

Research Data Management and Open Science

Lukas Pielsticker

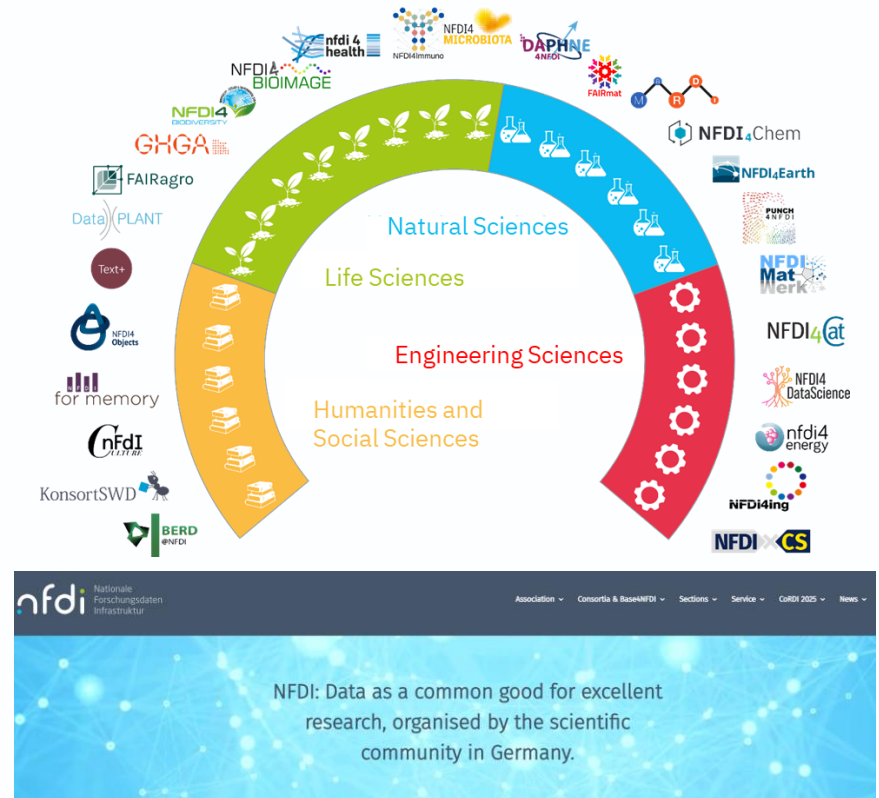
HU Berlin | FAIRmat | NOMAD



National research data infrastructure



- Founded in 2018 – funded by German federal government and states as part of federal government's digitalization strategy
- Main goal:
 - Improvement and creation of **Research Data Infrastructure & Management**
 - Provision of **tools, trainings and initiation of cultural change**
- 27 consortia (26 discipline-specific + Base4NFDI)
- 5 sections (metadata, infra, edutrain, ELSA, industry)

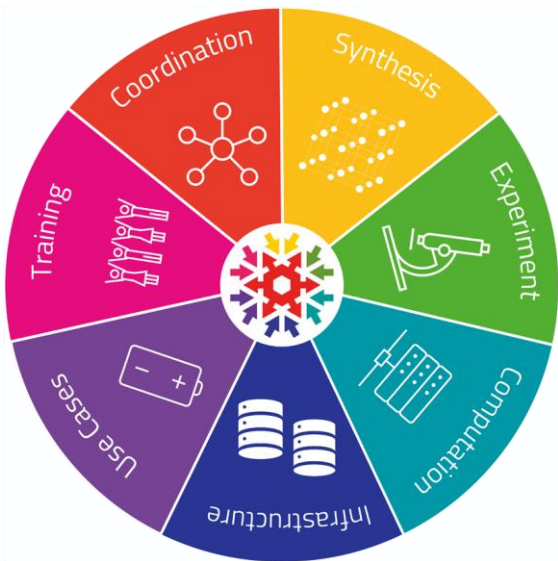


<https://www.nfdi.de/?lang=en>

What is FAIRmat?



FAIRmat is the NFDI consortium building a FAIR data infrastructure for **condensed-matter physics** and the **chemical physics of solids**



<https://fairmat-nfdi.eu>

The FAIRmat team



~60 PIs + ~30 central coworkers



Claudia Draxl
Spokesperson



Christoph T. Koch
Deputy Spokesperson

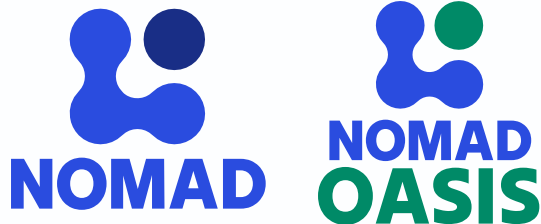
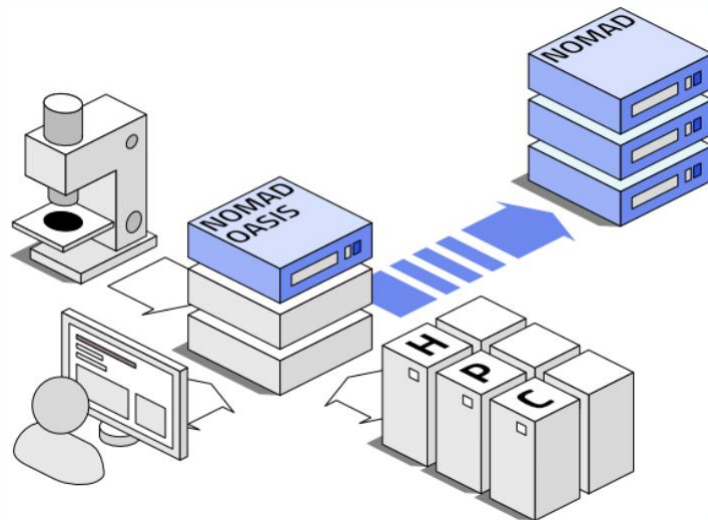


FAIRmat develops and operates NOMAD



NOMAD is a web-based software for FAIR research data management in materials science

NOMAD provides tools for data management, sharing, and publishing. The platform lets you structure, explore, and analyze your data and the data of others.



<https://nomad-lab.eu/>

Learning Objectives for this lecture



- Learn about the research data lifecycle.
- Become familiar with some recommended data management practices to enhance your current and future research.
- Understand the logic behind these practices to be able to adapt to your own research activities.

Contents



WHY

- Why is Research Data Management (RDM) important?

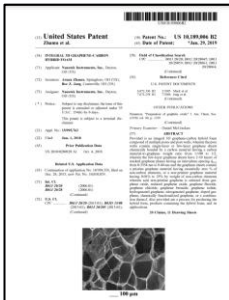
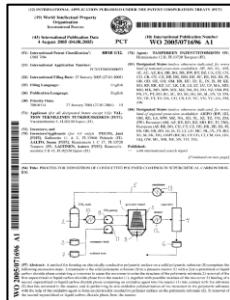
WHY

- What is Research Data Management?

HOW

- How to implement RDM practices?

Outcomes of scientific research

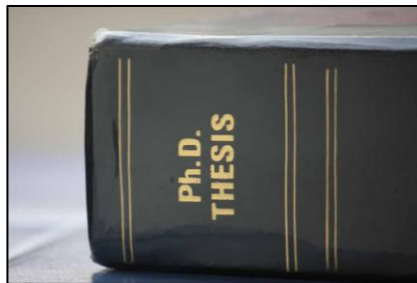


Journal papers

Patents



Dissertations



Outcomes of scientific research

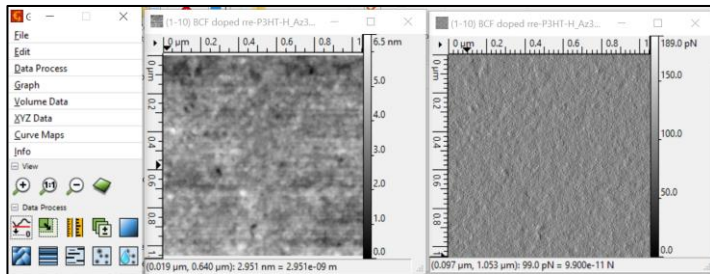
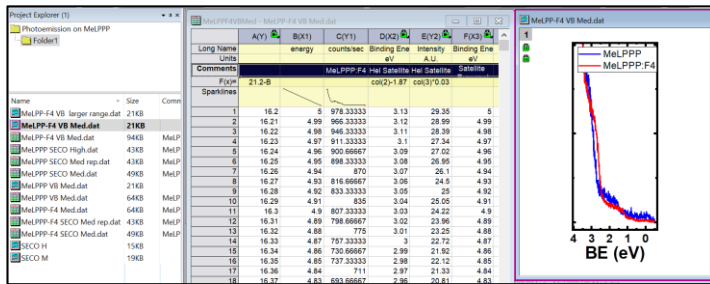


Research data

```
30 w d1 - Notepad
File Edit Format View Help
Date: 10/18/2021
Time: 5:28 PM
Delay time start (s): 1.000000
Delay time between points (s): 0.500000
range of current (A): 1.000000E+0
Auto-range = on

Comments:
IDT-BT 2000 rpm

Voltage Current
V A
-1.900000E+1 -5.370490E-6
-1.710010E+1 -4.772090E-6
-1.519970E+1 -4.178950E-6
-1.329970E+1 -3.592260E-6
-1.139990E+1 -3.013240E-6
-9.499560E+0 -2.440950E-6
-7.599690E+0 -1.880830E-6
```



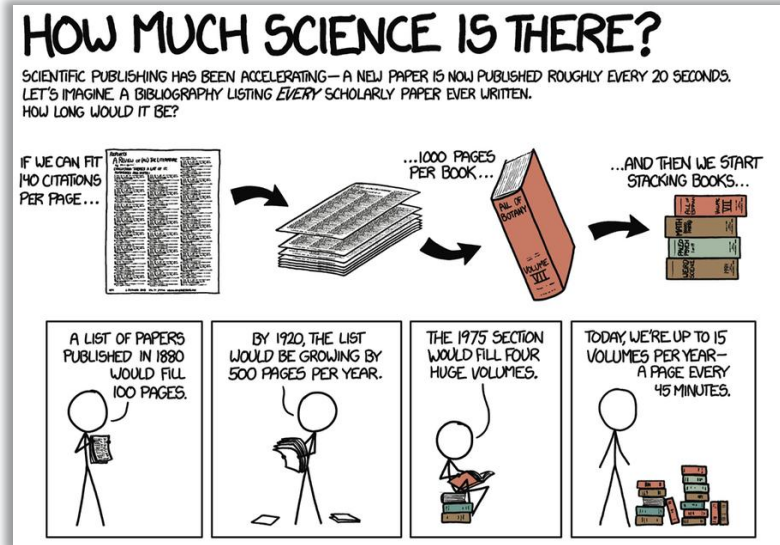
Diverse and vast amounts are being produced and at an increasing rate

Some data are never discussed in traditional publication channels

Stored in disconnected locations using different file formats

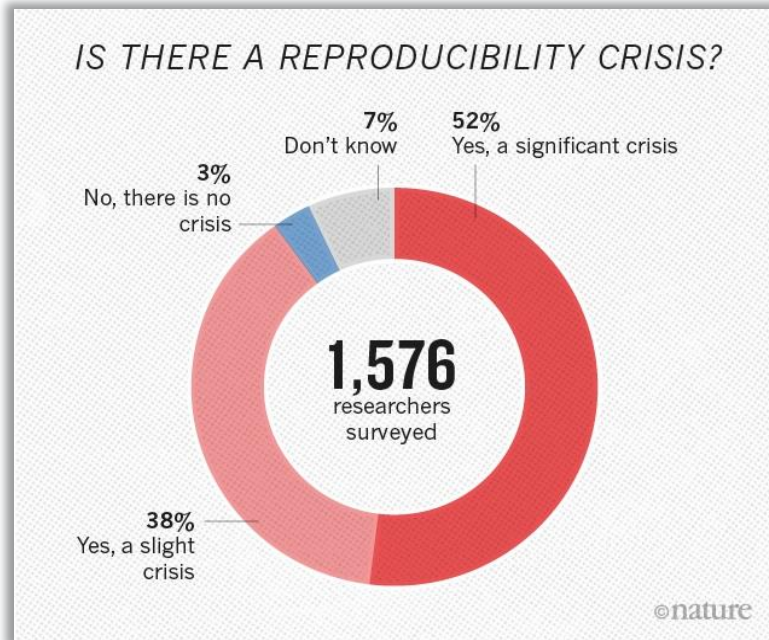
Varying levels of documentation and context using non-consistent terms

