

## FAIRmat Tutorial 8: Using NOMAD as an Electronic lab notebook (ELN) for FAIR data

Organized by FAIRmat Area A Synthesis



# Introduction

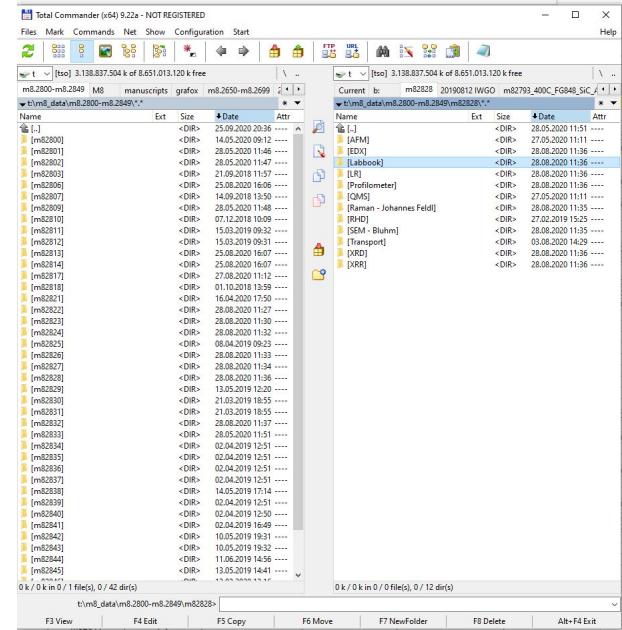
# Motivation

- Data are not accessible and findable
- Data are not well characterized
- Data are not interoperable and reusable.

## Spreadsheet Files

B4	AMT_0595												
1	batch	AMTwin-label	d <sub>layer</sub>	p <sub>L</sub>	t <sub>L</sub>	d <sub>L</sub>	d <sub>hatch</sub>	plane	porosity	density [%]	φ-density [%]	number of voids	max. pore size [mm]
2			μm	W	μs	μm	μm						
4	1	AMT_0595	30	200	65	55	105	x-y	0,00%	100,00%		12	0,04
5								x-z	0,01%	99,99%	100,00%	11	0,07
6	6	AMT_1140	25	200			100	x-y	0,00%	100,00%		2	0,01
7								x-z	0,00%	100,00%	100,00%	8	0,04
8													
9													
10	Batch 1 AMT_0595												
11	30 μm layer thickness / 200 W / 850 mm/s												
12	stress relieved (550 °C / 180 min / Ar / furnace cooling)												
13													
14	Batch 6 AMT_1140												
15	25 μm layer thickness / 200 W / 1200 mm/s												
16	HIP (920 °C, 1000 bar / 120 min / Ar) - 25 μm												
17													
18													
ENG_Porenanalyse_C1_C6			Porenanalyse_C1_C6										

## File Servers



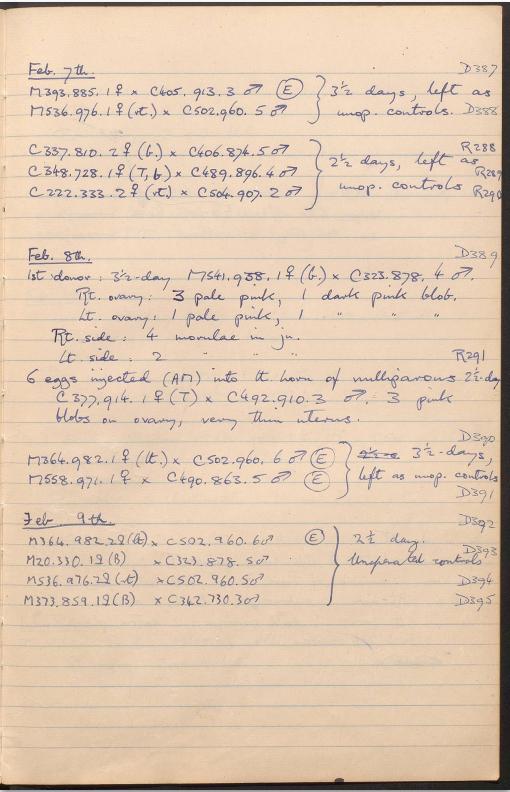
## Lab Notes

Feb. 7th. D387  
M 393.885.1♀ x C605.913.3♂ { } 3½ days, left as unop. controls. D388  
M 7536.976.1♀(t) x C502.960.5♂ { } unop. controls. D389  
C 337.810.2♀(b) x C406.874.5♂ { } 2½ days, left as R388  
C 348.728.1♀(T,b) x C489.896.4♂ { } unop. controls R390  
C 222.333.2♀(vt) x C504.907.2♂ { } unop. controls R390

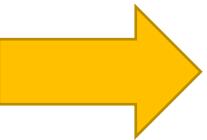
Feb. 8th. D389  
1st donor: 3½-day M 7641.958.1♀(b) x C323.878.4♂.  
Rt. ovary: 3 pale pink, 1 dark pink blob.  
Lt. ovary: 1 pale pink, 1 " "  
Rt. side: 4 monilae in ju.  
Lt. side: 2 " "  
6 eggs injected (AN) into lt. horn of nulliparous 2½-day C 377.914.1♀(T) x C492.910.3♂, 3 pink blobs on ovary, very thin uterus.

