

## **FAIRmat Tutorial 8:**

### **Using NOMAD as an Electronic lab notebook (ELN) for FAIR data**

Organized by FAIRmat Area A Synthesis



**FAIRmat**

# Introduction

# Present Situation in crystal growth – unstructured data

Spreadsheet Files

batch	AMTwin-label	d <sub>layer</sub> µm	P <sub>t</sub> W	t <sub>t</sub> µs	d <sub>t</sub> µm	d <sub>hatch</sub> µm	plane	porosity [%]	density [%]	Ø-density [%]	number of voids	max. pore size [mm]
1	AMT_0595	30	200	65	55	105	x-y	0,00%	100,00%		12	0,04
6	AMT_1140	25	200			100	x-y	0,00%	100,00%		2	0,01
							x-z	0,00%	100,00%		8	0,04

Batch 1: AMT\_0595  
30 µm layer thickness / 200 W / 850 mm/s  
stress relieved (550 °C / 180 min / Ar / furnace cooling)

Batch 6: AMT\_1140  
25 µm layer thickness / 200 W / 1200 mm/s  
HIP (920 °C, 1000 bar / 120 min / Ar) – 25 µm

File Servers

Name	Ext	Size	Date
[m2800]	<DIR>	25.09.2020 09:36	
[m2801]	<DIR>	14.05.2020 09:12	
[m2802]	<DIR>	28.05.2020 11:46	
[m2803]	<DIR>	28.05.2020 11:47	
[m2804]	<DIR>	21.09.2018 11:57	
[m2805]	<DIR>	25.08.2020 16:06	
[m2806]	<DIR>	14.09.2018 13:50	
[m2807]	<DIR>	15.03.2019 09:31	
[m2808]	<DIR>	25.08.2020 16:07	
[m2809]	<DIR>	25.08.2020 16:07	
[m2810]	<DIR>	27.08.2020 11:12	
[m2811]	<DIR>	01.10.2018 13:59	
[m2812]	<DIR>	16.04.2020 17:50	
[m2813]	<DIR>	28.08.2020 11:27	
[m2814]	<DIR>	28.08.2020 11:30	
[m2815]	<DIR>	28.08.2020 11:32	
[m2816]	<DIR>	08.04.2019 09:23	
[m2817]	<DIR>	28.08.2020 11:33	
[m2818]	<DIR>	28.08.2020 11:34	
[m2819]	<DIR>	28.08.2020 11:36	
[m2820]	<DIR>	13.05.2019 12:20	
[m2821]	<DIR>	21.03.2019 18:55	
[m2822]	<DIR>	21.03.2019 18:55	
[m2823]	<DIR>	28.08.2020 11:37	
[m2824]	<DIR>	28.05.2020 11:51	
[m2825]	<DIR>	02.04.2019 12:51	
[m2826]	<DIR>	02.04.2019 12:51	
[m2827]	<DIR>	02.04.2019 12:51	
[m2828]	<DIR>	14.05.2019 17:14	
[m2829]	<DIR>	02.04.2019 12:51	
[m2830]	<DIR>	02.04.2019 12:50	
[m2831]	<DIR>	02.04.2019 16:49	
[m2832]	<DIR>	10.05.2019 19:31	
[m2833]	<DIR>	10.05.2019 19:32	
[m2834]	<DIR>	11.06.2019 14:56	
[m2835]	<DIR>	13.05.2019 14:41	

Lab Notes

Feb. 7th  
M393.885.1♀ x C405.913.3♂ (E) } 3½ days, left as  
M536.976.1♀ (T) x C502.960.5♂ } unop. controls. D388

C337.810.2♀ (L) x C406.874.5♂ } 2½ days, left as  
C348.728.1♀ (T, b) x C489.896.4♂ } unop. controls. R288  
C222.333.2♀ (T) x C504.907.2♂ } unop. controls. R290

