

Research Data Management with FAIRmat and NOMAD

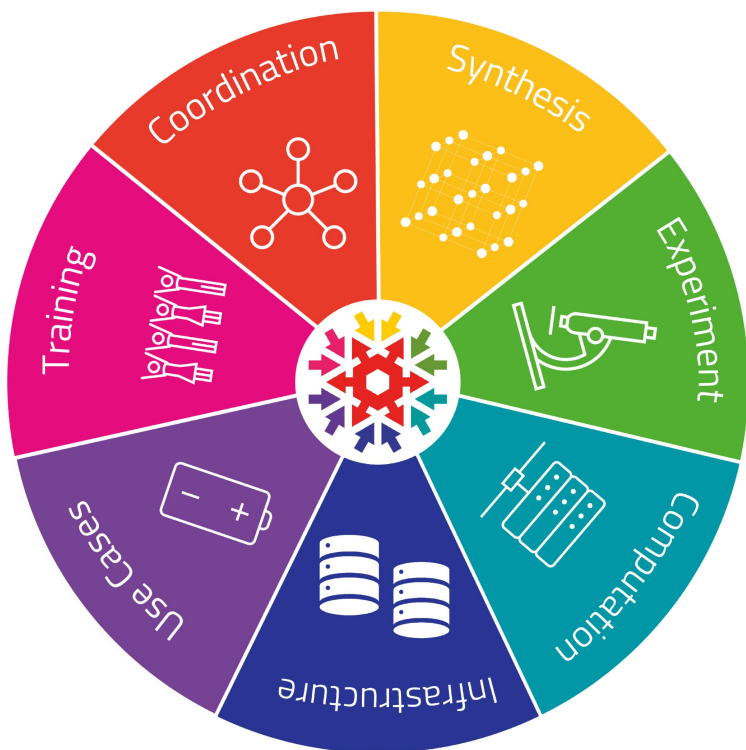
Sandor Brockhauser
Humboldt Universität zu Berlin



What are FAIRmat and NOMAD



FAIRmat is the german national consortium to build a FAIR federated data infrastructure for solid state physics/materials science



FAIR

Open source

Open science

Bottom up

The FAIRmat team

~60 PIs + ~30 central coworkers



Claudia Draxl
Spokesperson



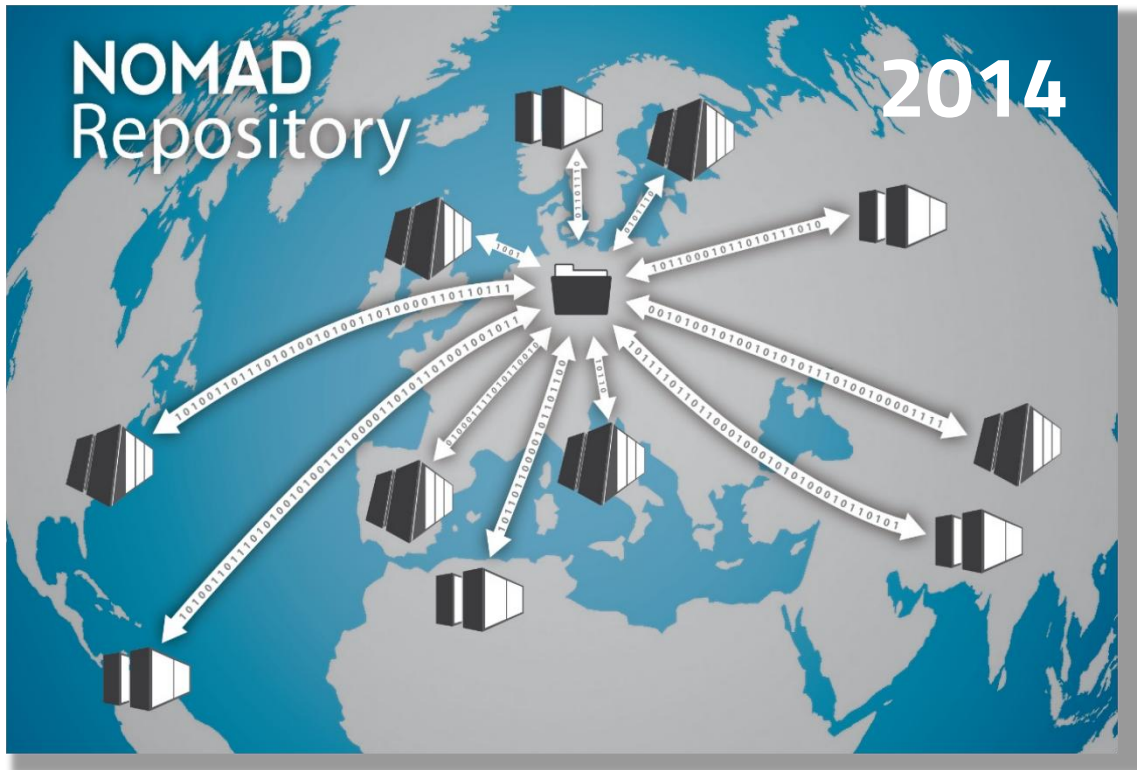
Christoph T. Koch
Deputy Spokesperson



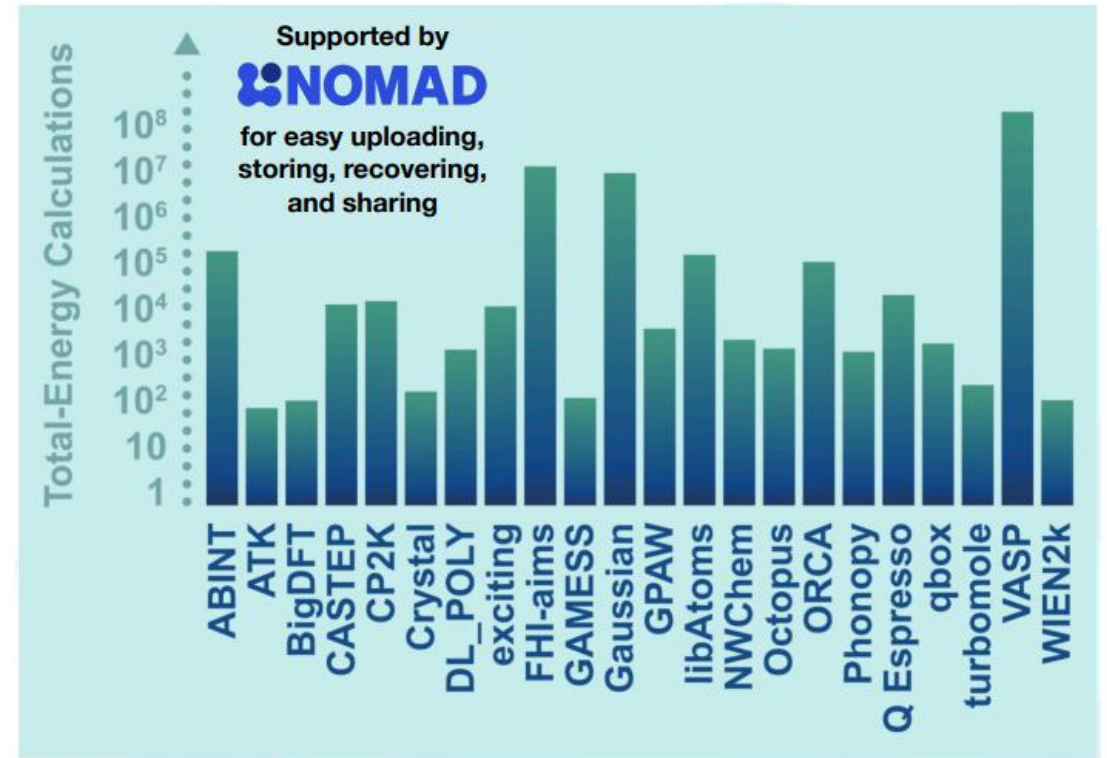
The NOMAD central hub



The **Novel Materials Discovery** Laboratory



- Support +60 codes
- GW, DMFT
- Drag and drop
- Publish
- FAIR
- AI ready



The NOMAD central hub



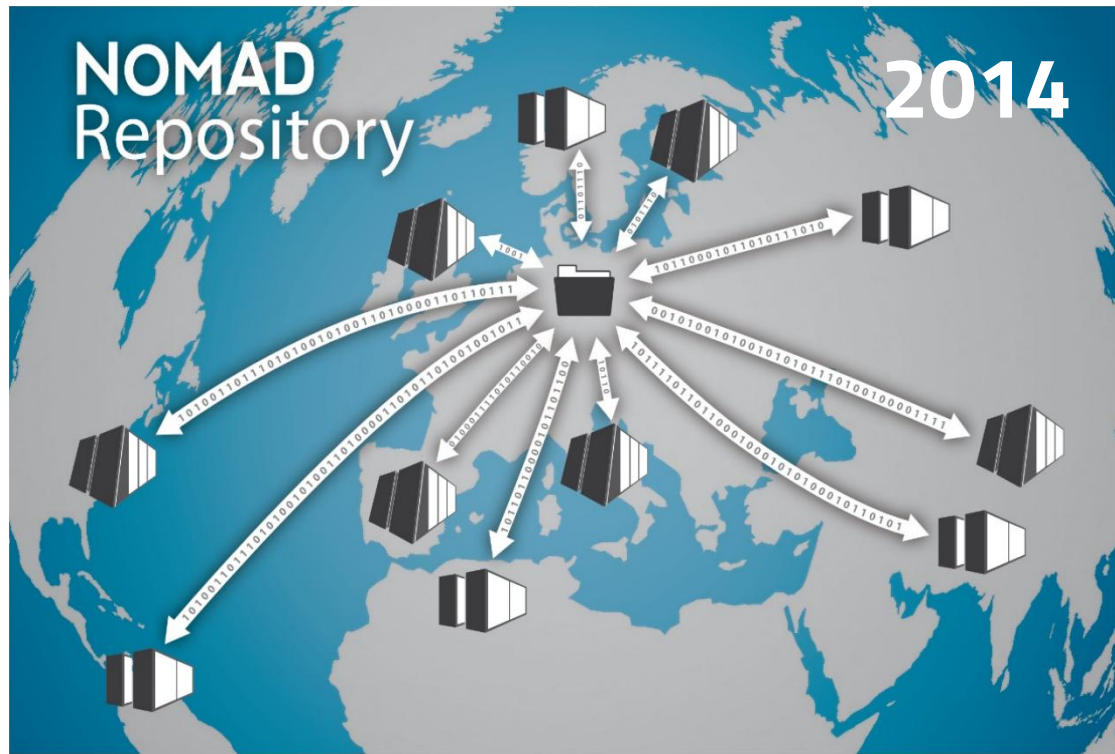
NOMAD

Materials data managed and shared

UPLOADED ENTRIES
16,469,399

REPRESENTED MATERIALS
3,751,773

UPLOADED FILES
110.6 TB



The NOMAD ecosystem

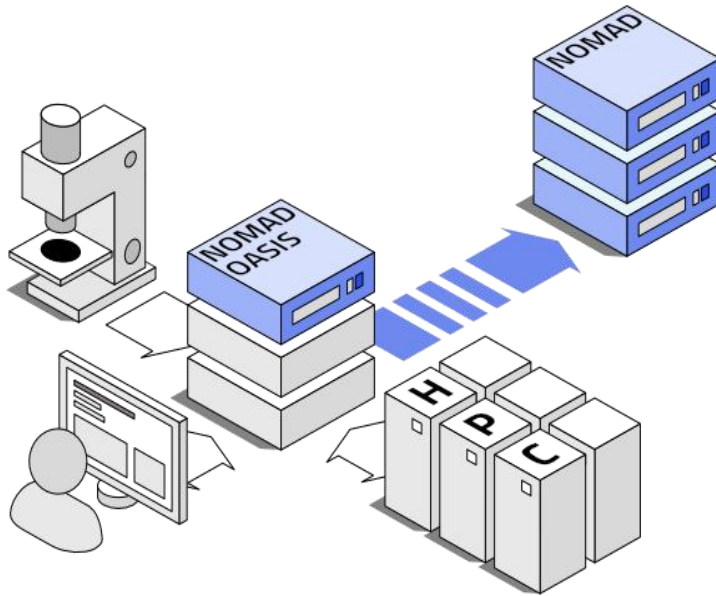


NOMAD Oasis

Local management tool for your lab, adapted to your data-management needs and lab policies.