M 366.982.1♀(lt) x C502.960.6♂ { } 3½ days, D390  
M 20.330.1♀(B) x C32.878.5♂ { } left as unop. controls D391  
M 336.976.2♀(t) x C502.960.5♂ { } unop. controls D392  
M 373.859.1♀(B) x C342.730.3♂ { } D393

Feb. 9th. D392  
M 364.982.2♀(A) x C502.960.6♂ { } 2½ day. D393  
M 20.330.1♀(B) x C32.878.5♂ { } unop. controls D394  
M 336.976.2♀(t) x C502.960.5♂ { } D395  
M 373.859.1♀(B) x C342.730.3♂ { } D395



## Digitizing data



Flow cytometry - OneNote

Sandra's eLabbook ▾ Getting Started Project information Methods Experimental reports Outputs Deadlines +

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

= $3 \times 10^5$ /ml      = $3.2 \times 10^5$ /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.

*Incubate & wash with DMEM instead of PBS*

Outcomes / results:

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

UNSTAINED      U251      U87      U87

Search (Ctrl+E)

+ Add Page

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

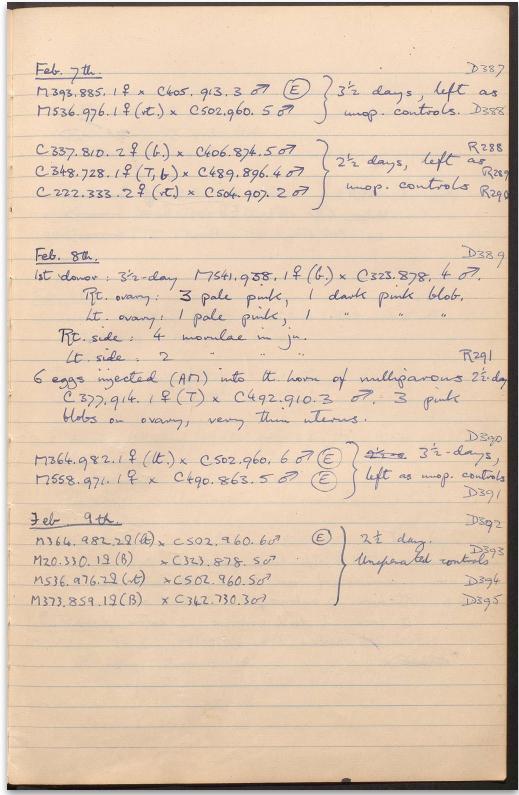
October 2017

Untitled page

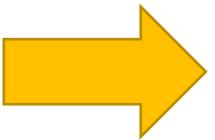
Dated lab work report

November 2017

December 2017



## Digitizing data



is not enough  
to make them  
machine-readable

Flow cytometry - OneNote

Sandra's eLabbook ▾ Getting Started Project information Methods Experimental reports Outputs Deadlines +

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

= $3 \times 10^5$ /ml      = $3.2 \times 10^5$ /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.

*Incubate & wash with DMEM instead of PBS*

Outcomes / results:

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

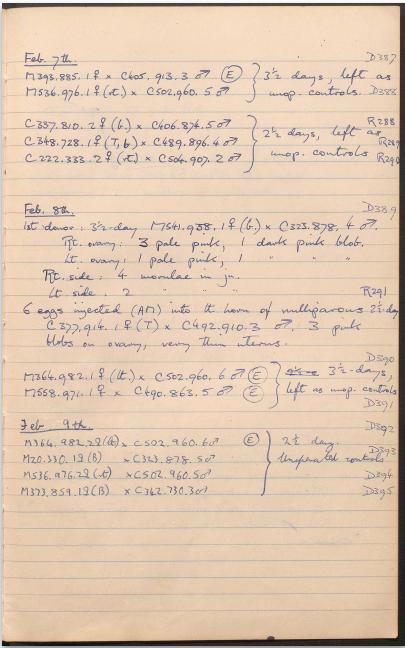
UNSTAINED      U251      U87      U87

Search (Ctrl+E)

+ Add Page

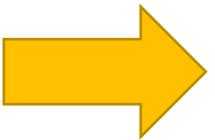
Lab work report guidelines  
 July-Dec 2016  
 Jan-Feb 2017  
 March 2017  
 WB: purified CSPG4-SNAP/An  
 Flow cytometry: CSPG4-SNAP  
 Confocal & flow: CSPG4-SNAP  
 Cytotoxicity assay: CSPG4-Ang  
 Flow cytometry: CSPG4-SNAP  
 Flow cytometry

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

Methods referred to:

- SNAP-fusion protein fluorescent labeling
- Flow cytometry

- Lifted & counted cells (diluted 10x):

U87	U251
31	27
21	37
21	24
32	34

$\times 10^6$  ml =  $3.2 \times 10^6$  ml

- Alligated cells for 4 tests:

U87	U251	U87	U251
+SFP	+SFP	-SFP	-SFP

- Incubated O/N at 4°C.
- Washed 3x with **F65**.
- Resuspended in 0.5ml PBS/DMEM in FACS tube.
- Counted 20,000 cells per sample.

