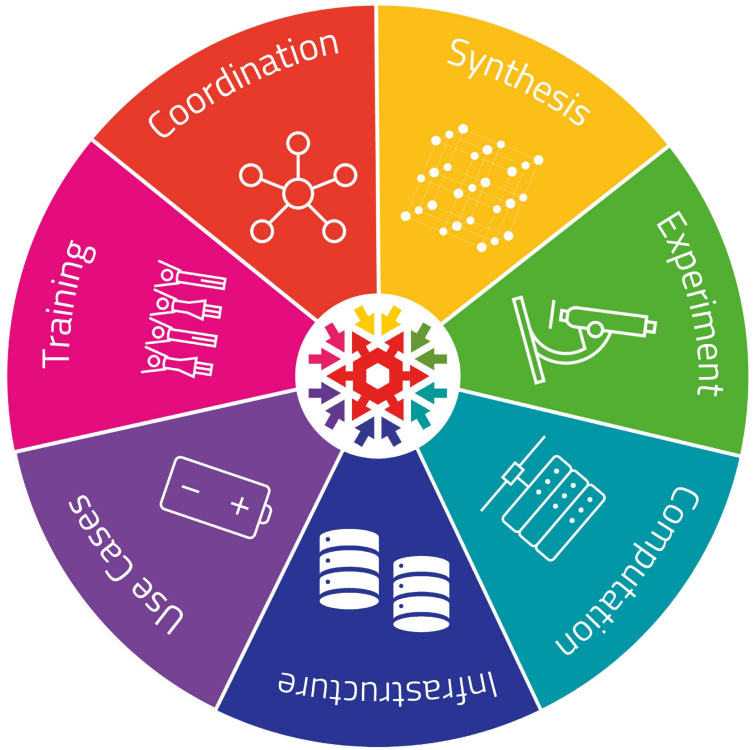
Research Data Management with FAIRmat and NOMAD