Reproducibility in science



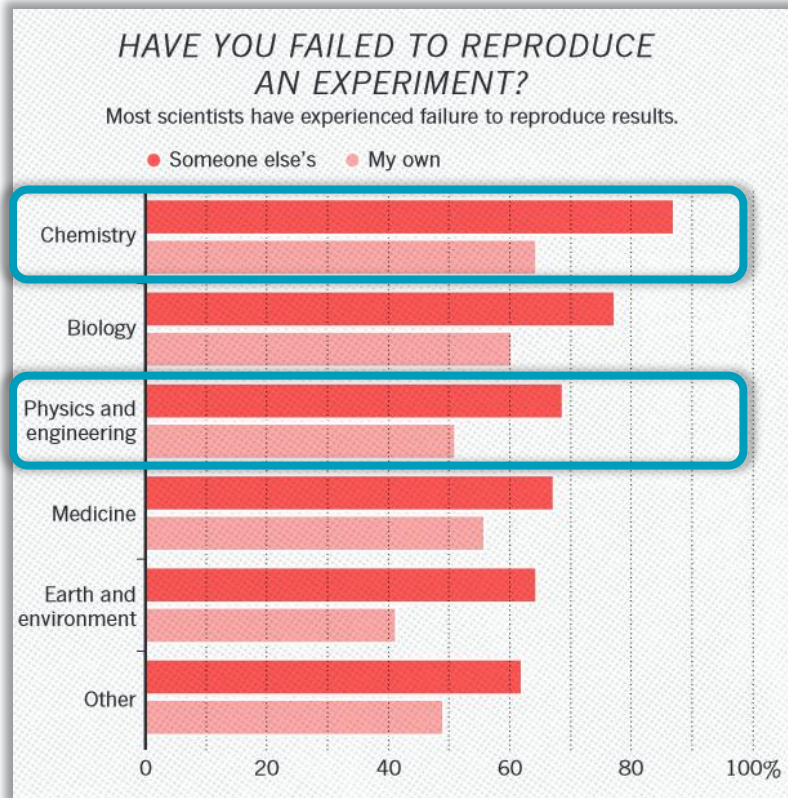
XKCD comic in Science Magazine:

<https://www.science.org/doi/epdf/10.1126/science.342.6154.58>



<https://www.nature.com/articles/533452a>

Reproducibility in science



<https://www.nature.com/articles/533452a>



**Research
Data
Management**

What are Research Data?



Research data are

- recorded factual materials used as primary sources for scientific research.
- commonly accepted in the scientific community as necessary to validate research findings or develop scientific hypotheses, models, or theories.

**Measurement
data**

**Audiovisual
information**

**Computations/
simulations**

**Methodological
test protocols**

**Objects/
samples**

Research Data Lifecycle



Traditional data lifetime



New data lifecycle



What is Research Data Management?



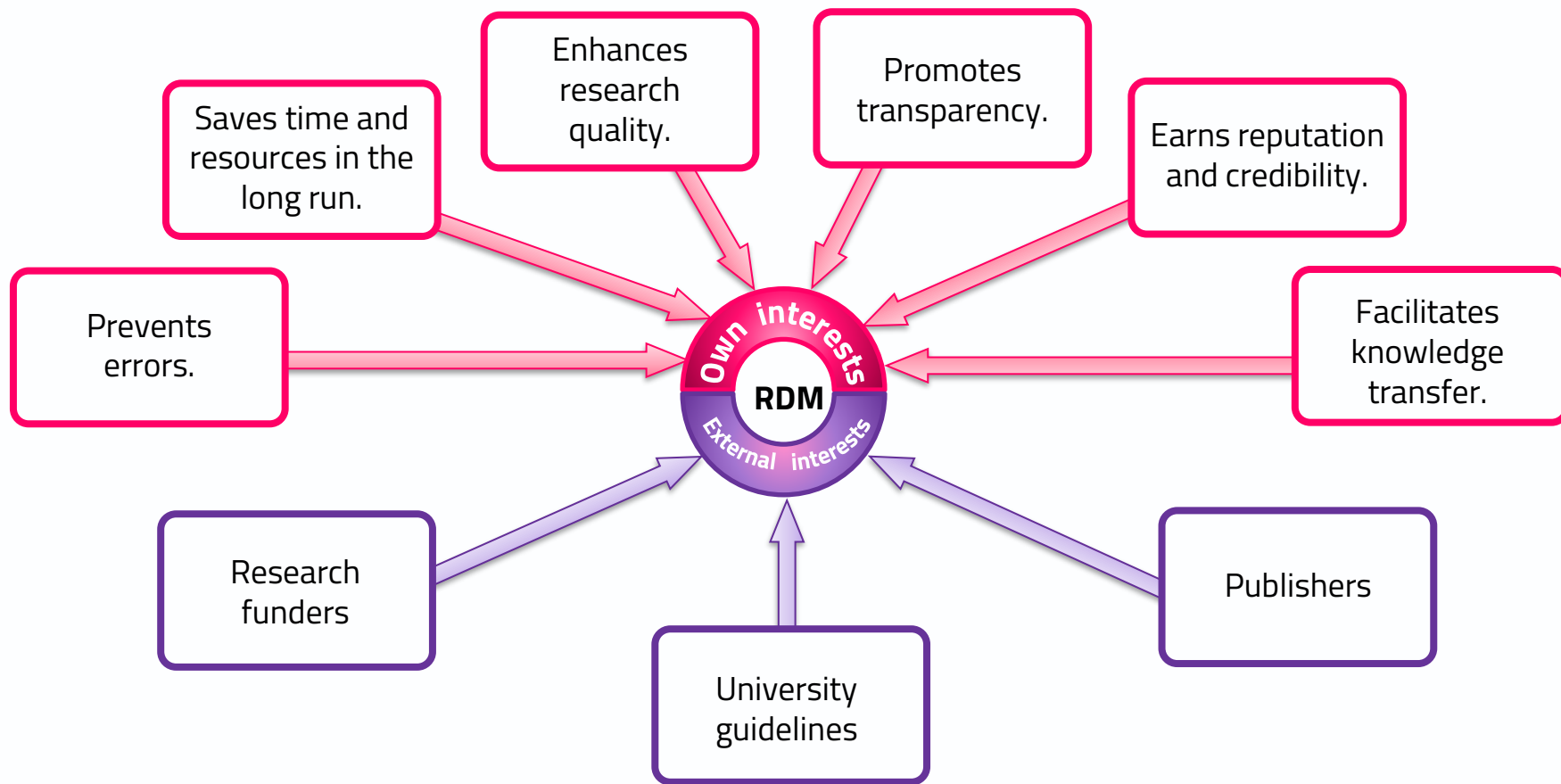
- Why didn't I write down that information?
- Where did I put that file?
- Is this the latest version?



Research Data Management (RDM) is the process of making *informed decisions* about the handling of research data and *implementing* those decisions throughout the data's *lifecycle*.



Why is RDM important?



Recommended RDM Practices in Data Lifecycle



1. Planning



At the planning stage, all RDM aspects are important.
Data Management Plans (DMP) are living documents that dynamically address RDM aspects.

<https://fairmat-nfdi.eu/fairmat/outreach-fairmat/dmp-resources>

2. Data Collection



2.1 Metadata Documentation

2.2 Data Quality Assurance

2.3 Use of Electronic Lab Notebooks (ELNs)

2.1 Metadata Documentation



Metadata are data about your data.

Metadata are structured and contextual information that describe, explain, locate data.

INPUT

Data Object 1

Metadata 1

What experiment?
What instrument?
What instrument settings?
How was it calibrated?
What sample?
What sample history (e.g. shelf time)?
Who, when, why collected the data ...

- Document rich and detailed metadata for your data.
- How much detail? At least enough for your peers to replicate your work.

2.2 Data Quality Assurance



Adhere to Protocols



- Understand the logic, ask if unsure.

Be Consistent



- Consistency ensures comparability and reproducibility.

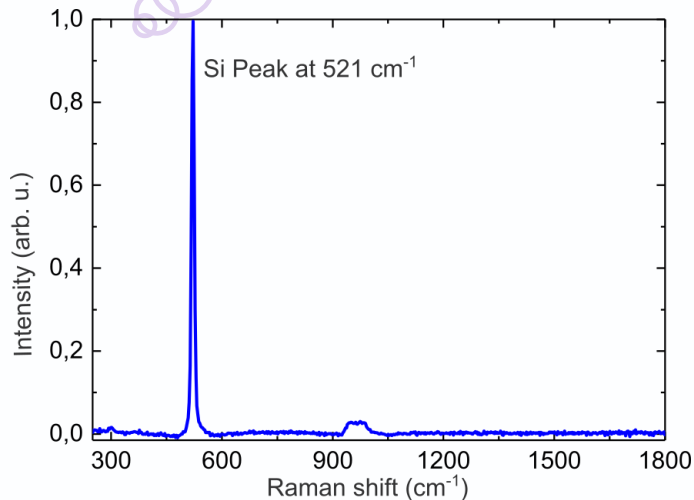
Perform Quality Checks



- Maintain instruments and calibrate regularly.

What, how, and when to calibrate?

- Balance/scale
- Thermometers
- Gas chromatographs
- Mass spectroemters
- Raman spectrometers



2.3 Electronic Lab Notebooks (ELNs)

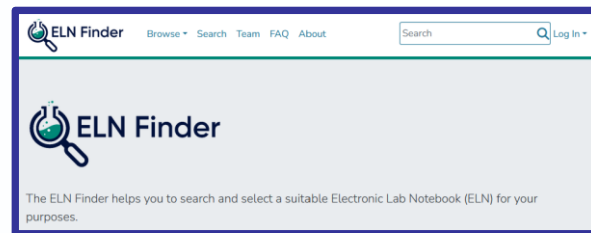


Advantages of using ELNs

- Ensures data integrity (e.g., by uniform data entry, or time stamps)
- Facilitates collaboration
- Integrated into other RDM tools

Which ELN?

- Ease of use
- Customizability
- Analytics support
- Recommendation of your group or research community



Tailored for the *condensed matter physics* and *materials science* communities

- Schema-based metadata collection;
- Built-in schemas for common use cases
- Interoperability with other materials ELNS and databases;



3. Data Analysis



3.1 Analysis Documentation

- Document every step of data transformation and analysis, including the logic behind each decision.

3.2 Intermediate Results Preservation

- Store intermediate results, if needed. (e.g., computationally intensive calculations)
- Use version control (e.g., by adding v1, v2, v3 ... to file names, or advance version control such as Git)

4. Data Preservation



4.1 Short-Term Preservation
(during the project)

4.2 Long-Term Preservation
(after the project)

4.1 Short-Term Data Preservation



4.1.1 Manage Short-Term Data Access

- Manage data access for collaborators, team members, supervisor etc.

4.1.2 Regular Backups

- Consistently back up data to prevent loss.

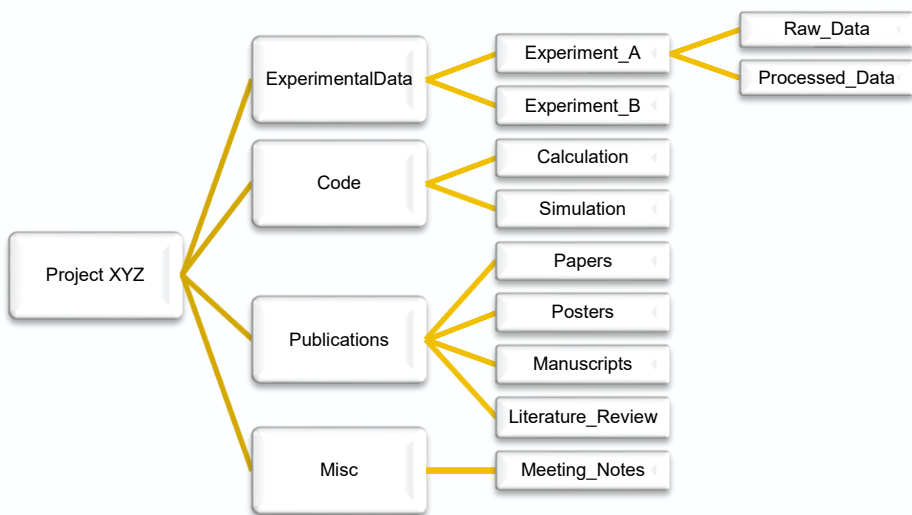
4.1.3 Organized Data Structure

- Maintain a clear and systematic folder hierarchy for easy data retrieval.

