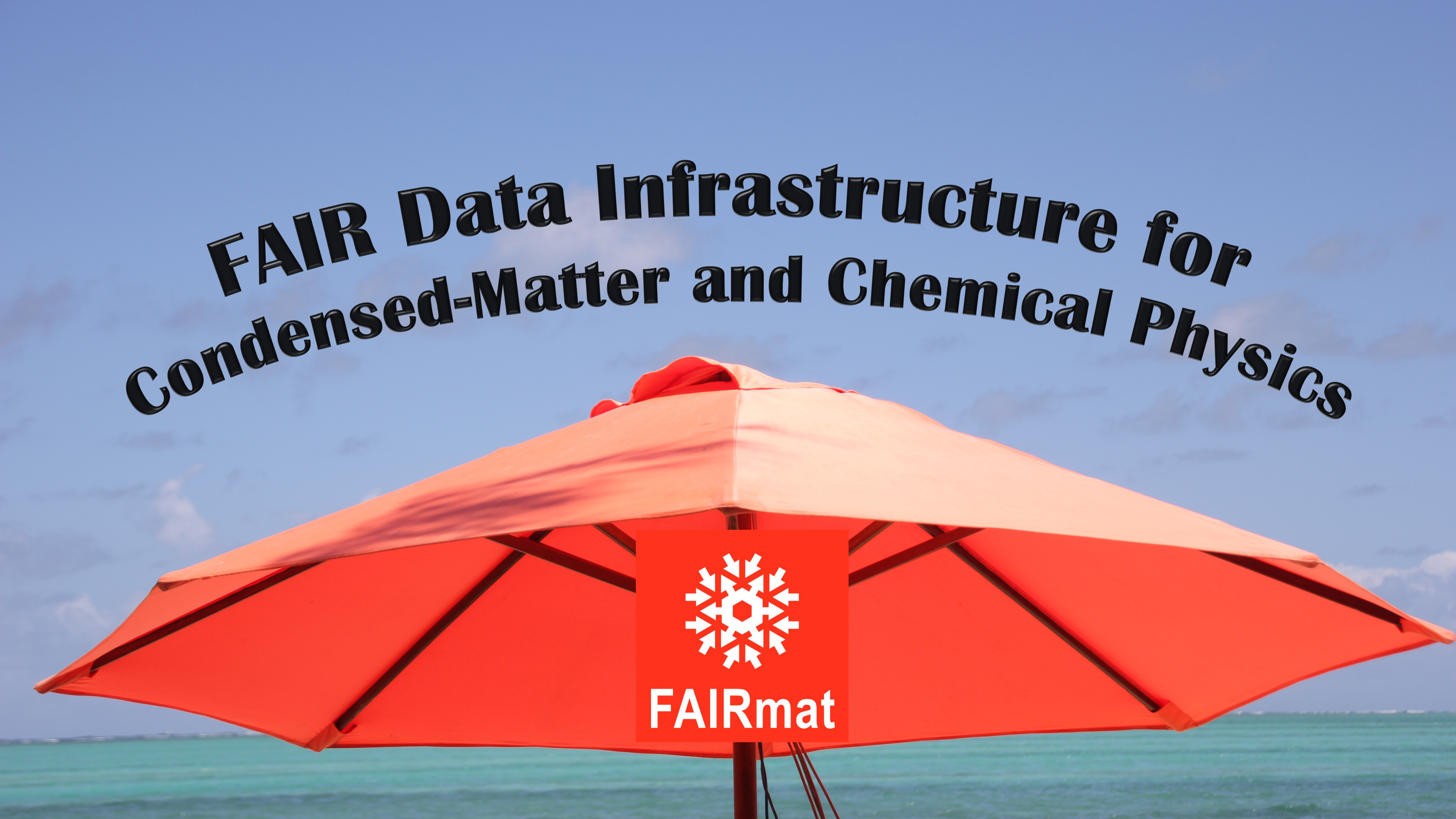
