.



# Application definitions and base classes

**ENTRY:** (required) [NXentry](#)

**title:** (required) [NX\\_CHAR](#)

**start\_time:** (required) [NX\\_DATE\\_TIME](#)

Datetime of the start of the measurement.

**INSTRUMENT**

**energy\_**

**SOURCE**

**ELECTRO**

desc

energ

E

fr

fast\_

slow\_

**COLL**

s

- Selective area
- Deflector
- PEEM
- Momentum Microscope

**mode:** (recommended) [NX\\_CHAR](#)

**projection:** (recommended) [NX\\_CHAR](#)

**description:** (optional) [NX\\_CHAR](#)

Free text description of the type of the detector

**name:** (optional) [NX\\_CHAR](#)

Name or model of the equipment

Detailed discussion of base classes and application definitions tomorrow morning.

(loading)

NUMBER}

These refer only to  
er variables such as  
n the data.

low\_axes

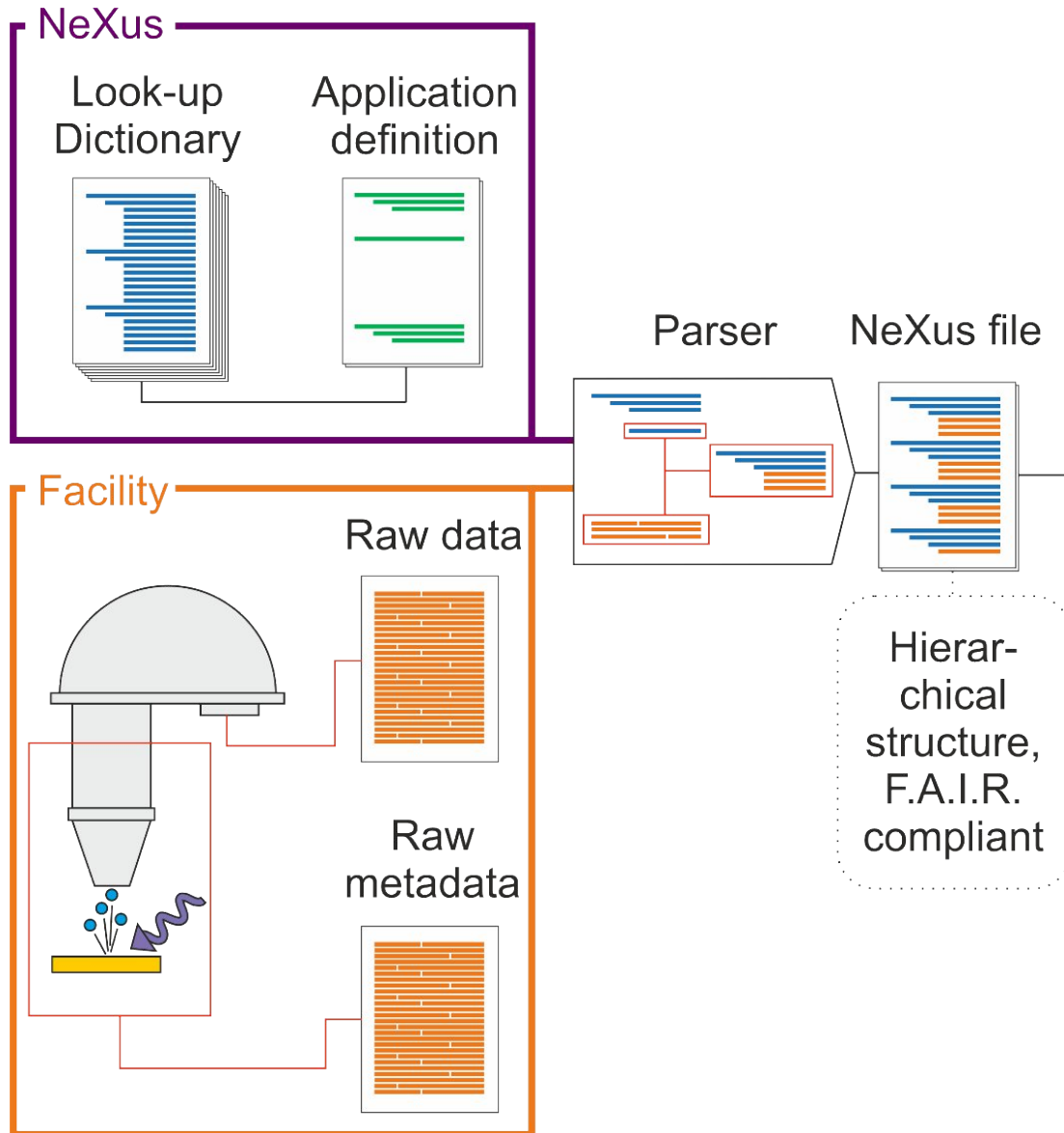
energy"]

sweeping energy mode		
Tof	['energy', 'kx', 'ky']	
Momentum microscope, spin-resolved	['energy', 'kx', 'ky']	['spin up-down', 'spin left-right']

Axes may be less abstract than this, i.e. ['detector\_x', 'detector\_y']. If energy\_scan\_mode=sweep, fast\_axes: ['energy', 'kx']; slow\_axes: ['energy'] is allowed.