4.1.3 Organized Data Structure



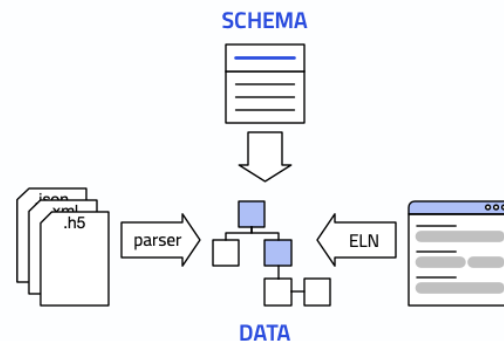
Example hierarchical folder structure



File naming recommendations

- Choose descriptive, content-reflective names
- Consistent naming conventions
- Avoid special characters (e.g., /, , , ;, *, ?)
- Include version if needed (e.g., v05)
- Date formatting (YYYYMMDD) is preferred.

NOMAD's structured data and Metainfo



4.2 Long-Term Data Preservation



4.2.1 Data Preservation Policy

4.2.2 Standard File Formats

4.2.3 Data Archiving and Repositories

4.2.4 Long-Term Data Access
Management



4.2.1 Data Preservation Policy

Meet the requirements, recommendations and conditions

- Institutional data policy
- Funding agencies (DPG, Horizon Europe ...)

Examples:

Deutsche Forschungsgemeinschaft



- Raw data archived for **ten** years.
- Location: Institution or cross-location repositories.
- Justify reasons when not archiving (e.g. personal info)

Horizon Europe



- "As open as possible - as restricted as necessary"
- Provide info for re-use about research outputs, tools and instruments.
- Location: trusted repositories



**Guidelines for Safeguarding
Good Research Practice**

Code of Conduct



4.2.2 Standard File Formats



Recommendations

- Pick common, open formats (e.g., CSV for tables).
- Avoid instrument-specific or software-specific formats as much as possible.
- Avoid encrypted, or compressed or proprietary formats.
- Avoid losing metadata (e.g., TIFF vs. JPG)

Category	Data type	Recommended	Non-recommended
General data formats	Text	TXT, HTML, RTF, PDF/A	DOC, PPT
	Tabular	CSV, TSV, SPSS portable (.por)	XLS
	Images	TIFF, JPEG2000, PNG	GIF, JPG
	Multimedia	Container: AVI, WAV, MP4, Ogg Codec: Theora, Dirac, FLAC	Windows Media Video, QuickTime, H264
	Structured	XML, RDF	RDBMS
Formats common in physics and materials science		HDF5, NeXus, JCAMP-DX	

4.2.3 Data Archiving and Repositories



Institutional Repositories

- edoc-Server (HU Berlin)
- DepositOnce (TU Berlin)
- Refubium (FU Berlin)
- ...

General Public Repositories

- Zenodo, Globus
- GitHub, GitLab
- ...

Discipline-Specific

- List available **re3data.org**
- NOMAD

The screenshot shows the Zenodo website interface. At the top, there's a search bar with 'Graphene' entered. Below the search bar, a message states: 'Zenodo.org will be unavailable for 2 hours on September 29th from 06:00-08:00 UTC. See announcement.' The main content area displays search results for 'CVD graphene SEM image segmentation'. It lists two versions of the dataset, both uploaded on September 9, 2022. The first version is by Aagam Shah, Joshua Schiller, Isiah Ramos, Sameh Tawfik, and Elf Enskin. The second version is also by the same authors. Both versions have a 'Download RDF Package' button. The interface includes navigation links like 'All versions', 'Access Right', and 'File Type'.

The screenshot shows the re3data.org website. The header features the 're3data.org' logo and the tagline 'REGISTRY OF RESEARCH DATA REPOSITORIES'. Below the header, there's a search bar with the placeholder text 'Search...'. The background is a light blue gradient with a subtle pattern.

The screenshot shows the NOMAD Materials science data managed and shared interface. The header features the 'NOMAD' logo and the tagline 'Materials science data managed and shared'. Below the header, there's a search bar with the placeholder text 'Type your query in the search bar'. The main content area displays a periodic table of elements, with various data points and charts overlaid. At the bottom, there are three boxes showing statistics: 'UPLOADED ENTRIES 12,901,910', 'REPRESENTED MATERIALS 3,217,197', and 'UPLOADED FILES 115.6 TB'.

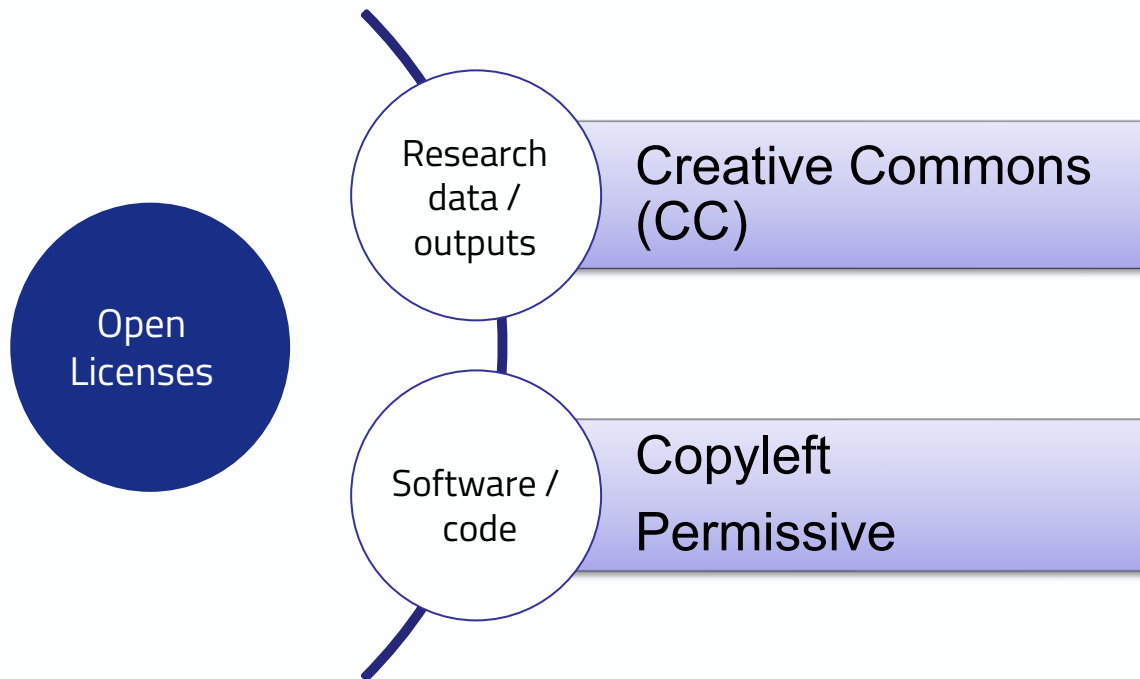
5. Data Sharing



5.1 Licensing

5.2 Persistent Identifiers (PIDs)









5.1 Licensing



5.1 Licensing – Research Outputs



- Most widely used for licensing research outputs, e.g., publications.
- NOT suitable for code/software. They lack necessary provisions.
- Outline restrictions on **commercial use**, **derivative works**, and **licensing of derivative works**.
- 'CC0', 'CC BY', and 'CC BY-SA' meet *Open Access* requirements, with 'CC BY' established as standard.

	Description, permissions and restrictions
	Identified as copyright-free. Use, adapt, share.
	Dedicated to public domain. Use, adapt, share.
	Protected by copyright. Use, adapt, share. Must credit author.
	Like CC BY. Derivatives must also be "ShareAlike."
	Like CC BY. No adaptations allowed.
	Like CC BY. No commercial use.
	Like CC BY-NC. No adaptations allowed.
	Like CC BY-NC. Derivatives must also be "ShareAlike."

5.1 Licensing – Code & Software



Open-source licenses

- Pick a proper open-source license for your code, e.g., **GPL**, **LGPL**, **MIT**, or **Apache 2.0**.
- Copyleft ensures derivative works remain free, open, and copyleft-licensed.
- Most widely used is the MIT license. Anyone can do anything with the source code.

	License	Integration rules	Compatibility	IPR provisions	Distribution obligations
Copyleft	GPL	Strict	Limited	No	Share entire source code
	LGPL	Moderate	Moderate	No	Share LGPLed changes
Permissive	MIT	Flexible	High	No	None
	Apache 2.0	Flexible	Moderate	Yes	None

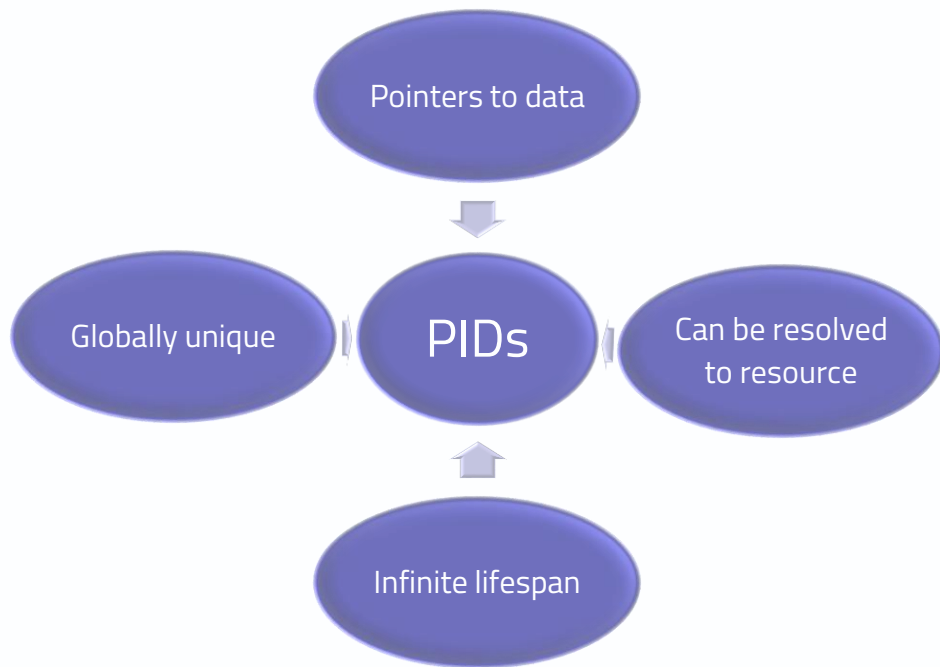
Increasing
permissiveness

5.2 Persistent Identifiers (PIDs)



How do you address your research outputs?
Why are URLs not good enough?

PIDs are pointers to various types of resources:
data files, metadata files, documents, etc.