REGISTERED INSTALLATIONS

> 70



HZB Helmholtz
Zentrum Berlin



**universität
innsbruck**

I ILLINOIS



**TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN**

li.u LINKÖPING
UNIVERSITY



ikz

Technische
Universität
Berlin

WARWICK
THE UNIVERSITY OF WARWICK



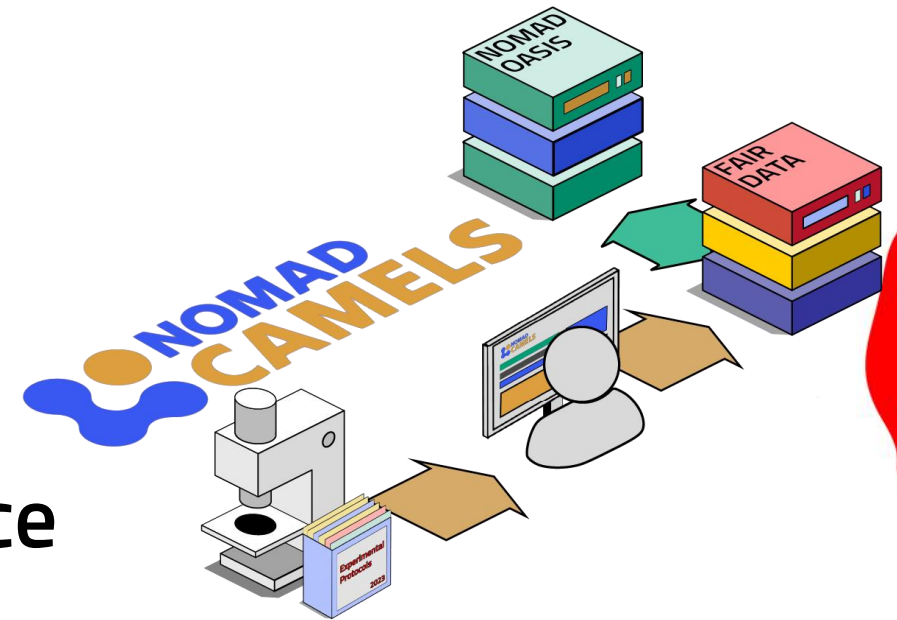
NOMAD CAMELS

Control instruments, run measurement protocols, and record FAIR data!

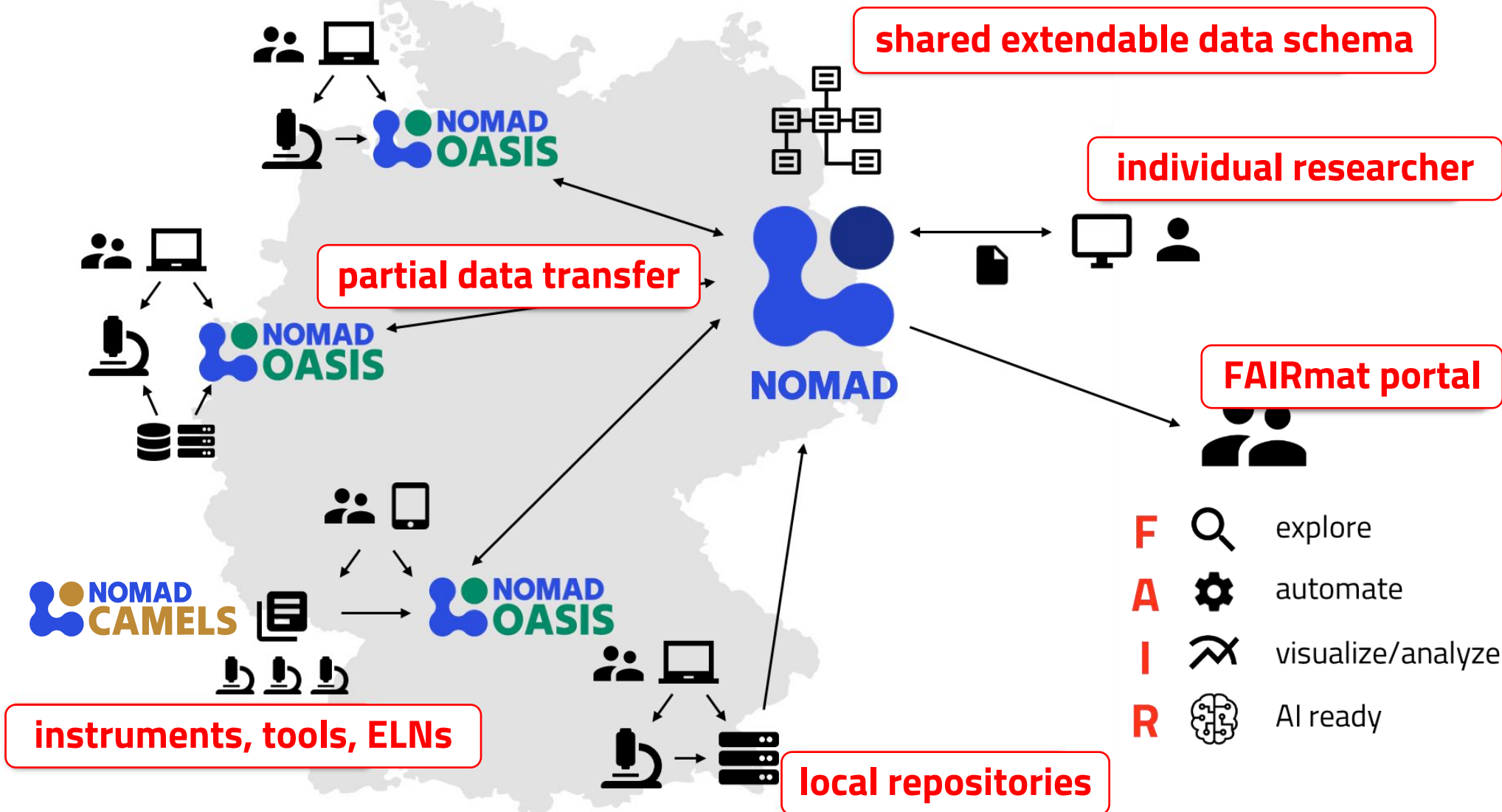
- Store all available data and metadata.
- Define measurement protocols using an intuitive graphical user interface.
- Connect to instruments via local interface or EPICS.

SUPPORTED INSTRUMENTS

> 200



FAIRmat's federated data infrastructure



NOMAD core functionalities

Query the repository by element, material, or method, using NOMAD's customized GUI dashboard or programmatically.

The screenshot displays the NOMAD web interface. At the top, navigation links include PUBLISH, EXPLORE, ANALYZE, and ABOUT. A user is logged in as 'Pepe Marquez'. The main section is titled 'Entries' and shows 662 results. On the left, a 'FILTERS' sidebar lists categories: Material, Elements / Formula, Elements, Structure, Method, DFT, GW, Projection, DMFT, EELS, Workflow, and Molecular Dynamics. The 'Elements / Formula' filter is active, showing 'Se' and 'Zn' selected. The central area features a periodic table where elements Se (34) and Zn (30) are highlighted in green. Below the table, there are input fields for 'Chemical Formula Hill' and 'Chemical Formula IUPAC'. On the right, an 'Authors' list includes Markus Scheidgen, Hieu Ngo, Niloofar Hadaeghi, and Honghui Shang, each with a right-pointing arrow.

NOMAD core functionalities

Access effortlessly essential (meta)data visualized in the overview page of each entry.

NOMAD core functionalities

Navigation: PUBLISH EXPLORE ANALYZE ABOUT

Entries / Entry / Data

Welcome [Pepe Marquez](#) LOGOUT UNITS

Tabs: OVERVIEW FILES DATA LOGS

Search: Type your keyword here

Filters: ☐ code specific ☐ all defined ☐ definitions <> ↺

Browse the detailed (meta)data of each entry within a standardized structured schema.

System

section

QUANTITIES

- type = bulk
- configuration_raw_gid = _OT2z27hNuJyVSPqchZr2Im6szSC
- is_representative = true
- chemical_composition = KKSeZn
- chemical_composition_hill = K2SeZn
- chemical_composition_reduced = K2SeZn


SUB SECTIONS

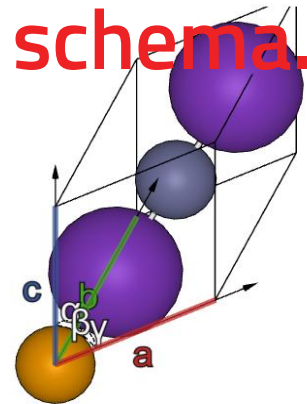
- atoms**
- prototype
- symmetry

REFERENCED BY [closed](#)

Atoms

section





QUANTITIES

- species = 4 vector
- labels = 4 list
- positions = 4 x 3 matrix

lattice_vectors

quantity

VALUE

0	4.02990	4.02990
4.02990	0	4.02990
4.02990	4.02990	0

(3 x 3)
Å

REFERENCED BY [closed](#)