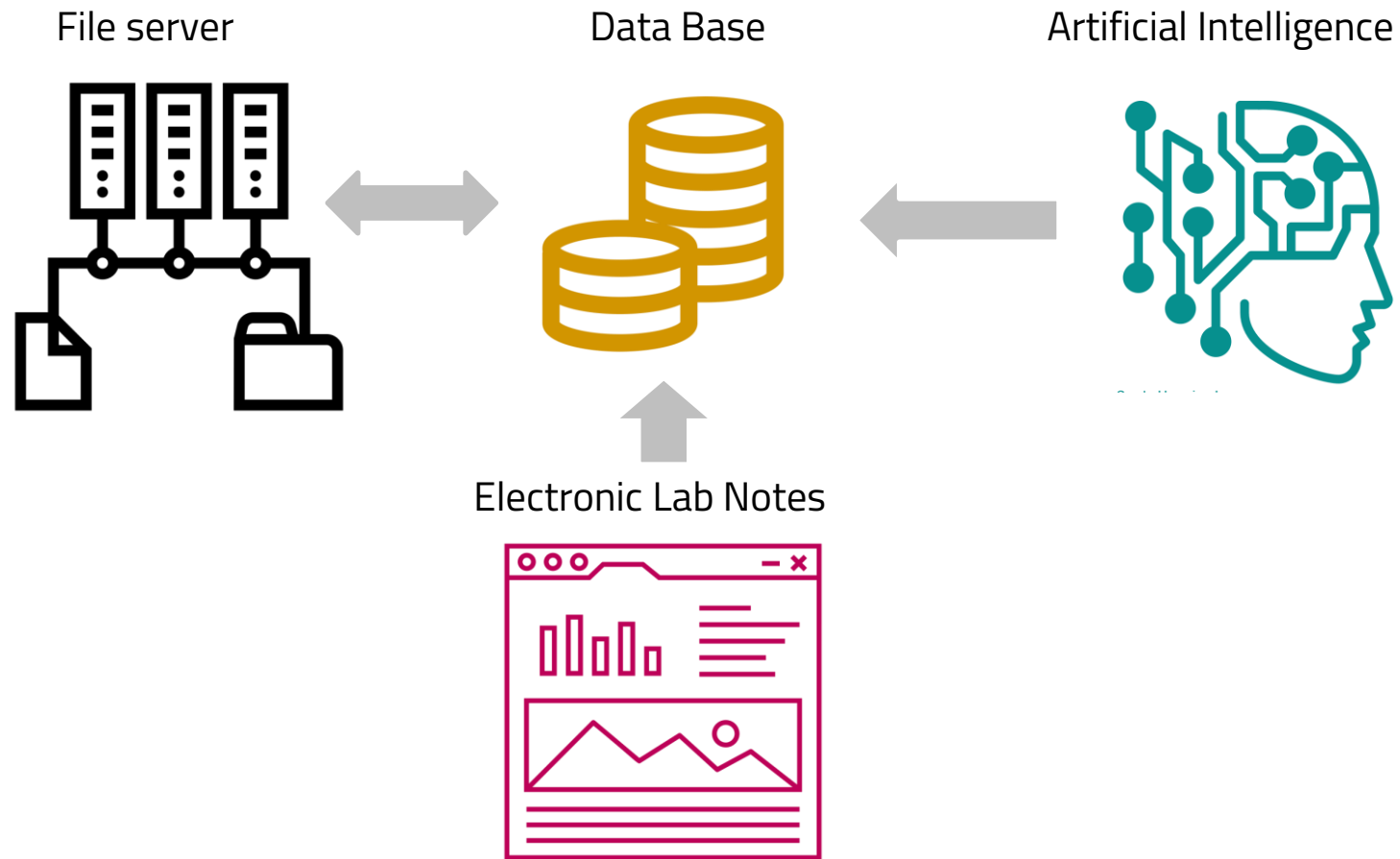
Feb. 8th  
1st. donor: 3½-day M7541.938.1♀ (L) x C223.878.4♂.  
Rt. ovary: 3 pale pink, 1 dark pink blob.  
Lt. ovary: 1 pale pink, 1 " " "  
Rt. side: 4 monilae in ju.  
Lt. side: 2 " " "  
6 eggs injected (AM) into the horn of multiparous 2½ day  
C377.914.1♀ (T) x C492.910.3♂. 3 pink  
blobs on ovary, very thin uterus.

D390  
M364.982.1♀ (L) x C502.960.6♂ (E) } 2½ days, left as unop. controls  
M558.971.1♀ x C490.863.5♂ (E) } unop. controls. D391

Feb. 9th  
M364.982.2♀ (L) x C502.960.6♂ (E) } 2½ days, left as unop. controls  
M20.330.13♀ (B) x C343.878.5♂ } unop. controls. D393  
M536.976.2♀ (L) x C502.960.5♂ } unop. controls. D394  
M373.859.12♀ (B) x C362.730.3♂ } unop. controls. D395

- Data are not accessible and findable
- Data are not well characterized
- Data are not linked to other data
- Data are not interoperable and reusable
- Data are not machine-readable.

# Research data management (RDM) with Electronic Lab Notebooks (ELN) in NOMAD

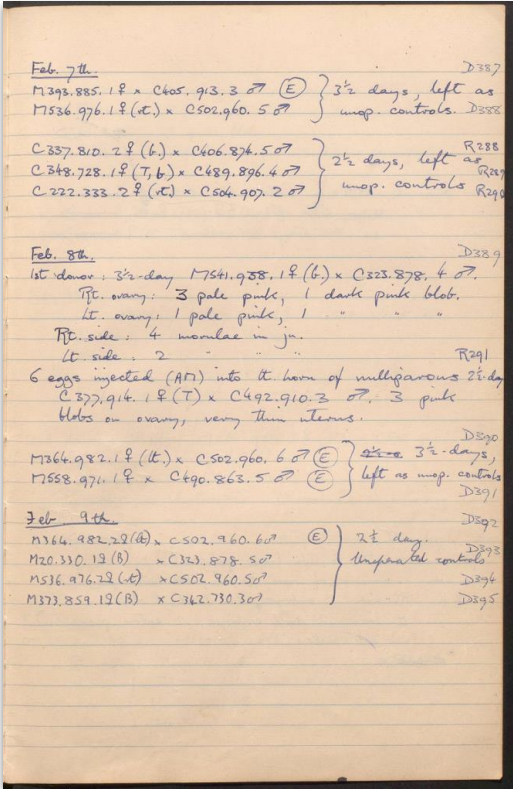


The ELN is an interface to a database

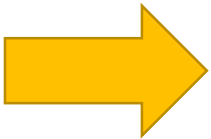
→ It is more than the simple electronic substitution of written lab notes!