Outcomes / results:

Blank U251 +SFP U87 -SFP U87 +SFP

Flow cytometry plots showing cell counts versus fluorescence intensity for each sample.

PUBLISH EXPLORE ANALYZE ABOUT

Your uploads / Upload / Entry

LOGOUT UNITS

OVERVIEW FILES DATA LOGS

Metadata

type Substance name gallium arsenide

comment no comment

references authors Andrea Albino

datasets no datasets

mainfile gallium\_arsenide.archive.json

entry id EVV1ovP1T7Y58Gr6UxVefcvl-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=JBRZTFJDHDCESZ-UHFFFAOYSA-N

Smile [As]#[Ga]

Canonical smile [Ga]#[As]

Molecular formula AsGa

Molecular mass Unit U

# Data-driven Benefits to the Research

## 1. By policy



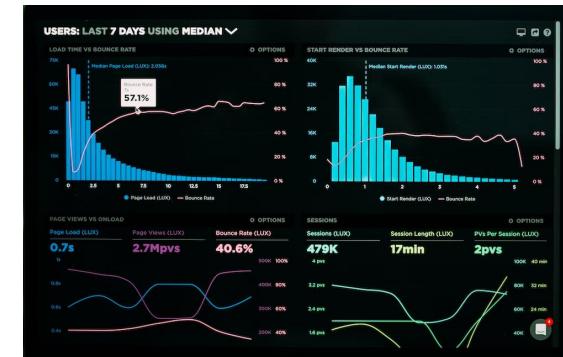
## 2. Reproducibility



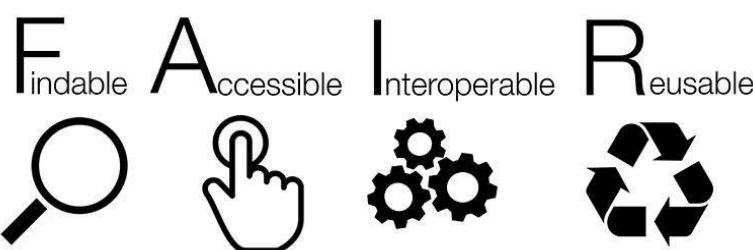
## 3. Knowledge retention

A photograph of a handwritten page from a lab notebook. The page contains various data tables and handwritten notes. At the top, it says 'Stonenhöhe: 34 mm'. Below that is a table with columns for 'Bearbeiter/-in', 'Soll WIR', 'Versuch Nr.', and '984'. Another table below shows 'Material: Wacker' with 'AusgangsØ [mm]' and 'Soll Ø [mm]' both listed as '103'. A third table details 'Dotierung' with values like '100 [ppm]' and 'Widerstand [Ωcm]'. Further down, there's a section for 'Induktor' with '42 DIFF.2 - 19' and 'SG Ar 15 [min] + N2 33 [%]'. The bottom part of the page is filled with handwritten text and calculations related to an experiment.