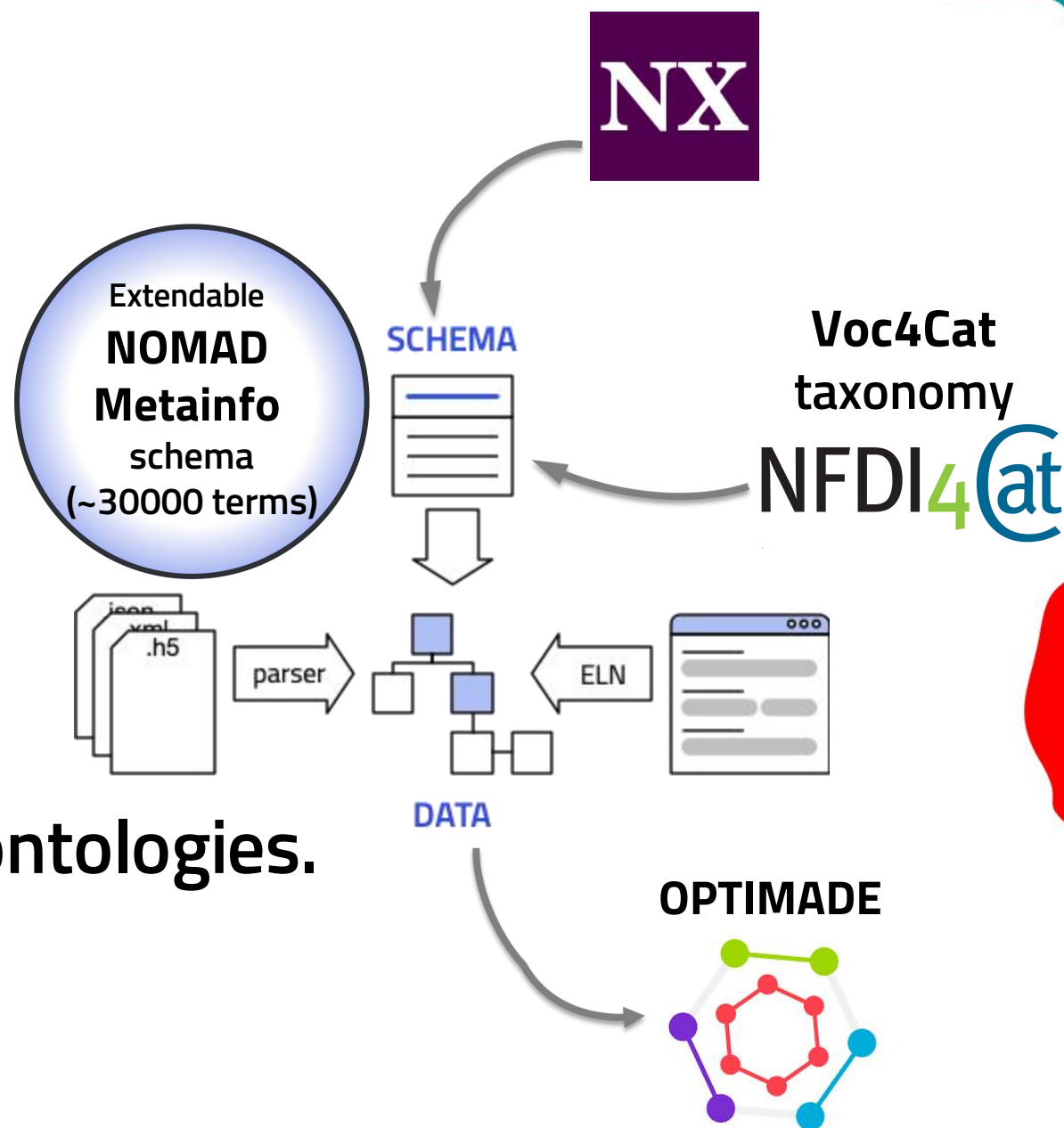
NOMAD core functionalities

The screenshot displays the NOMAD web application interface. At the top, there is a navigation bar with links for PUBLISH, EXPLORE, ANALYZE, and ABOUT. Below this, a header section shows 'Your uploads / Examples for demonstrations' and a 'Si_gw' logo. On the right, a user is logged in as 'Ahmed Mansour' with options to LOGOUT or adjust UNITS. The main content area features a blue bar with the text 'DROP FILES HERE OR CLICK TO OPEN DIALOG' and a 'CREATE FROM SCHEMA' button. A file upload dialog is open, showing a list of files with their names and types. The files include BAND.OUT, BAND-QP.OUT, bandstructure, bandstructure, bandstructure-qp, dos, EIGVAL.OUT, EIGVAL_GW.OUT, EVALQP, GW_INFO.OUT, info, INFO.OUT, input, TDOS.OUT, and TDOS-QP.OUT. The types are categorized as OUT File, DAT File, or XML Document. A red text overlay is positioned over the center of the image.

Upload your research data into NOMAD, share them with collaborators, and publish data sets with DOI.

Structured data

- NOMAD is built on structured data based on well-defined schema (*NOMAD Metainfo*).
- *NOMAD Metainfo* can be extended and enhanced with standardized vocabularies and ontologies.



CrystalLLM: a FAIR full-cycle example



Crystal Structure Generation with Autoregressive Large Language Modeling

Luis M. Antunes¹, Keith T. Butler^{1,2}, and Ricardo Grau-Crespo¹

¹*Department of Chemistry, University of Reading, Whiteknights, Reading RG6 6DX, United Kingdom.
l.m.antunes@pgr.reading.ac.uk*

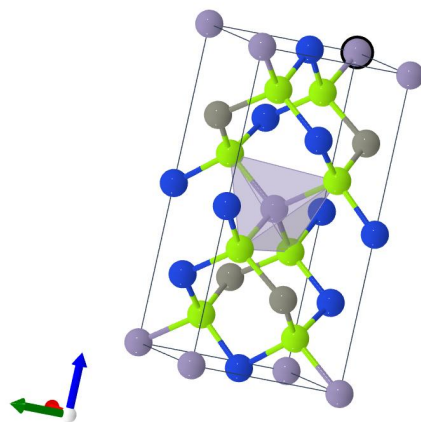
²*School of Engineering and Materials Science Queen Mary University of London Mile End Road
London E1 4NS, United Kingdom.*

Generate a crystal structure from a composition *

Composition:

▼ Advanced options

Model size:

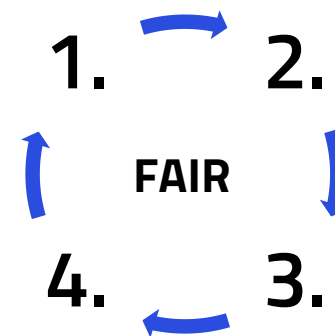


Zn Cu Sn Se

I-4 a=5.719Å b=5.719Å c=11.619Å α=90.000° β=90.000° γ=90.000°

<https://crystallm.com/>

1. Users upload the INPUT and OUTPUT files to NOMAD from different codes from their publications
2. NOMAD makes the data FAIR
3. Antunes et al. build CrystalLLM with data from NOMAD
4. Users can use CrystalLLM to draft their inputs for their new calculations



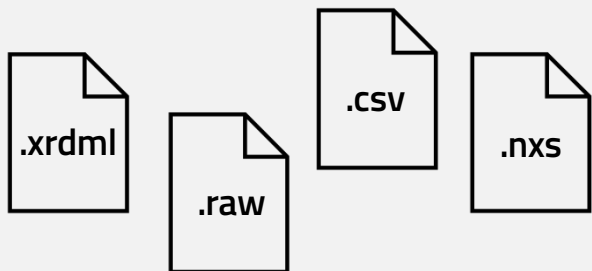
NOMAD Oasis



Share selectively

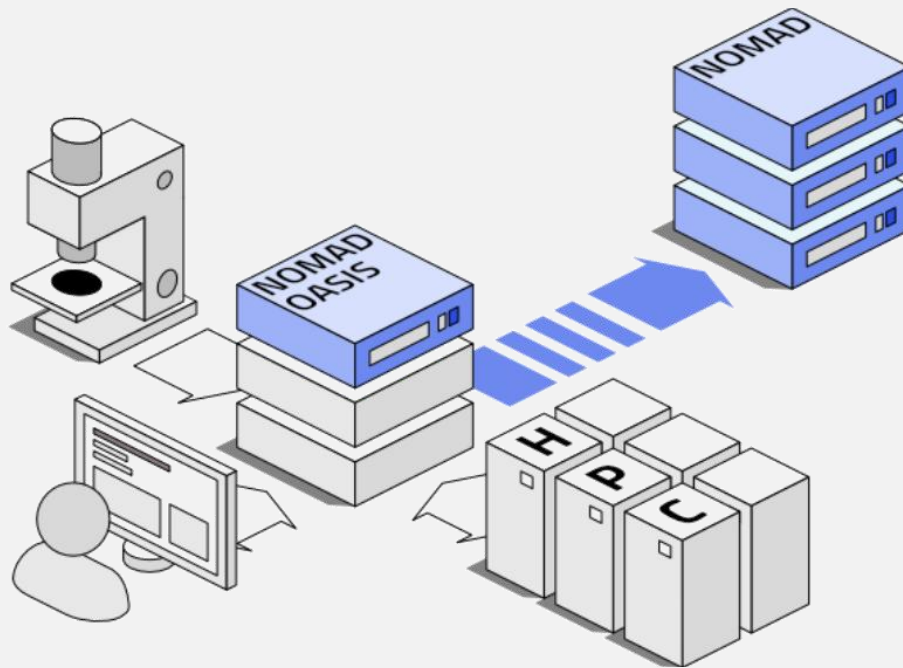


Automate data



**Search apps
database in a
database**

Manage your lab



ELN



API



NORTH
Cloud based analysis



Large files



NOMAD Oasis



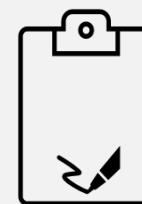
Share selectively



Manage your lab



ELN



API



Automate data



Search apps
database in a
database

NORTH
Cloud based analysis



Large files



NOMAD Oasis

OVERVIEW

FILES

DATA

LOGS

Metadata

type
ExampleSolarCell
name
solar cell

comment
no comment
references
<https://doi.org/10.1016/j.solmat.201...>
authors
Pepe Marquez
datasets
no datasets

mainfile
solar_cell.archive.json
entry id
[7hJA-uxMhJUC8N9_-rKyQyyNszP](#)
upload id
[r75jMoZzTcOFPPZd63BIew](#)
upload create time
3/23/2023, 12:27:21 PM
last processing time
3/24/2023, 4:52:39 PM
processing version
1.1.9.dev76+gc24f16631/

API

ExampleSolarCell

Description

Paragraph B I

This is an example of an ELN for a single solar cell.