ELN entries require a **structured** approach of collecting data and a metadata structure

# Electronic Lab Notebooks (ELN) in NOMAD: from analogue data to digitized data



Digitizing data



is not enough  
to make them  
machine-readable

Flow cytometry - OneNote

File Home Insert Draw History Review View

Sandra's eLabbook

Getting Started Project information Methods Experimental reports Outputs Deadlines

Search (Ctrl+E)

Methods referred to:

- SNAP-fusion protein fluorescent labeling
- Flow cytometry

1. Lifted & counted cells (diluted 10x):

U87		U251	
31	27	41	27
21	37	24	34

$= 3 \times 10^5/\text{ml}$   $= 3.2 \times 10^5/\text{ml}$

2. Aliquoted cells for 4 tests:

U251 blank U251 + SFP U87 -SFP U87 + SFP

3. Incubated O/N at 4°C.

4. Washed 3x with PBS.

5. Resuspend in 0.5ml PBS/DMEM in FACS tube.

6. Counted 20,000 cells per sample.

Outcomes / results:

Blank	U251 +SFP	U87 -SFP	U87 +SFP
UNSTAINED	U251	U87	U87

Incubate & wash with DMEM instead of PBS

Lab work report guidelines

July-Dec 2016

Jan-Feb 2017

March 2017

WB: purified CSPG4-SNAP/An

Flow cytometry: CSPG4-SNAP-Confocal & flow: CSPG4-SNAP

Cytotoxicity assay: CSPG4-Ang

Flow cytometry: CSPG4-SNAP

Flow cytometry

April 2017

May 2017

June 2017

July 2017

August 2017

September 2017

October 2017

Untitled page

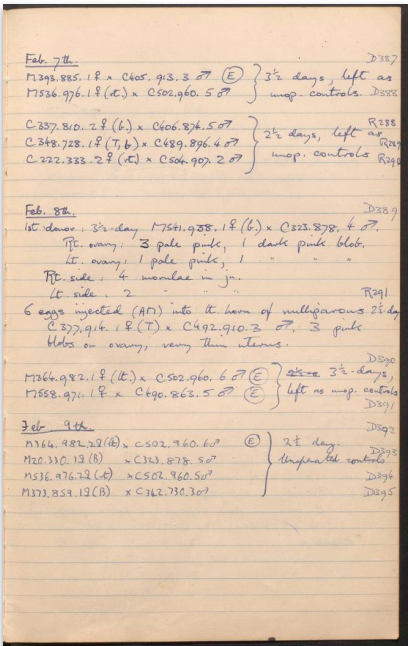
Dated lab work report

November 2017

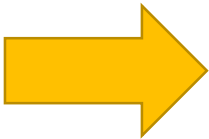
December 2017



# Electronic Lab Notebooks (ELN) in NOMAD: from analogue data to structured data



Structuring data



is required

**Metadatas**

type  
Substance  
name  
gallium arsenide

comment  
no comment

references  
authors  
Andrea Albino

datasets  
no datasets

mainfile  
gallium\_arsenide.archive.json

entry id  
EVVlovP1T7Y58Gr6UxVefcvl-N2x

upload id  
Ab4GufVbRTKd3onG8UpUhw

upload create time  
3/13/2023, 4:52:52 PM

last processing time  
3/13/2023, 4:59:23 PM

processing version  
1.1.8/

API

**Substance**

Substance name  
gallium arsenide

Safety reactivity

Datetime  
13/03/2023 16:59

Substance ID

State of matter

CAS uri  
substance/pt/1303000

CAS number  
1303-00-0

CAS name  
Gallium arsenide

image  
cas\_1303-00-0\_image.svg

Inchi  
InChI=1S/As.Ga

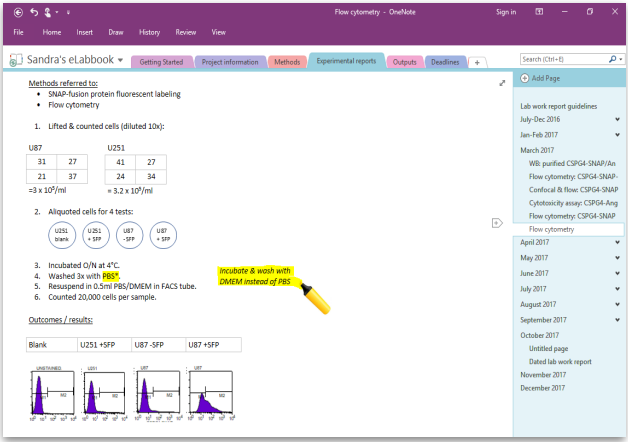
Inchi key  
InChIKey=JBR2TJJDHDCESZ-UHFFFAOYSA-N

Smile  
[As]#[Ga]