# Practical examples



Jupyter NXpy\_full (autosaved)

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Extended NeXus hierarchy for ARPES experiments

This notebook is built as a demonstrator, it collects data and metadata from a Beamtime performed at FLASH by the Structural and Electronic Surface Dynamics and Dynamics of Correlated Materials groups of the Physical Chemistry department of the Fritz-Haber-Institute.

It is designed to:

1. showcase the capabilities of NeXus hierarchy in a real-world application
2. create a shared dictionary of items in the hierarchy for metadata of ARPES experiments
3. provide a tool for the conversion to NeXus of data and metadata from experiments where metadata cannot be automatically parsed.

For this, I hand picked the parameters from a combination of a logbook, a preprocessed data file, a raw file containing the data from FLASH and two files containing the metadata from the data processing software. If NeXus format is accepted higher forms of automation will be implemented.

I tried to create an entry for every piece of information that might be available and relevant for an ARPES experiment. Where I could not find information in the metadata, I still added the field, but filled with NaN or "Not found" strings.

```
In [1]: import h5py
import numpy as np
import os
import six
import pytest

from nexustools.nexus import *

def printname(name):
    print(name)

# Packages needed to parse naive date in unix timestamp univocally

import pytz
from timezonfinder import TimezoneFinder
tf = TimezoneFinder()
from datetime import datetime as dt
from datetime import date as d
from datetime import timedelta as td
```

## Converter / parser

- reads data + metadata from experiment
- Assigns them to the correct NXmpes paths
- Creates a verified NeXus file according to a provided application definition



# Reader + parser structure

Data.h5



Volumetric data  
+ axes



Tree



of



metadata

config\_file.json

```
"/ENTRY[entry]/definition": "NXmpes",  
"/ENTRY[entry]/definition/@version": "None",  
"/ENTRY[entry]/title": "@attrs:metadata/entry_title",  
"/ENTRY[entry]/start_time": "@attrs:metadata/timing/acquisition_start",  
"/ENTRY[entry]/experiment_institution": "Fritz Haber Institute - Max Planck Society",  
"/ENTRY[entry]/experiment_facility": "Time Resolved ARPES",  
"/ENTRY[entry]/experiment_laboratory": "Clean Room 4",  
"/ENTRY[entry]/entry_identifier": "@attrs:metadata/entry_identifier",  
"/ENTRY[entry]/end_time": "@attrs:metadata/timing/acquisition_stop",  
"/ENTRY[entry]/duration": "@attrs:metadata/timing/acquisition_duration",  
"/ENTRY[entry]/duration/@units": "s",  
"/ENTRY[entry]/collection_time": "@attrs:metadata/timing/collection_time",  
"/ENTRY[entry]/collection_time/@units": "s",  
  
"/ENTRY[entry]/USER[user]/name": "@attrs:metadata/user0/name",  
"/ENTRY[entry]/USER[user]/role": "@attrs:metadata/user0/role",  
"/ENTRY[entry]/USER[user]/affiliation": "@attrs:metadata/user0/affiliation",  
"/ENTRY[entry]/USER[user]/address": "@attrs:metadata/user0/address",  
"/ENTRY[entry]/USER[user]/email": "@attrs:metadata/user0/email",
```



Python-based reader interprets data file, based on the information of the config file.  
Additional information from other meta data sources.

A custom reader for each file type/technique, but a common infrastructure.