PID examples

- **DOI** (Digital Object Identifier) for Research Articles, Datasets
- **ISBN** (International Standard Book Number) for Books
- **PURL** (Persistent Uniform Resource Locator) for Digital Resources, Web Pages
- **ORCID** (Open Researcher and Contributor ID) for Researchers, Academics
- **ISSN** (International Standard Serial Number) for Journals, Periodicals

6. Data Reuse

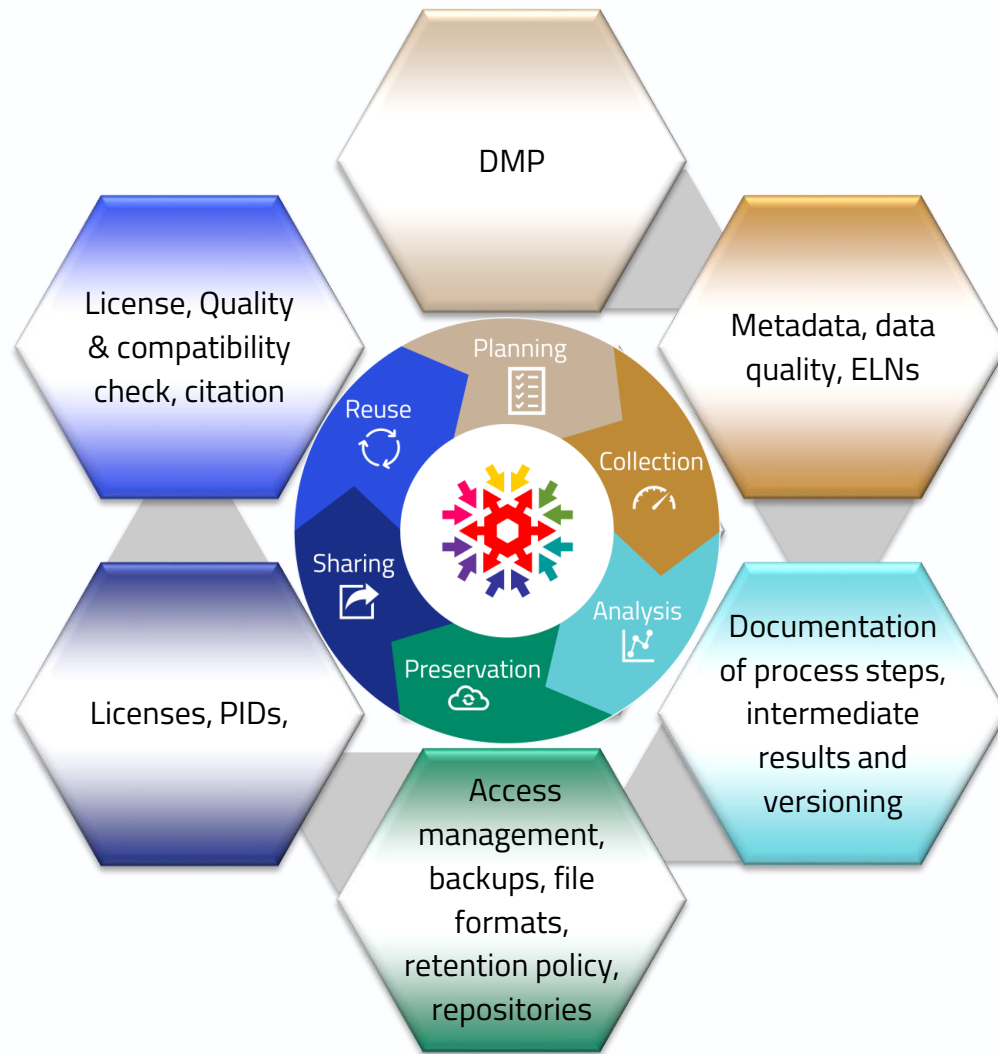


6.1 License Check

6.2 Quality and Compatibility Check

6.3 Proper Citation

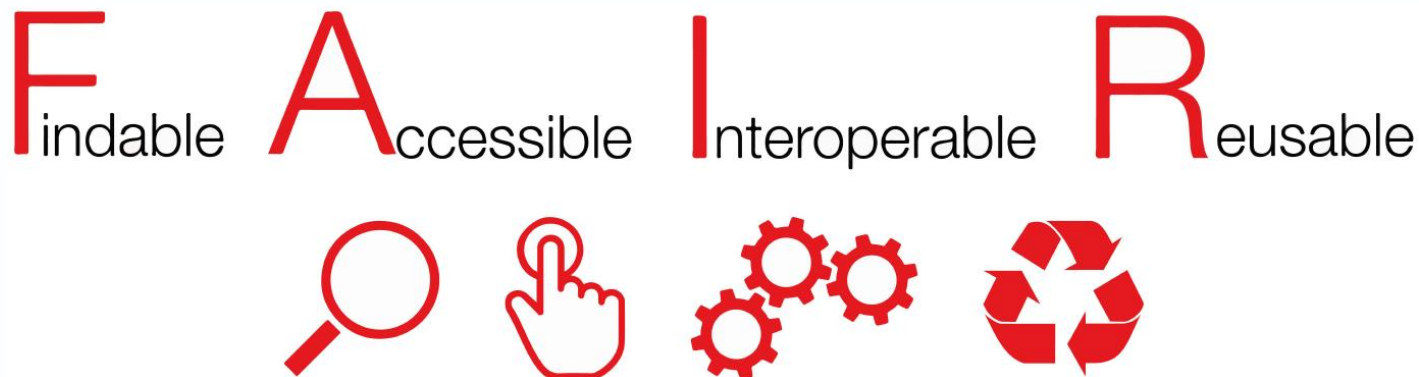
Summary





Break

FAIR data



Wilkinson, M. *et al.* *The FAIR Guiding Principles for scientific data management and stewardship*. *Sci Data* **3**, 160018 (2016).
<https://doi.org/10.1038/sdata.2016.18>

FAIR data



F
indable 

It should be possible for others to discover your data.

A
ccessible 

It should be possible for humans and machines to gain access to your data.

I
nteroperable 

It should be possible to combine and exchange data from different sources.

R
eusable 

Data should be sufficiently documented to support their interpretation and reuse.

The FAIR data principles



- A set of *guiding* principles to enable digital resources to become more Findable, Accessible, Interoperable, and Reusable
- Concise, domain-independent, high-level
- They aim to extend the notion of machine-actionability

Wilkinson, M. *et al.* *The FAIR Guiding Principles for scientific data management and stewardship*. *Sci Data* **3**, 160018 (2016).
<https://doi.org/10.1038/sdata.2016.18>

C. Draxl. *et al.* *FAIR data enabling new horizons for materials research* *Nature* **604**, 635 (2022). <https://www.nature.com/articles/s41586-022-04501-x>

Making data findable



F1 : (Meta)data are assigned a globally unique and persistent identifier

F2 : Data are described with rich metadata

F3 : Metadata clearly and explicitly include the identifier of the data they describe

F4 : (Meta)data are registered or indexed in a searchable resource



























Making data findable



F1: (Meta)data are assigned a globally unique and persistent identifier

F2: Data are described with rich metadata

By Part + Color (Element)	By Part	By Color	By Category
 300524	 407024	 24. Bright Yellow (Yellow)	 Bricks
 300521	 407021	 21. Bright Red (Red)	 Plates
 300523	 407023	 23. Bright Blue (Blue)	 Plates
 300501	 407001	 1. White	 Slope
 300526	 407026	 26. Black	 Slope
	 4070		 SNOT
			 Tiles
			 Technic

Making data findable



F4: (Meta)data are registered or indexed in a searchable resource

The screenshot shows the Bricklink website interface. At the top, there's a navigation bar with the Bricklink logo, a search bar, and various icons for Market, Studio, Programs, Community, Sell, and a Log in / Register button. Below this is a secondary navigation bar with links like Catalog, View, Search, Price Guide, Color Guide, Inventories, Appears In, Relationships, Download, Add or Change, Logs, Credits, and a Stores button.

The main content area displays the product page for a **Brick 1 x 1** (Item No: **3005**, Alternate Item No: **30071, 35382**). The color is set to **White**. A large image of the brick is shown, along with smaller thumbnail images of the brick in different colors (white, blue, red, green, yellow, black). Below the main image are links for **Color Images** and **3D/Lrg Images**.

The product details section includes:

- Item Info:** Years Released: 1954 - 2023, Weight: 0.44g, Stud Dim.: 1 x 1 x 1 in studs, Pack. Dim.: 0.8 x 0.8 x 1.15 cm.
- Item Consists Of:** N/A.
- Item Appears In:** [3928 Sets](#), [47 Minifigures](#), [21 Parts](#), [22 Books](#), [23 Gear](#).
- My Store Inventory:** [Add to My Store Inventory](#), 167044 Lots For Sale.
- My Wanted List:** [Add to My Wanted List](#), On 3064115 Wanted Lists.
- My Collection:** [Add to My Collection](#), In 3446 Collections.

<https://www.bricklink.com>

Making data accessible



A1: (Meta)data are retrievable by their identifier using a standardized communication protocol

A1.1: The protocol is open, free and universally implementable

A1.2: The protocol allows for an authentication and authorization procedure where necessary

A2: Metadata should be accessible even when the data is no longer available



Accessibility



A1: (Meta)data are retrievable by their identifier using a standardized communication protocol

A2: Metadata should be accessible even when the data is no longer available

What is a communication protocol?

A set of rules and standards that allow for exchange of data between two or more entities in a consistent and predictable manner.

For example: hypertext transfer protocol (HTTP).

<https://www.bricklink.com/v2/catalog/catalogitem.page?P=3005>

Catalog: [Parts](#) **Brick:** [3005](#)

Brick 1 x 1
Item No: [3005](#) Alternate Item No: [30071](#), [35382](#)



[View Price Guide](#) [Buy](#)

Select Color ▼

Item Info
Years Released: 1954 - 2023
Weight: 0.44g
Stud Dim.: 1 x 1 x 1 in studs
Pack. Dim.: 0.8 x 0.8 x 1.15 cm

Item Consists Of
N/A

Item Appears In
[4073 Sets](#)
[49 Minifigures](#)
[22 Parts](#)
[25 Books](#)
[23 Gear](#)


My Store Inventory
[Add to My Store Inventory](#)
183489 Lots For Sale

My Wanted List
[Add to My Wanted List](#)
On 2735504 Wanted Lists

My Collection
[Add to My Collection](#)
In 6590 Collections

If the data are no longer available, the metadata should still be accessible in order to at least offer an understanding of the data nature and provenance.

Making data interoperable



I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation

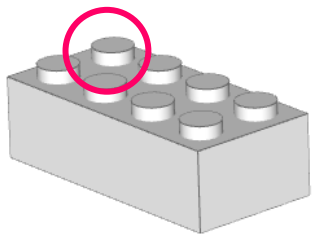
I2: (Meta)data use vocabularies that follow the FAIR principles

I3: (Meta)data include qualified references to other (meta)data.

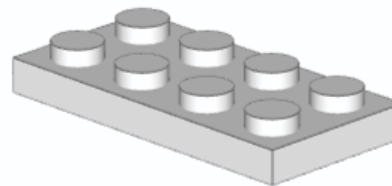
Making data interoperable



I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation



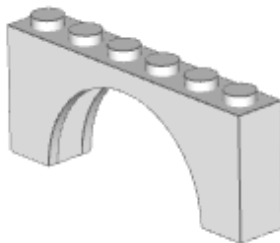
This is a **brick** with 8 studs



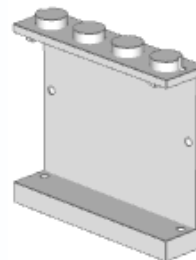
This is a **plate** with 8 studs



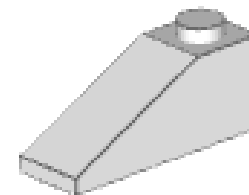
Tile



Arch



Panel



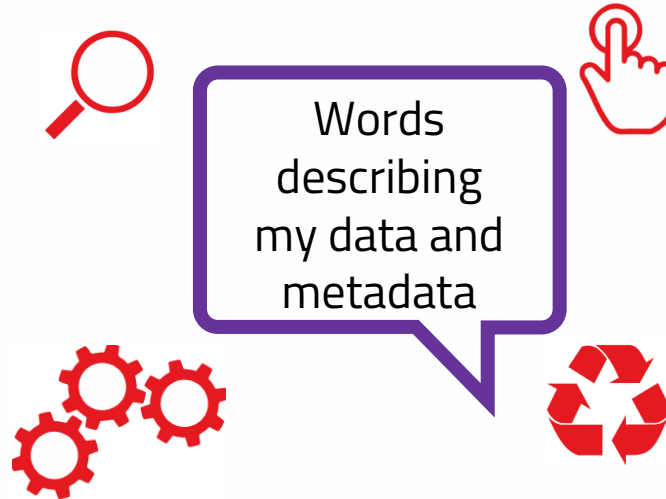
Slope

Making data interoperable



I2: (meta)data use vocabularies that follow FAIR principles.