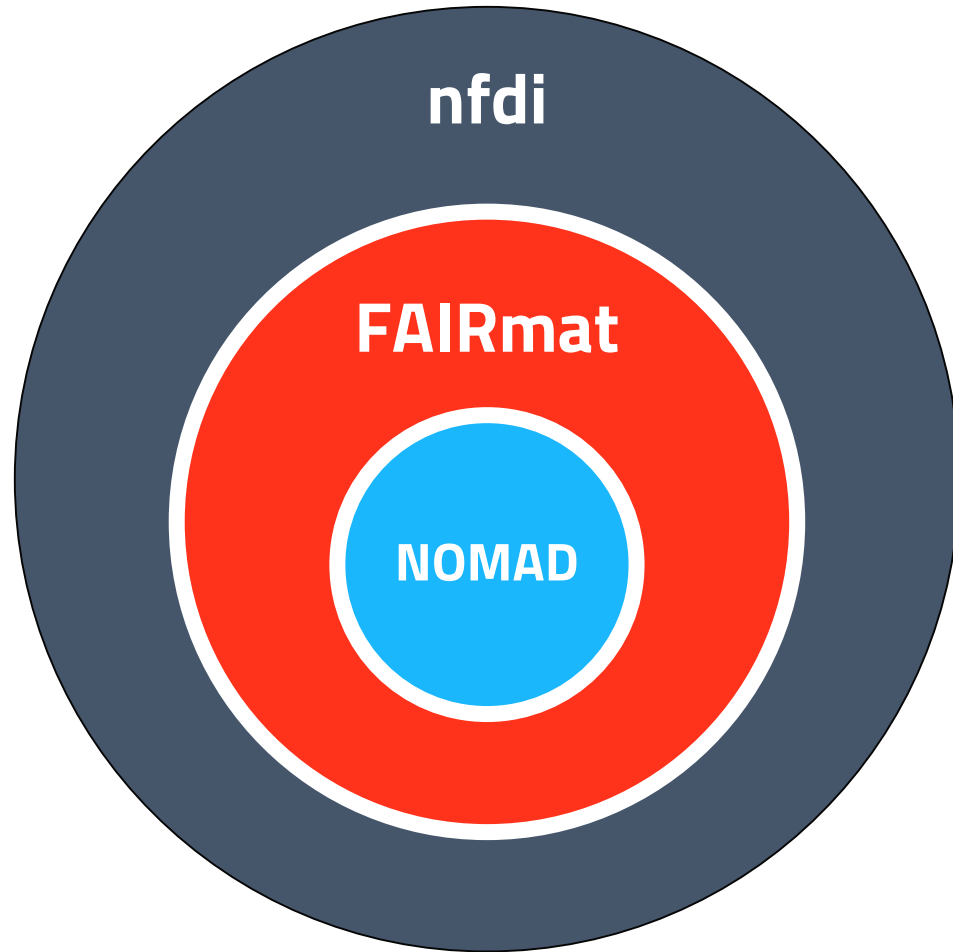
Canonical smile  
[Ga]#[As]

Molecular formula  
AsGa

Molecular mass  
Unit: u



# What are NFDI / FAIRmat / NOMAD



**nfdi:** Nationale Forschungsdaten Infrastructure, [link](#)  
(national research data infrastructure)

**FAIRmat:** NFDI consortium for FAIR materials science data, [link](#)  
(FAIR: findable, accessible, interoperable, re-usable)

**NOMAD:** A web-based service and software for managing FAIR materials science data, [link](#)  
FAIRmat uses NOMAD to build a federated infrastructure of connected NOMAD installations

# The FAIRmat Values

## **FAIR**

*Findable, Accessible, Interoperable, Re-usable*

FAIR principles can transform every field of science

## **Bottom-up approach**

*Embracing the community*

Development driven by the needs of scientists and already enjoys strong support from the community.

## **Open access**

*Use open processes to support a wide community*

FAIRmat advocates for an urgently needed culture shift towards data sharing, and stands for open access to scientific materials data and tools.