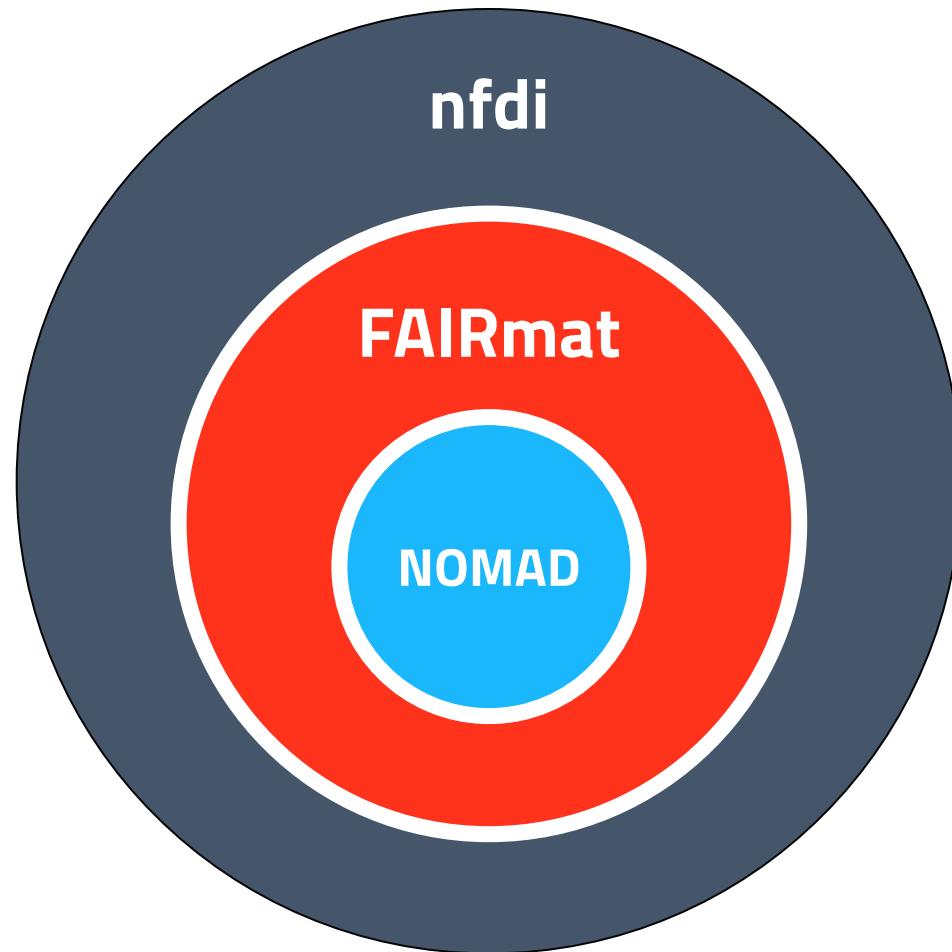
## 4. Automation, access to ML/AI tools



## 5. Prepare for the future:



# What are NFDI / FAIRmat / NOMAD



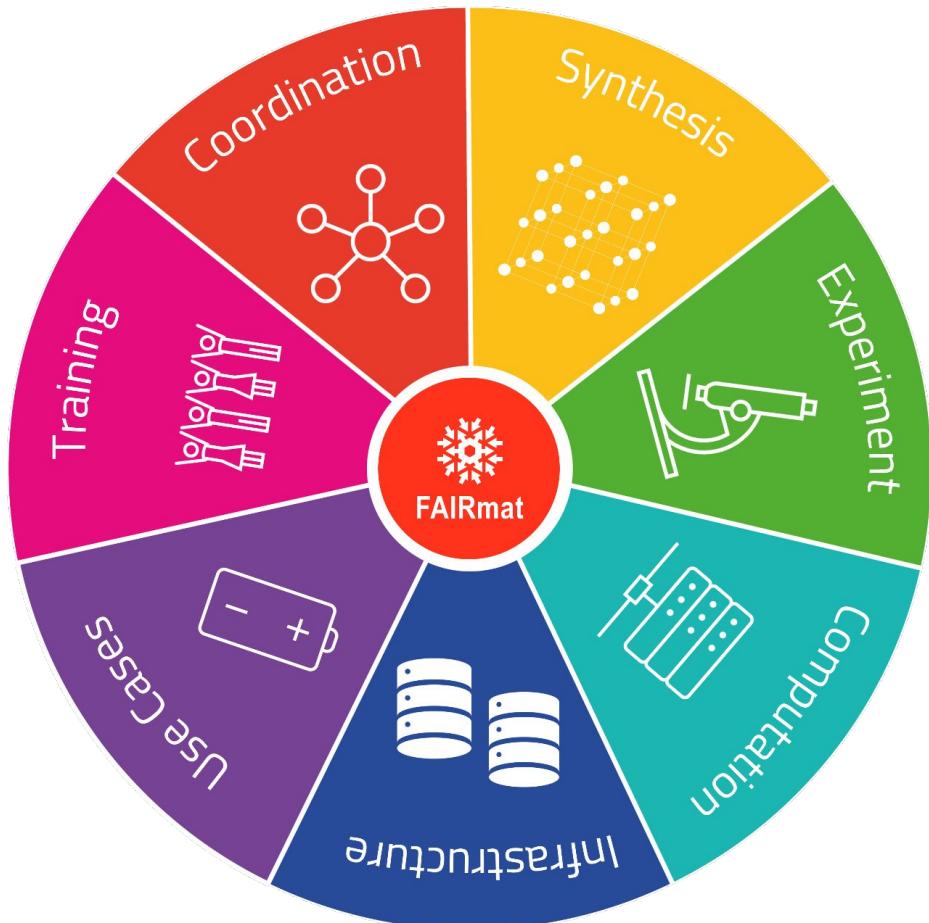
**nfdi:** Nationale Forschungsdaten Infrastructure, [link](#)  
(national research data infrastructure)

**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

# FAIRmat Project

Areas & Goals:



**A:** reproducible growth of materials from various synthesis routes

**B:** characterization methods used by the materials science communities

**C:** from voluminous classical simulations to highly sophisticated quantum-mechanical many-body techniques

**D:** federated data infrastructure development

**E:** demonstration that our tools can enhance science in daily life

**F:** outreach and community support

**G:** coordination and administration

# 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

Strategy

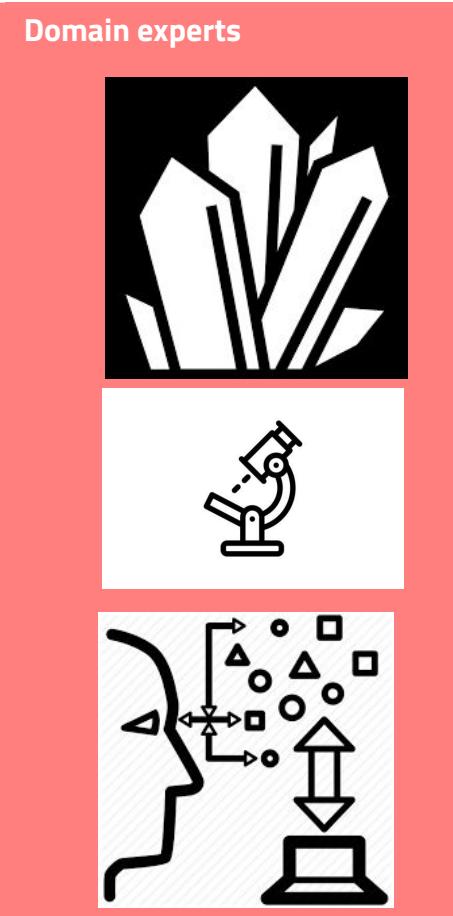


**Data scientists & Data stewards**  
→ experts in data, metadata ontologies



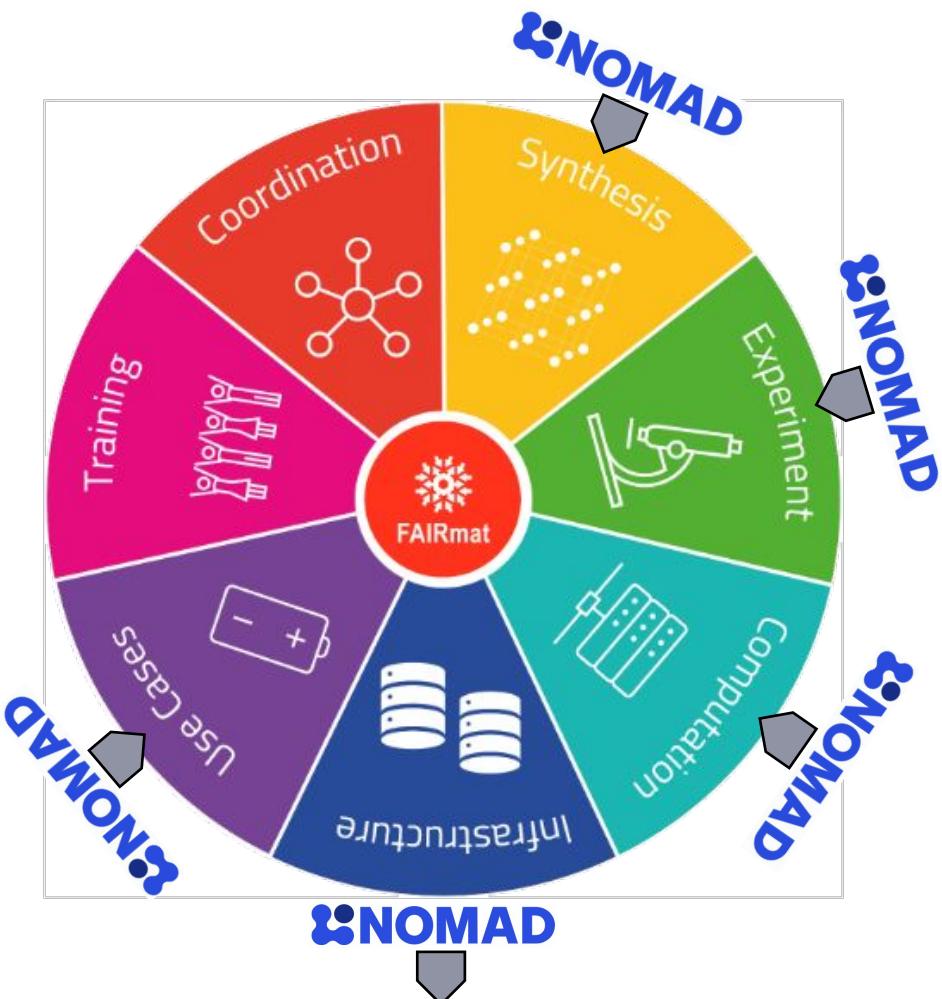
**Local Domain experts**

→ define domain specific needs for data management

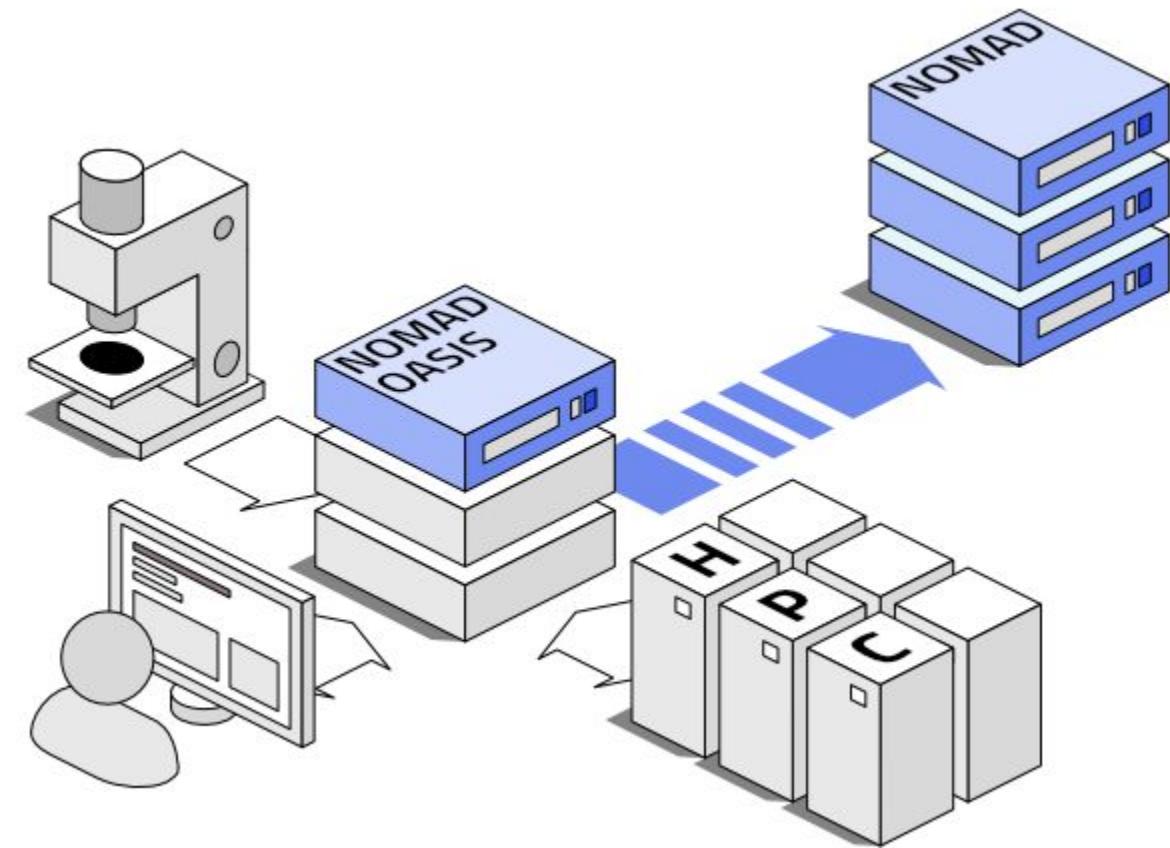


# What are NFDI / FAIRmat / NOMAD

**FAIRmat** is the NDFI 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**



## 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 ELNs

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.