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

NFDI: National (German) Research Data Infrastructure

The FAIRmat team

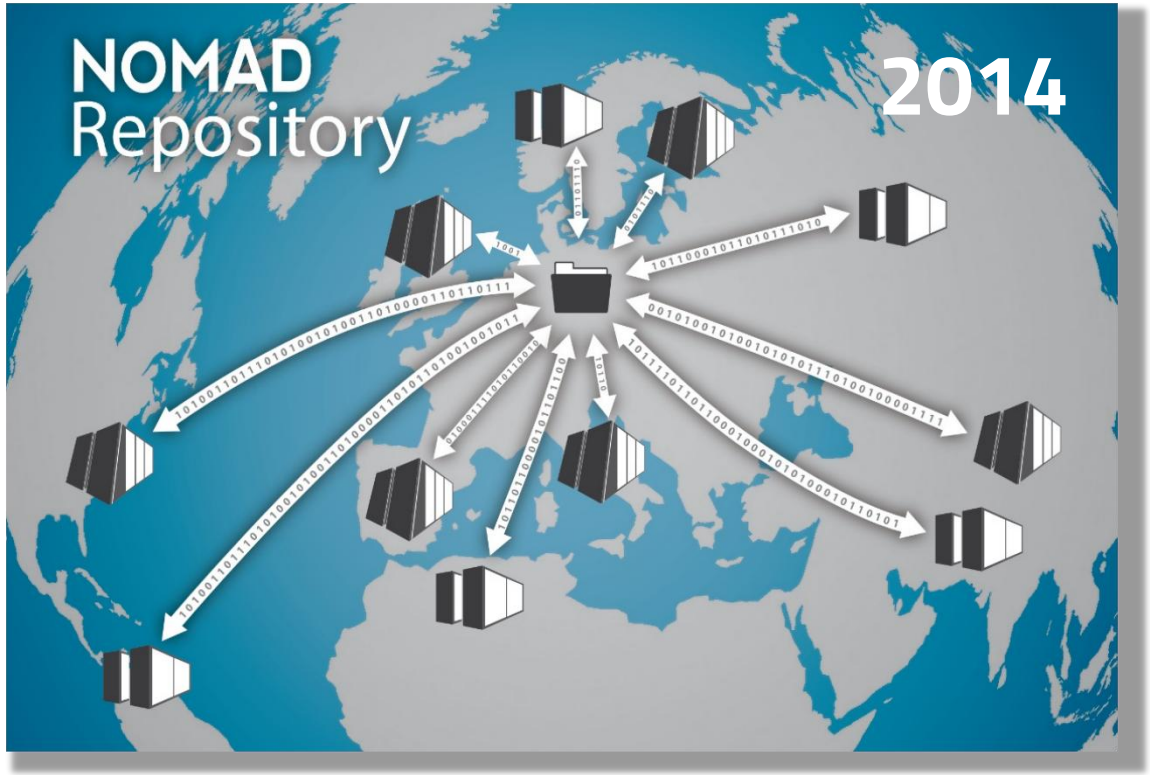
~60 PIs + ~30 central coworkers

Claudia Draxl
Spokesperson

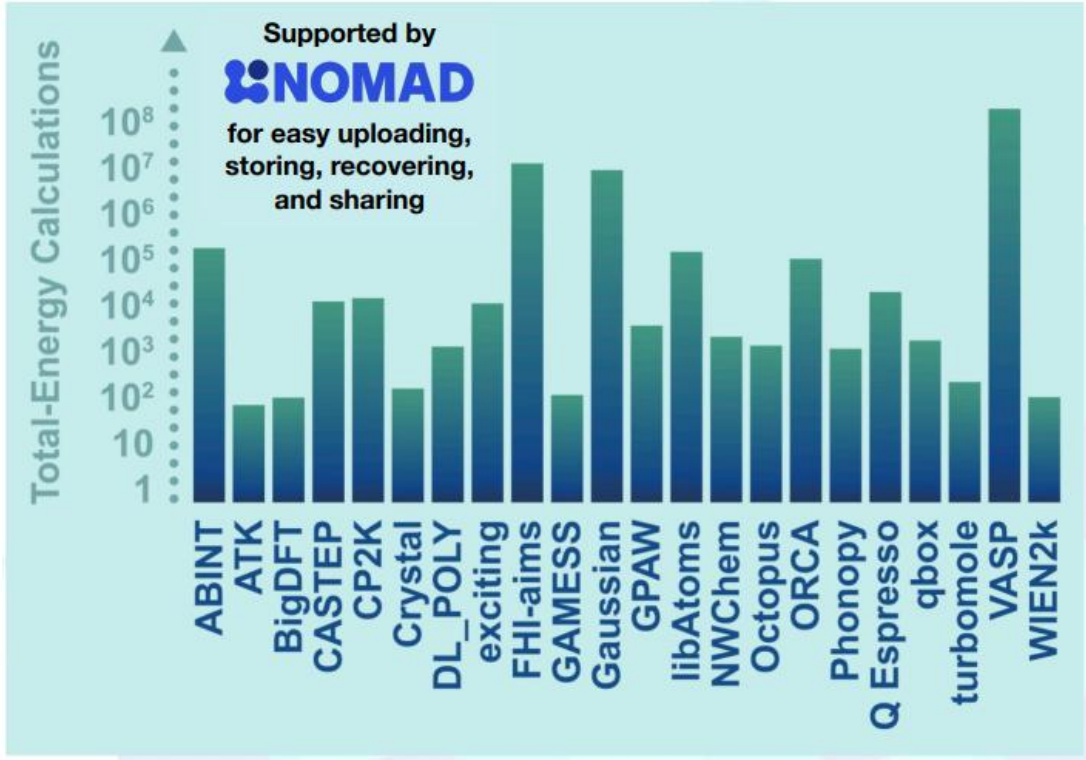
Christoph T. Koch
Deputy Spokesperson

The NOMAD central hub

The Novel Materials Discovery Laboratory



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

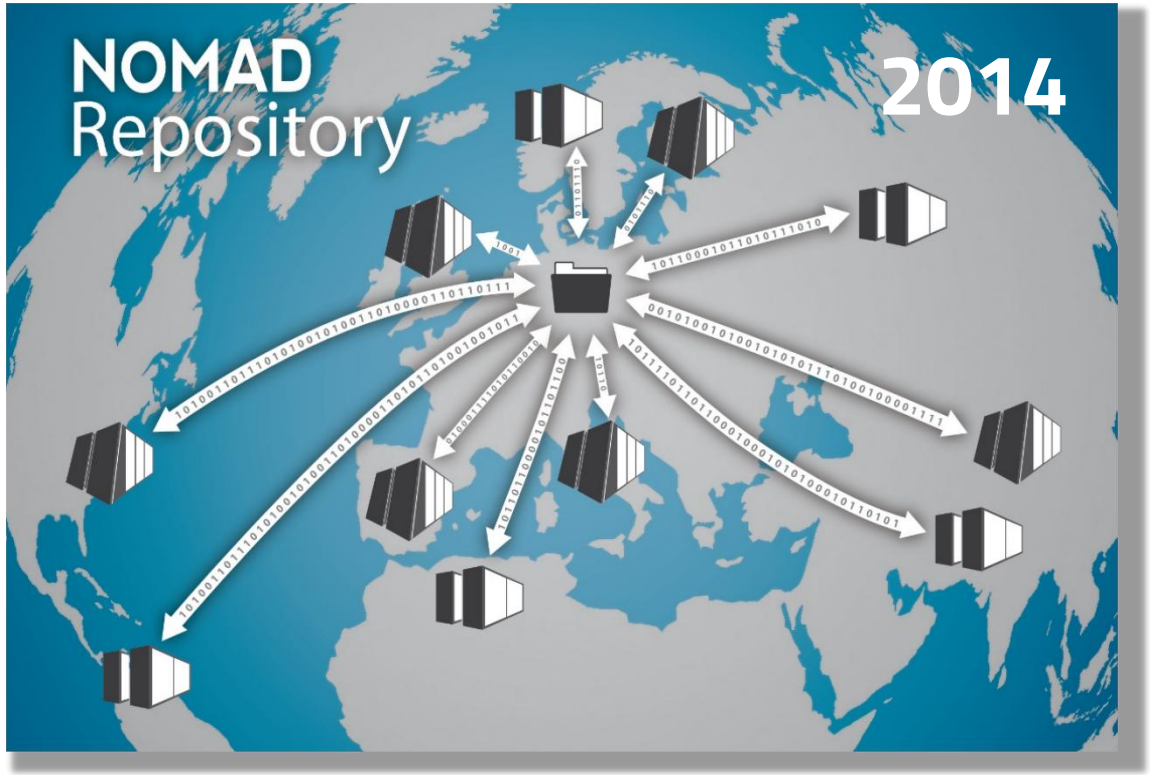


The NOMAD central hub

NOMAD

Materials data managed and shared

UPLOADED ENTRIES	REPRESENTED MATERIALS	UPLOADED FILES
16,469,399	3,751,773	110.6 TB

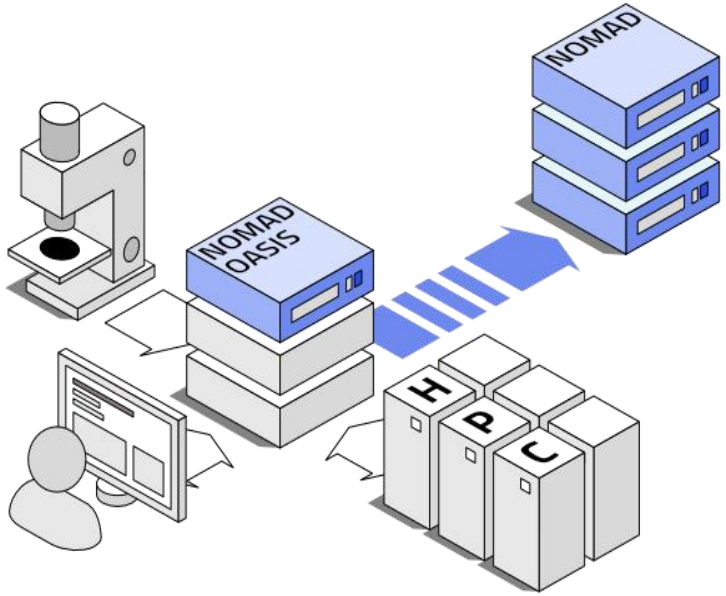


The NOMAD ecosystem

NOMAD Oasis

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