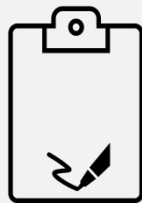
P 12 WORDS POWERED BY TINY

Datetime
23/03/2023 12:28

SolarCellParameters



ELN



API



NORTH

Cloud based analysis



Large files



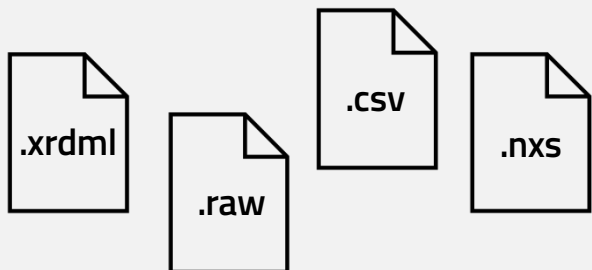
NOMAD Oasis



Share selectively

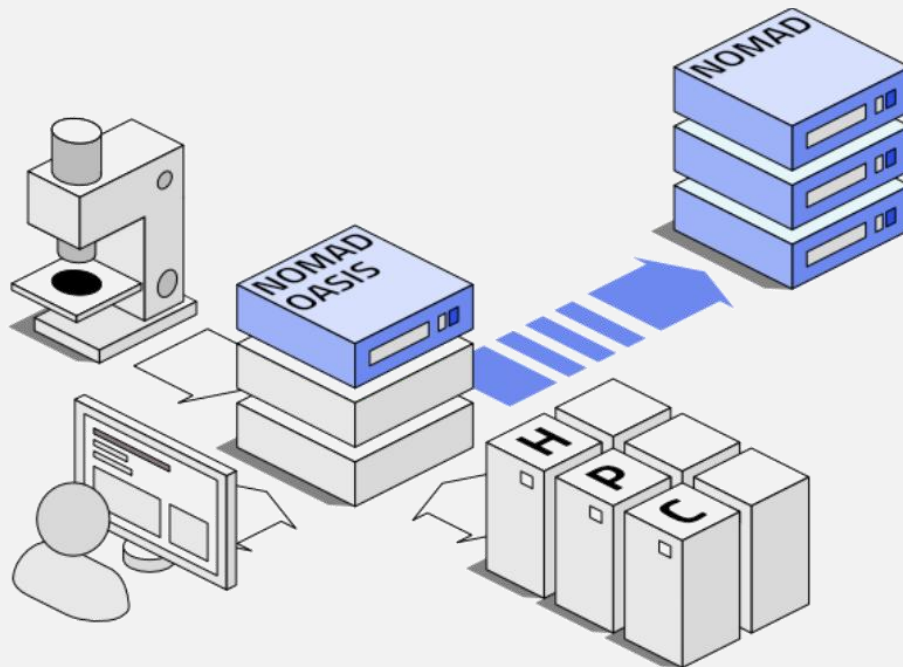


Automate data



**Search apps
database in a
database**

Manage your lab



ELN



API



NORTH
Cloud based analysis



Large files



NOMAD Oasis



Share selectively



Manage your lab



ELN



API



Automate data



Search apps
database in a
database

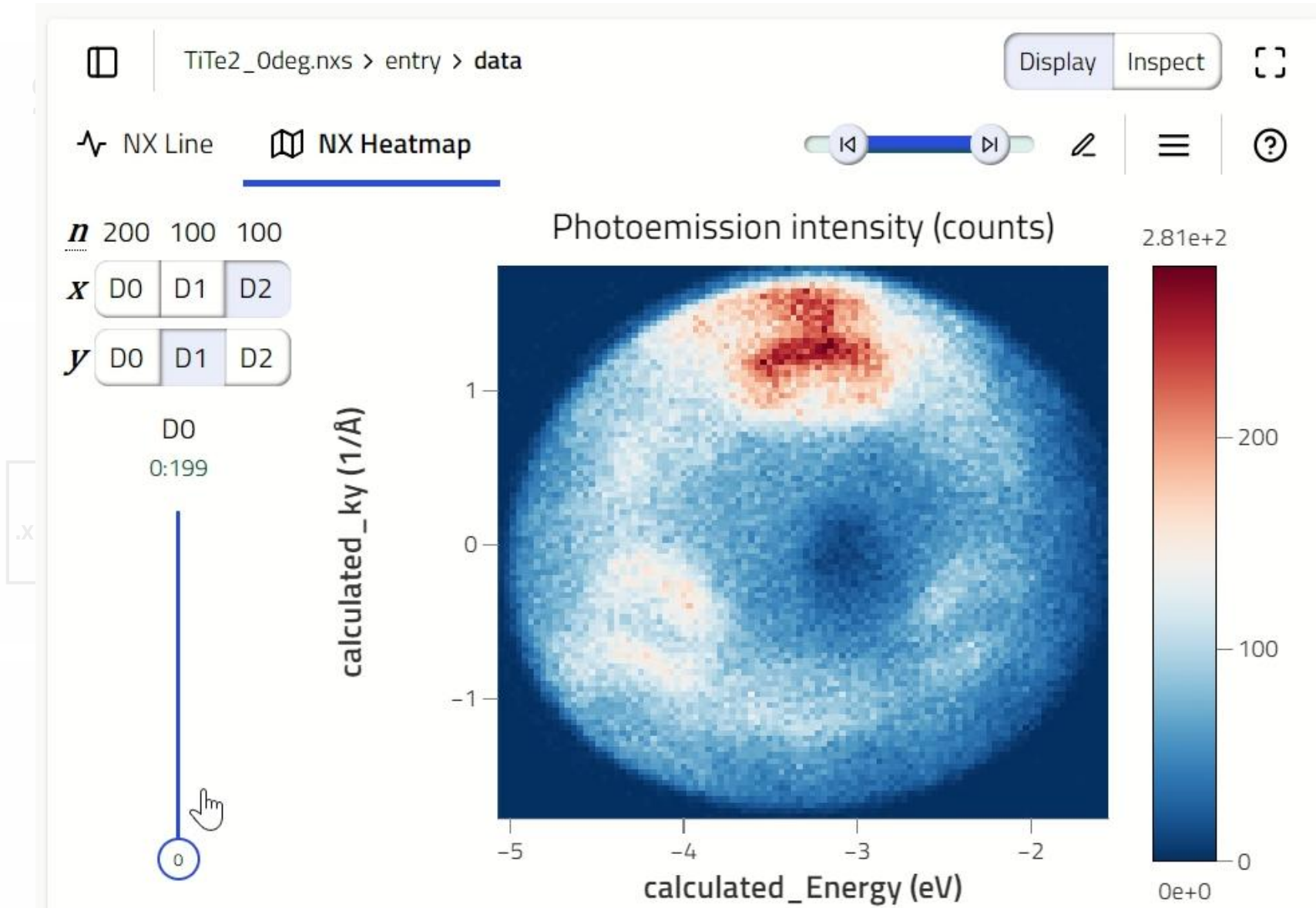
NORTH
Cloud based analysis



Large files



NOMAD Oasis



ELN



API



NORTH

Cloud based analysis



Large files



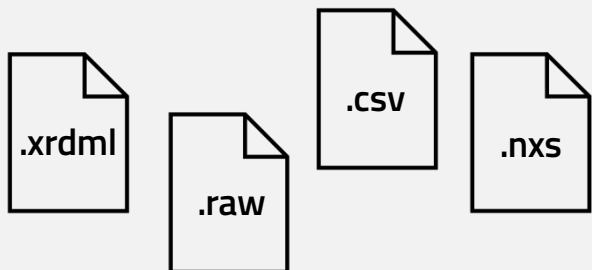
NOMAD Oasis



Share selectively

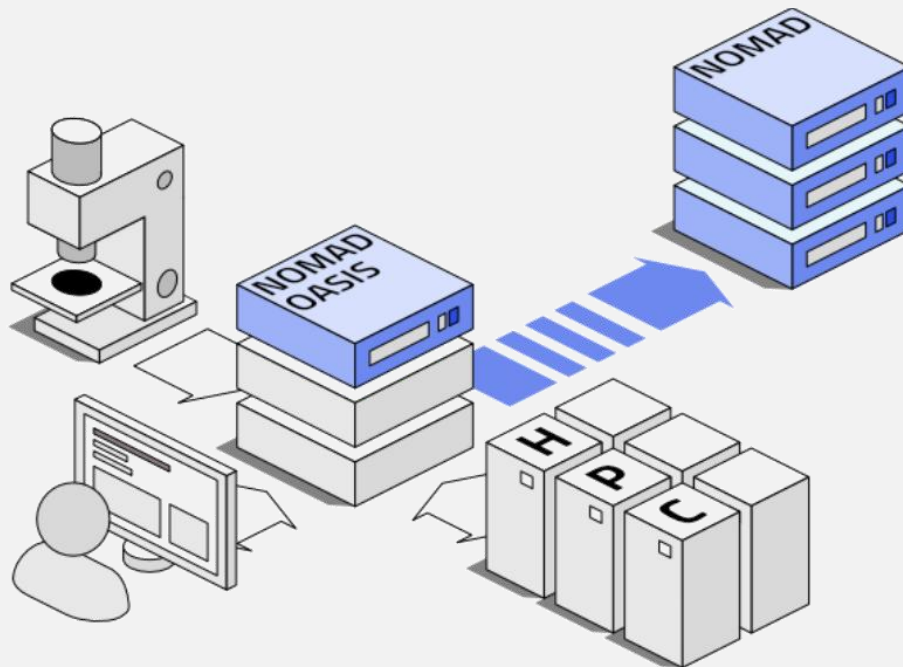


Automate data

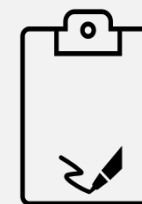


**Search apps
database in a
database**

Manage your lab



ELN



API



NORTH
Cloud based analysis



Large files



NOMAD Oasis



Share selectively



Manage your lab



ELN



API



Automate data



Search apps
database in a
database

NORTH
Cloud based analysis



Large files



NORTH (NOMAD Remote Tools Hub)



Share selectively



Manage your lab

ELN



API



Automate data



Search apps
database in a
database



NORTH
Cloud based analysis



Large files



NORTH (NOMAD Remote Tools Hub)



Share selectively



Manage your lab

ELN



API



Authenticate data



Search apps
database in a
database



NORTH
Cloud based analysis



Large files



NORTH (NOMAD Remote Tools Hub)




PUBLISH ▾ EXPLORE ▾ ANALYZE ▾ ABOUT ▾

NOMAD Remote Tools Hub ?

Welcome [Pepe Marquez](#)

 LOGOUT  UNITS

 **jupyter** Basic jupyter run with an empty notebook or on given notebook file. ^

Jupyter Notebook: The Classic Notebook Interface

The Jupyter Notebook is the original web application for creating and sharing computational documents. It offers a simple, streamlined, document-centric experience.

Maintainer: [Markus Scheidgen](#)


File extensions: ipynb


LAUNCH

 **pyiron** Jupyterlab with pyiron installed.

 **nionswift** Run Nion Swift to analyze data as well as prepare focus series reconstructions in NOMAD.

 **nexustools** Analyse your NeXus files in NOMAD with several NeXus-compatible tools. ▾

 **fiji** Use FIJI to analyze and visualize your images in your NOMAD files. ▾

 **vesta** Run VESTA to analyse and visualize your crystal structures in your NOMAD files. ▾

NORTH
Cloud based analysis



Large files



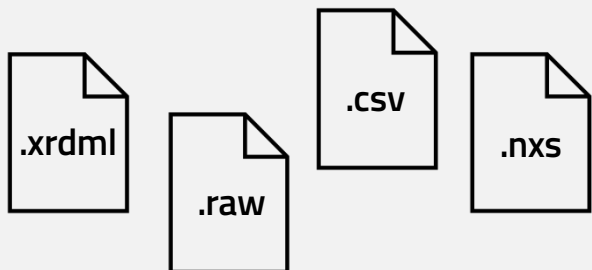
NOMAD Oasis



Share selectively

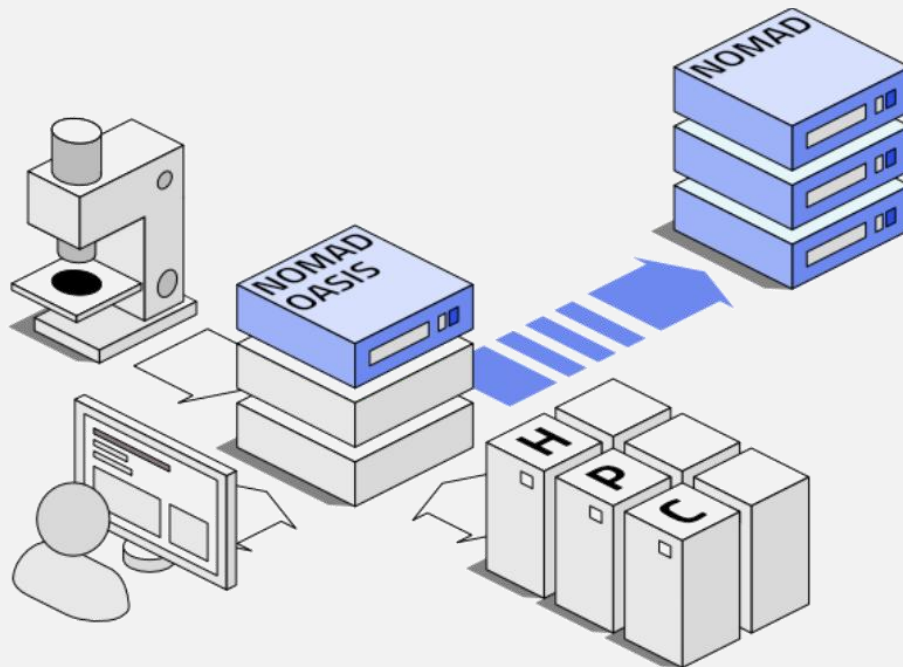


Automate data



**Search apps
database in a
database**

Manage your lab



ELN



API



NORTH
Cloud based analysis



Large files



NOMAD Oasis



Share selectively



Manage your lab

ELN



API



Automate data



**Search apps
database in a
database**



NORTH
Cloud based analysis



Large files



TADF Molecules ?

 LOGOUT
  UNITS

↑ FILTERS

+ TERMS

HISTOGRAM

SCATTER PLOT

 PERIODIC TABLE

↶ ↷ ↵

linear ▾ [illegible]

59 Pr 0	60 Nd 0	61 Pm 0	62 Sm 0	63 Eu 0	64 Gd 0	65 Tb 0	66 Dy 0	67 Ho 0	68 Er 0	69 Tm 0	70 Yb 0	71 Lu 0
91 Pa 0	92 U 0	93 Np 0	94 Pu 0	95 Am 0	96 Cm 0	97 Bk 0	98 Cf 0	99 Es 0	100 Fm 0	101 Md 0	102 No 0	103 Lr 0

☒ autorange

A scatter plot showing the relationship between Peak emission wavelength (nm) on the x-axis and Delayed lifetime (s) on the y-axis. The x-axis ranges from 350 to 850 nm, and the y-axis is a logarithmic scale from 100 ns to 0.001 s. Data points are colored according to their PLQY, with a color bar on the right ranging from 0.2 (dark blue) to 0.8 (yellow). The plot shows a general trend where delayed lifetime decreases as peak emission wavelength increases, with higher PLQY values generally corresponding to longer delayed lifetimes.