The vocabulary used to describe the data and the metadata should also be FAIR. Documenting vocabularies, using unique identifiers, standardized protocols, and knowledge representation languages are key steps in achieving FAIRness.



Making data Reusable



R1: (meta)data are richly described with a plurality of accurate and relevant attributes

R1.1: (meta)data are released with a clear and accessible data usage license

R1.2: (meta)data are associated with detailed provenance

R1.3: (meta)data meet domain-relevant community standards

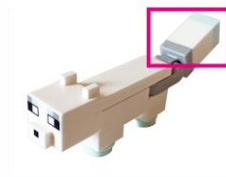
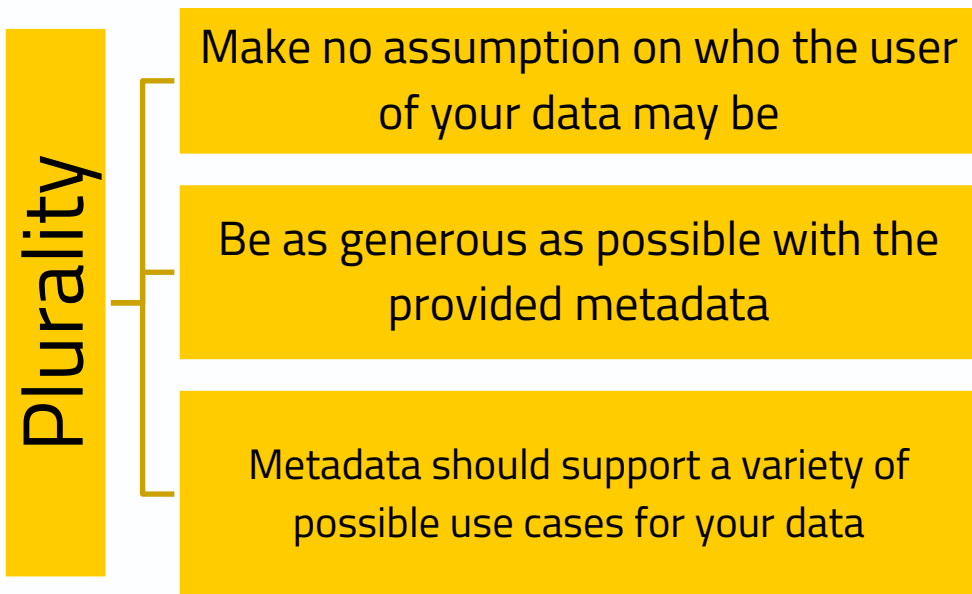


Making data reusable



R1: (meta)data are richly described with a plurality of accurate and relevant attributes

Both the metadata and data should be described in a manner that enables assessment of the resource in a certain study.

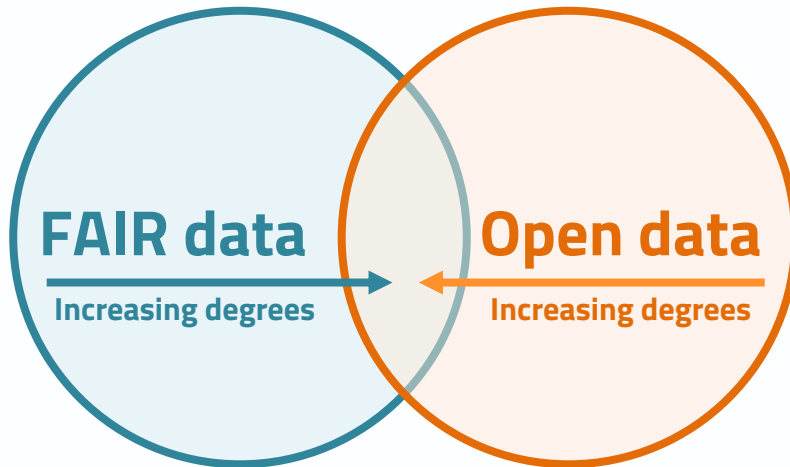


FAIR data and open data



Not all open data are FAIR, not all FAIR data are necessarily open

Findable
Accessible
Interoperable
Reusable



Structured data



Unstructured data (or unstructured information) is information that either does not have a pre-defined **data model** or is not **organized** in a **pre-defined** manner.

Unstructured information is typically text-heavy, but may contain data such as dates, numbers, and facts as well. This results in irregularities and ambiguities that make it difficult to understand using traditional programs as compared to data stored in **fielded form** in databases or **annotated** (semantically tagged) in documents.

en.wikipedia.org/wiki/Unstructured_data

Tables, trees, and graphs



\$id	datetime	temperature	source_peak _wavelength	data
xrd_0	2026-04-09T 08:24:21.629	298	1.54060	data_0
xrd_1

- + Human actionability
- + Easy analysis / ML
- + Fast to query
- Flat, no nesting
- Low extensibility
- Data description



Tables, trees, and graphs



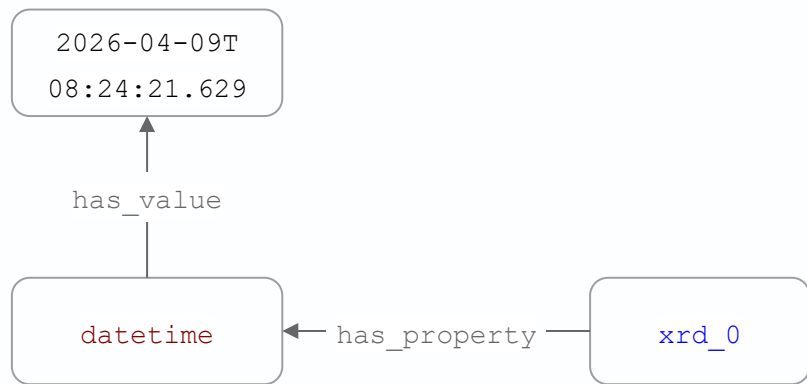
```
$id: xrd_0
datetime: 2026-04-09T08:24:21.628816+00:00
temperature: 278
source:
  peak_wavelength: 1.54060
  tube_material: Cu
  tube_current: 40
  tube_voltage: 4e4
intensity:
  - 2086
  - 2053
  - 2118
  - ...
two_theta:
  - 10.007
  - 10.020
  - 10.033
  - ...
```

- + Human readable
- + Extensible and modular
- + OOP and Web integration
- Linked data



MongoDB®

Tables, trees, and graphs



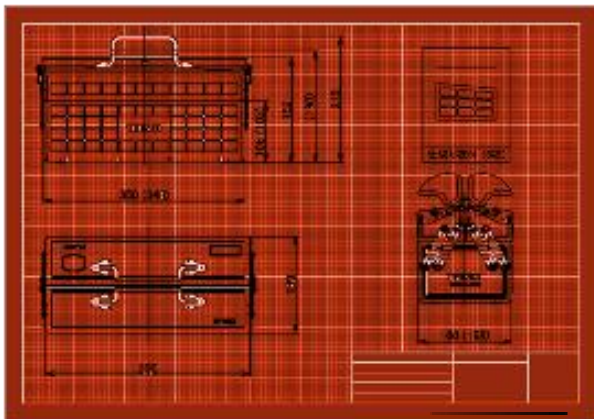
- + Linked data
- + Flexible query
- Complex to query
- Poor scaling



Data schema



Schema



Template



+



Data

Structured
Data Archive



Data schema example



- Interoperable web services
- JSON transport format
- Schema for responses

JSON Schema (in YAML)



```
intensity:
  - 2086
  - 2053
  - 2118
  - ...
two_theta:
  - 10.007
  - 10.020
  - 10.033
  - ...
temperature: 278
datetime: 2026-04-09T08:24:21.628816+00:00
source:
  peak_wavelength: 1.54060
  tube_material: Cu
  tube_current: 40
  tube_voltage: 4e4
```

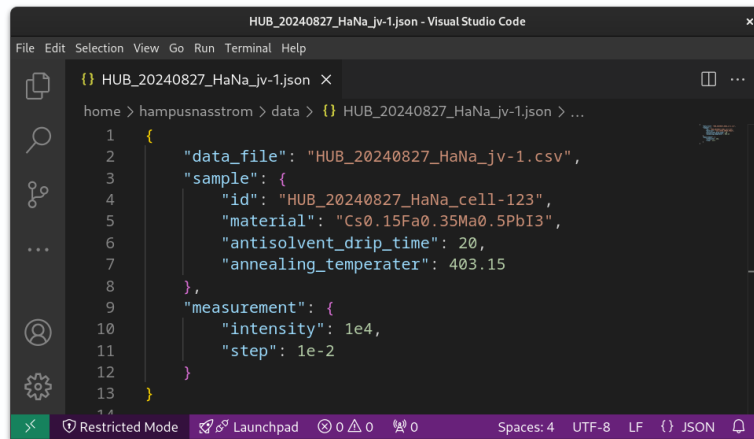
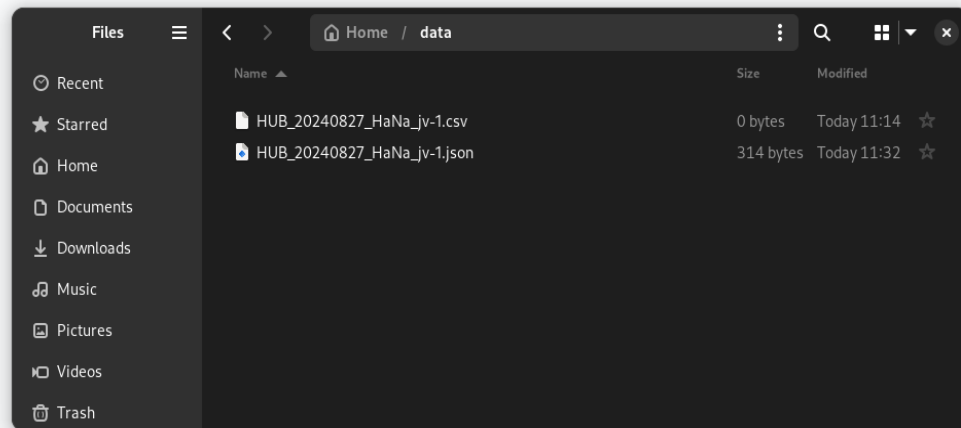
```
$schema: https://json-schema.org/draft/2020-12/schema
$id: https://example.org/schemas/xrd-measurementschema.json
title: XRD Measurement
description: Schema for XRD measurement record.
type: object
properties:
  intensity:
    type: array
    description: Intensity counts at each two-theta angle.
    minItems: 1
    items:
      type: integer
      minimum: 0
  two_theta:
    type: array
    description: Two-theta angles in degrees.
    minItems: 1
    items:
```


Where do I start?



Somewhere!

Don't let perfect be the enemy of good

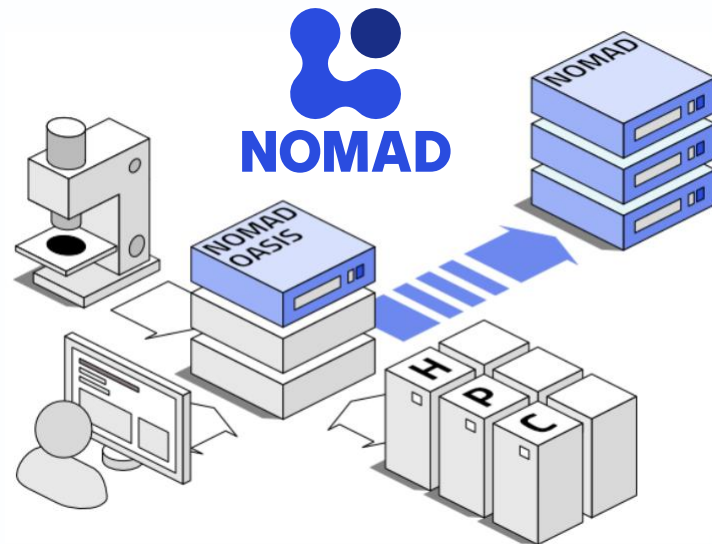


RDM in practice: FAIRmat operates NOMAD



NOMAD is a web-based software for FAIR research data management in materials science

NOMAD provides tools for data management, sharing, and publishing. The platform lets you structure, explore, and analyze your data and the data of others.