REGISTERED INSTALLATIONS
> 70



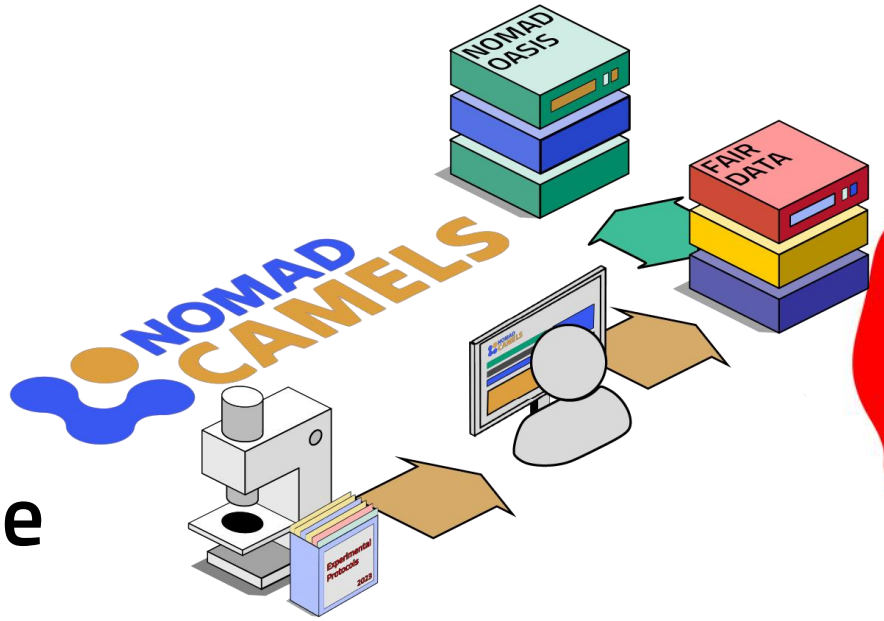
https://nomad-lab.eu/

NOMAD CAMELS

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

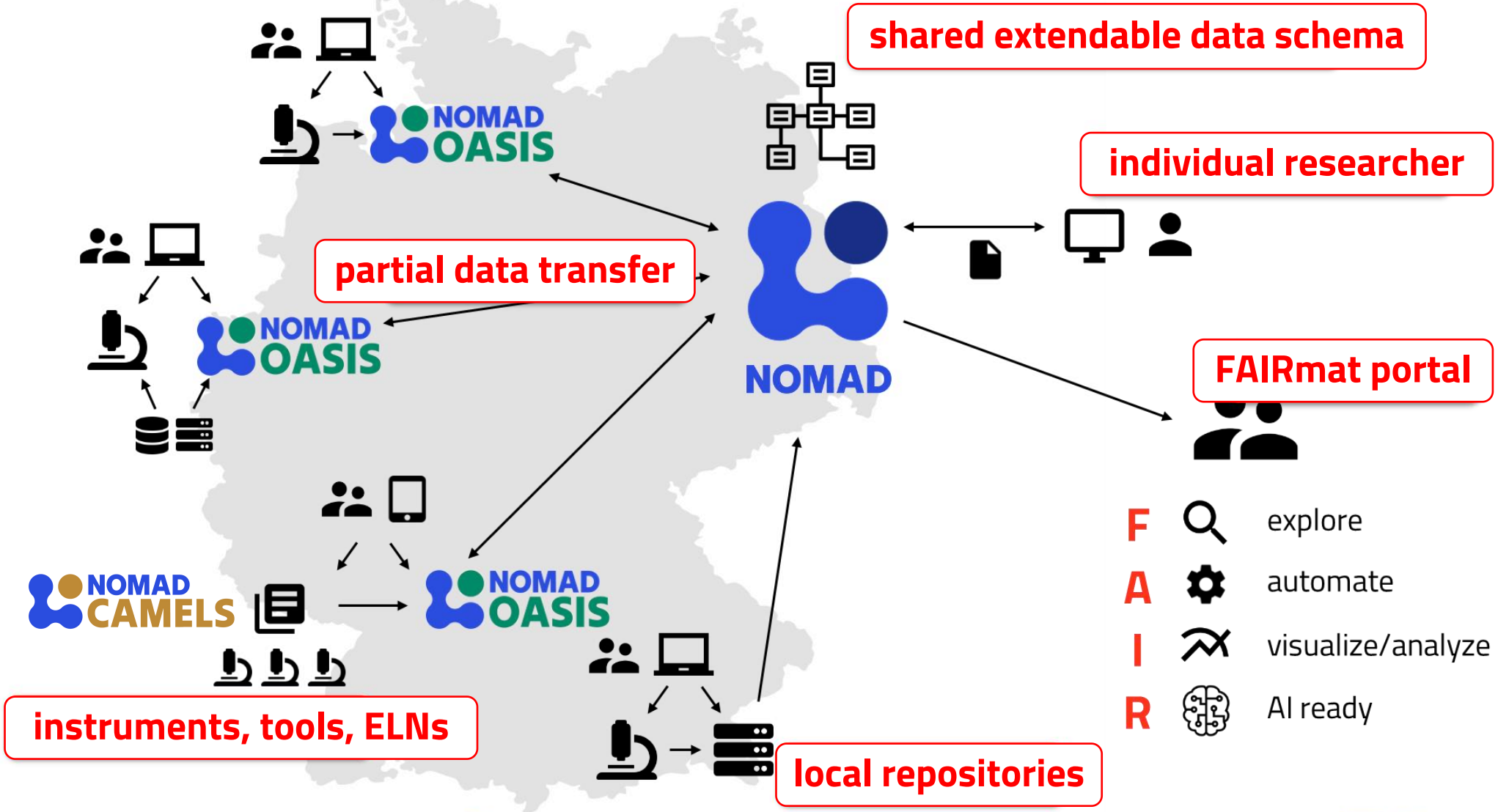
- Store all available data and metadata.
- Define measurement protocols using an intuitive graphical user interface.
- Connect to instruments via local interface or EPICS.

SUPPORTED INSTRUMENTS
> 200

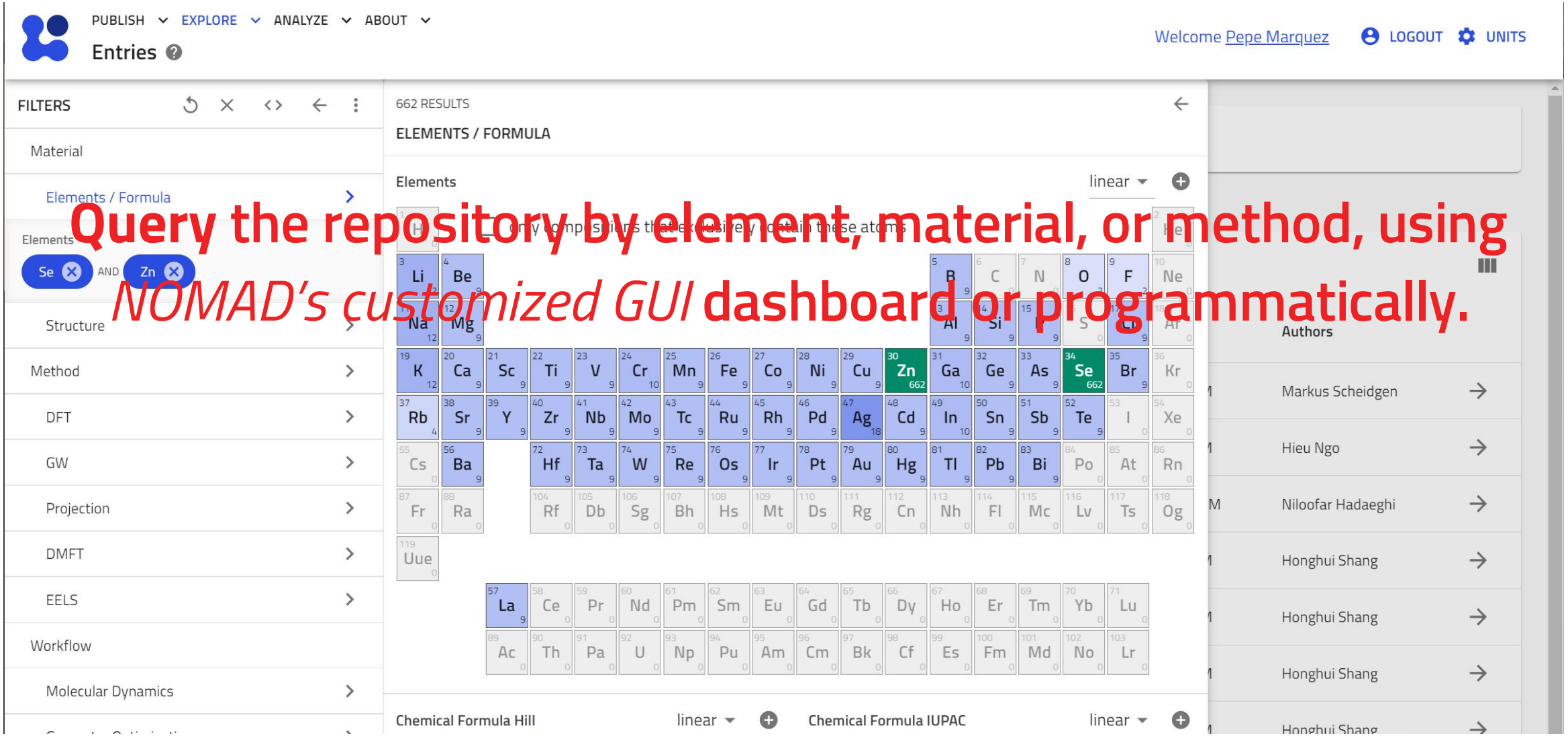


https://nomad-lab.eu/

FAIRmat's federated data infrastructure

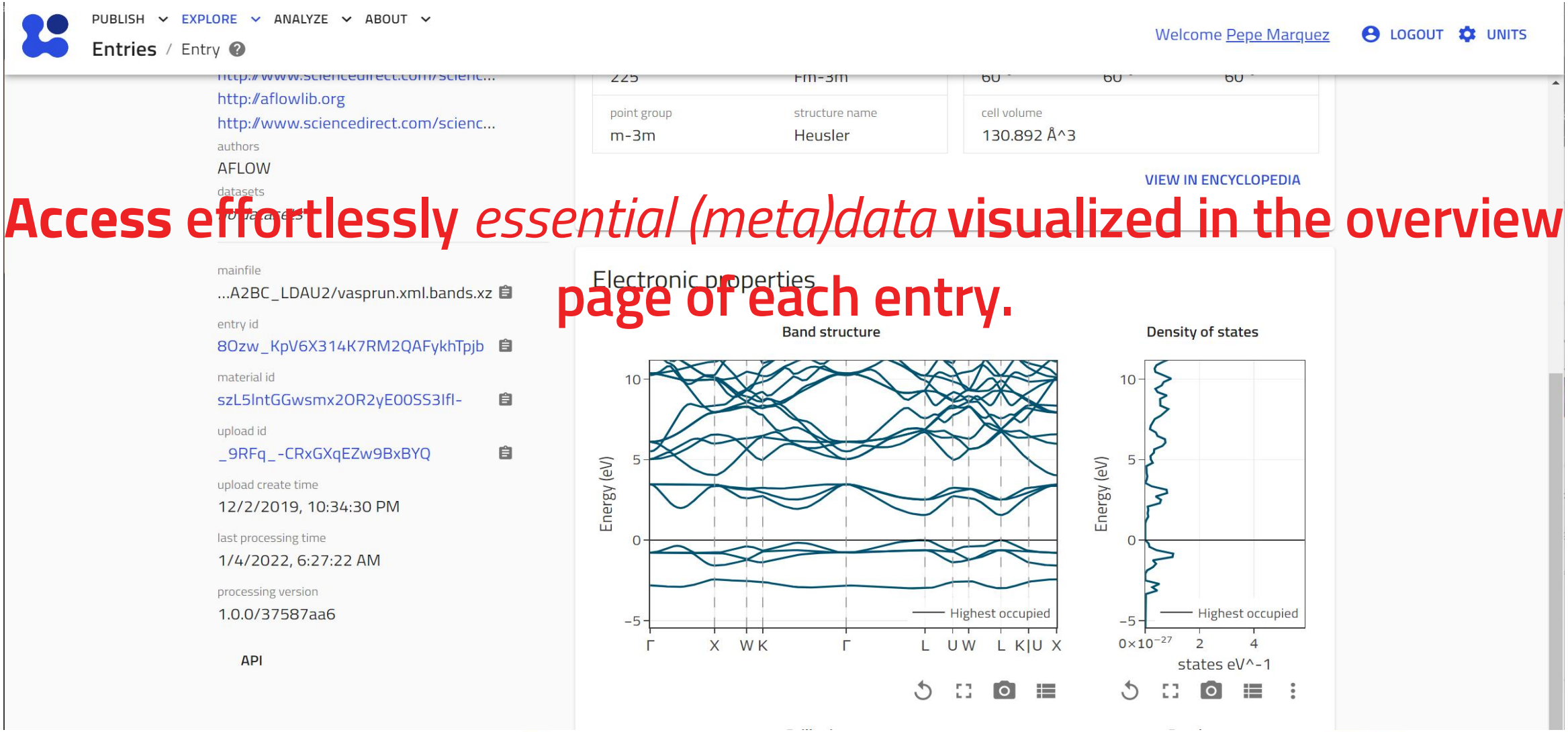


NOMAD core functionalities



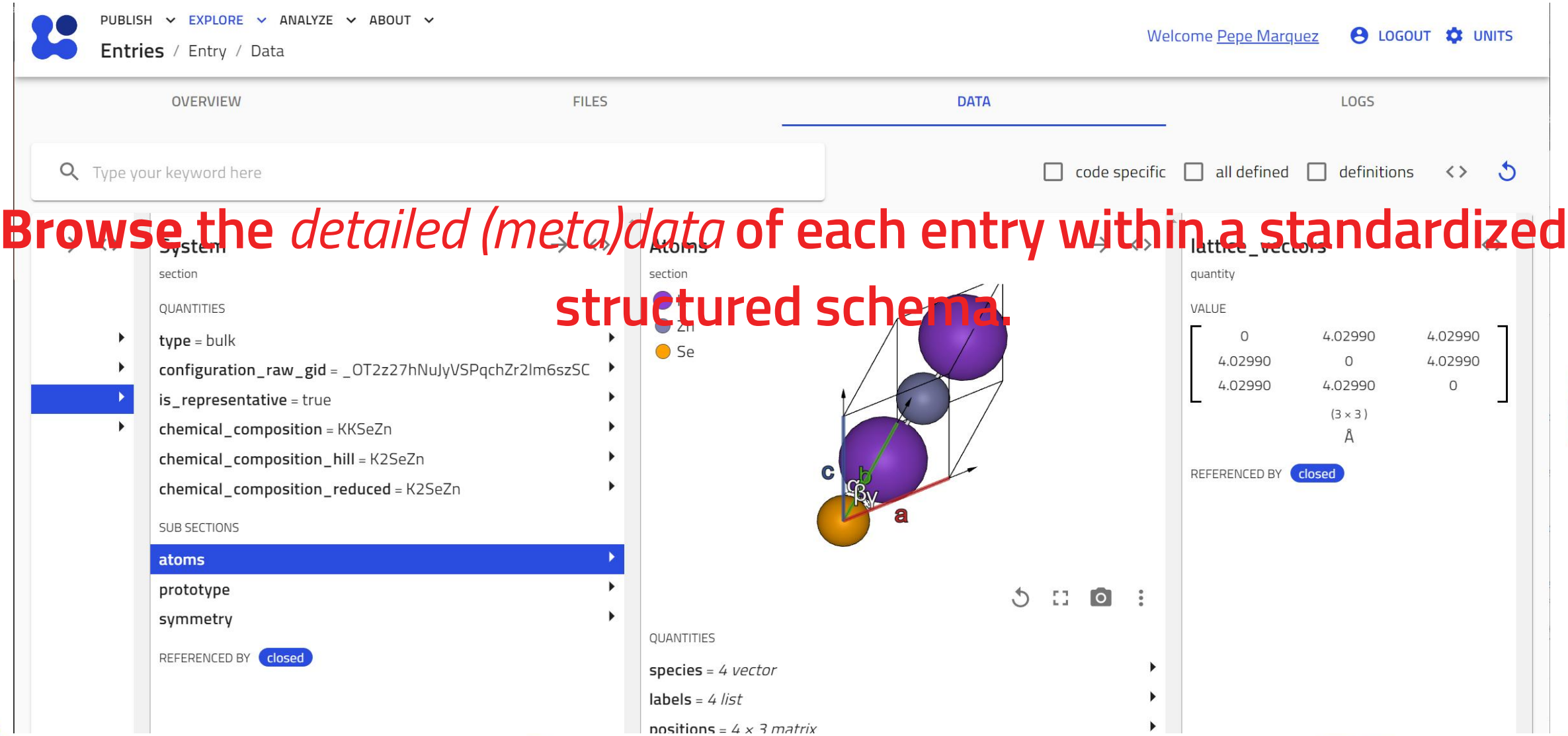
Query the repository by element, material, or method, using *NOMAD's customized GUI* dashboard or programmatically.