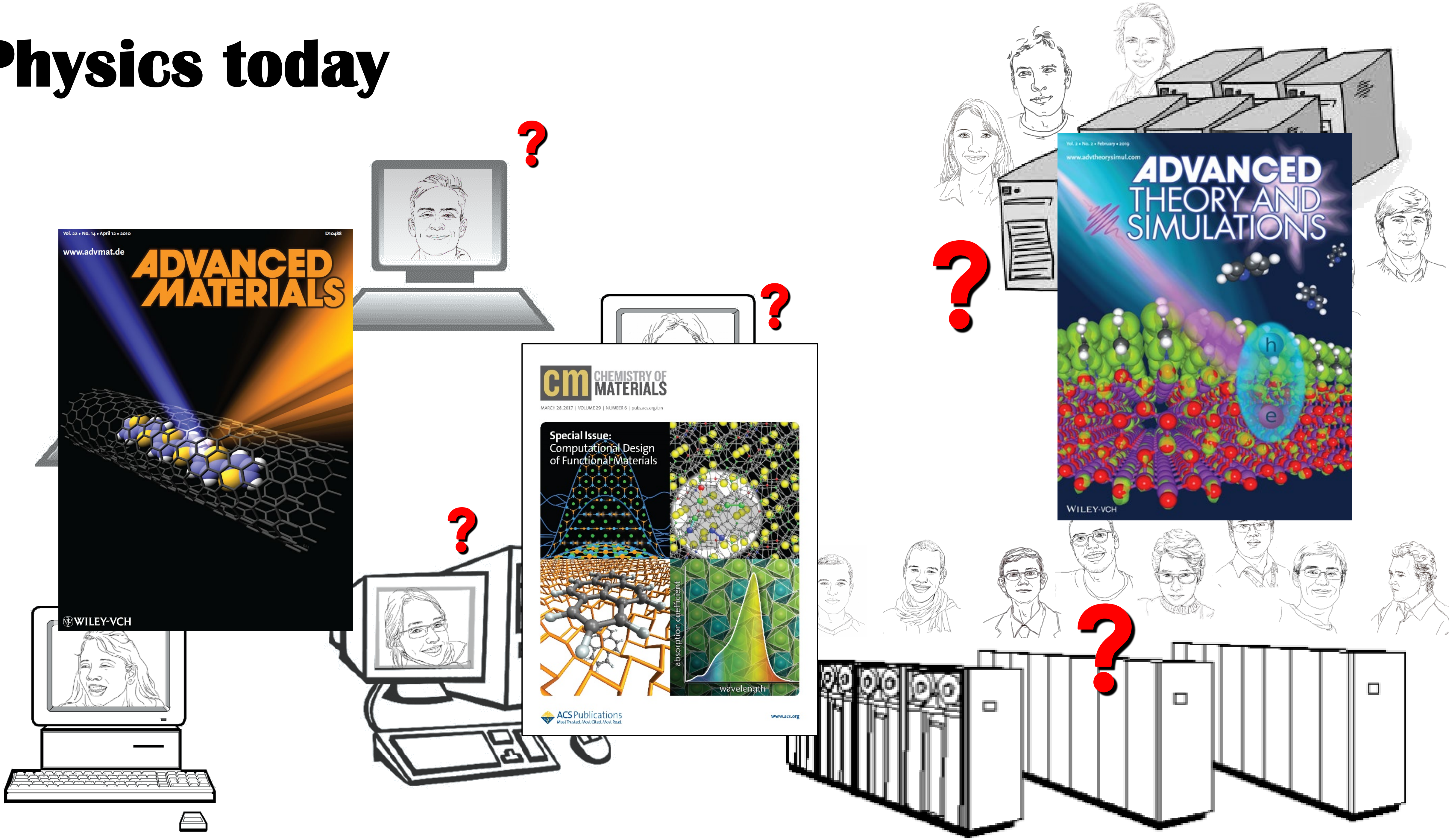