# FAIRmat Approach

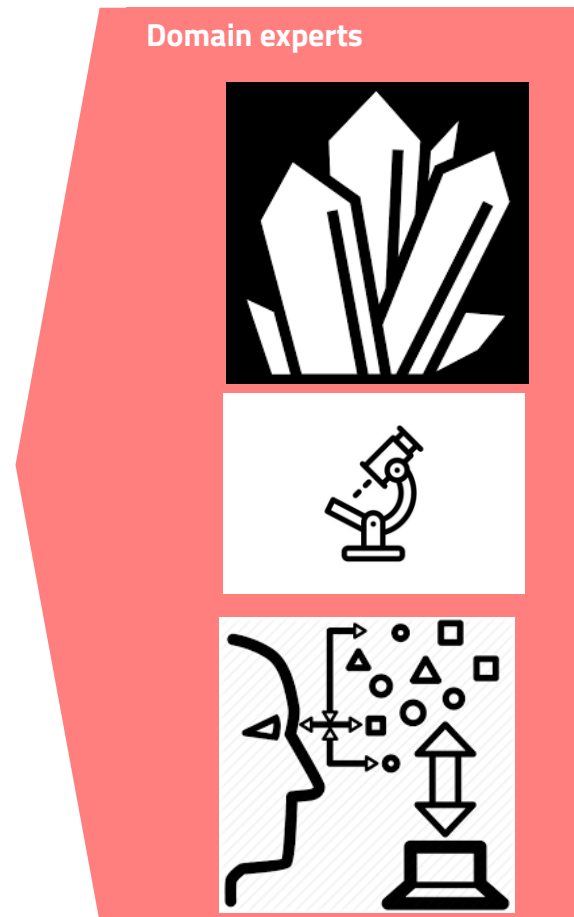
## Local Domain experts

→ define domain specific needs for data management



### Data scientists & Data stewards

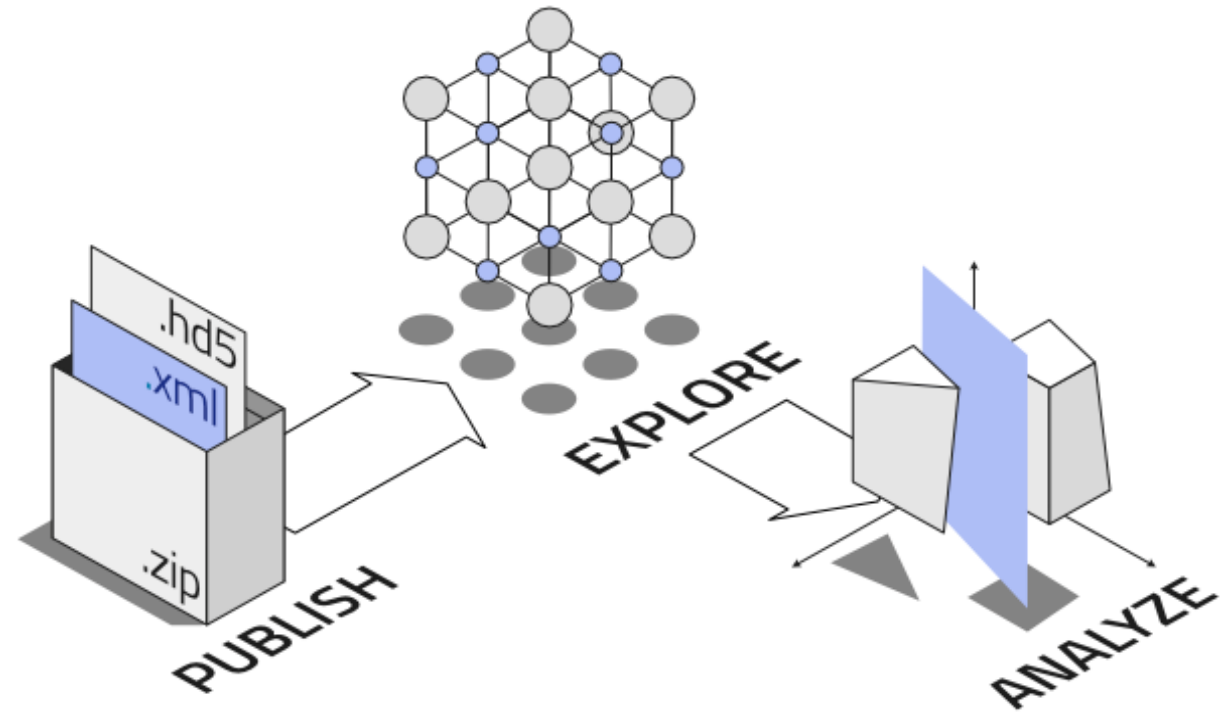
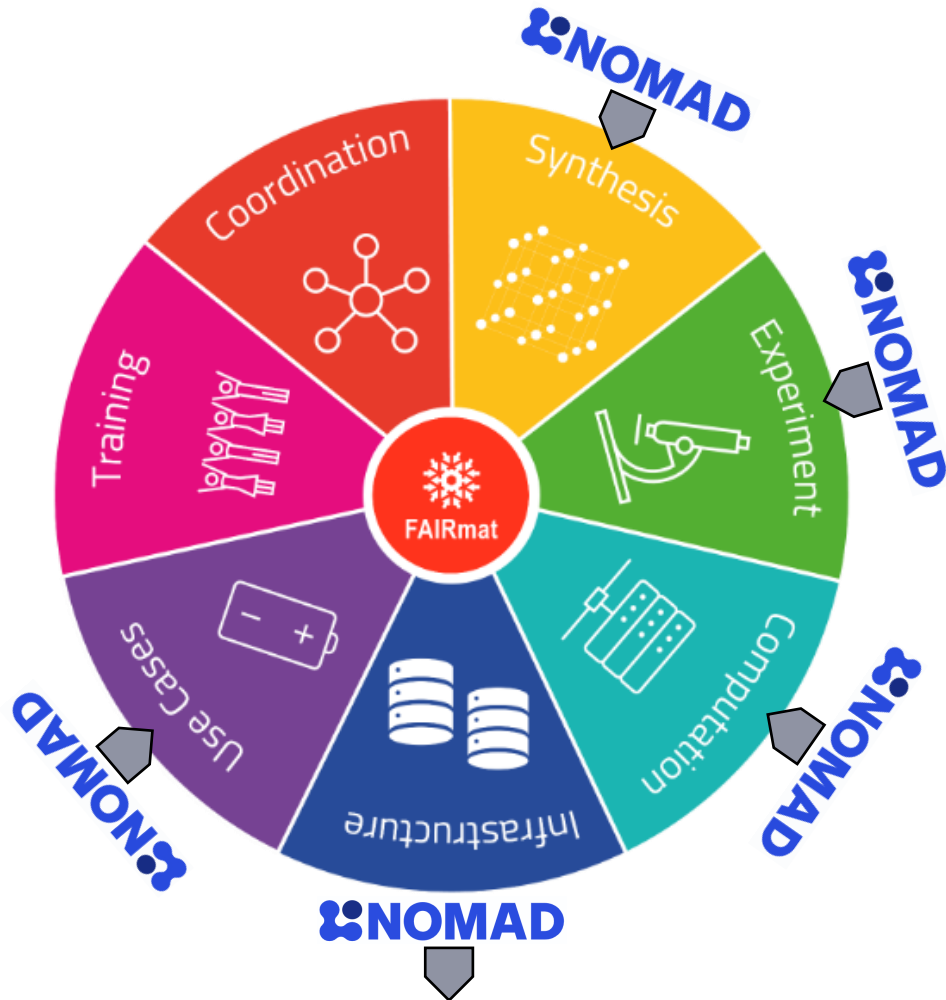
→ experts in data, metadata ontologies