NOMAD core functionalities



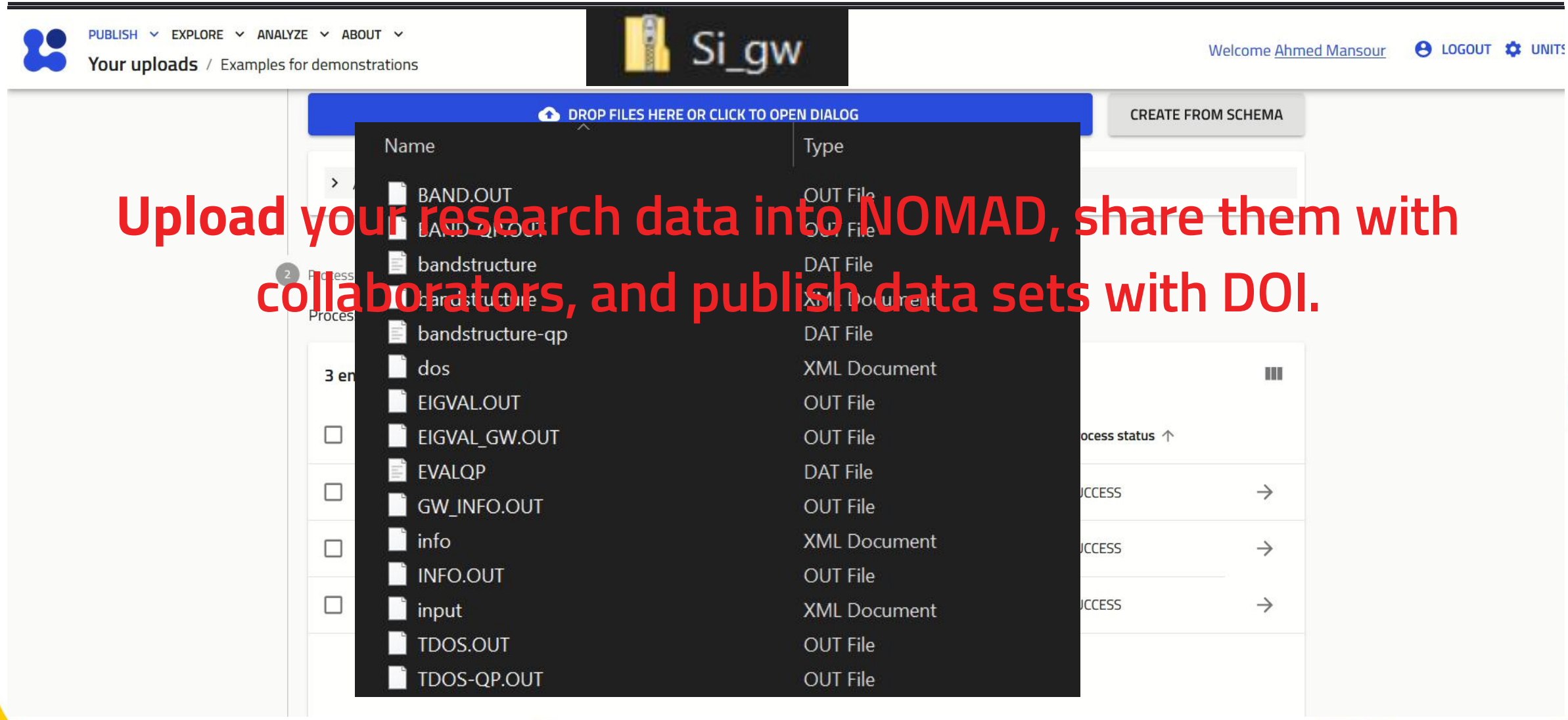
Access effortlessly *essential (meta)data* visualized in the overview page of each entry.

NOMAD core functionalities



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

NOMAD core functionalities

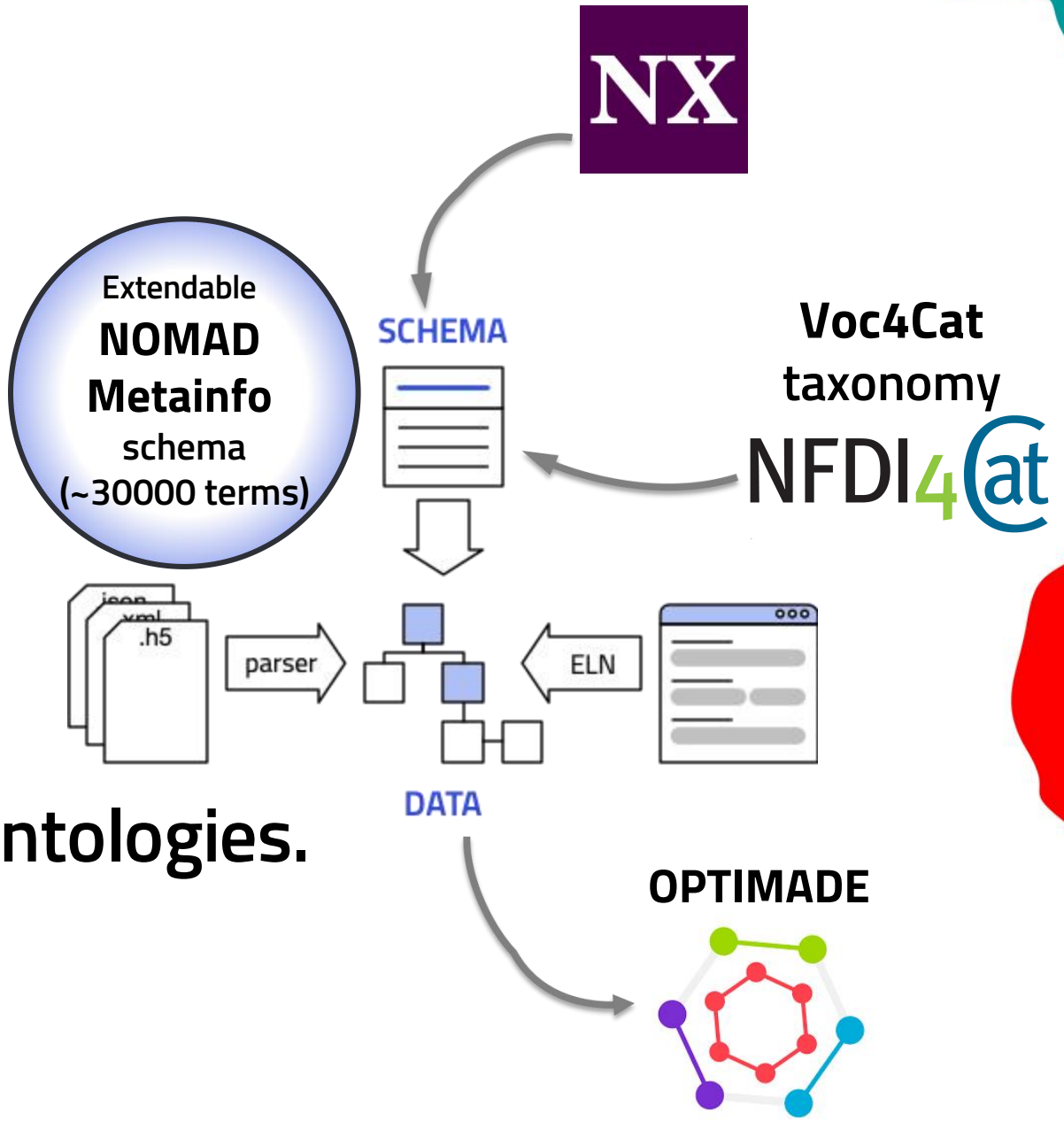


Upload your research data into NOMAD, share them with collaborators, and publish data sets with DOI.

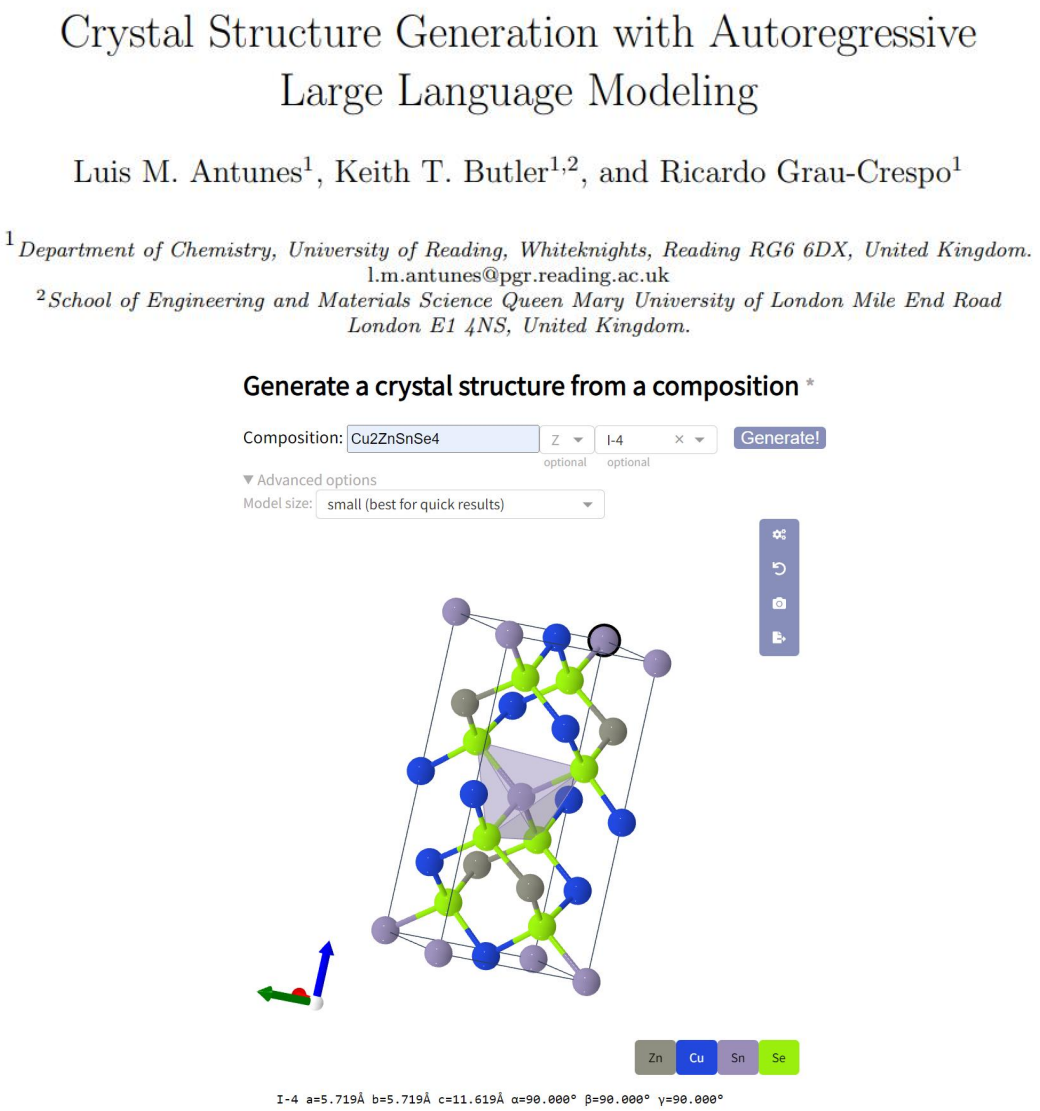
Structured data

- **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.**

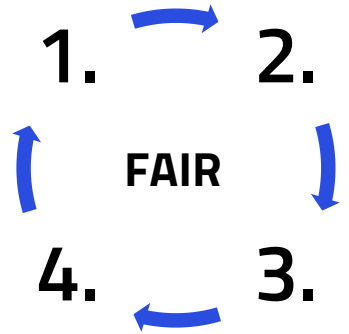


CrystalLLM: a FAIR full-cycle example



https://crystallm.com/

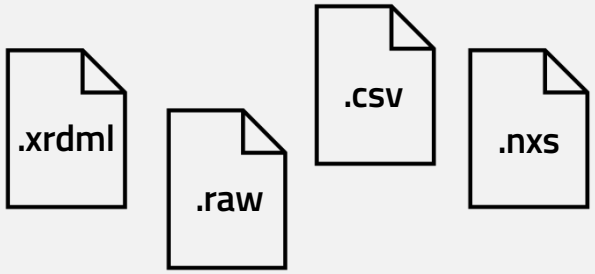
1. Users upload the INPUT and OUTPUT files to NOMAD from different codes from their publications
2. NOMAD makes the data FAIR
3. Antunes et al. build CrystalLLM with data from NOMAD
4. Users can use CrystalLLM to draft their inputs for their new calculations