UPLOADED ENTRIES

19,229,014

UPLOADED FILES

113.7 TB

REGISTERED USERS

> 4000

<https://nomad-lab.eu/>


NOMAD core functionalities






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






NOMAD core functionalities




 PUBLISH ▾ EXPLORE ▾ ANALYZE ▾ ABOUT ▾

Entries 





Welcome [Pepe Marquez](#)  LOGOUT  UNITS


FILTERS     


Material


Elements / Formula 


Elements


  AND  


Structure 


Method 

DFT 


GW 

Projection 

DMFT 


EELS 

Workflow

Molecular Dynamics 



662 RESULTS

ELEMENTS / FORMULA


Elements linear ▾ 


☐ only compositions that exclusively contain these atoms


1 H																	2 He
3 Li	4 Be									5 B	6 C	7 N	8 O	9 F	10 Ne		
11 Na	12 Mg									13 Al	14 Si	15 P	16 S	17 Cl	18 Ar		
19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
55 Cs	56 Ba	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn	
87 Fr	88 Ra	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Cn	113 Nh	114 Fl	115 Mc	116 Lv	117 Ts	118 Og	
119 Uue																	
57 La	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu			
89 Ac	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr			


Chemical Formula Hill linear ▾  Chemical Formula IUPAC linear ▾ 


Authors


M Markus Scheidgen 


M Hieu Ngo 

M Niloofar Hadaeghi 

M Honghui Shang 


M Honghui Shang 


M Honghui Shang 



M Honghui Shang 

NOMAD core functionalities



 PUBLISH ▾ EXPLORE ▾ ANALYZE ▾ ABOUT ▾

Entries / Entry 

Welcome [Pepe Marquez](#)  LOGOUT  UNITS

OVERVIEW

FILES

DATA

LOGS

Metadata

method name
DFT

program version
4.6.35 3Apr08 complex parallel LinuxlFC

program name
VASP

basis set type
plane waves

core electron treatment
pseudopotential

jacob's ladder
GGA

xc functional names
GGA_C_PBE, GGA_X_PBE

comment
no comment

references
<http://www.sciencedirect.com/scienc...>
<http://aflowlib.org>
<http://www.sciencedirect.com/scienc...>

authors

Material

Composition

formula	K ₂ SeZn
dimensionality	bulk
elements	K, Se, Zn
number of elements	3 (ternary)

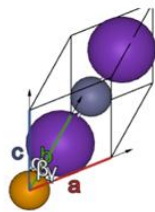
Symmetry

crystal system	bravais lattice
cubic	cF
space group number	space group symbol
225	Fm-3m
point group	structure name
m-3m	Heusler

Lattice parameters

a	b	c
5.699 Å	5.699 Å	5.699 Å
α	β	γ
60 °	60 °	60 °
cell volume		
130.892 Å ³		

Structure



Original ▾

Legend: K (purple), Zn (blue), Se (orange)

Structure visualization showing the arrangement of atoms in the unit cell.


NOMAD core functionalities



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



NOMAD core functionalities





PUBLISH ▾ EXPLORE ▾ ANALYZE ▾ ABOUT ▾

Entries / Entry / Data

Welcome [Pepe Marquez](#)  LOGOUT  UNITS

OVERVIEW

FILES

DATA

LOGS

☐ code specific ☐ all defined ☐ definitions <> ↺

→ <> **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

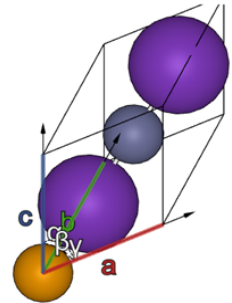
→ <> **Atoms**

section

 K

 Zn

 Se



QUANTITIES

species = 4 vector

labels = 4 list

positions = 4 × 3 matrix

→ <> **lattice_vectors**

quantity

VALUE

0	4.02990	4.02990
4.02990	0	4.02990
4.02990	4.02990	0

(3 × 3)
Å

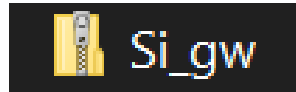
REFERENCED BY closed
















NOMAD core functionalities



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

NOMAD core functionalities



Name	Type
 BAND.OUT	OUT File
 BAND-QP.OUT	OUT File
 bandstructure	DAT File
 bandstructure	XML Document
 bandstructure-qp	DAT File
 dos	XML Document
 EIGVAL.OUT	OUT File
 EIGVAL_GW.OUT	OUT File
 EVALQP	DAT File
 GW_INFO.OUT	OUT File
 info	XML Document
 INFO.OUT	OUT File
 input	XML Document
 TDOS.OUT	OUT File
 TDOS-QP.OUT	OUT File

NOMAD core functionalities



PUBLISH ▾ EXPLORE ▾ ANALYZE ▾ ABOUT ▾

Your uploads / Examples for demonstrations

Welcome [Ahmed Mansour](#)

LOGOUT UNIT5

OVERVIEW

FILES



Examples for demonstrations

upload id: 5VEbvP_GQKuNikao-uusXA



1 Prepare and upload your files

Here you can upload files. Top-level .zip/.tar files will be uncompressed automatically. For more information, see our documentation on [uploading files](#) or view the [supported codes](#). Optionally, you can also create an entry from built-in or uploaded schemas. Please take a look at our documentation on [schemas](#).

DROP FILES HERE OR CLICK TO OPEN DIALOG

CREATE FROM SCHEMA

> /

2 Process data

3 Edit visibility and access



Enabling this will allow all users, including guests without an account, to view the upload even before it is published.

You can edit the access to the upload by adding or removing users as upload members.

EDIT UPLOAD MEMBERS

NOMAD core functionalities




[PUBLISH](#) ▾ [EXPLORE](#) ▾ [ANALYZE](#) ▾ [ABOUT](#) ▾

Your uploads / [Examples for demonstrations](#)

Welcome [Ahmed Mansour](#)

[LOGOUT](#) [SETTINGS](#)

 DROP FILES HERE OR CLICK TO OPEN DIALOG

CREATE FROM SCHEMA




> /

2 Process data

Processing completed, 3/3 entries processed

3 entries



<input type="checkbox"/>	Name	Type	Mainfile	Process status ↑	
<input type="checkbox"/>	Si exciting GW SinglePoint simulation	exciting GW SinglePoint	INFO.OUT 	SUCCESS	→
<input type="checkbox"/>	Si exciting DFT SinglePoint simulation	exciting DFT SinglePoint	INFO.OUT 	SUCCESS	→
<input type="checkbox"/>	Si exciting DFT+GW simulation	exciting DFT+GW	INFO.OUT 	SUCCESS	→

The NOMAD Metainfo



NOMAD is built on structured data with a well-defined schema (*NOMAD Metainfo*).

Extendable
**NOMAD
Metainfo**
schema
(~30000 terms)



Hierarchical data structures and cross-references



A domain-independent superstructure



Detailed, method-specific data description

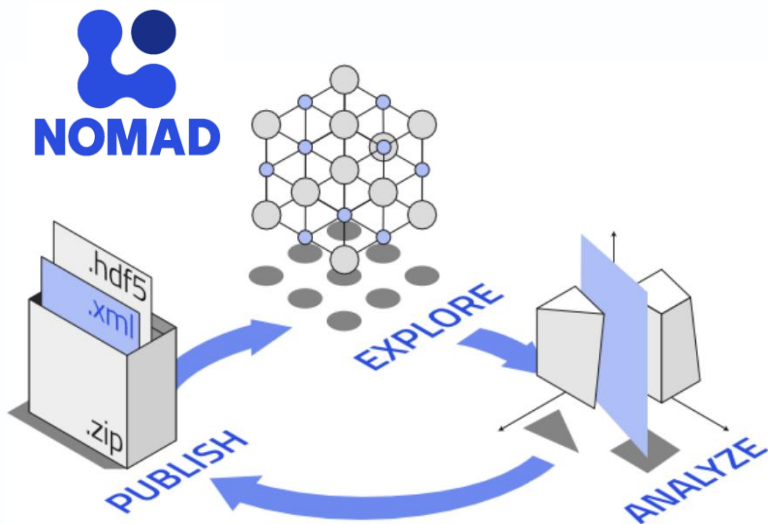


Shared models across domains

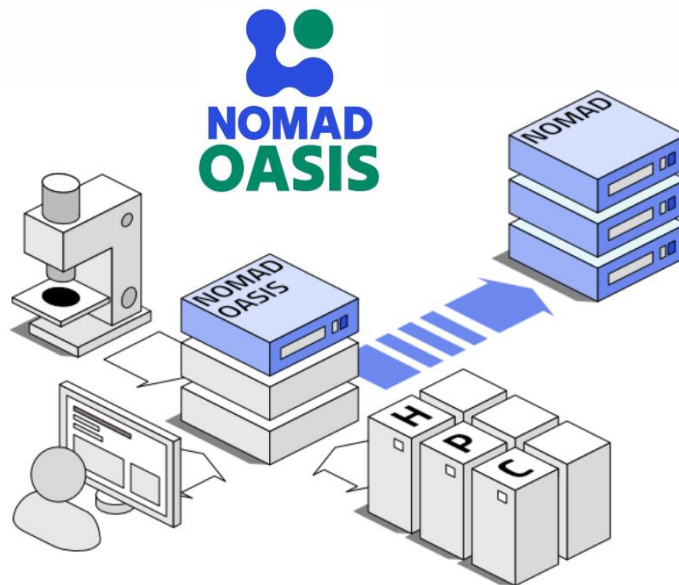
NOMAD comes in two flavors



Publish your data and analysis
Archive Repository

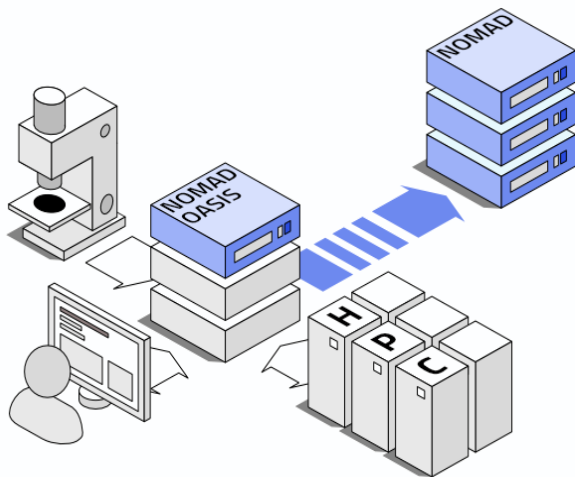


Manage your lab
Local management tool | ELN





Standalone, locally deployable instance of NOMAD that enables the management of research data within labs, groups, and institutions.