Search apps database in a database

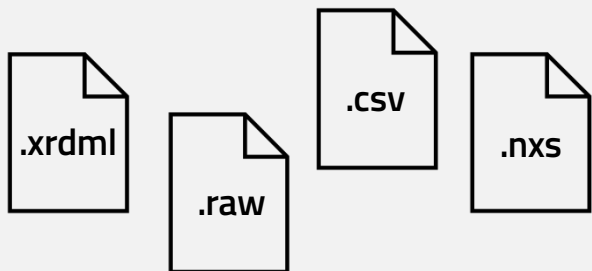
NOMAD Oasis for Experiments



Share selectively

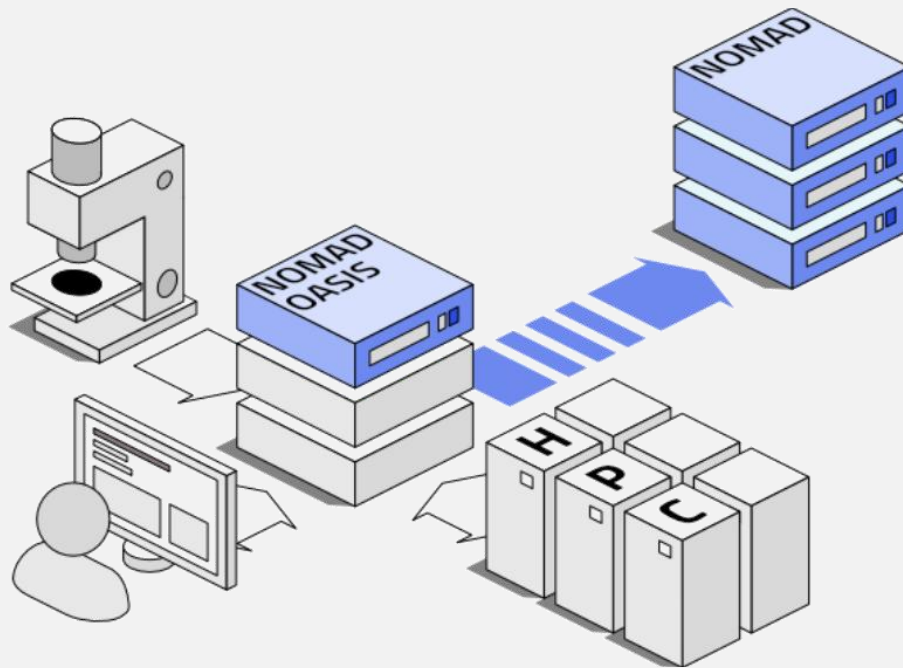


Automate data

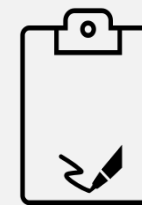


**Search apps
database in a
database**

Manage your lab



ELN



API



NORTH
Cloud based analysis



Large files

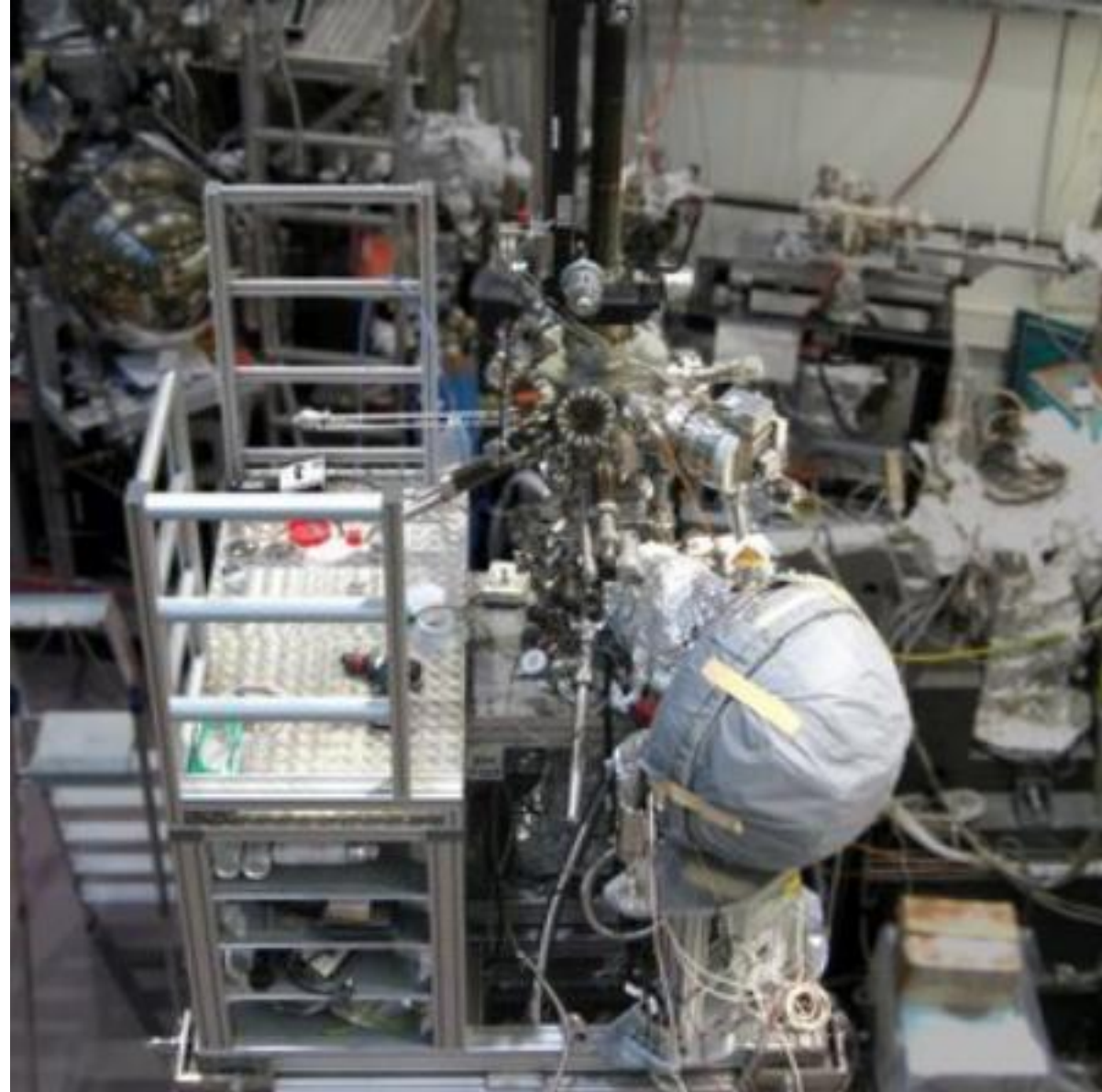


Experiments

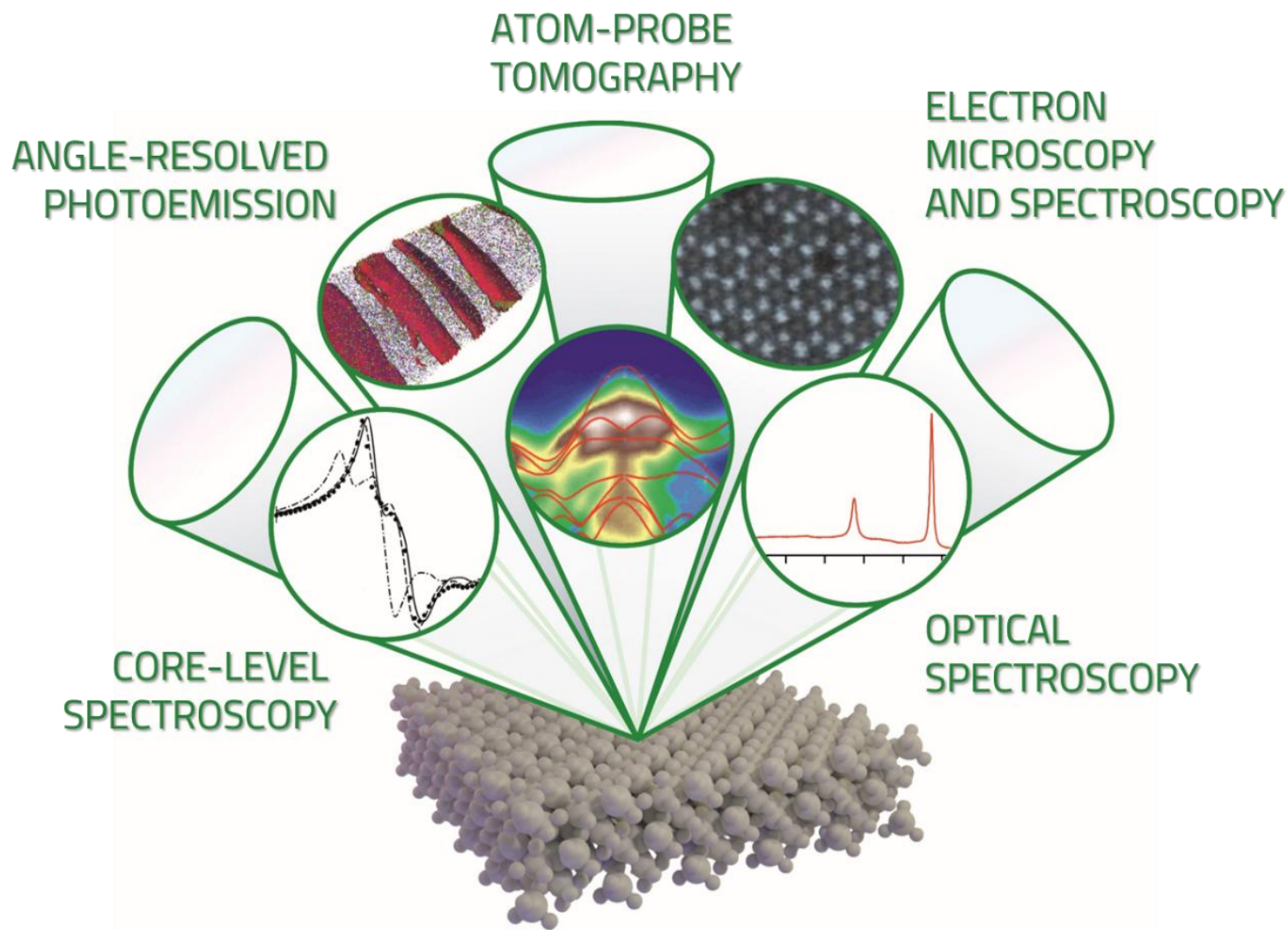
Ever-new methodology

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

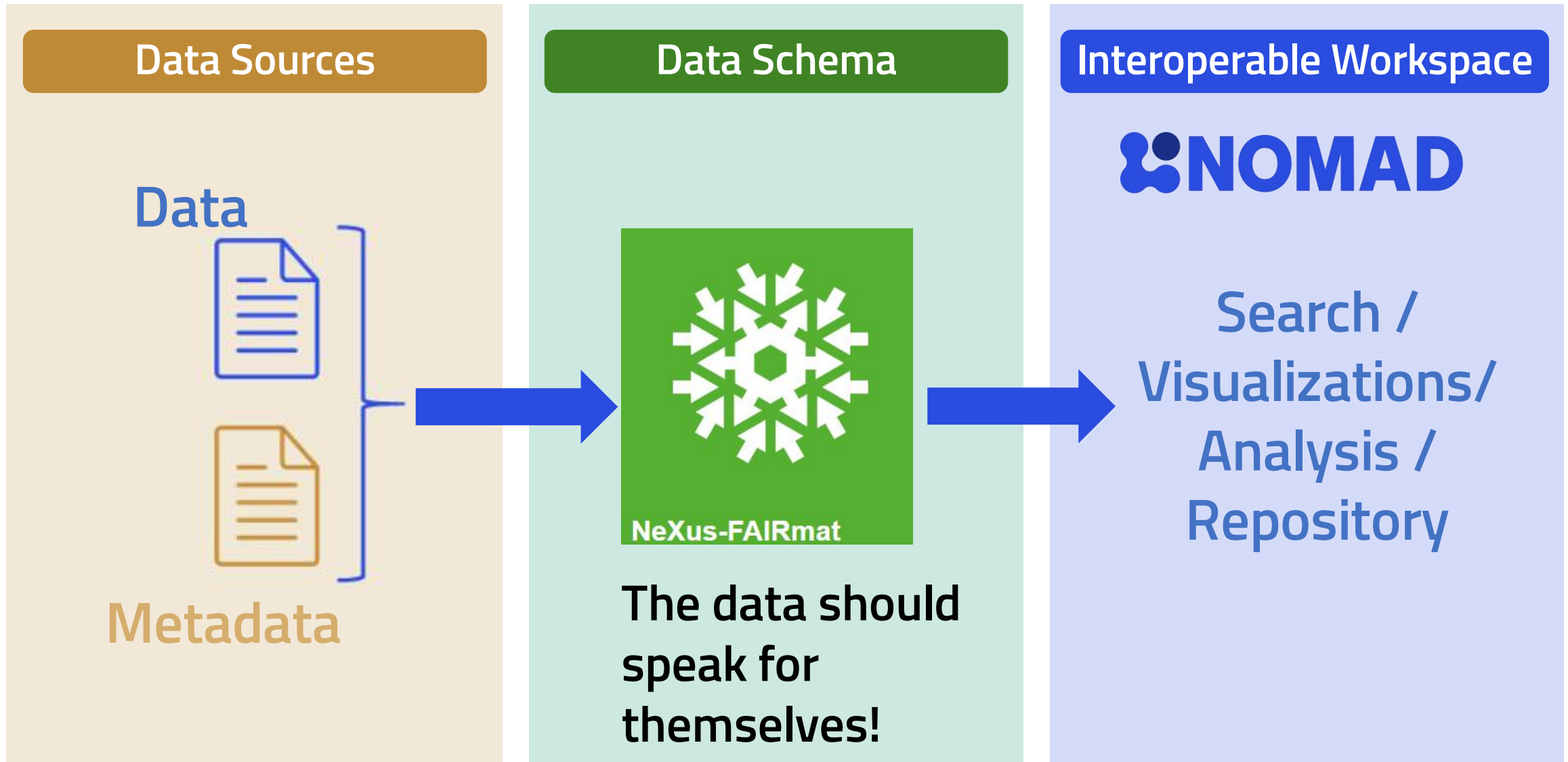
Broad and structured data are barely available.