NOMAD Oasis

Share selectively

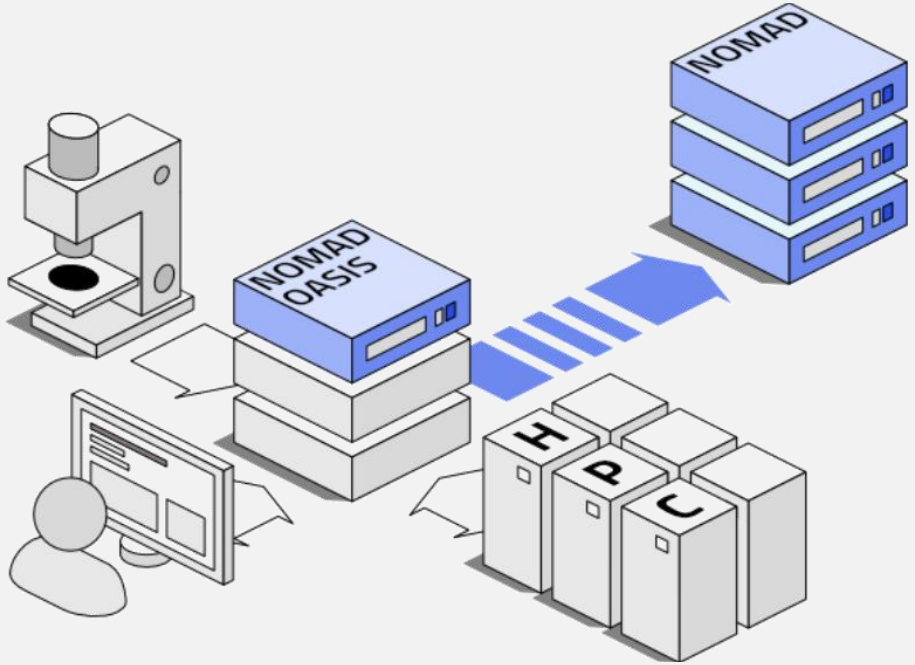
Automate data



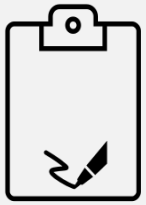
Search apps

database in a database

Manage your lab



ELN



API

NORTH

Cloud based analysis

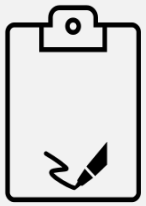
Large files

NOMAD Oasis

Share selectively

Manage your lab

ELN



API

Automate data

NORTH

Cloud based analysis

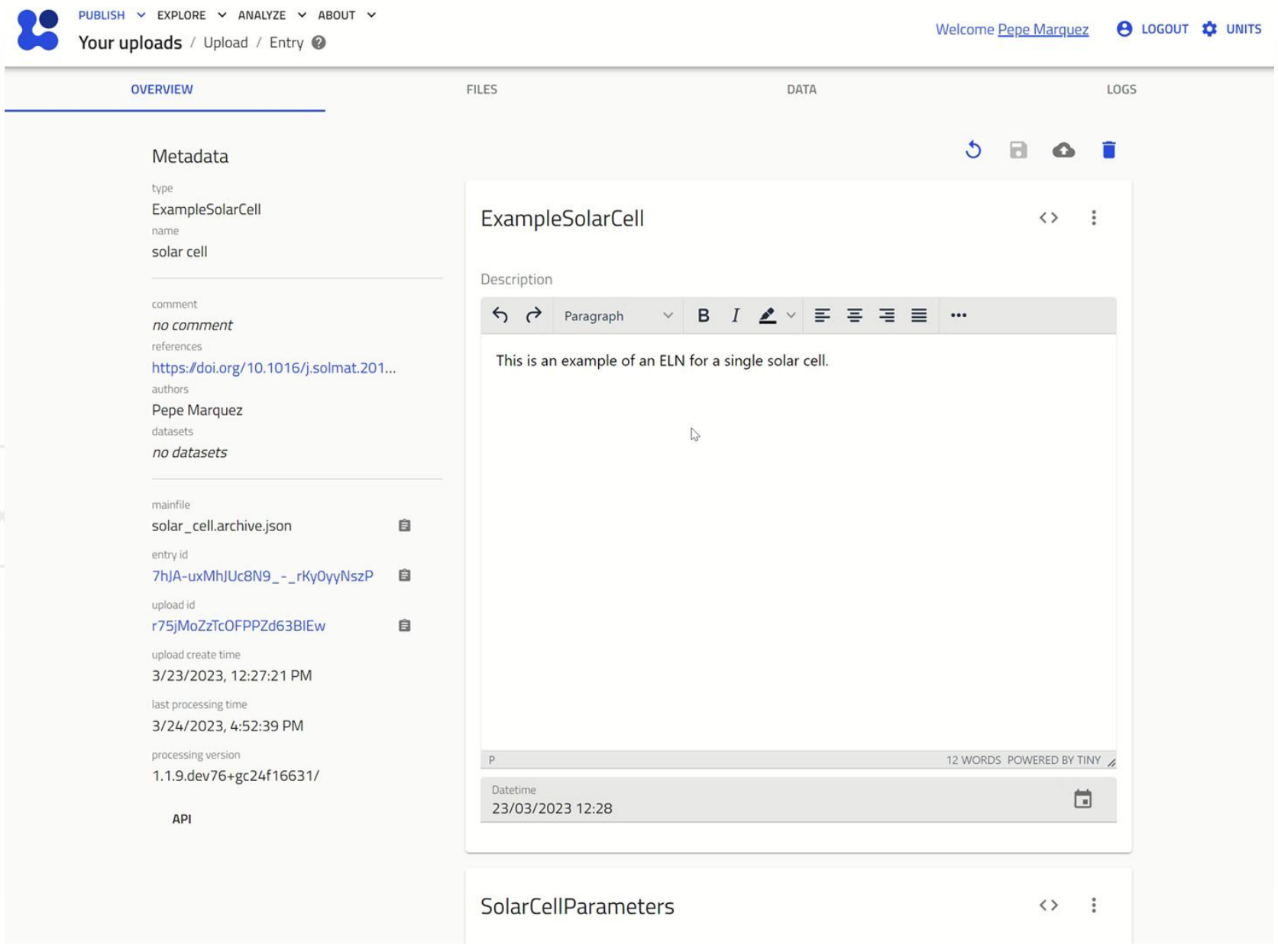
Search apps

database in a

database

Large files

NOMAD Oasis



ELN

API

NORTH

Cloud based analysis

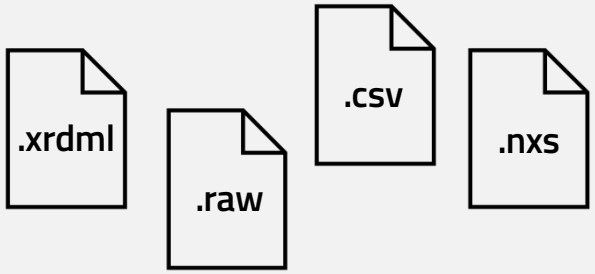
Large files

.hdf5

NOMAD Oasis

Share selectively

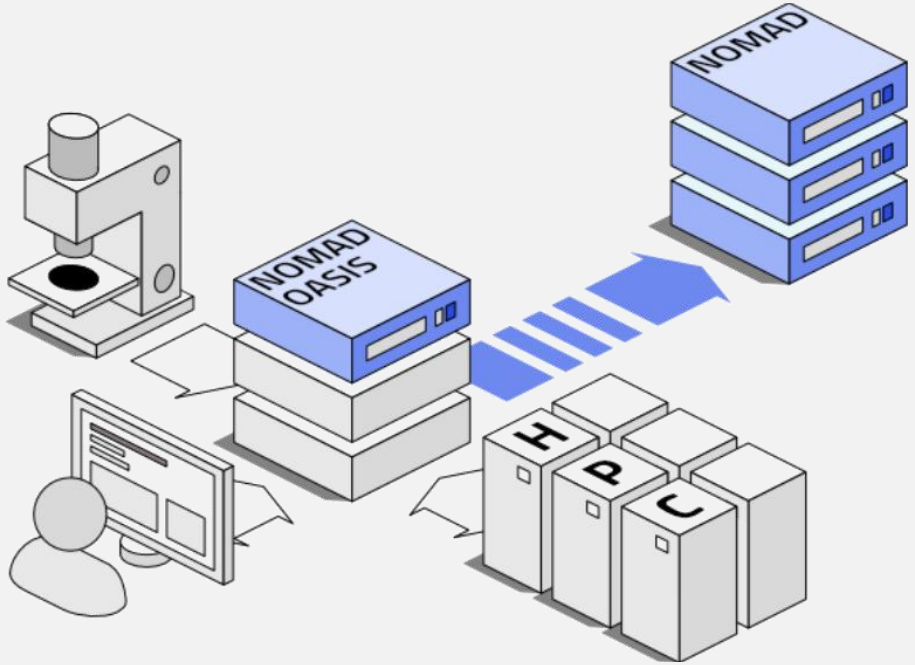
Automate data



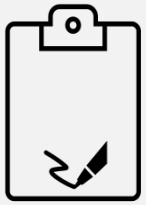
Search apps

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

NORTH

Cloud based analysis

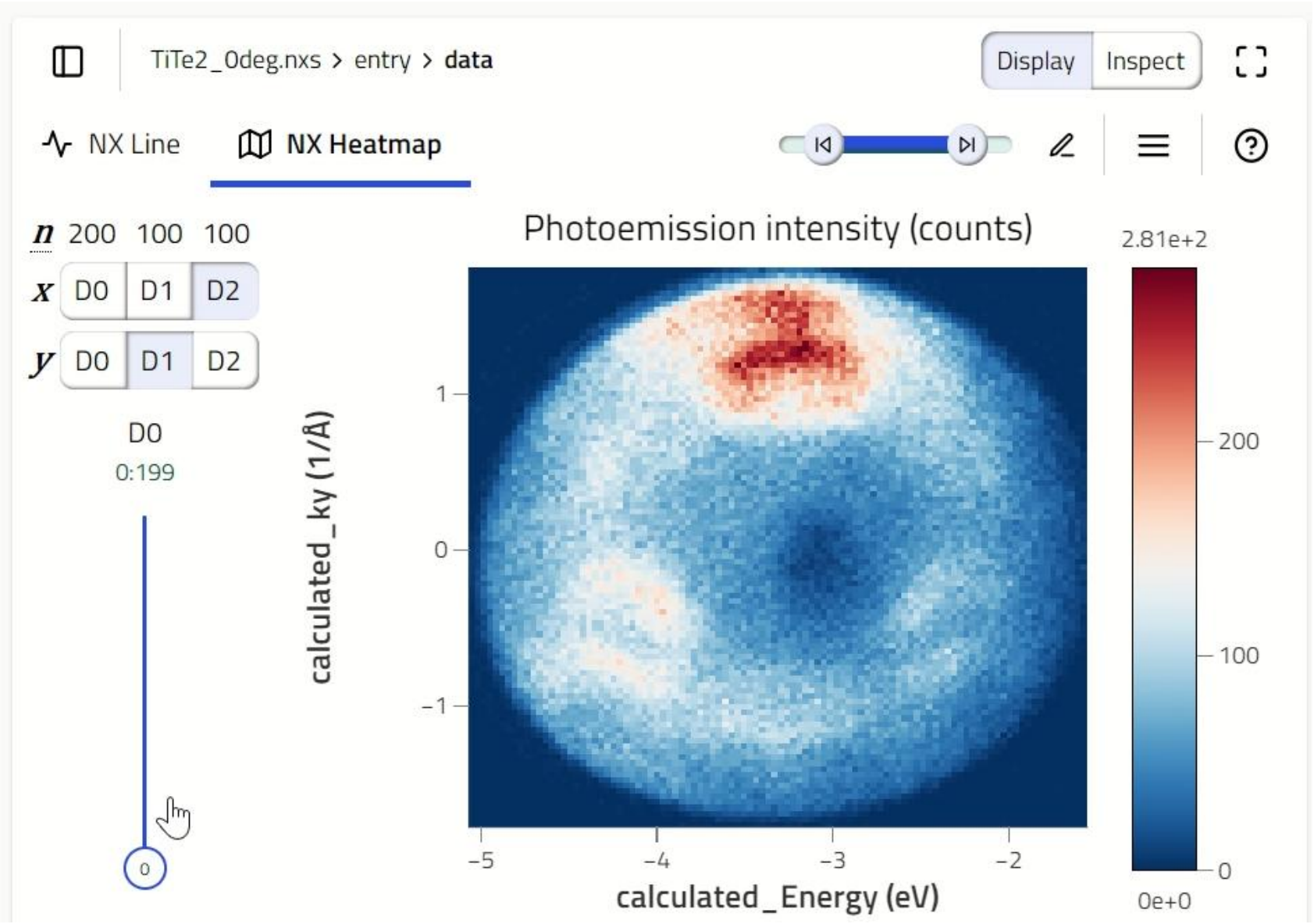
Search apps

database in a

database

Large files

NOMAD Oasis



ELN

API

NORTH

Cloud based analysis

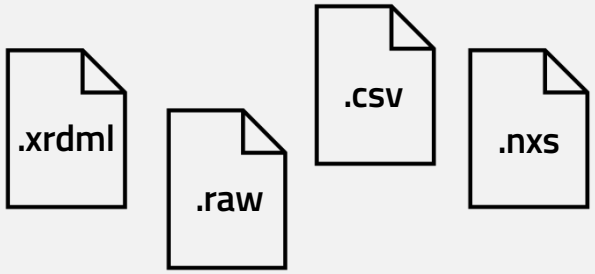
Large files

.hdf5

NOMAD Oasis

Share selectively

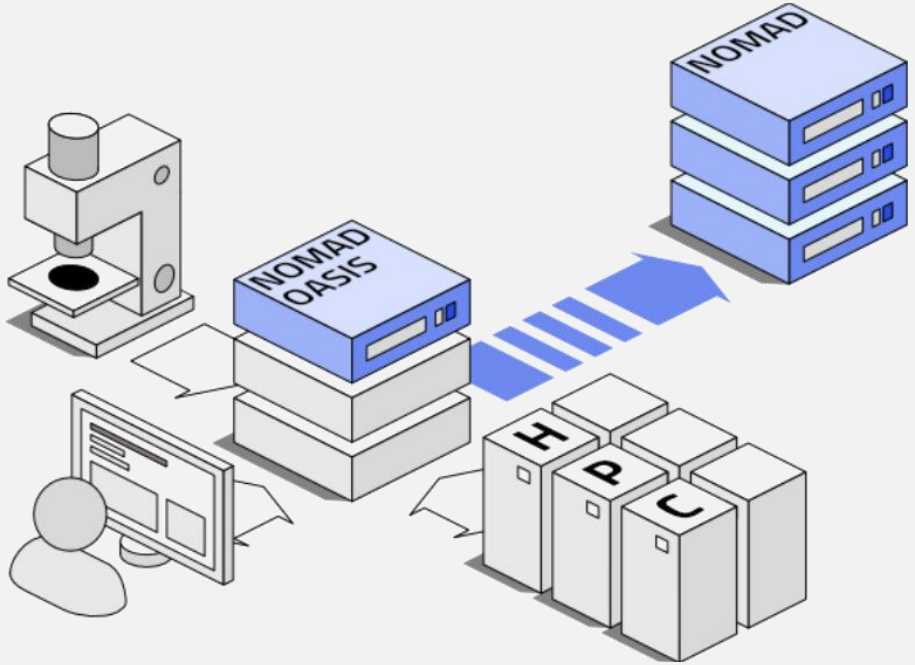
Automate data



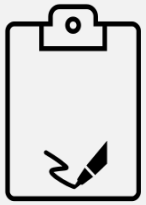
Search apps

database in a database

Manage your lab



ELN



API

NORTH

Cloud based analysis

Large files

NOMAD Oasis

Share selectively

Manage your lab

ELN

API

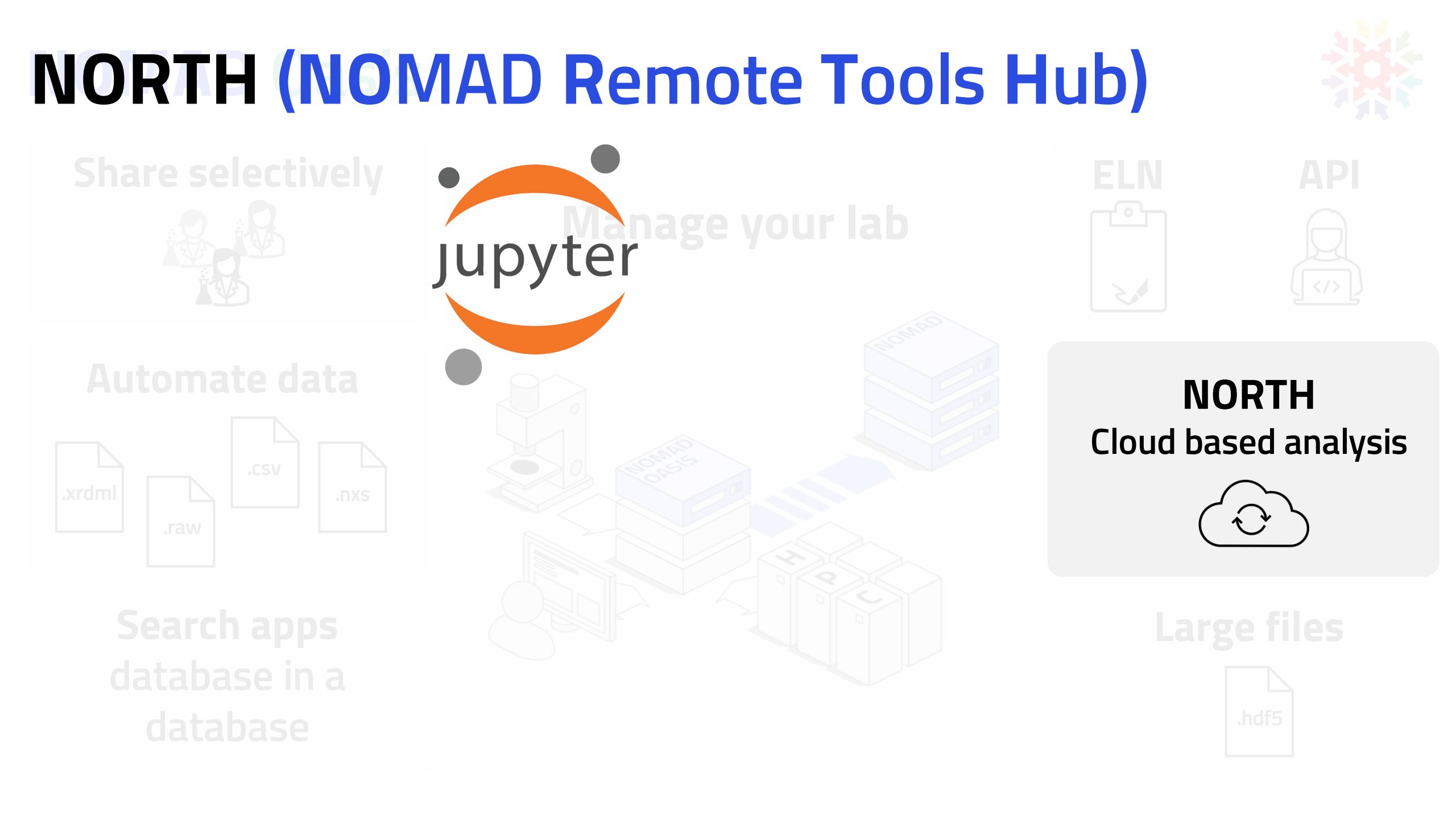
Automate data

NORTH
Cloud based analysis

Search apps
database in a
database

Large files

NORTH (NOMAD Remote Tools Hub)



jupyter

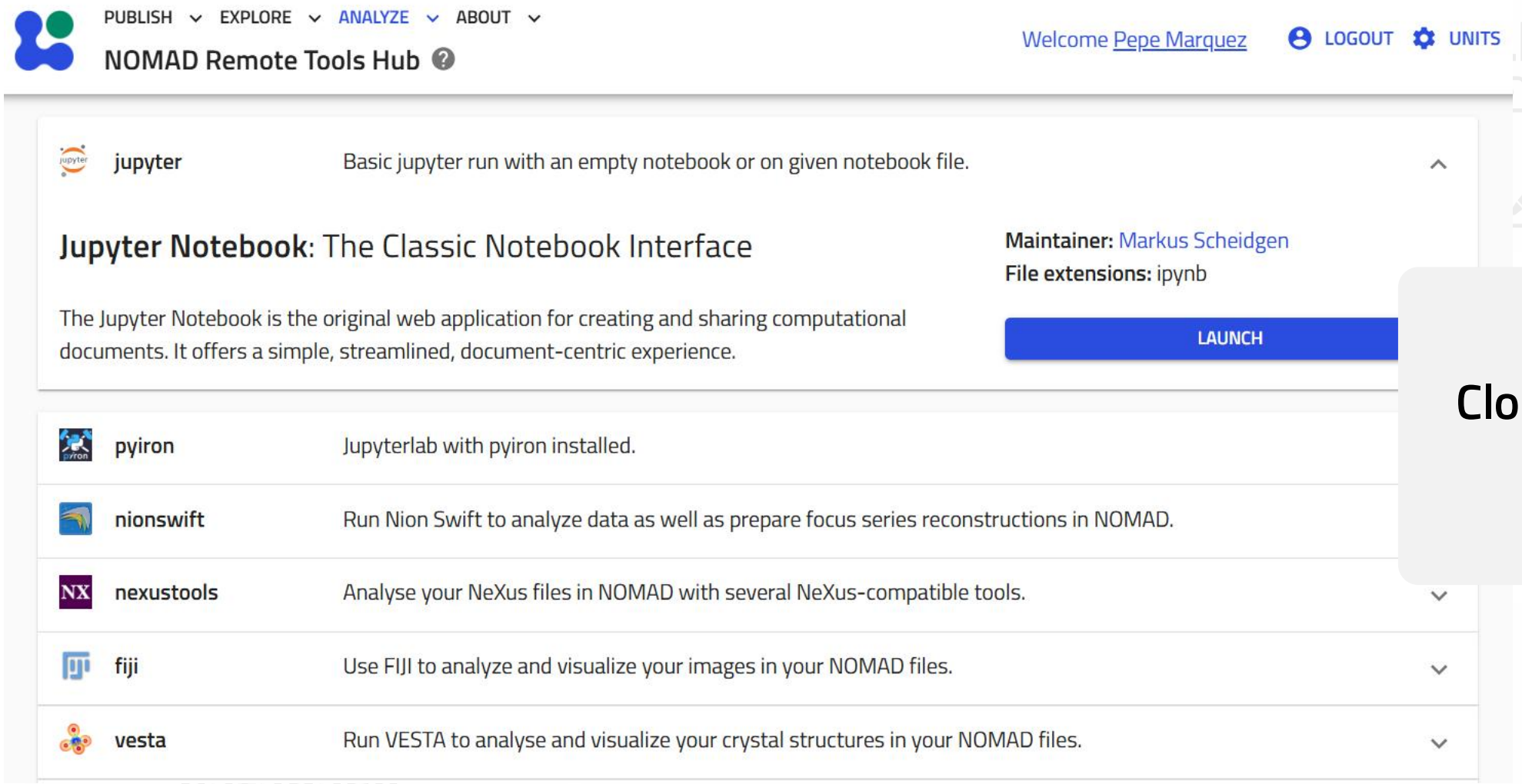
NORTH
Cloud based analysis

NORTH (NOMAD Remote Tools Hub)

NORTH

Cloud based analysis

NORTH (NOMAD Remote Tools Hub)