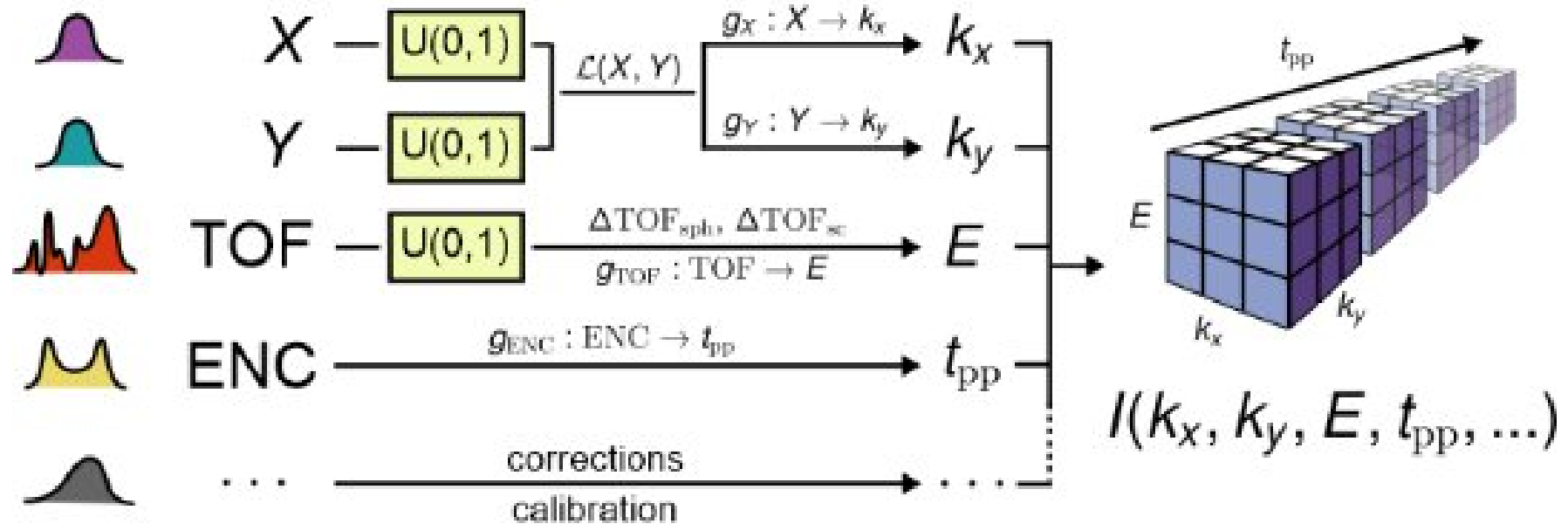




# Single event data processing

Time-resolved Photoemission data, recorded with SPECS METIS 1000 momentum microscope.

Single-event data calibrated and converted into multidimensional volumetric data.



Xian, et al., Scientific Data 7, 442 (2020)

<https://mpes.science/>



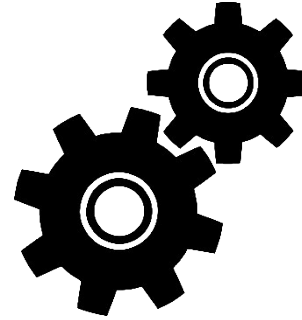
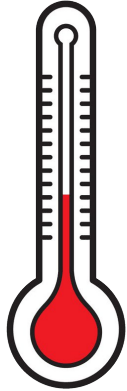
# Practical examples

LIVE Example on NOMAD



# Sources of metadata

Machine-generated meta data:



Automated recording and assigning to respective fields

“Soft” meta data need structured user input

e.g.:

- User who does the experiment
- Name and composition of the sample
- ...

Requires structured electronic lab notebook or user interfaces

Structured and configurable ELN capabilities available in NOMAD



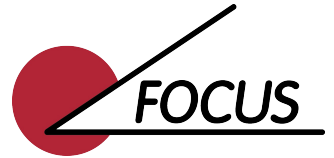


# MPES reader infrastructure

Lukas Pielsticker and the FAIRmat  
team

[lukas.pielsticker@cec.mpg.de](mailto:lukas.pielsticker@cec.mpg.de)

## Storage of photoemission data



...

Problem:

- Each technology partner has its own file format
- Within a technology partner the file format can be different

→ How to go from vendor-specific format to structured NXmpes data?