FAIRmat's pilot experiments



Experimental data workflow



Harmonization: the NeXus proposal



NeXus: \approx 30 year-old description schema for experiments in the electron/neutron/photon beam community.

FAIRmat contributes most actively to new standards and plays an important role in the NeXus committee NIAC.

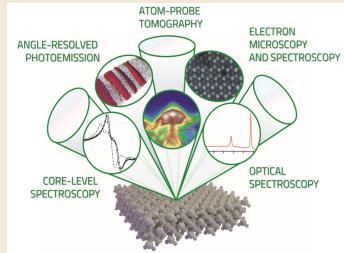
Harmonized structure of experimental data

https://fairmat-nfdi.github.io/nexus_definitions/

Experimental data workflow



Data Sources



Pynxtools

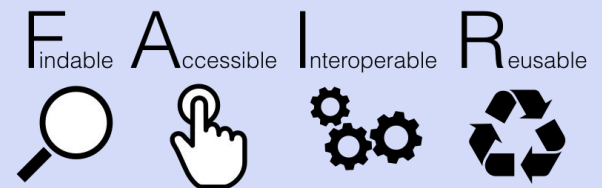
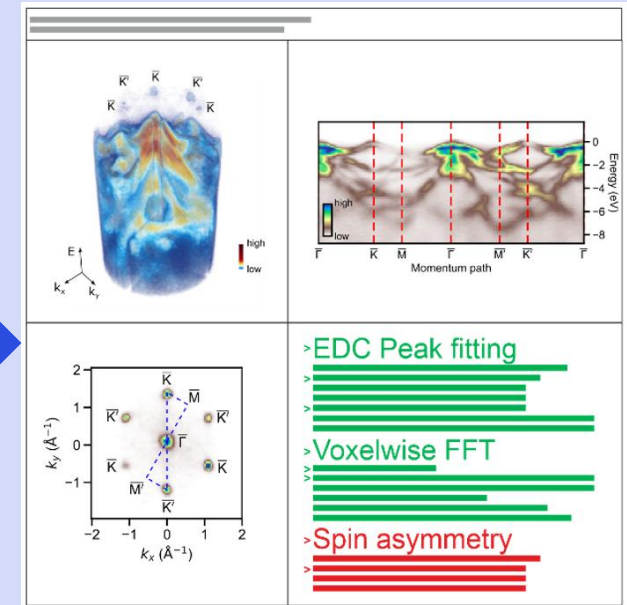
Data Schema

Community agreement!