# What are NFDI / FAIRmat / NOMAD

**FAIRmat** is the NFDI consortium to build a FAIR federated data infrastructure for solid state physics

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



**NOMAD**

# NOMAD webpage



SOLUTIONS ▾

LEARN ▾

GET INVOLVED ▾

ABOUT ▾

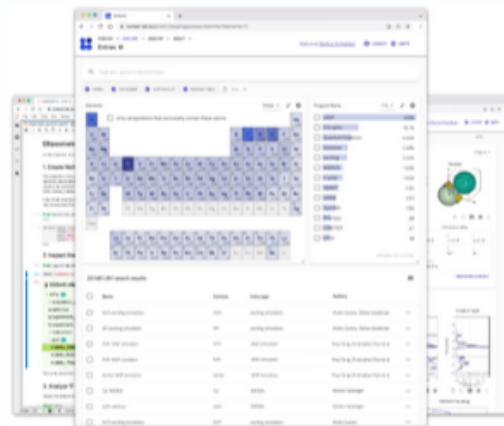
OPEN NOMAD

## 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.

Open NOMAD →



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

USED BY THOUSANDS OF MATERIALS SCIENTISTS

UPLOADED ENTRIES  
**12,460,881**

REPRESENTED MATERIALS  
**2,976,441**

UPLOADED FILES  
**108.5 TB**



## Drag & Drop

Upload file-by-file or zip  
and upload whole directory structures.



## Get a DOI

NOMAD allows you to publish and archive  
your data for free. Assign a DOI to uploaded  
datasets and reference your papers.



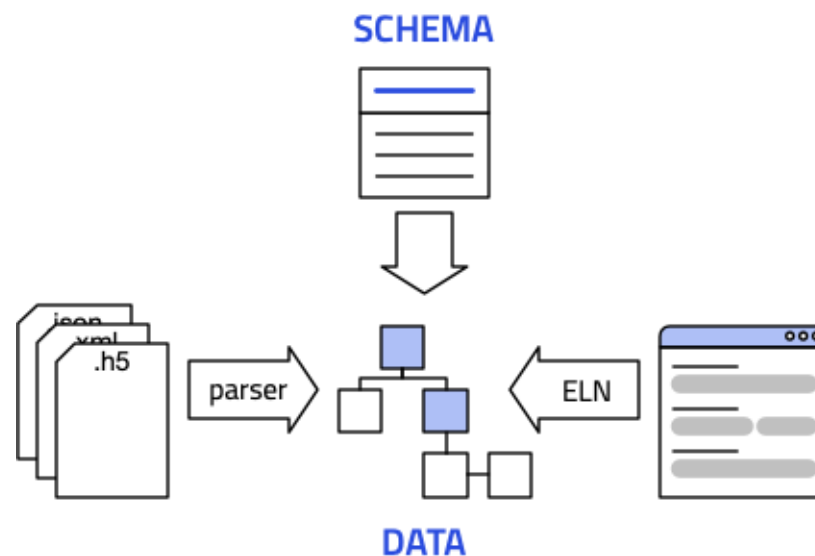
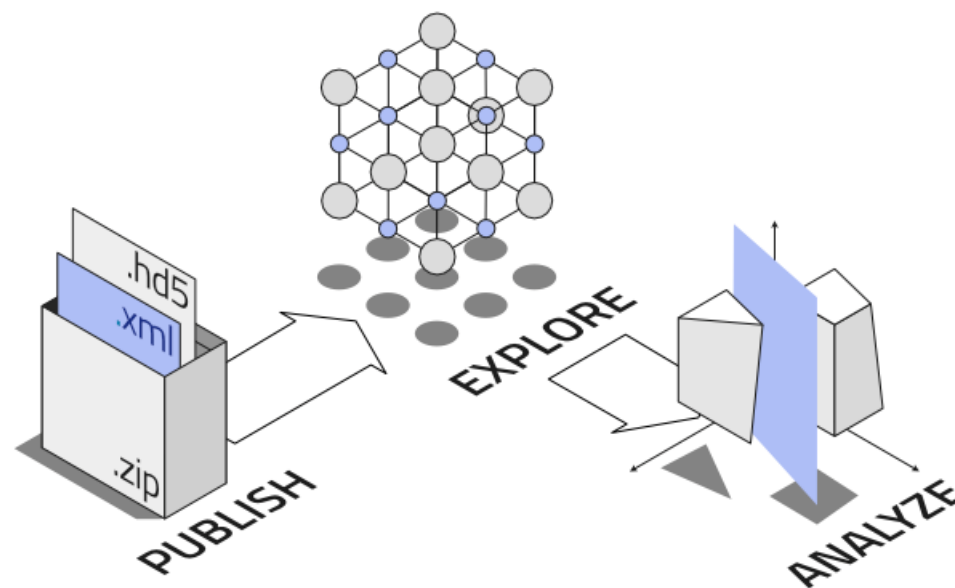
## You Control Your Data

Not ready to publish? Organize data and  
collaborate in private first.



## Ready to use Tools

Run Jupyter and other tools directly on NOMAD.





## Drag & Drop

Upload file-by-file or zip  
and upload whole directory structures.