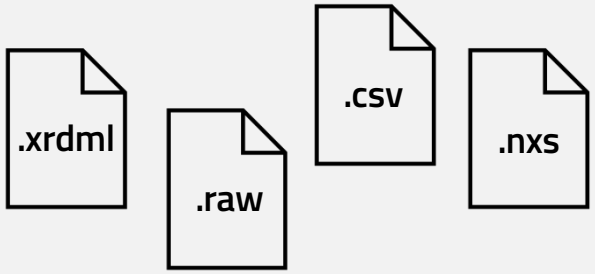
NORTH

Cloud based analysis

NOMAD Oasis

Share selectively

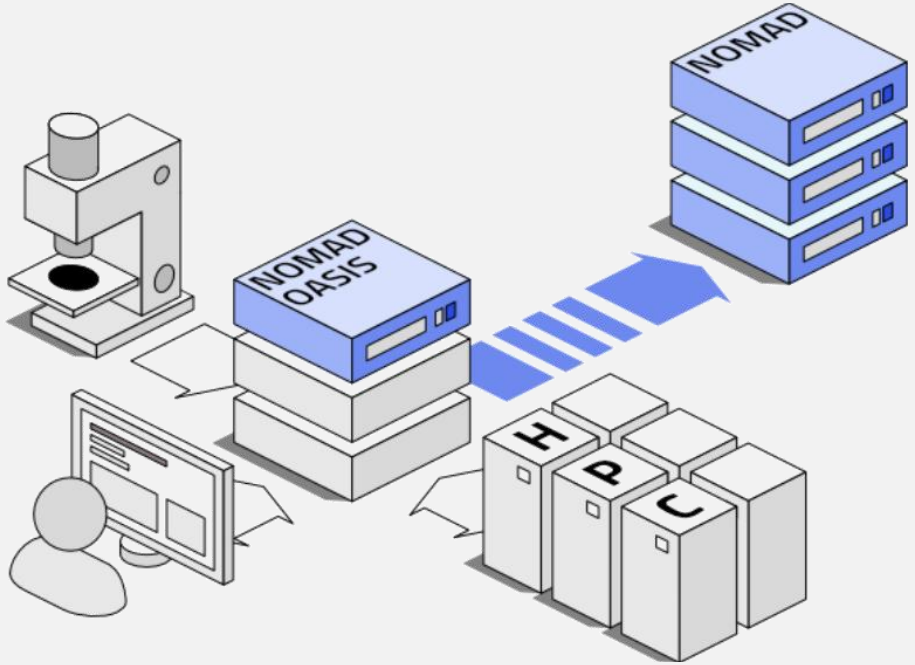
Automate data



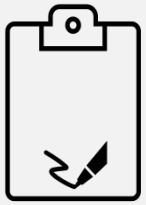
Search apps

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

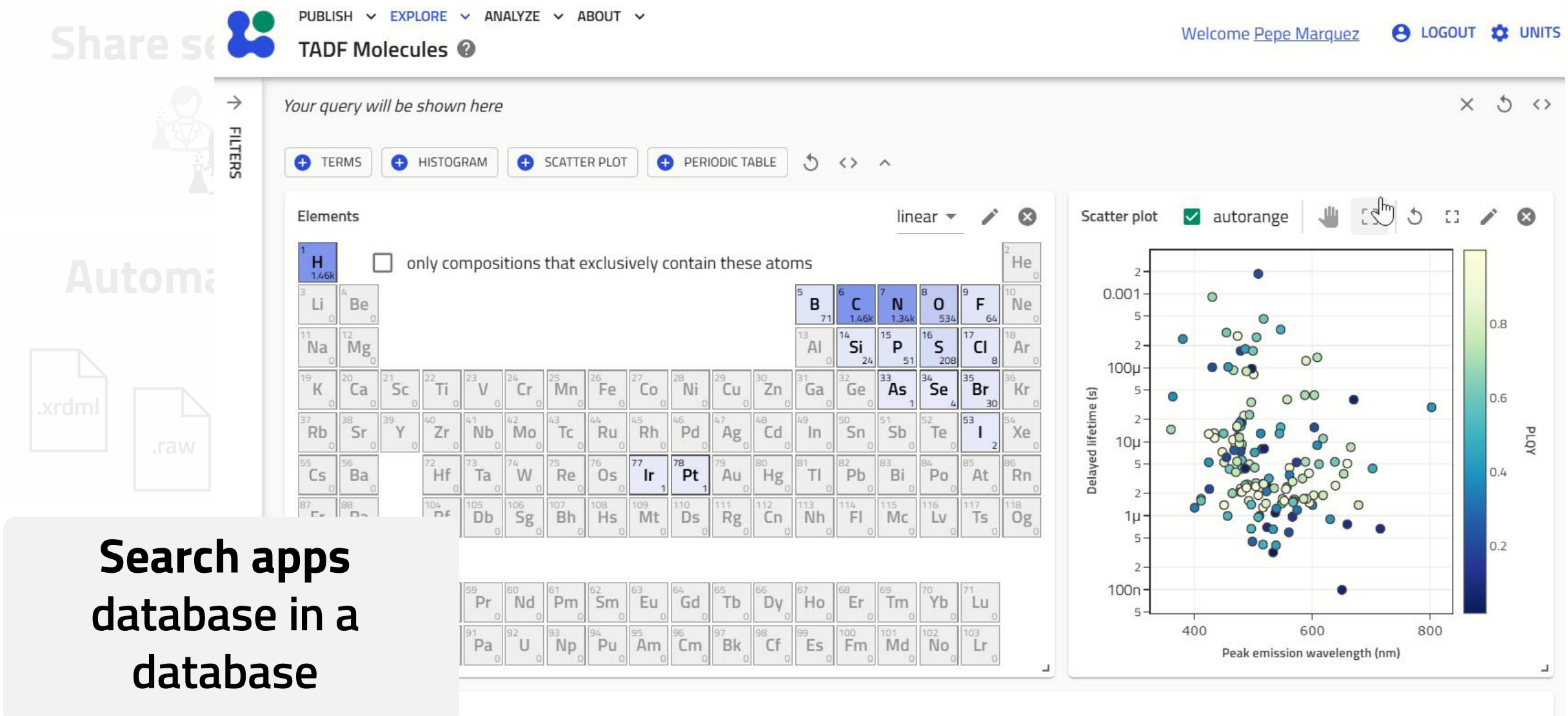
NORTH

Cloud based analysis

Search apps
database in a
database

Large files

NOMAD Oasis

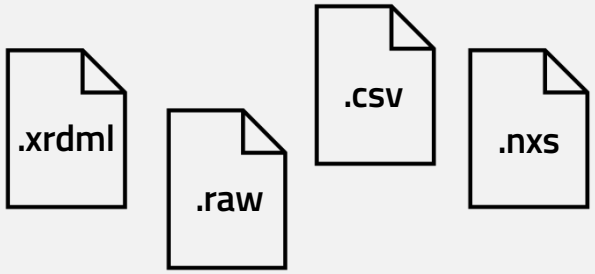


Search apps
database in a
database

NOMAD Oasis for Experiments

Share selectively

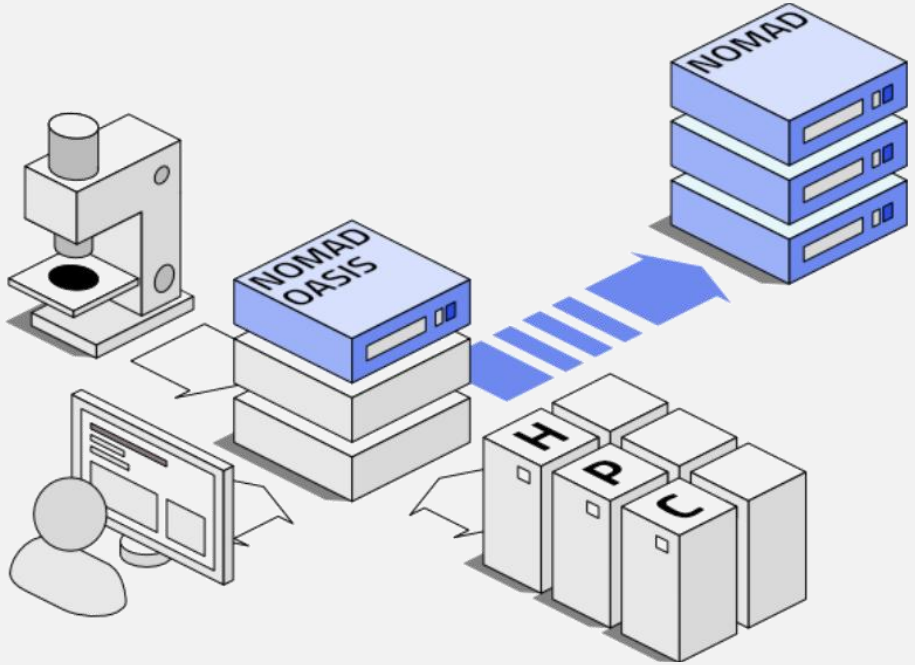
Automate data



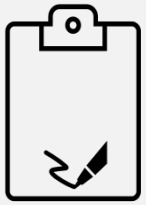
Search apps

database in a database

Manage your lab



ELN



API

NORTH

Cloud based analysis

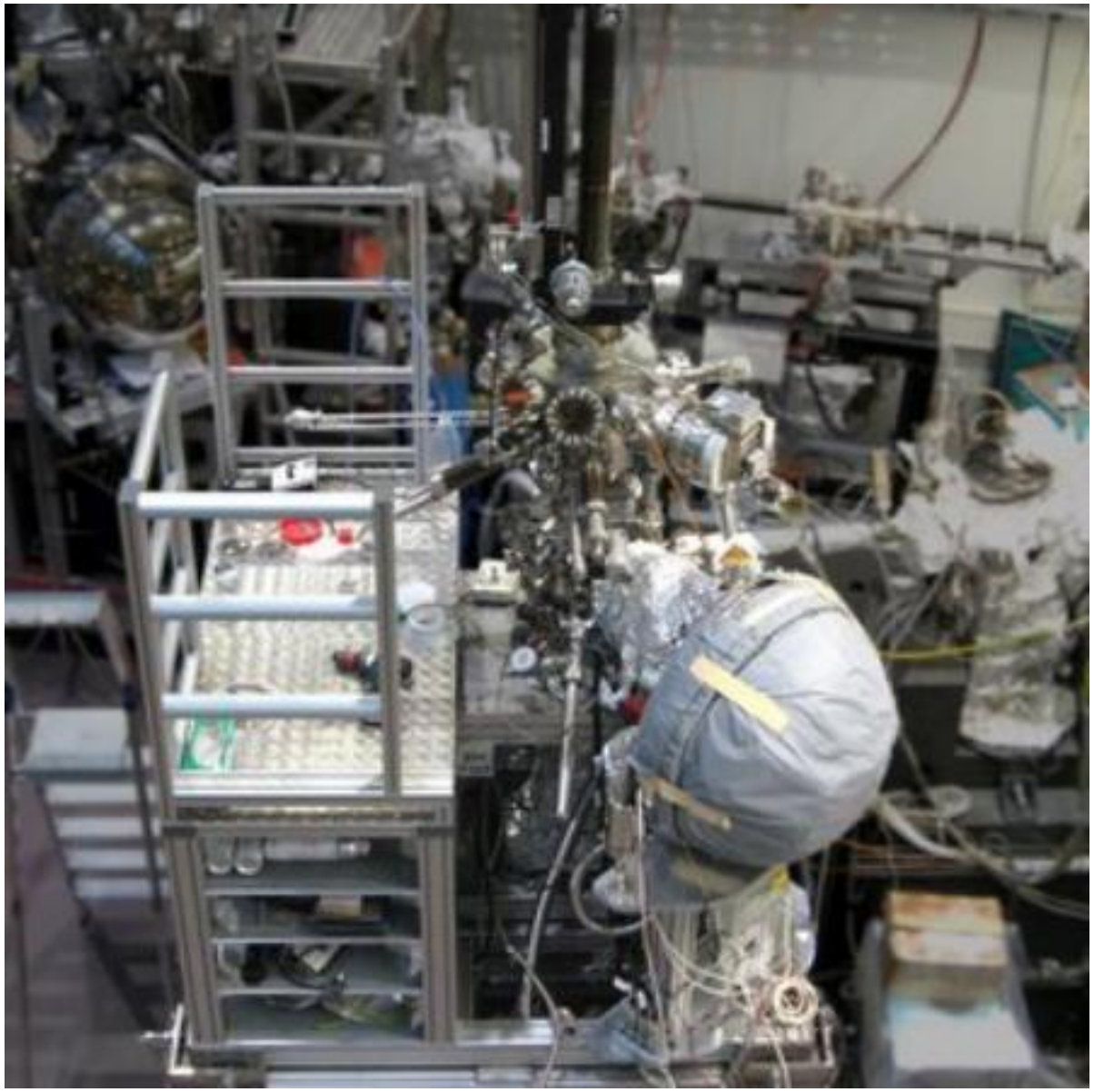
Large files

Experiments

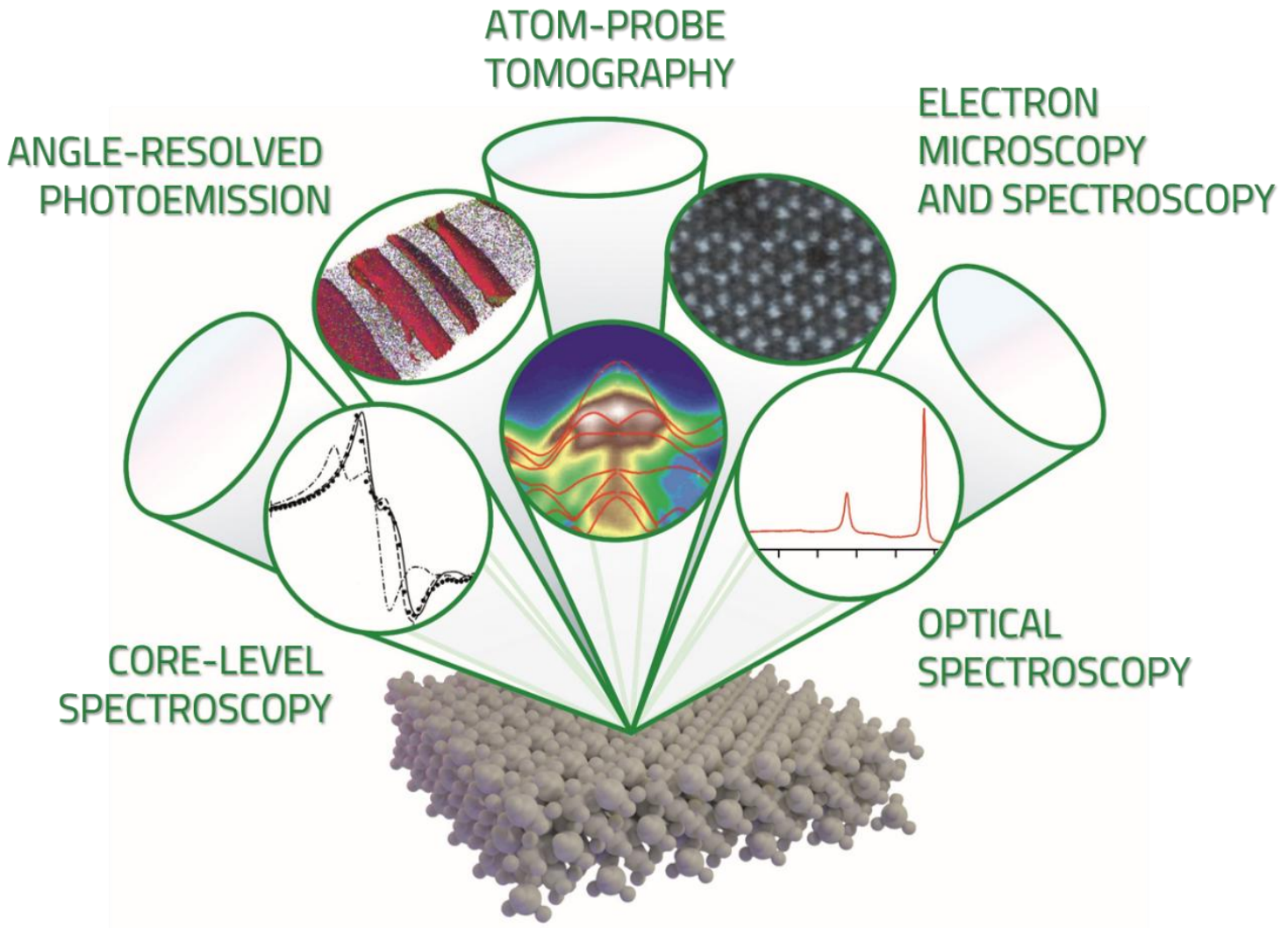
Ever-new methodology

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

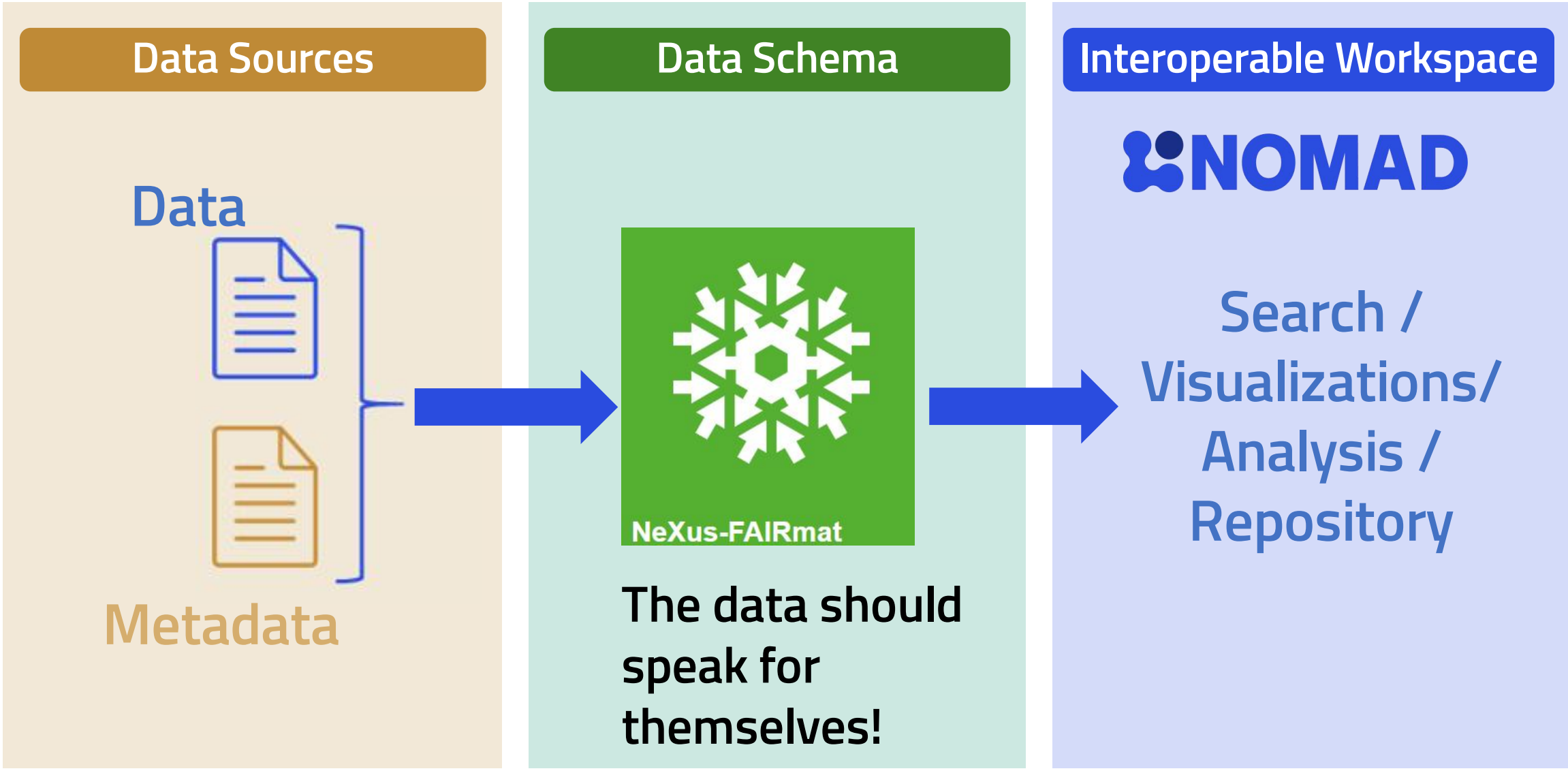
Broad and structured data are barely available.



FAIRmat's pilot experiments



Experimental data workflow



Harmonization: the NeXus proposal

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

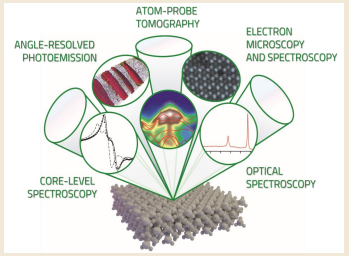
FAIRmat contributes most actively to new standards and plays an important role in the NeXus committee NIAC.

Harmonized structure of experimental data

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

Experimental data workflow

Data Sources



Data Schema

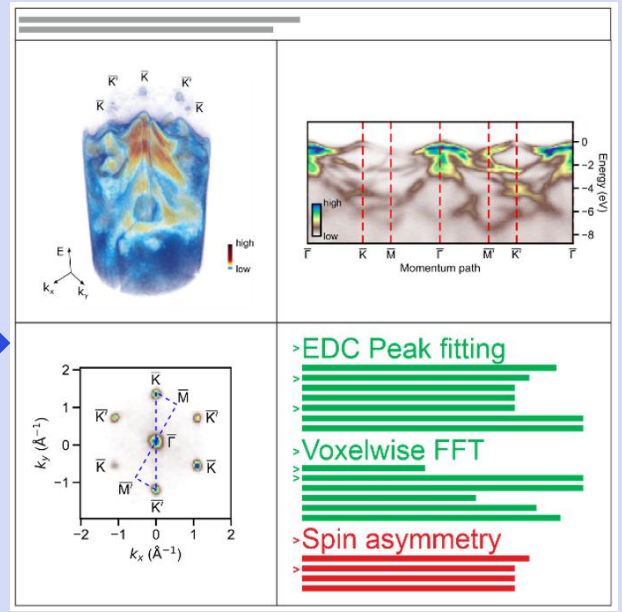
Community agreement!