Interoperable Workspace

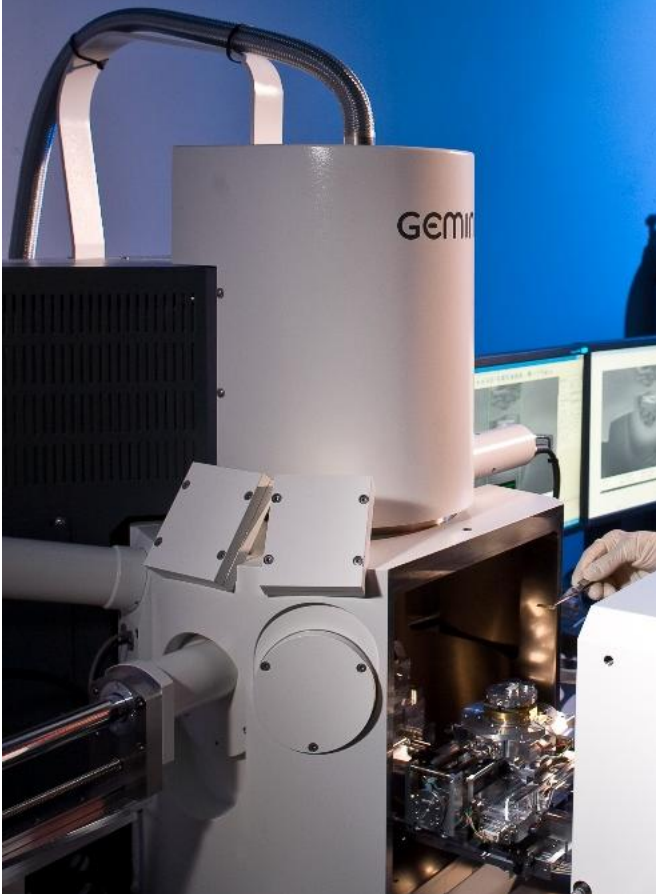
NOMAD



Experimental data sources

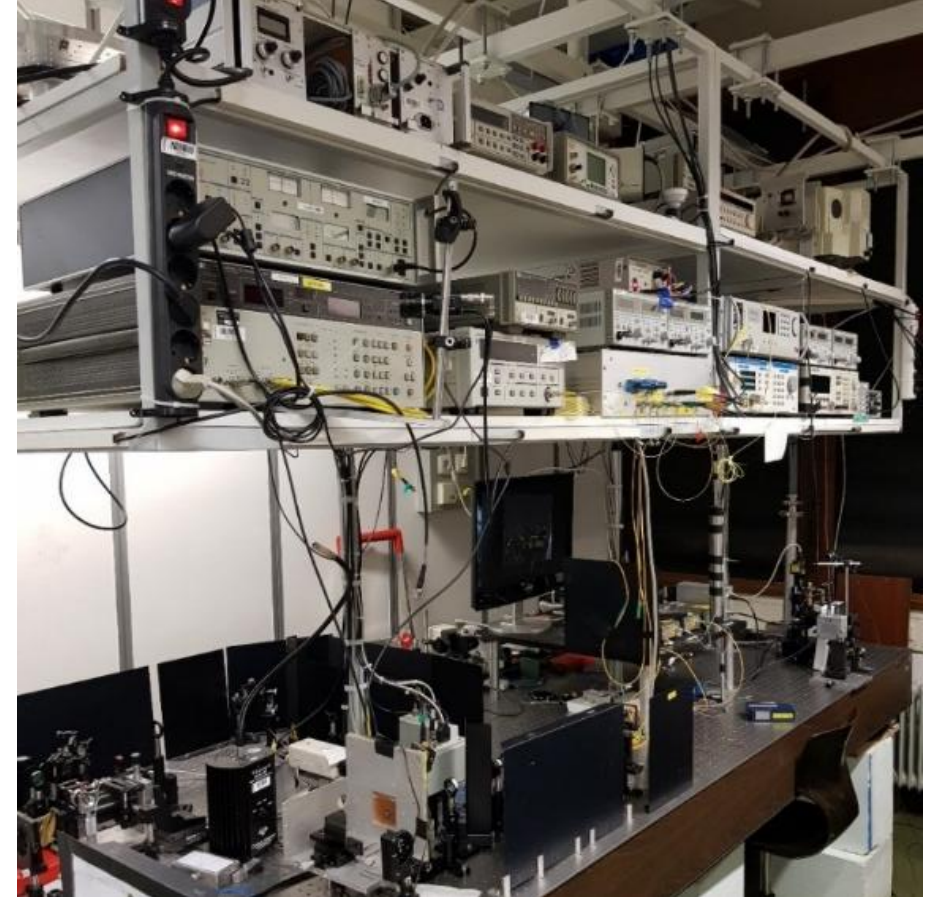


Commercial integrated systems



Vendor-provided software

Specialized custom systems

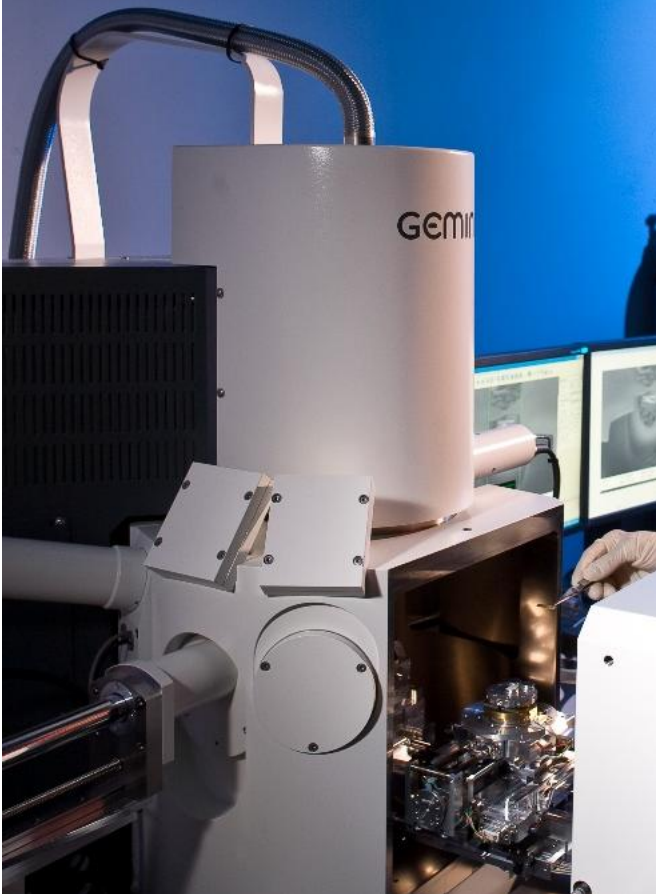


Self-written software

Experimental data sources

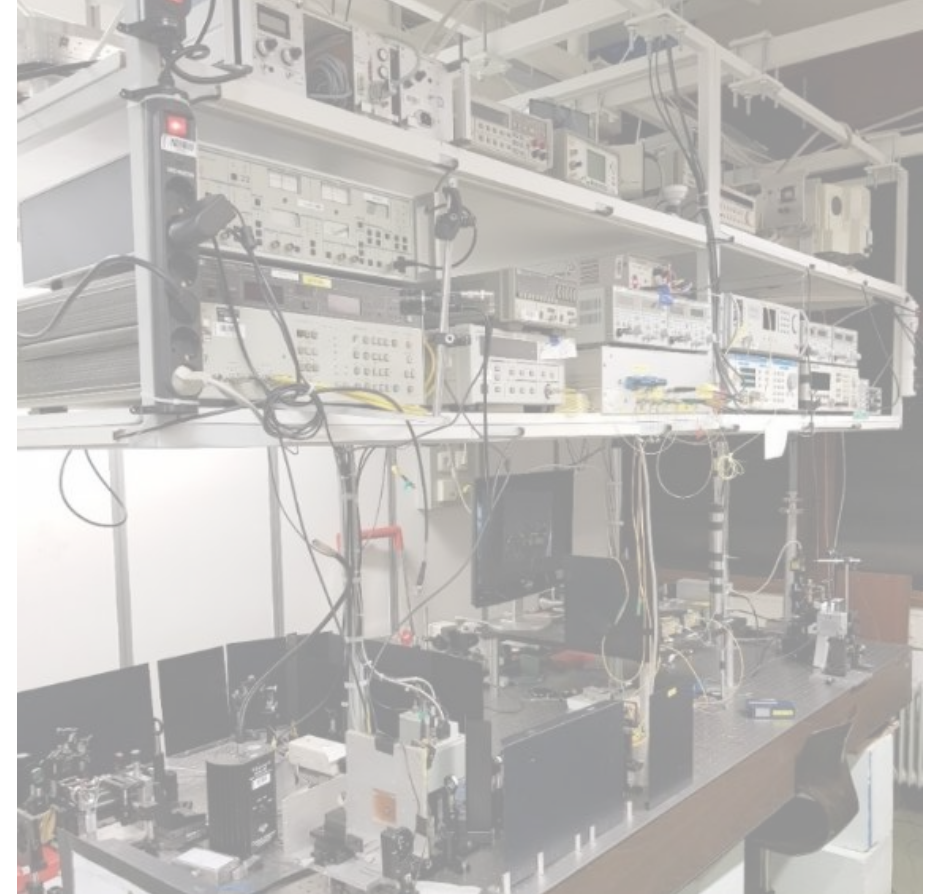


Commercial integrated systems



Vendor-provided software

Specialized custom systems



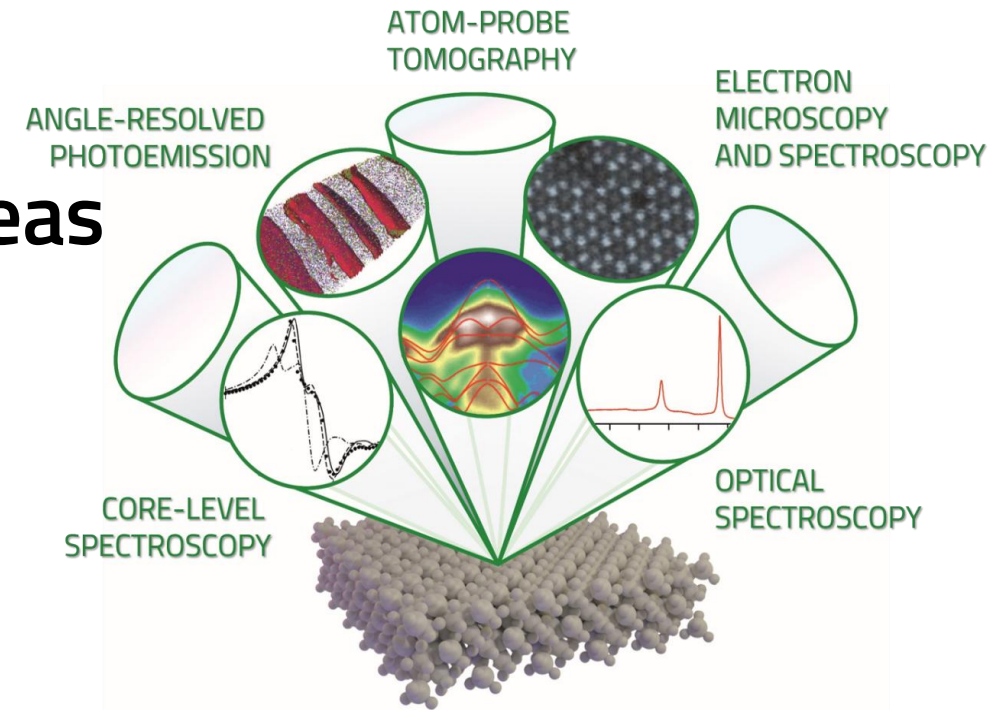
Self-written software

Organizing community standards



FAIRmat approach:

- Elaborate data schemas in selected areas
- Bring scientists, manufacturers and data experts together
- Handle concepts openly



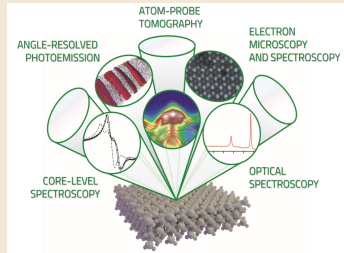
Manufacturers (Technology Partners) are teaming up for providing NeXus output



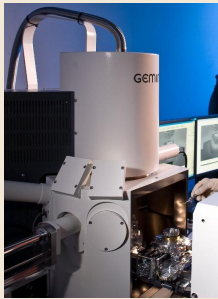
Experimental data workflow



Data Sources



Technology
partners



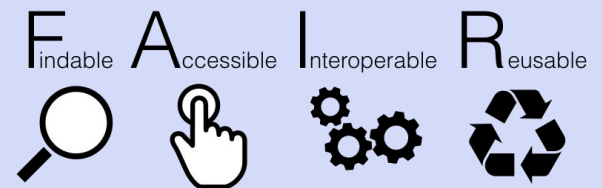
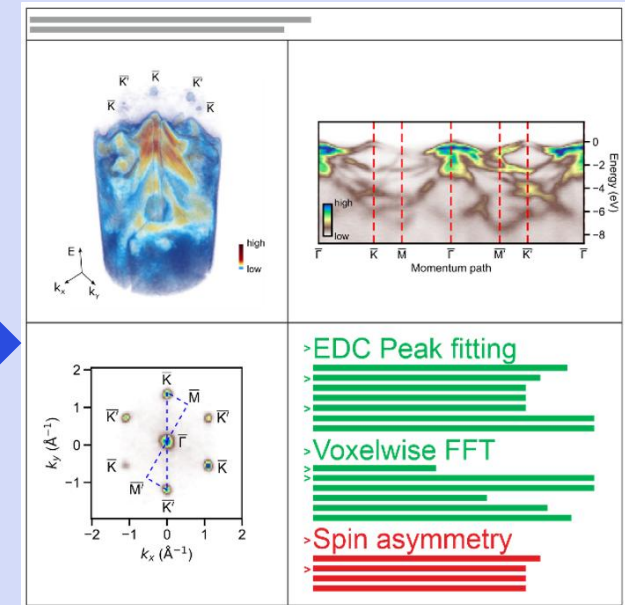
Data Schema

Community agreement!