## Get a DOI

NOMAD allows you to publish and archive  
your data for free. Assign a DOI to uploaded  
datasets and reference your papers.



## You Control Your Data

Not ready to publish? Organize data and  
collaborate in private first.



## Ready to use Tools

Run Jupyter and other tools directly on NOMAD.



## Runs on your premises

Runs behind your firewall and inside your  
VPN. Use your own resources for processing  
and running analysis tools.



## Custom ELMs

Extend and customize NOMAD's schema to  
create specialized editors to document your work.



## Your own parsers

Support your own file formats. Add parsers  
and normalization routines.



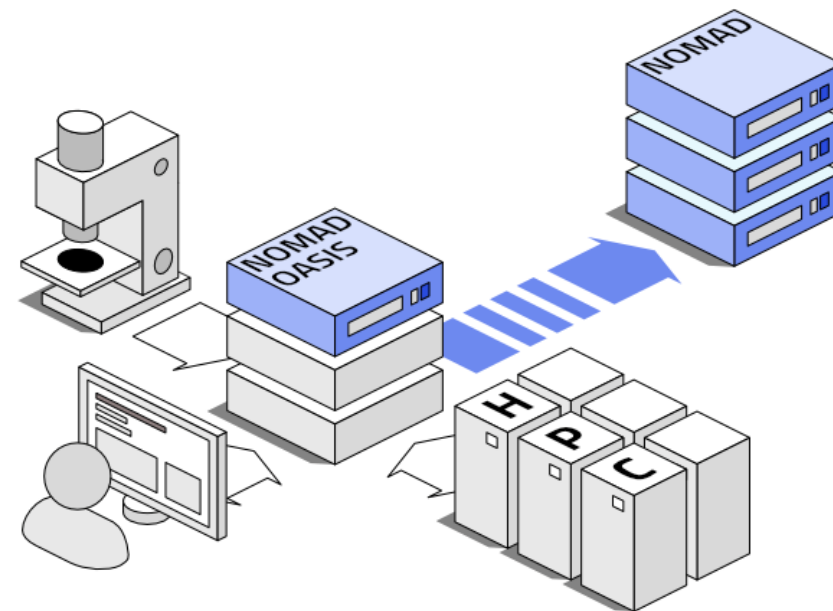
## Use your own tools

Add tools and notebooks that directly access and  
manipulate your data in NOMAD.