REGISTERED INSTALLATIONS

> 90



HZB Helmholtz
Zentrum Berlin



**universität
innsbruck**

ILLINOIS



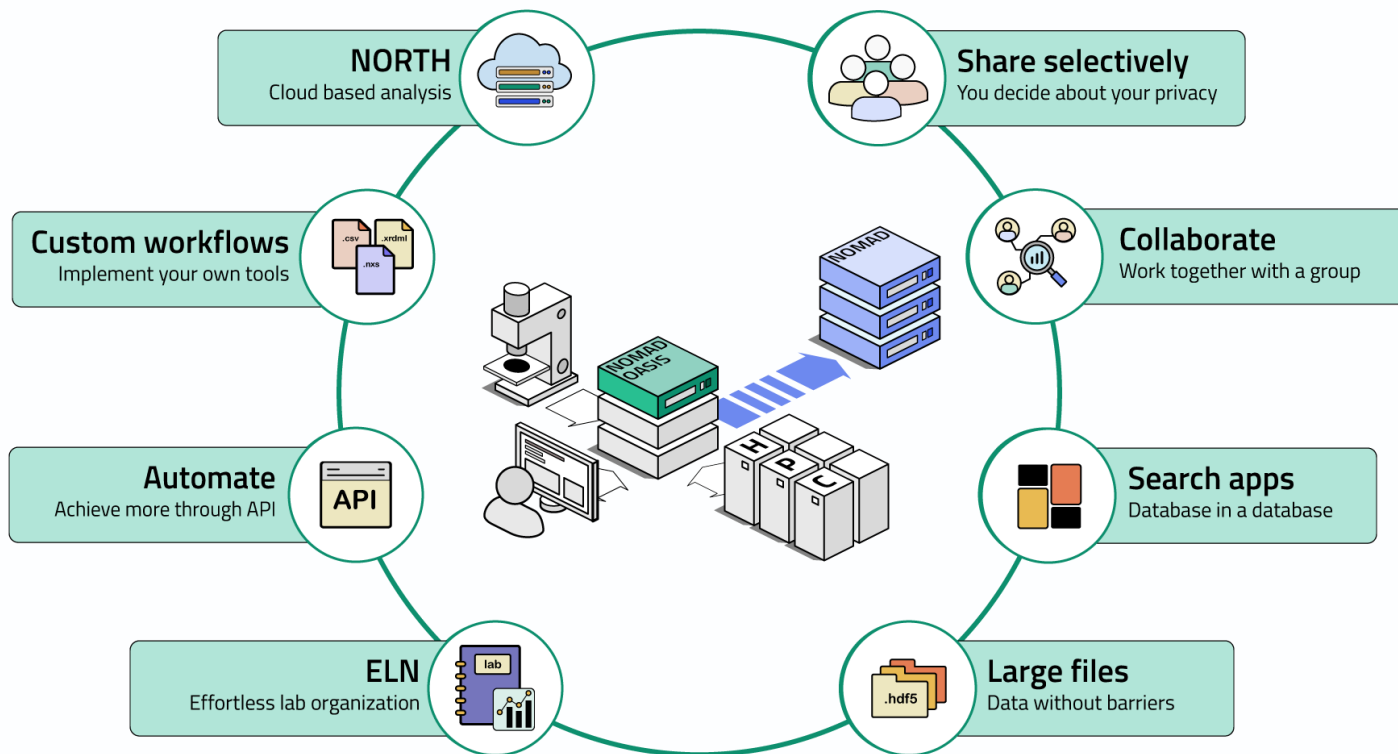
**TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN**

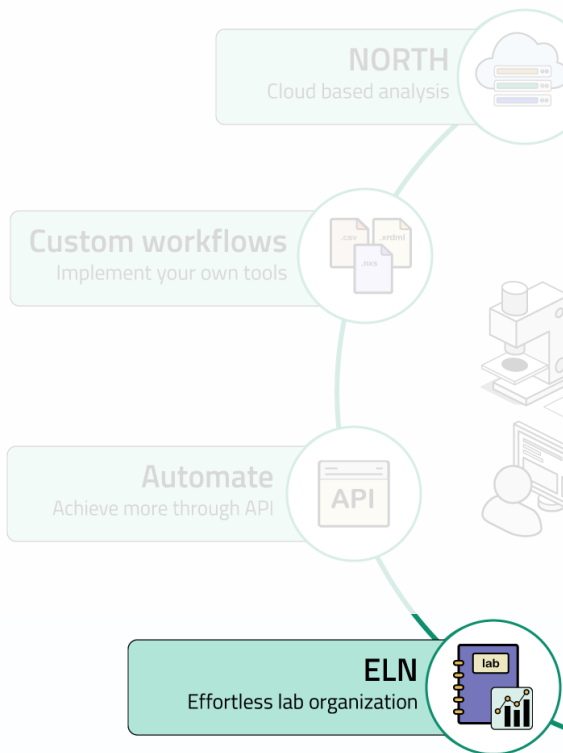
li.u LINKÖPING
UNIVERSITY



WARWICK
THE UNIVERSITY OF WARWICK







PUBLISH ▾ **EXPLORE** ▾ **ANALYZE** ▾ **ABOUT** ▾

Welcome [Pepe Marquez](#) **LOGOUT** **UNITS**

Your uploads / Upload / Entry

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-uxMhJUCBN9_-rKyOyyNszP](#)

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

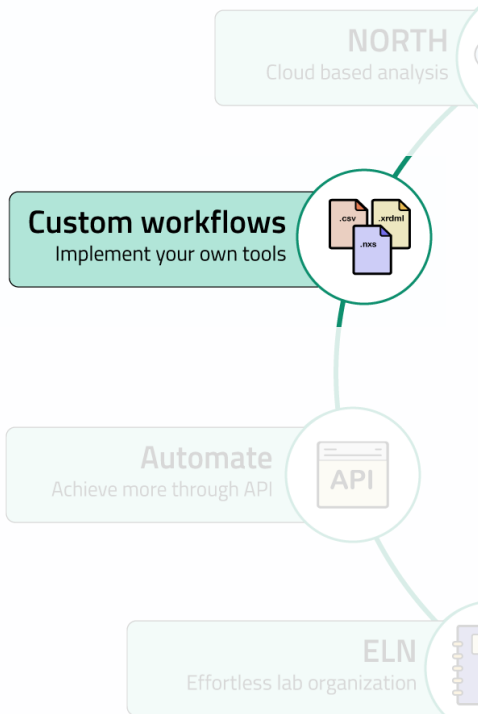
Paragraph

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



PUBLISH ▾ **EXPLORE** ▾ **ANALYZE** ▾ **ABOUT** ▾

Welcome [Pepe Marquez](#) [LOGOUT](#) [UNITS](#)

Your uploads / Upload / Entry / Data

OVERVIEW FILES **DATA** LOGS

search

☐ code specific ☐ all defined ☐ definitions [cloud](#) [<>](#) [refresh](#) [save](#)

Entry ... <>

section

SUB SECTIONS

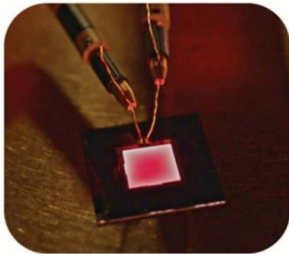
results

metadata

data

REFERENCED BY [closed](#)

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



12 WORDS. POWERED BY TINY

Datetime
23/03/2023 12:28 [calendar](#)

SUB SECTIONS

users

publication_reference

solar_cell_definition

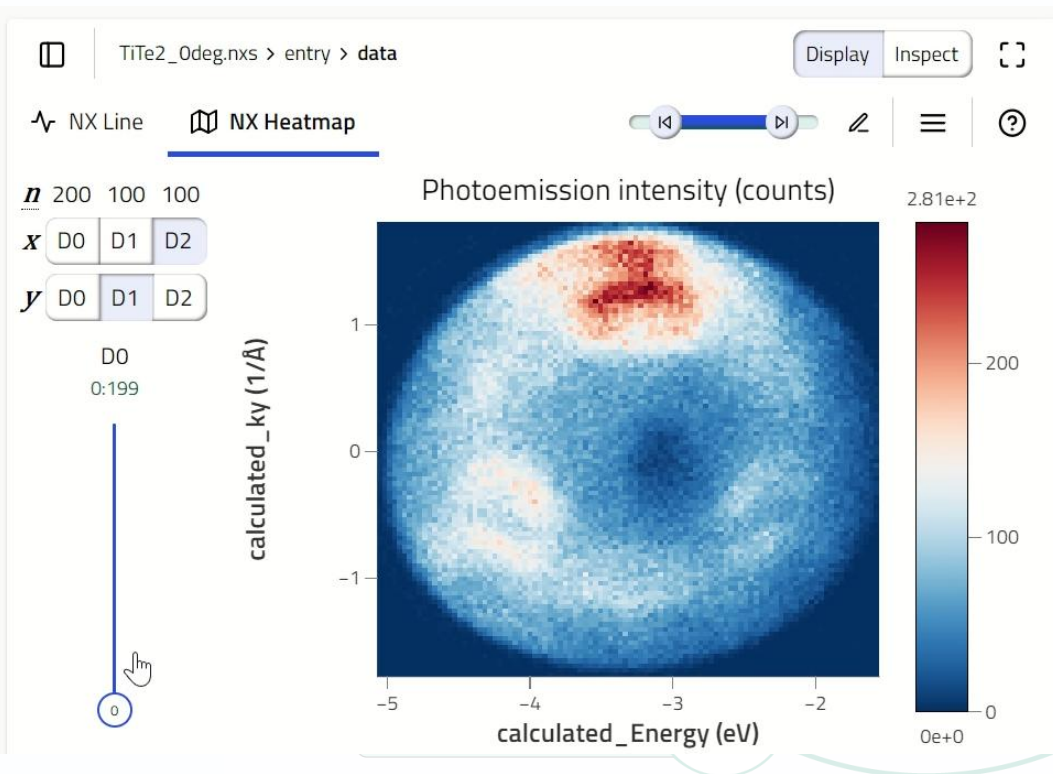
solar_cell_parameters

solar_cell_eqe

absorber_layer

REFERENCED BY [closed](#)

NOMAD Oasis



Share selectively
You decide about your privacy

Collaborate
Work together with a group

Search apps
Database in a database

Large files
Data without barriers



NORTH
Cloud based analysis



Share selectively
You decide about your privacy



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

Maintainer: [Markus Scheidgen](#)

File extensions: ipynb

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

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.



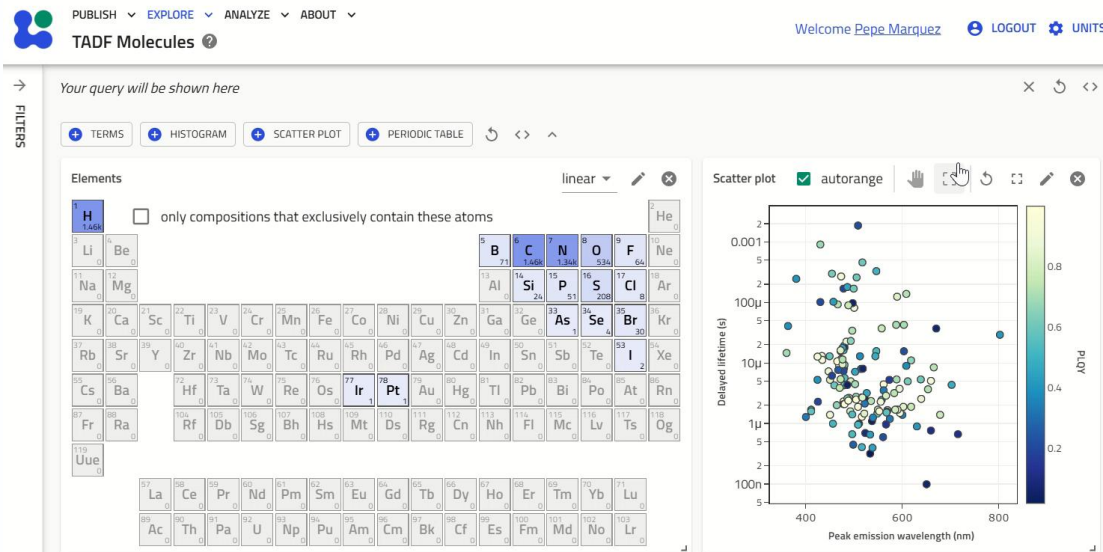
vesta

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



1 a group

5
abase



Share selectively
You decide about your privacy

Collaborate
Work together with a group

Search apps
Database in a database

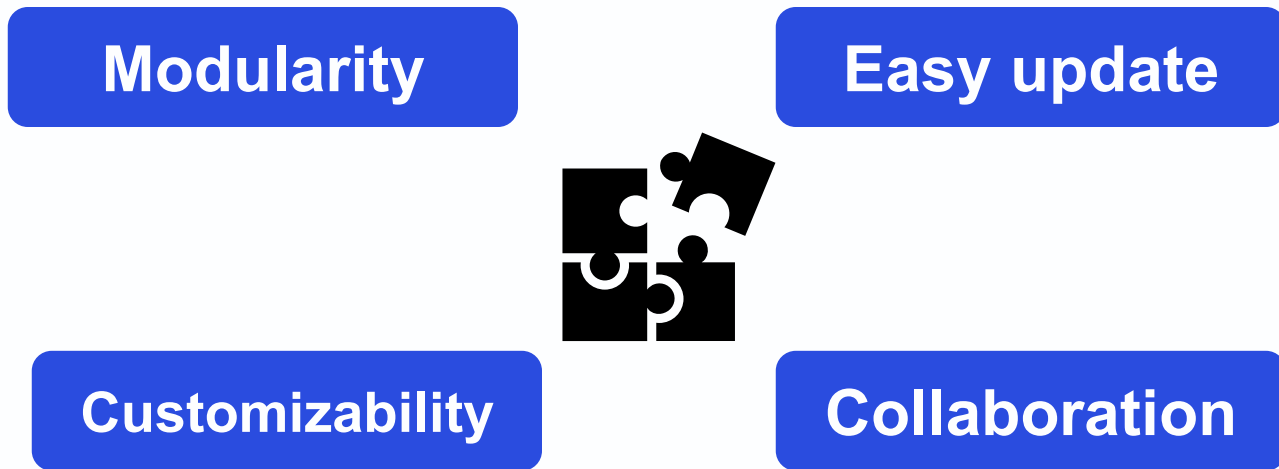
ELN
Effortless lab organization

Large files
Data without barriers

NOMAD Plugins



Framework that extends NOMAD's support for file formats, scientific methods, workflows, and domain-specific data models.