Pynxtools

NeXus-FAIRmat

Interoperable Workspace

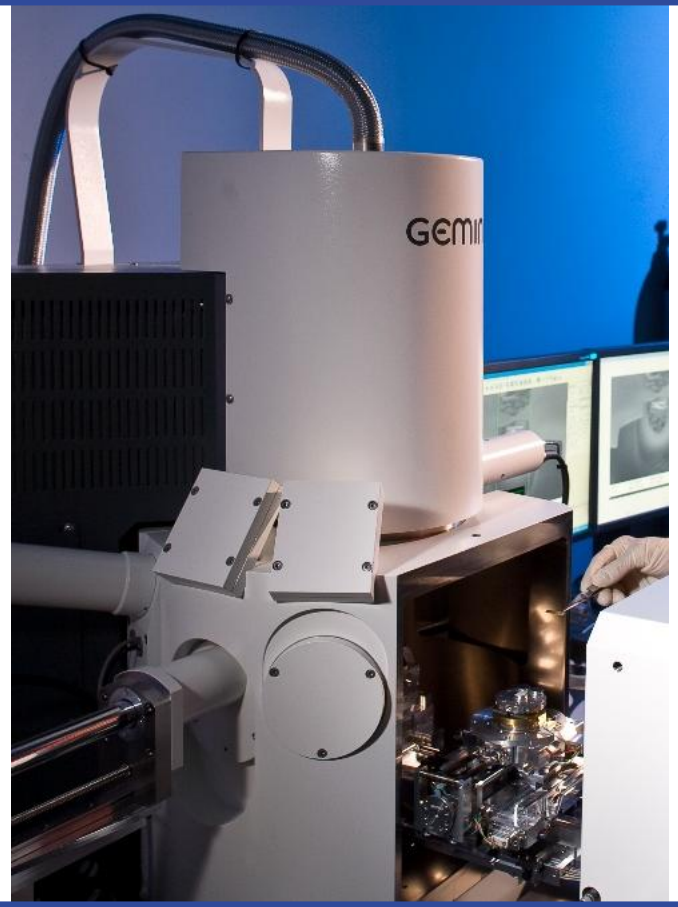
NOMAD



Findable Accessible Interoperable Reusable

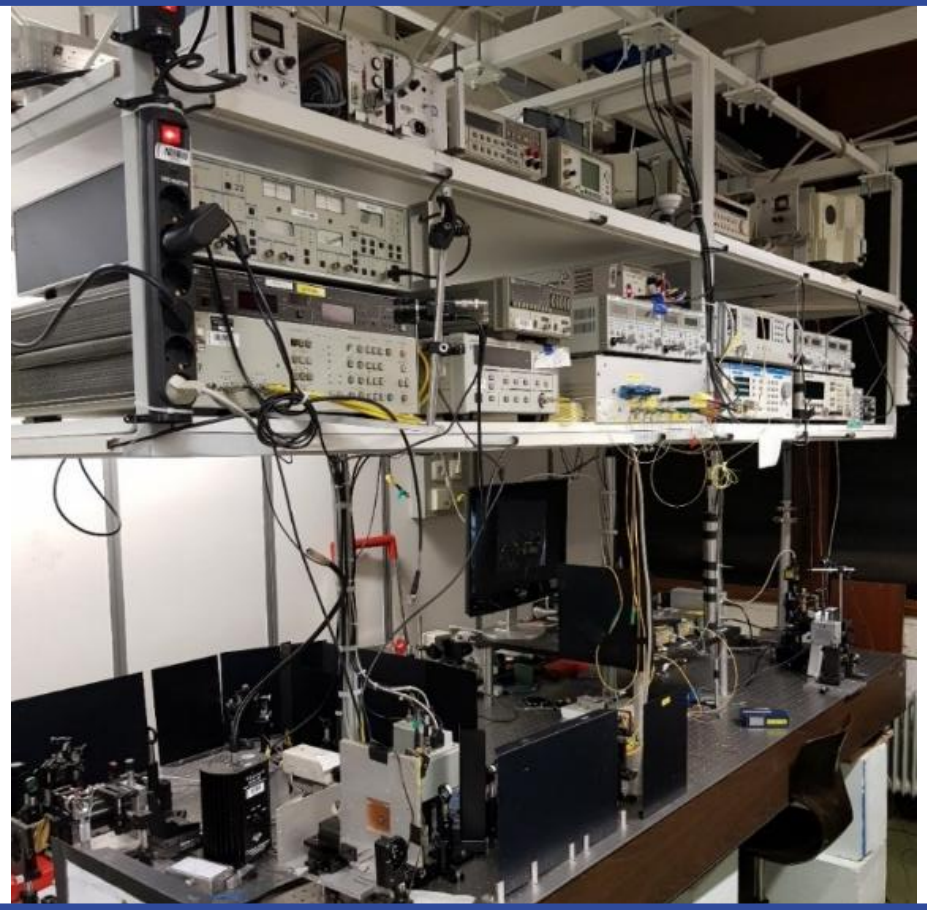
Experimental data sources

Commercial integrated systems



Vendor-provided software

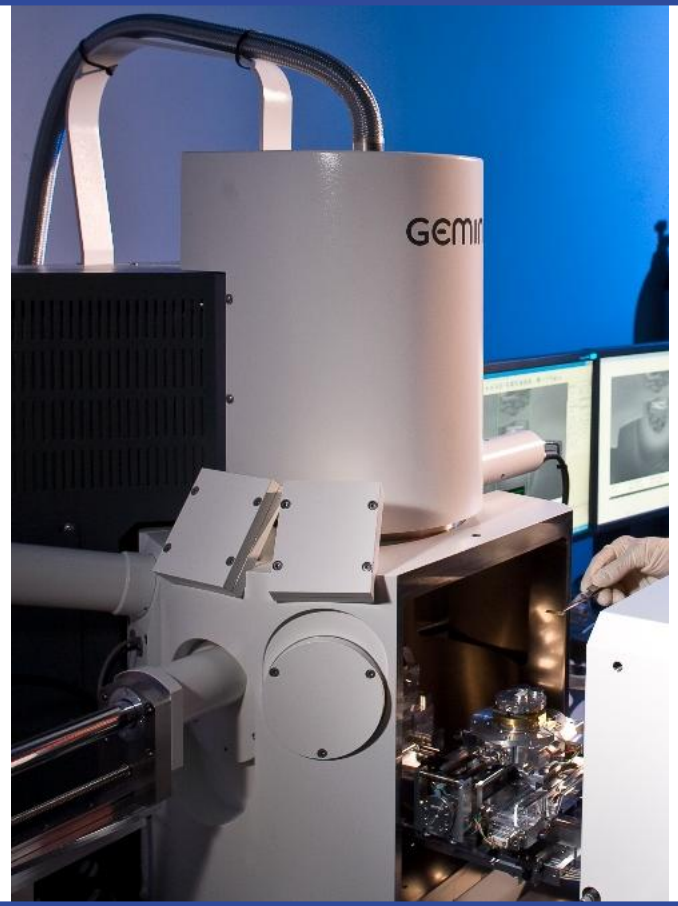
Specialized custom systems



Self-written software

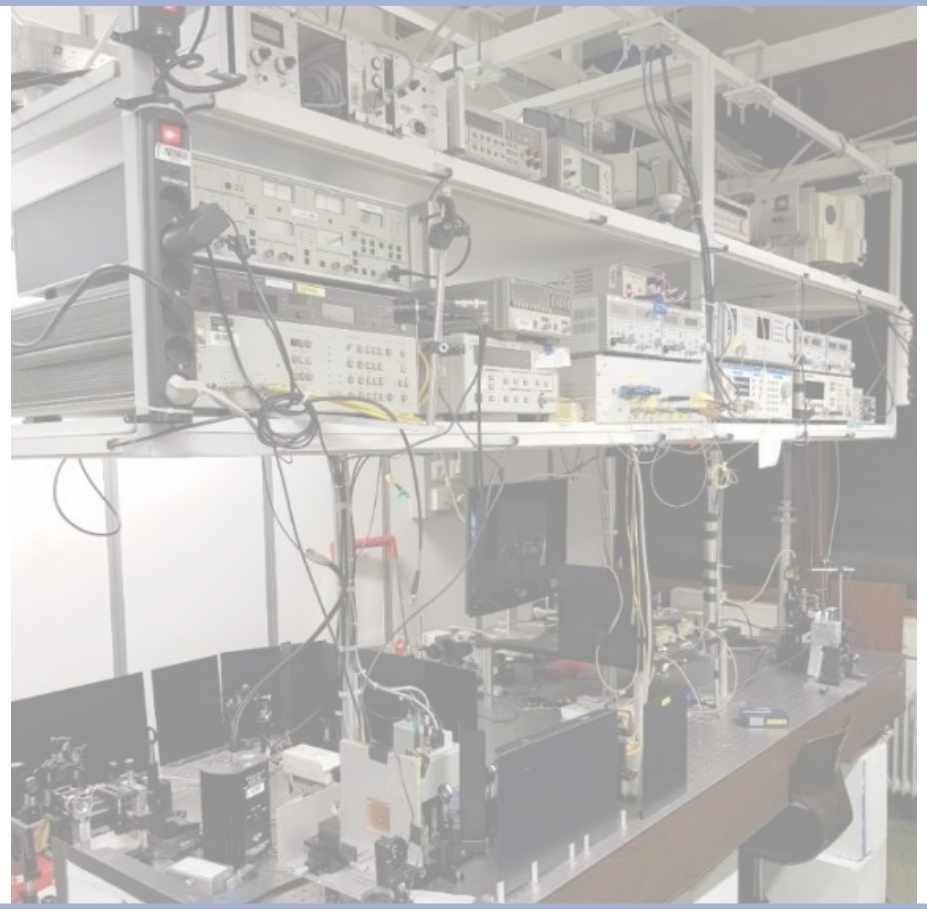
Experimental data sources

Commercial integrated systems



Vendor-provided software

Specialized custom systems

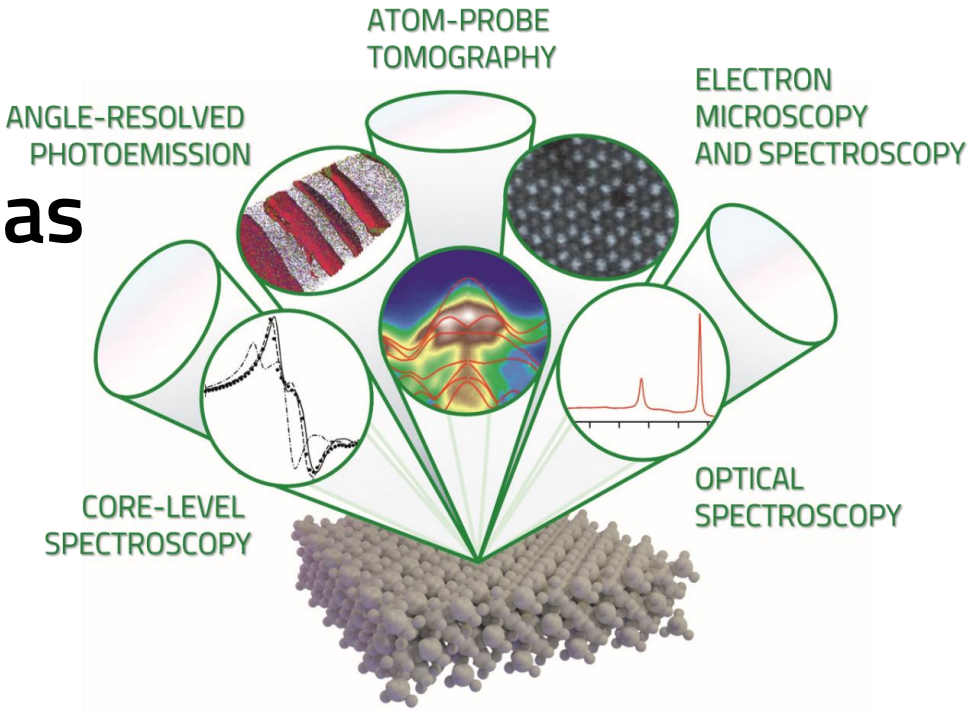


Self-written software

Organizing community standards

FAIRmat approach:

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



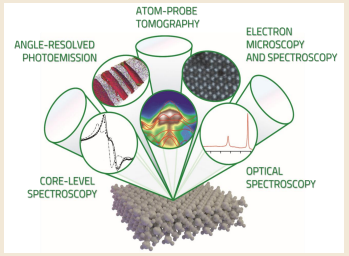
Manufacturers (Technology Partners) are teaming up for providing NeXus output

SPECS™ FOCUS J.A. Woollam Ellipsometry Solutions

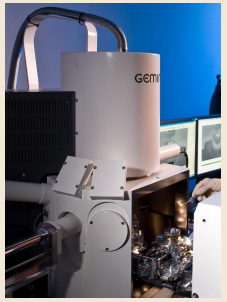
ThermoFisher SCIENTIFIC SURFACE CONCEPT scientaomicron SENTECH

Experimental data workflow

Data Sources



Technology partners



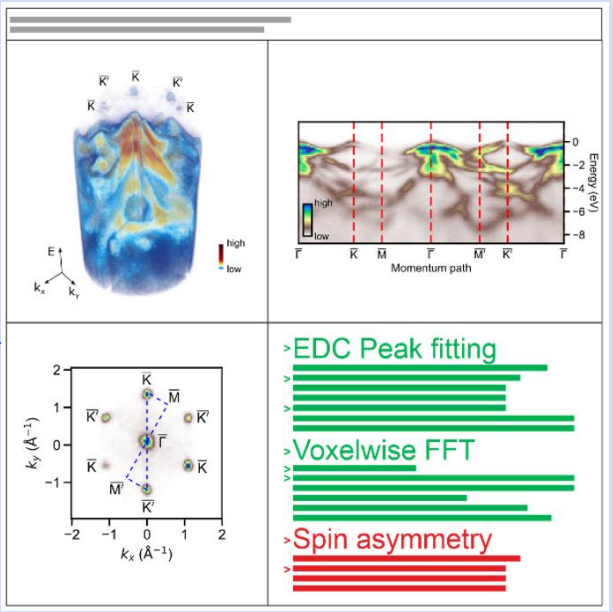
Data Schema

Community agreement!