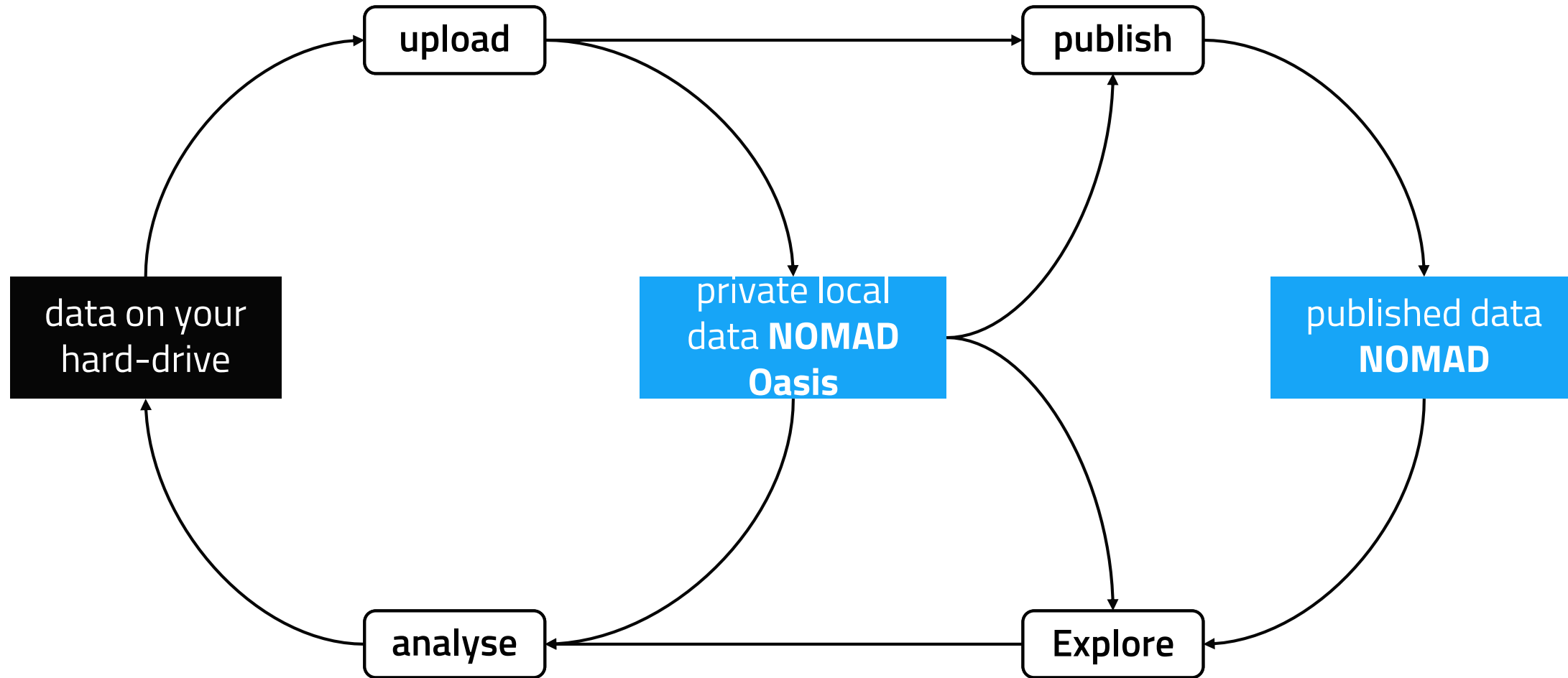


## Publish selected data

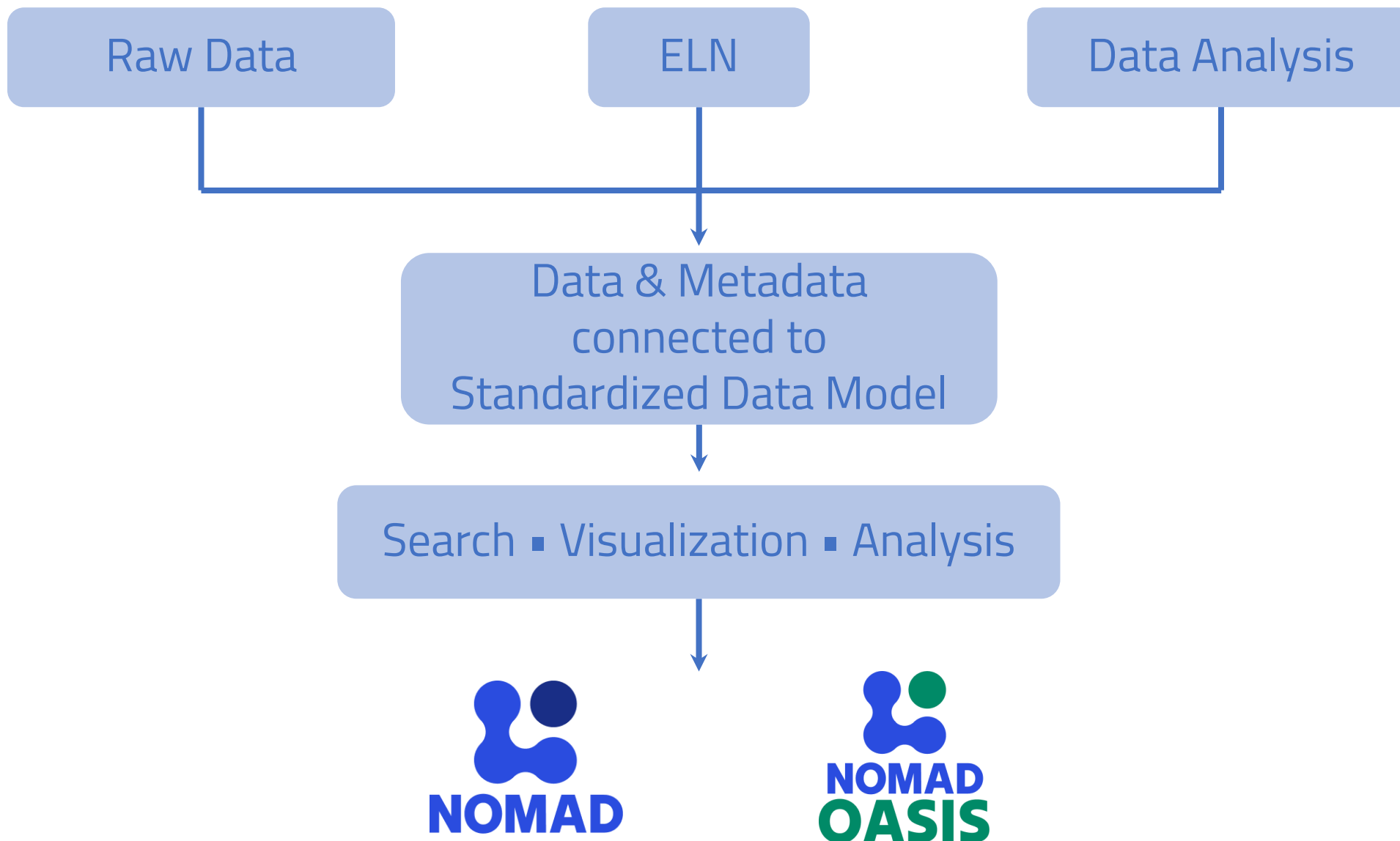
Oasis will soon be connected to the public NOMAD service.



# What is NOMAD Oasis?



# Research Data Management in NOMAD



# Area A Team

Sebastian Brückner



Coordinator

Andrea Albino



Data Modelling  
Implementation

Hampus Näsström



Data Modelling

Jose Marquez-Prieto



Implementation  
Link to **Area E**

Florian Dobener



Infrastructure  
Link to **Area B**



# Tutorial Outline

Time	Session	Speaker
14:00	1. Introduction	Sebastian Brückner
14:10	2. Vocabulary and Key Concepts	Andrea Albino
14:30	3. NOMAD usage (interactive)	Florian Dobener
15:00	4. Writing a Custom Schema (interactive)	Hampus Näsström
15:30	5. Reading Files with the Tabular Parser and Adding Plots (interactive)	Andrea Albino
16:00	6. Using Base Classes and References (interactive)	Hampus Näsström
16:30	7. Search your ELN data (interactive)	José Marquez





Thank you!