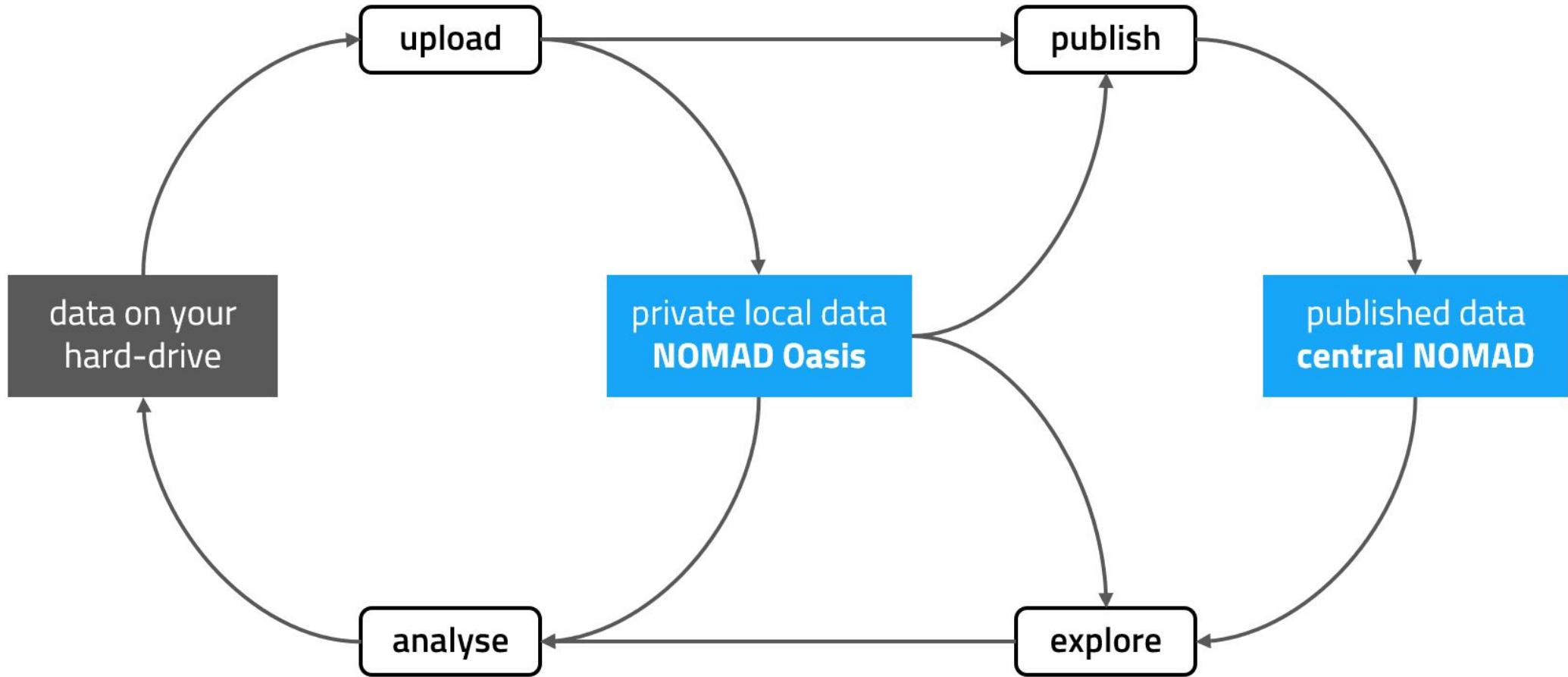


## Publish selected data

Oasis will soon be connected to the public NOMAD service.