NeXus-FAIRmat

Interoperable Workspace

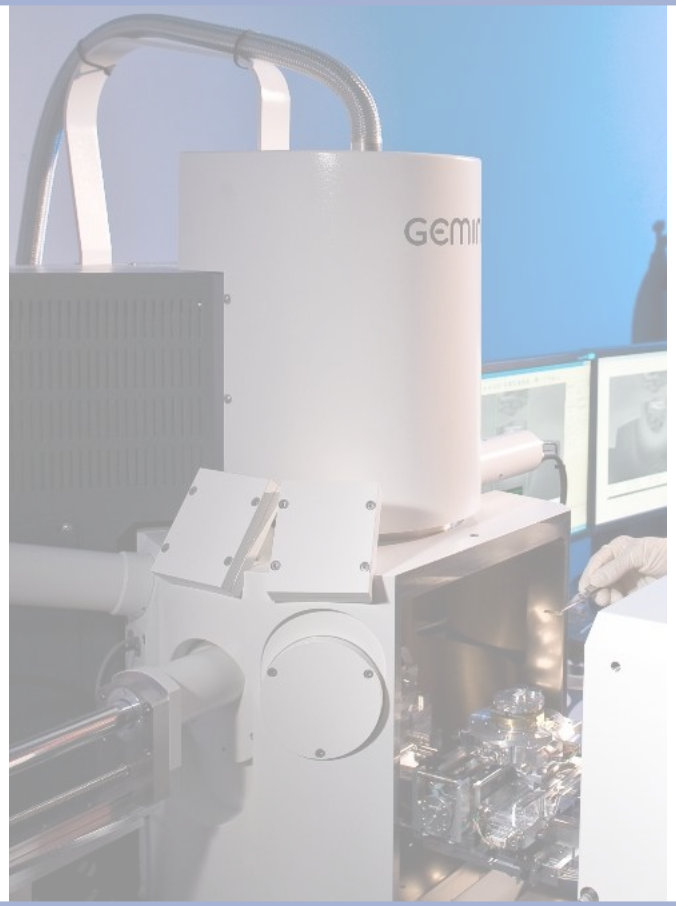
NOMAD



Findable Accessible Interoperable Reusable

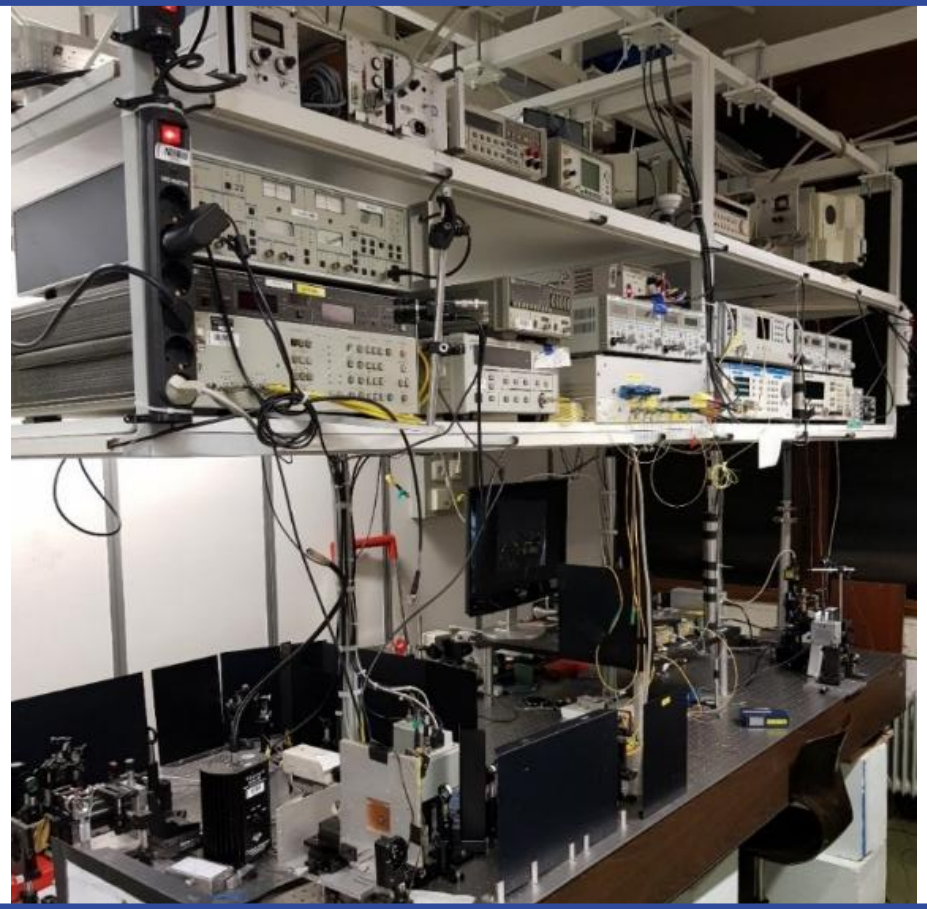
Experimental data sources

Commercial integrated systems



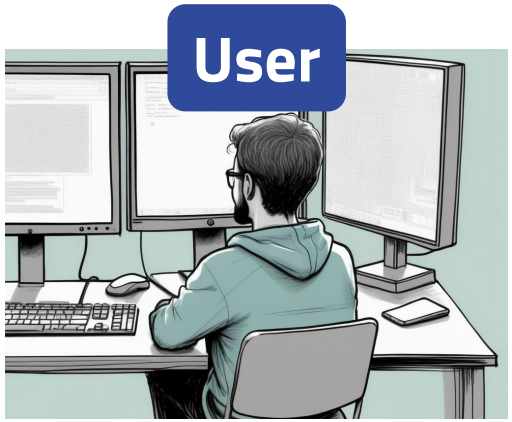
Vendor-provided software

Specialized custom systems



Self-written software

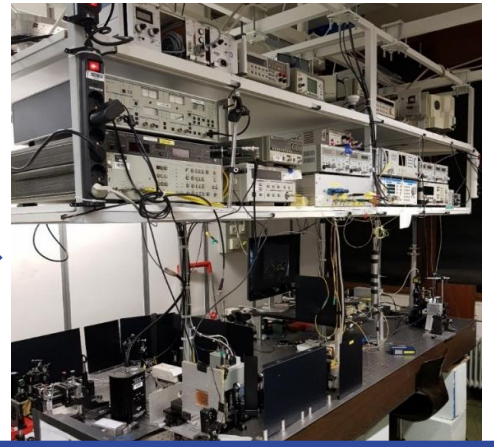
Workflow



CAMELS

1 Configure

2 Measure



Experimental setup

3 Save

NeXus Data

4 Upload

NOMAD

Repository / ELN

optional

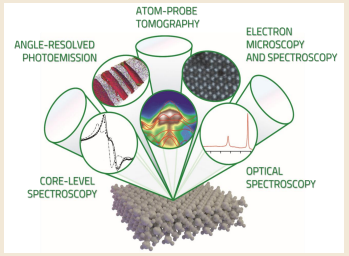
eLabFTW

5 Connect

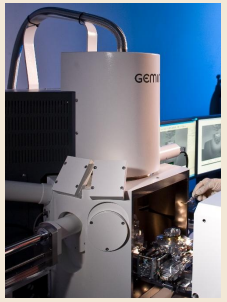
Connect to ELN entries

Experimental data workflow

Data Sources



Technology partners



Any Experiment

NOMAD CAMELS

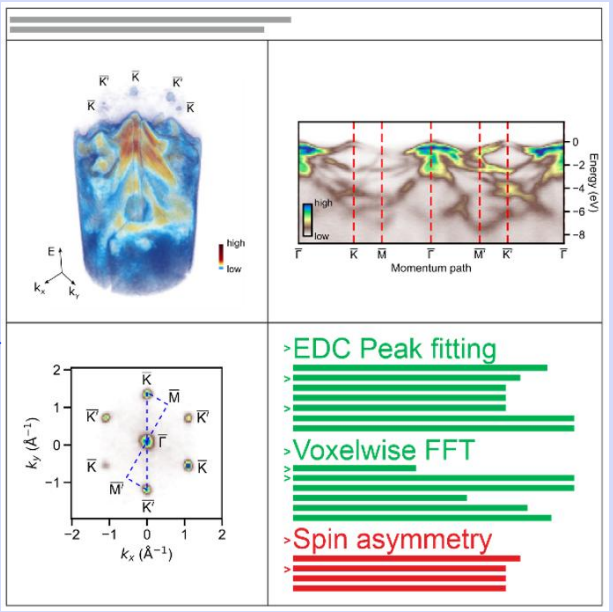
Data Schema

Community agreement!

NeXus-FAIRmat

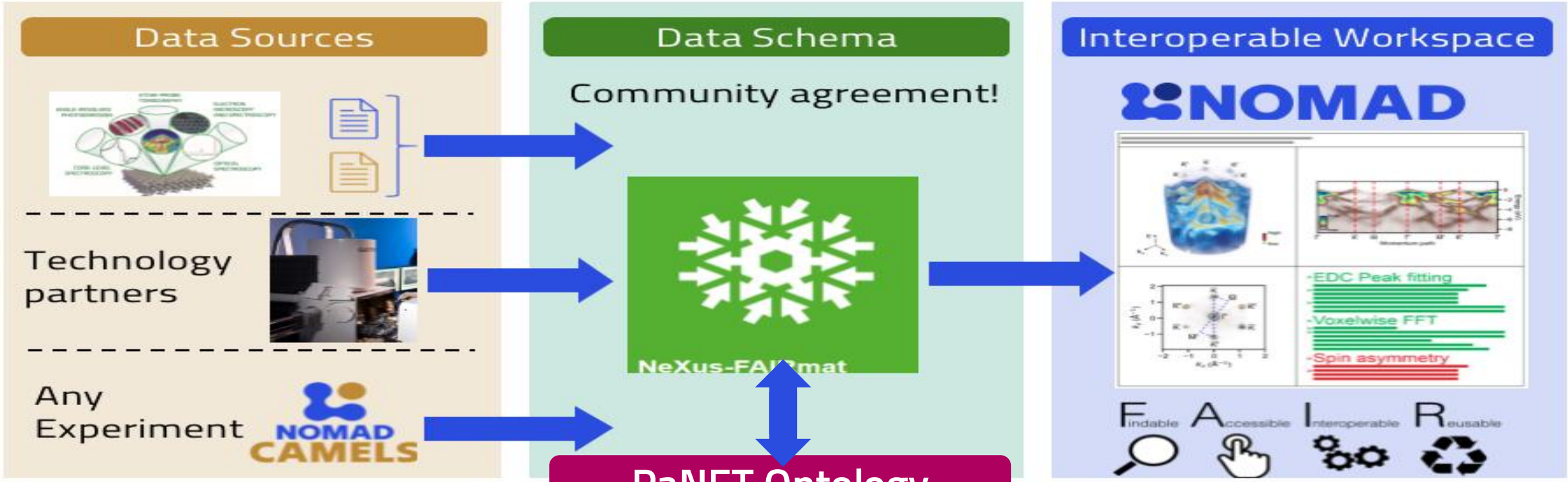
Interoperable Workspace

NOMAD



Findable Accessible Interoperable Reusable

Experimental data workflow



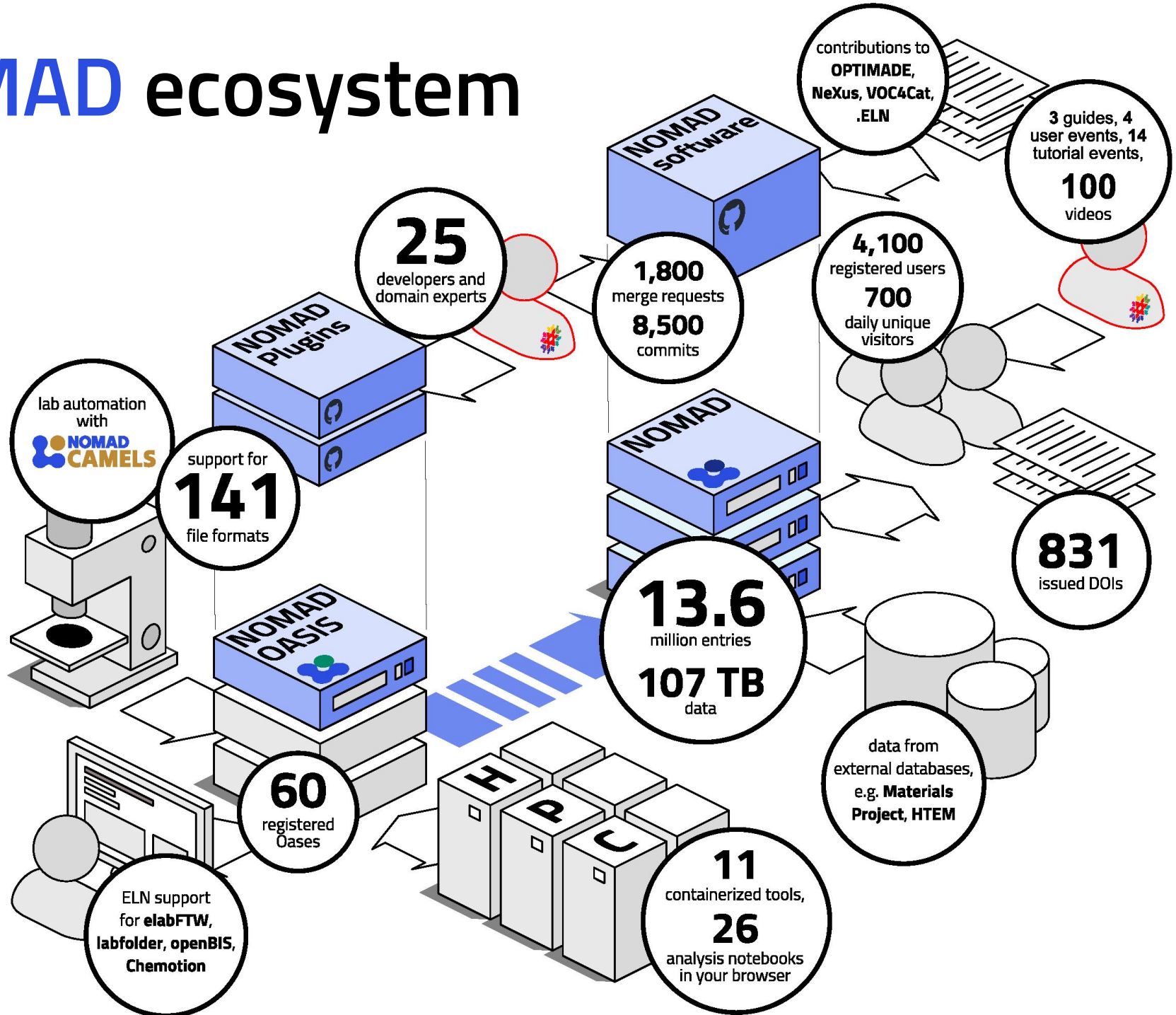
DAPHNE 4NFDI

ESRF The European Synchrotron

OSCARS Open Science Clusters' Action for Research & Society

https://www.oscars-project.eu/projects/findable-big-data-various-material-characterisation-techniques-0

The NOMAD ecosystem



Thank you!

www.fairmat-nfdi.eu

fairmat@physik.hu-berlin.de

company/fairmat-nfdi

www.nomad-lab.eu