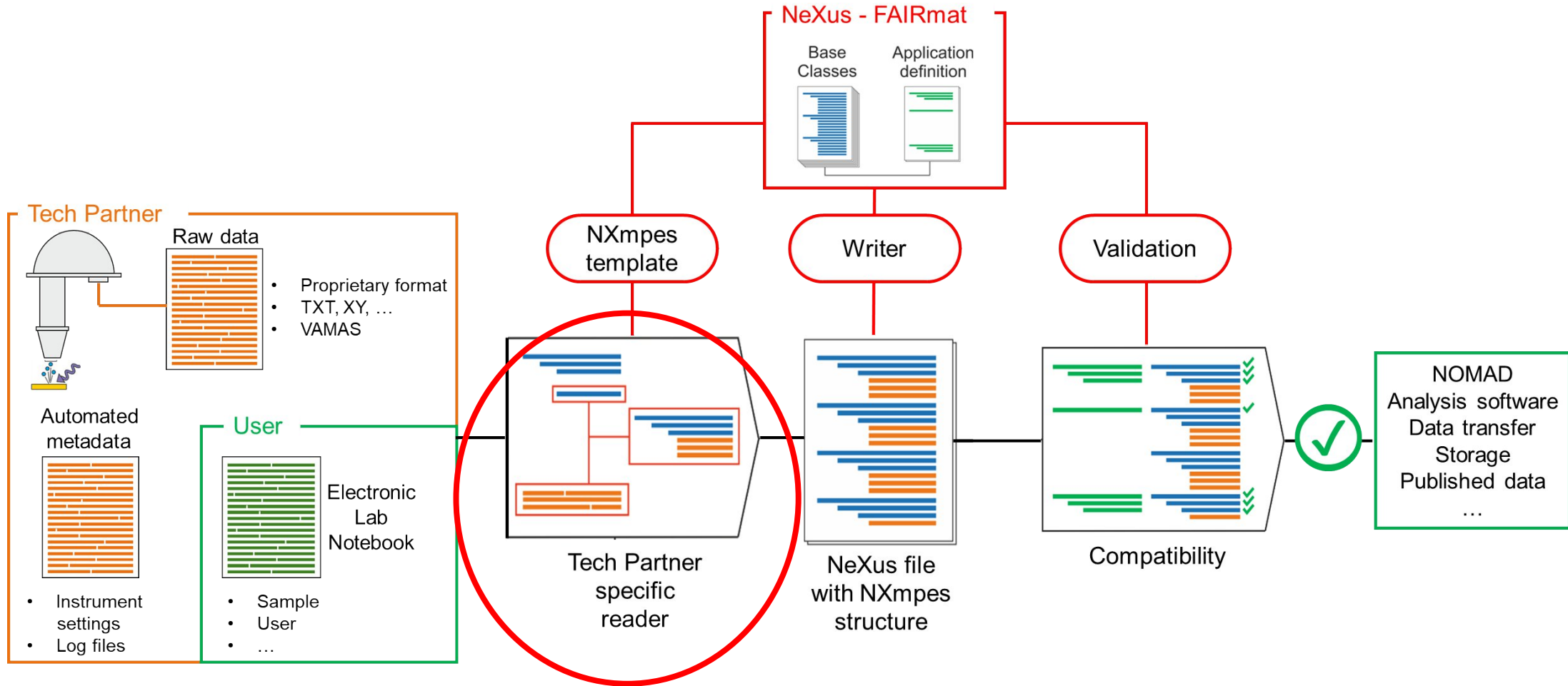


## NXmpes readers - objectives

- Data- and vendor agnostic Nxmpes files
  - **Parsers, normalizers, and converters**
- Data curation and quality assessment
  - **Validation** against NeXus-FAIRmat application definitions
- Capture all **metadata** to produce valid NXmpes files
  - Instrument settings, ELN data, temperature/pressure logs, ...
- Interoperability and repeatability
  - Calibration, transmission functions, etc.
  - Storing **raw, processed, and analysed** data



## From vendor-specific data to NXmpes



# Link to technology partners and the community

## Choice for technology partners

1. Full integration with native support to read/write NXmpes
2. Integrated option to directly export to NXmpes
3. Provide external parser to convert to NXmpes

## What needs to be discussed

- Which (meta-)data would we like to store in NXmpes?
  - Discussion of MPES application definition (tomorrow, 9 am)
- Which are the limits of open data?
- Metadata sources
- Do the technology partners need support?

## What we can provide

- Detailed description of application definitions
- Reader examples
- Validation infrastructure

NeXus-FAIRmat application definitions:

<https://fairmat-nfdi.github.io/nexus-fairmat-proposal/>



Example readers and validators:

<https://github.com/FAIRmat-NFDI/pynxtools>





# How to contribute further

beam in a synchrotron and so on.

type: (required) NX\_CHAR

Any Annotate Highlight





- Synchrotron X-ray Source
- Rotating Anode X-ray
- Fixed Tube X-ray
- UV Laser
- Free-Electron Laser
- Optical Laser
- UV Plasma Source
- Metal Jet X-ray
- HHG laser

**rettig** (edited Just now) 1 min ago

Public

Synchrotron X-ray Source Rotating Anode X-ray Fixed Tube X-ray UV  
Laser Free-Electron Laser Optical Laser UV Plasma Source Metal Jet  
X-ray HHG laser

Dual anode source is missing.

We are very happy to integrate your input into the Nxmpes and base classes definitions!



## Discussion points and questions:

- Do our proposed definitions fit your use case? What is missing, needs to be changed?
- Requirement/optionality of parameters
- Which parameters can be provided, which not?
- How are “soft” metadata acquired e.g. in vendor software?
- What are legal constraints, IP restrictions?
- Interplay of user-specific data and instrumental metadata
- ...