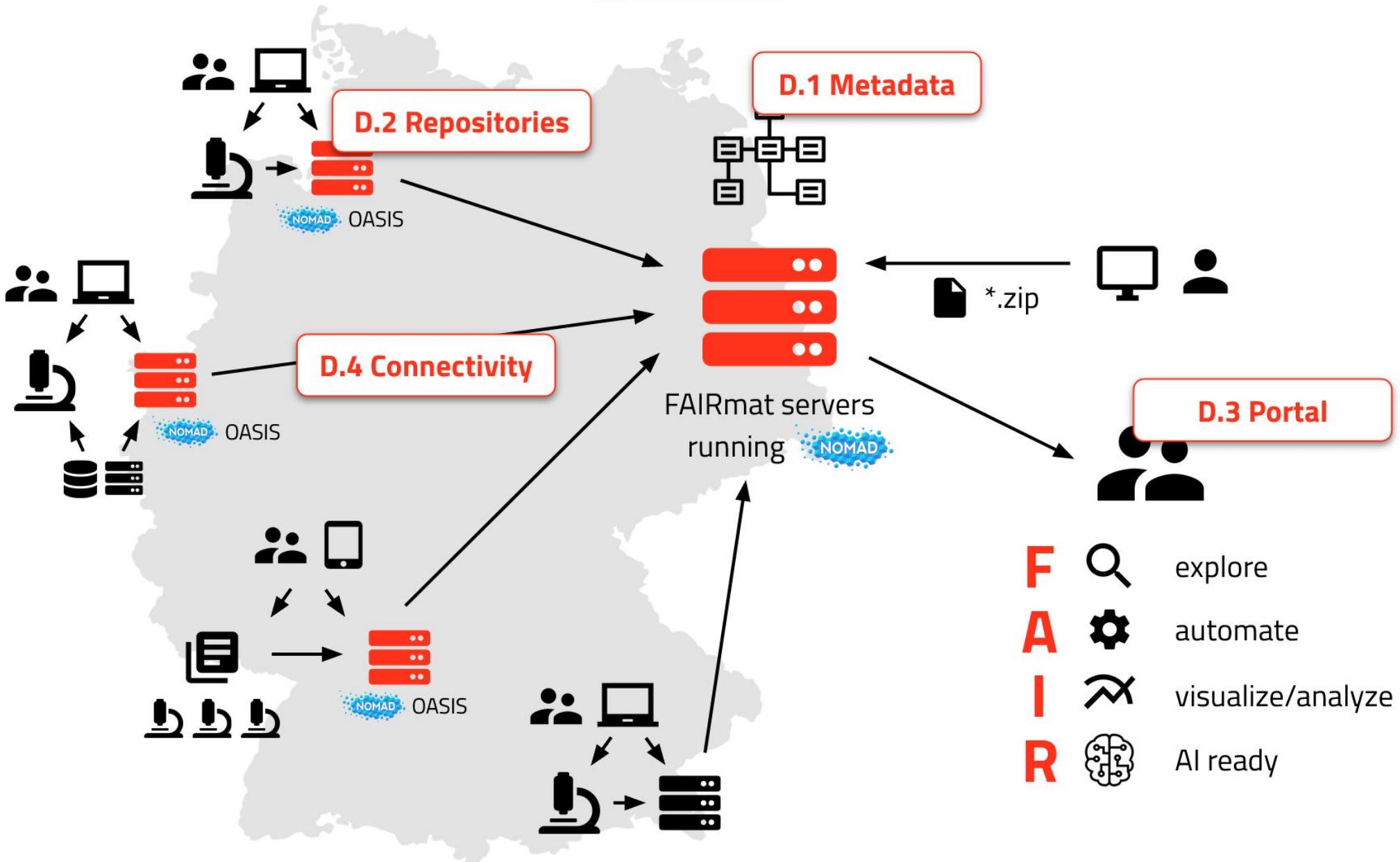


# What is NOMAD Oasis?

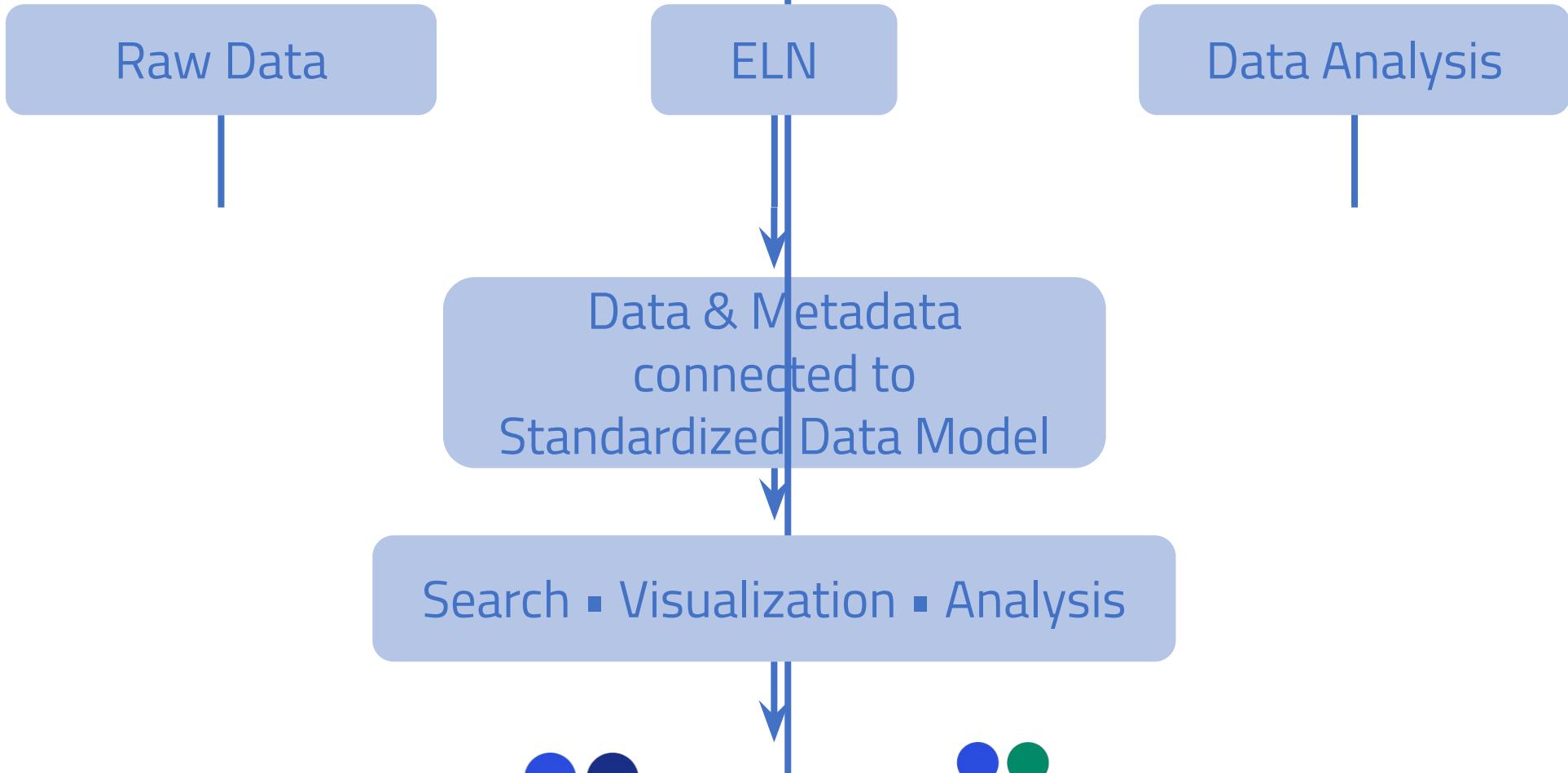




# What is NOMAD Oasis?



# Research Data Management in NOMAD





# Challenges & Goals

Worldwide, synthesis recipes are collected for the personal use of scientists, often documented in handwritten lab notebooks. Log files created by synthesis instruments are often not kept.

**Goal 1:** Establish **metadata** (standards), ontologies, and tools

**Goal 2:** Harmonize metadata **schemas** of synthesis and experimental characterization

**Goal 3:** Towards **computer-aided** development of synthesis recipes - interweaving experiment & theory

# 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

1. Introduction to FAIRmat's Area A  
and NOMAD

Sebastian Brückner

2. Vocabulary and Key Concepts

Andrea Albino

3. NOMAD usage

Florian Dobener

4. Writing a Custom Schema

Hampus Näsström

5. Reading Files with the Tabular Parser  
and Adding Plots

Andrea Albino

6. Using Base Classes and References

Hampus Näsström

7. Search your ELN data

José Marquez



Thank you!