Customize your **Oasis** with **Plugins**



Schema packages

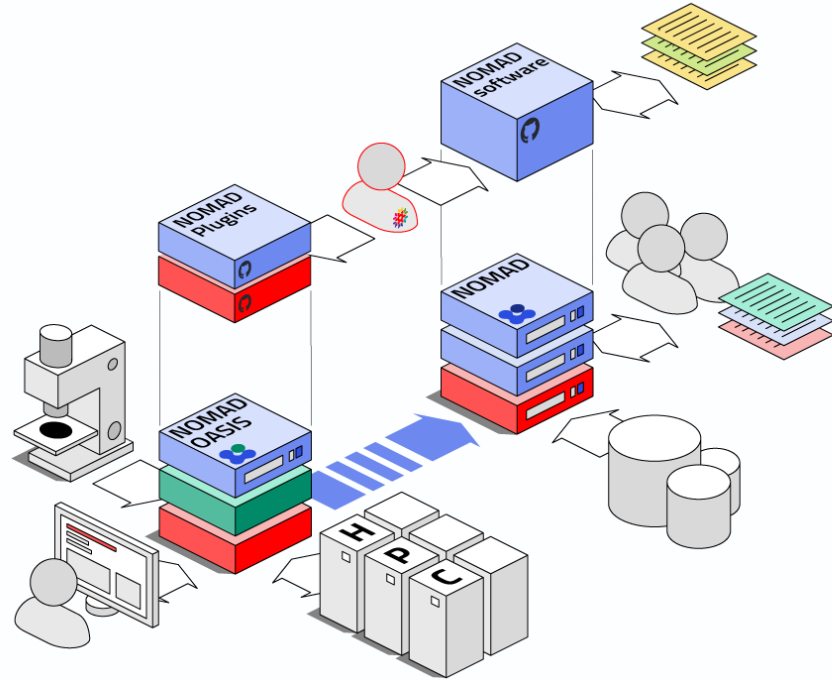
Parsers

Apps

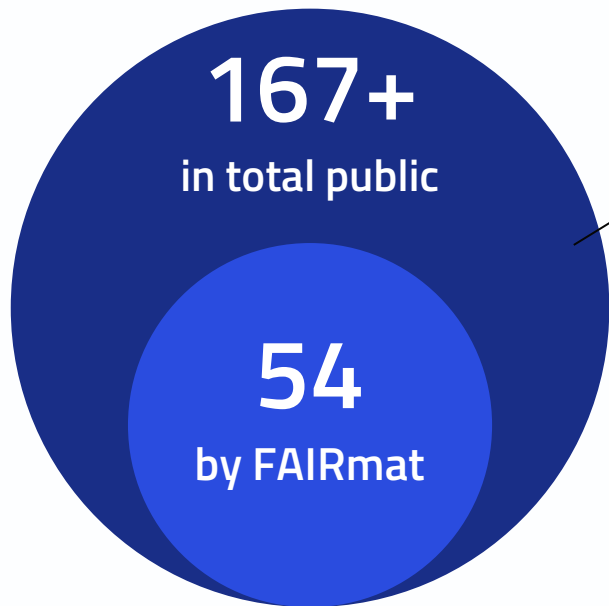
Example uploads

Normalizers

Actions

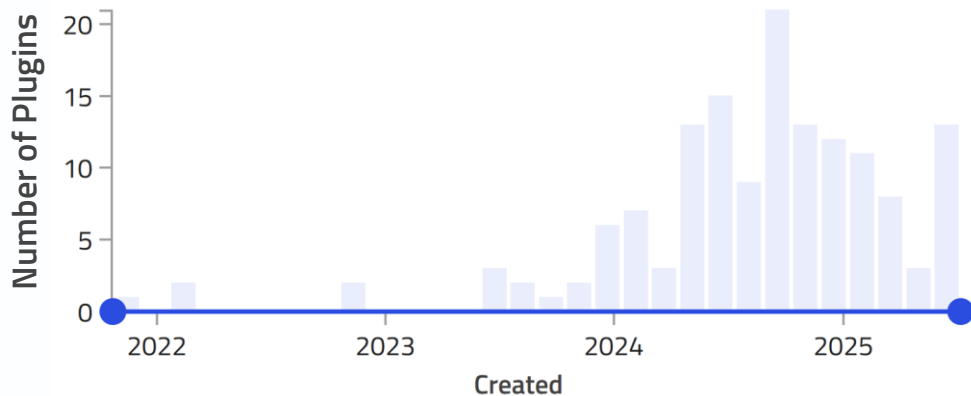


Plugins allow the community
to extend NOMAD
functionality



70 Unique
plugin authors

Schema package	317
Parser	282
App	112
Example upload	54
Normalizer	36
API	2



NOMAD plugins and supported methods/files



nomad-measurements

XRD

UV-Vis-NIR



pynxtools

EM

APM

PES

optical spectroscopy



nomad-materials-processing

MBE

PLD

CVD

MOVPE

sputtering

thermal evaporation

solution processing

nomad-simulations





**Want to learn more about RDM,
open science, FAIRmat, NOMAD?**



SOLUTIONS ▾

LEARN ▾

GET INVOLVED ▾

ABOUT ▾

OPEN NOMAD

Tutorials

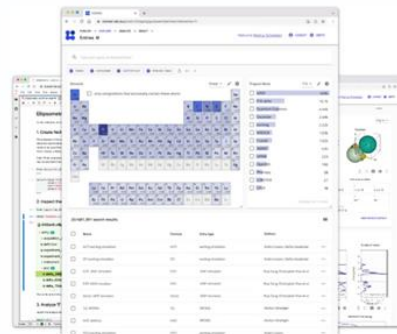
Documentation

Support

NOMAD Materials science data managed and shared

NOMAD lets you manage and share your materials science data in a way that makes it truly useful to you, your group, and the community. **Free and open source.**

Open NOMAD →



NOMAD Discord

Join our brand new NOMAD Discord, where you can connect with fellow researchers, share ideas, and get answers.

Open Discord



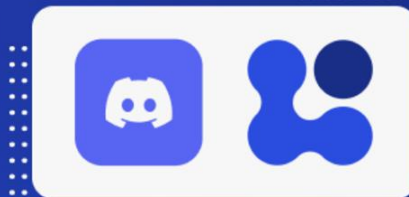
Join the FAIRmat | NOMAD community



NOMAD Discord

Join our brand new NOMAD Discord, where you can connect with fellow researchers, share ideas, and get answers.

[Open Discord](#)



Receive guidance from our domain experts
and active community of > 650 members

www.nomad-lab.eu

FAIRmat Users Meeting

Public event and workshops on data infrastructure and research data management practices

June 16-17

CSMB (IRIS) Adlershof



Program and registration



<https://events.fairmat-ntdi.eu>

Thank you for your attention!

 www.fairmat-nfdi.eu

 [@fairmat.bsky.social](https://bsky.social/@fairmat)

 fairmat@physik.hu-berlin.de

 [company/fairmat-nfdi](https://www.linkedin.com/company/fairmat-nfdi)

 www.nomad-lab.eu



Direct contact: lukas.pielsticker@physik.hu-berlin.de



Appendix

Identify Your Role



User

Upload datasets, search dataset,
use defined applications



Application Administrator

Implement and create
use-case-based applications

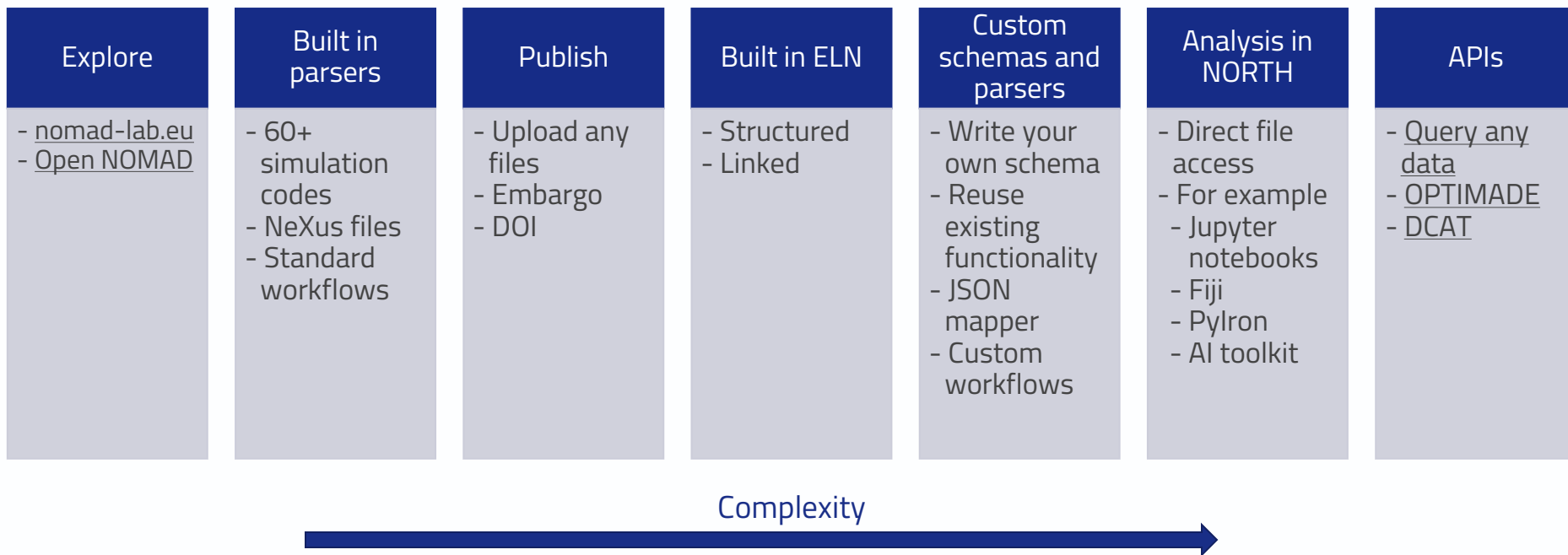


System Administrator

Setup, configure and maintain server
instance



Using Central Installation



The Plugin System



Existing plugins

- github.com/FAIRmat-NFDI/nomad-lab.eu/prod/v1/osis

Develop plugins

- [Tutorial in docs](#)
- [Example upload](#)
- [Cookiecutter](#)

Schema packages

- Add automatization
- Custom plots

Parsers

- Read any files
- Match by type/name/content
- Link to other entries

Apps

- Dashboard
- Custom search
- Inventory

Complexity



Hosting Your Own Installation



Default NOMAD

- Docker image
- Base Linux

Custom image with plugins

- Template repository
- GitHub Actions
- End of [Tutorial 13](#)

NORTH images

- Custom images
- For example
 - Igor Pro
 - Casa XPS

Local file system

- Mount existing drives
- Supports very large files

User management

- Configure user groups
- Use internal keycloak

Complexity