Interoperable Workspace

NOMAD



Experimental data sources

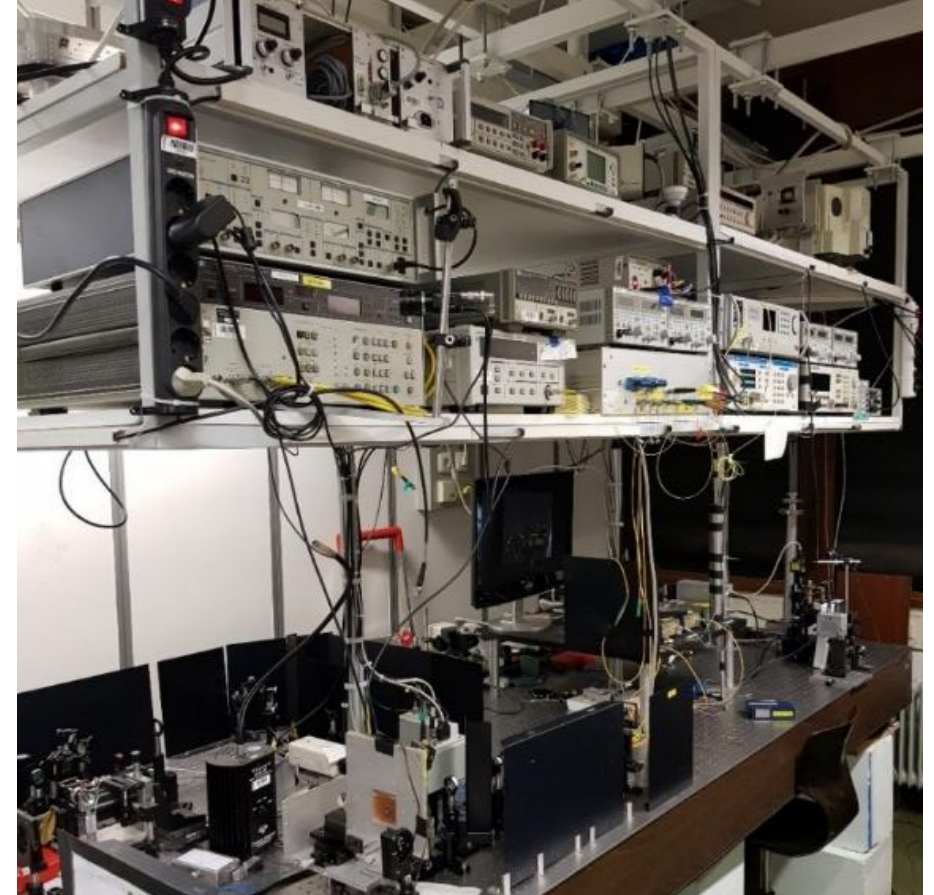


Commercial integrated systems



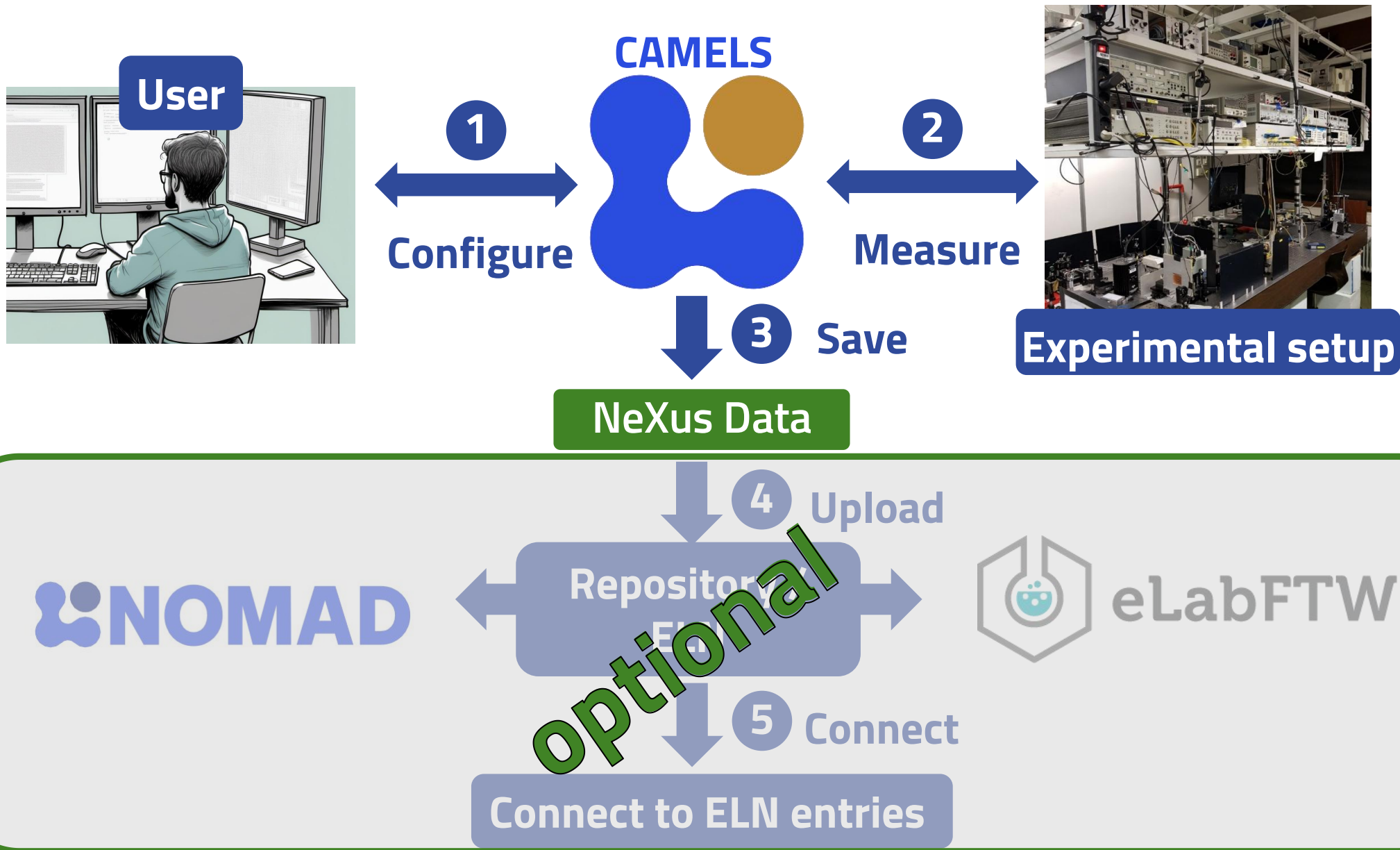
Vendor-provided software

Specialized custom systems



Self-written software

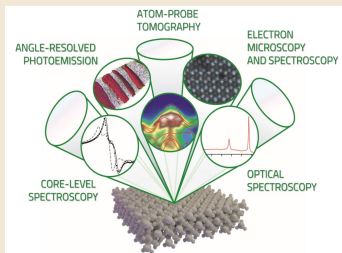
Workflow



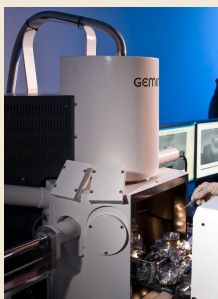
Experimental data workflow



Data Sources



Technology
partners



Any
Experiment



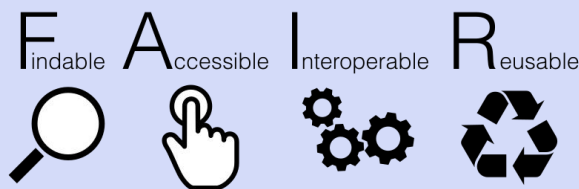
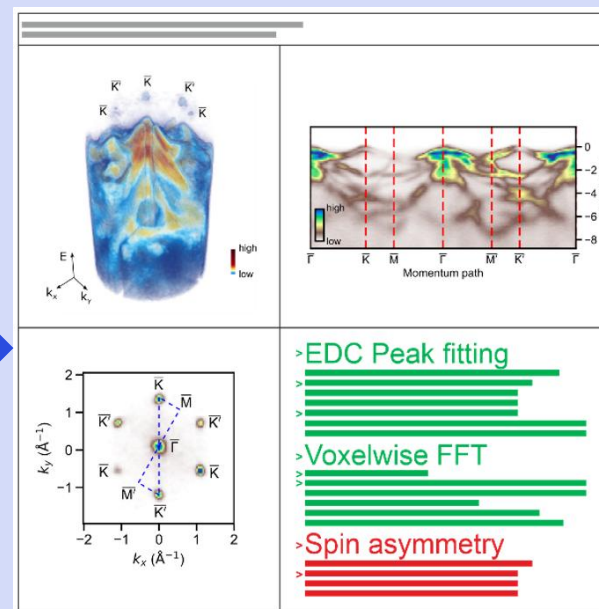
Data Schema

Community agreement!

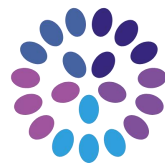
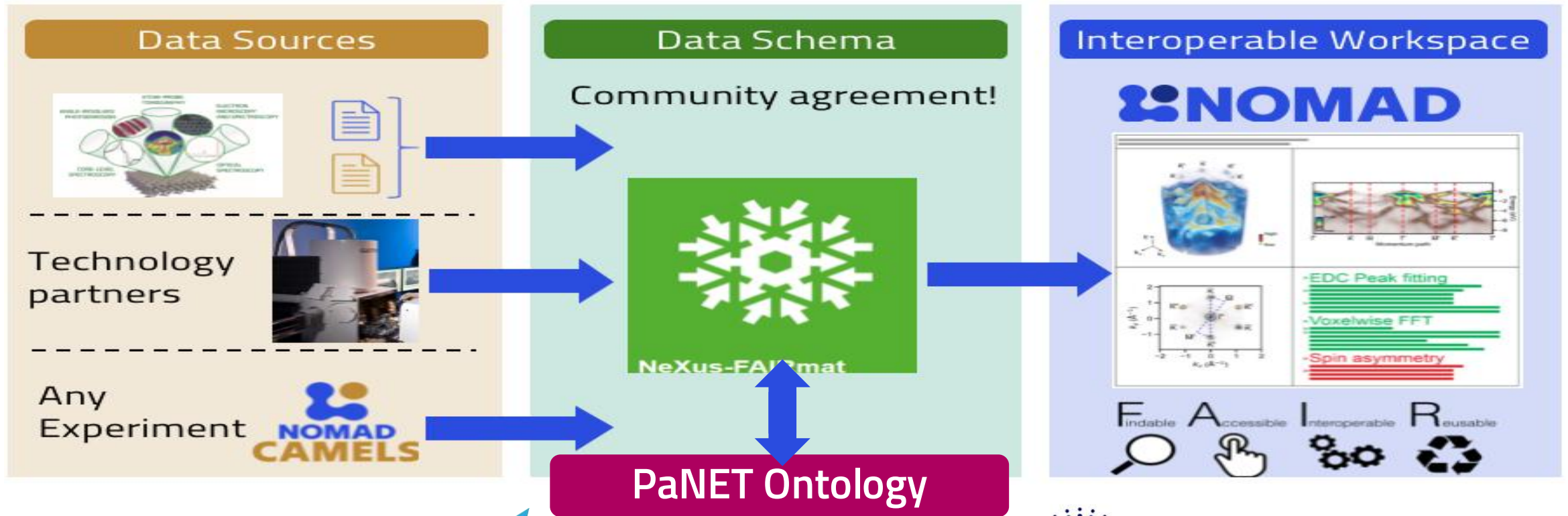


Interoperable Workspace

NOMAD



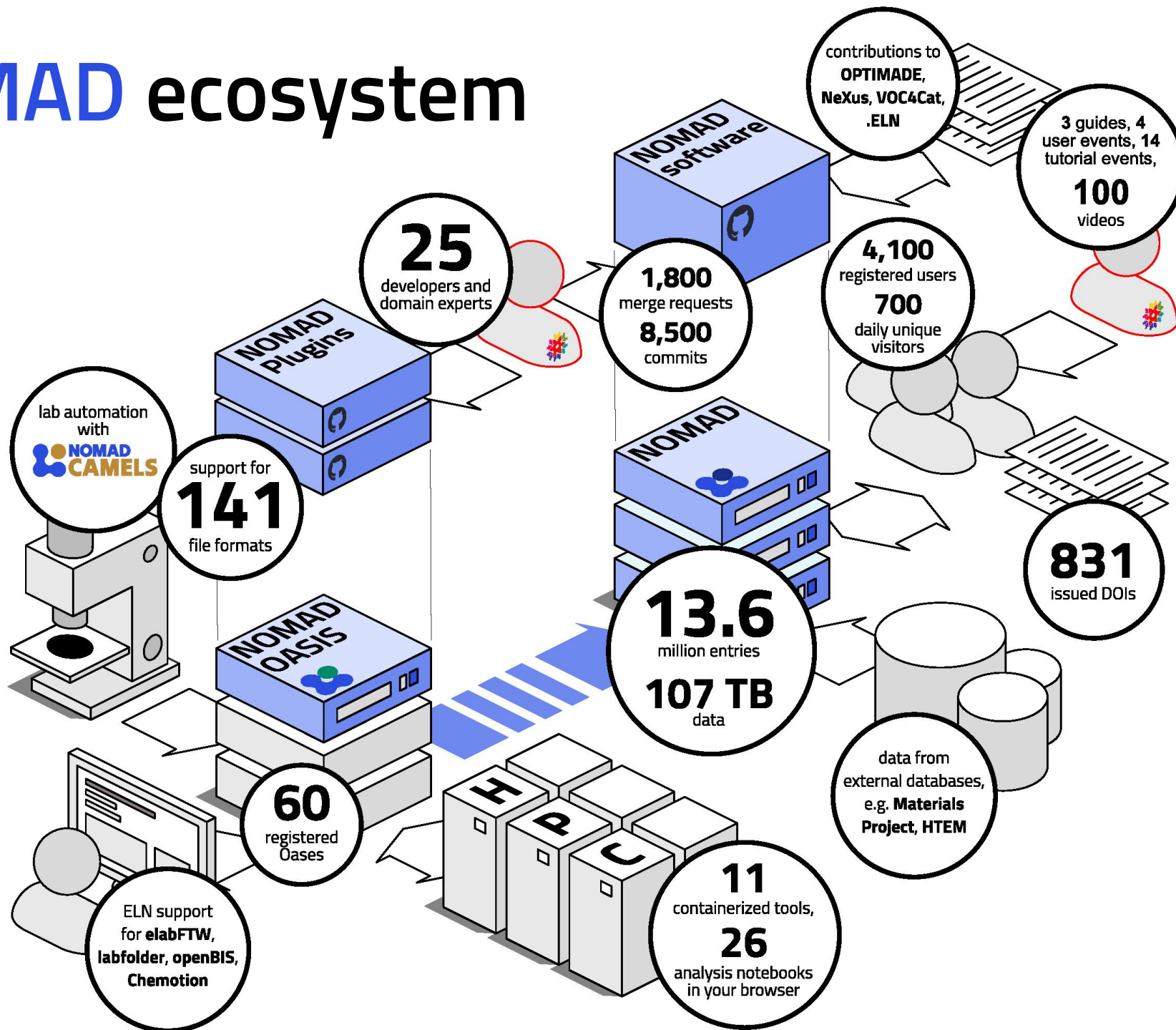
Experimental data workflow



OSCARS
Open Science Clusters' Action
for Research & Society



The **NOMAD** ecosystem



Thank you!

 www.fairmat-nfdi.eu

 fairmat@physik.hu-berlin.de

 company/fairmat-nfdi

 www.nomad-lab.eu