FAIR Data Infrastructure for
Condensed-Matter and Chemical Physics
FAIRmat

FAIRmat

A proposed consortium for the **German** program NFDI (National Research Data Infrastructure)

International in nature

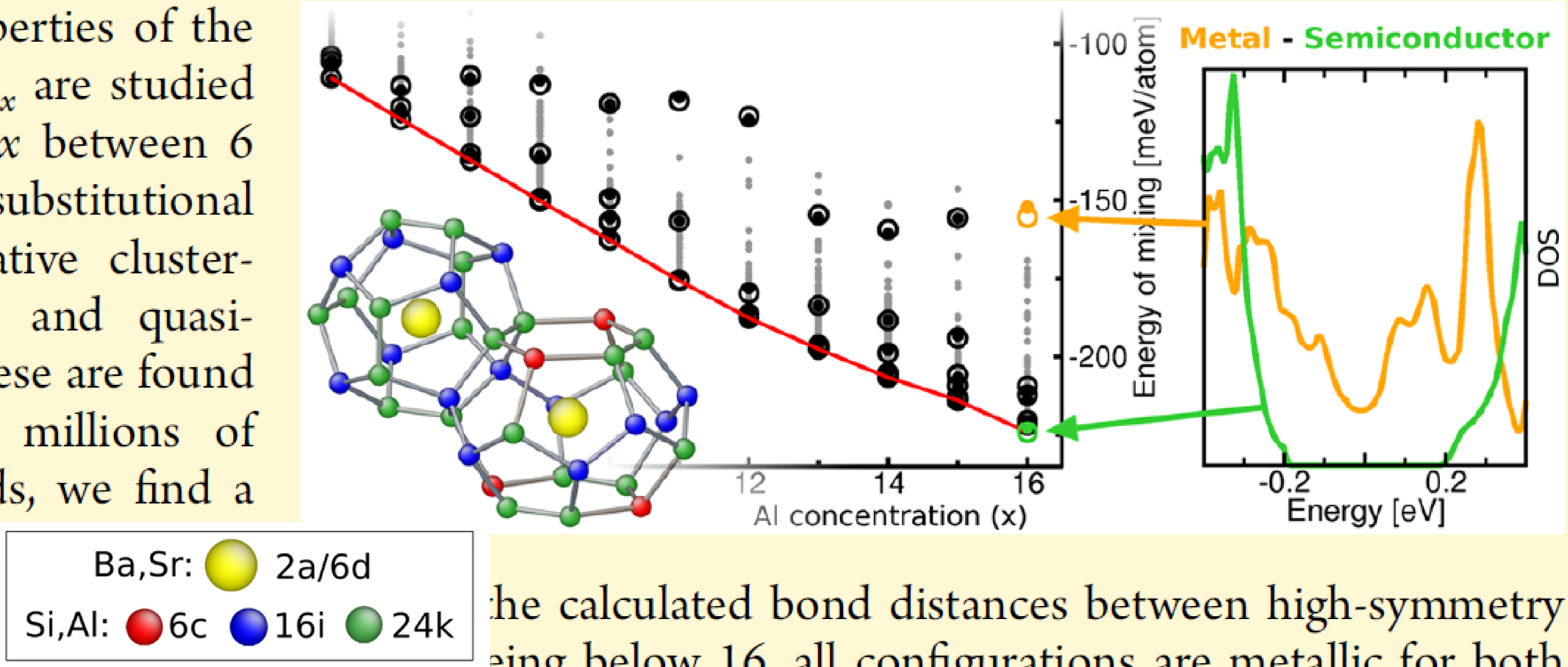
Physics today



Let us zoom in ...

M. Troppenz, S. Rigamonti, and C. Draxl, Chem. Mater. 29, 2414 (2017).

perties of the
are studied
x between 6
substitutional
ative cluster-
s and quasi-
ese are found
millions of
ds, we find a



the calculated bond distances between high-symmetry
...eing below 16, all configurations are metallic for both
... = 16), the substitutional ordering leads to a metal–semiconductor transition

What do we find in publications?

A table?

A figure?

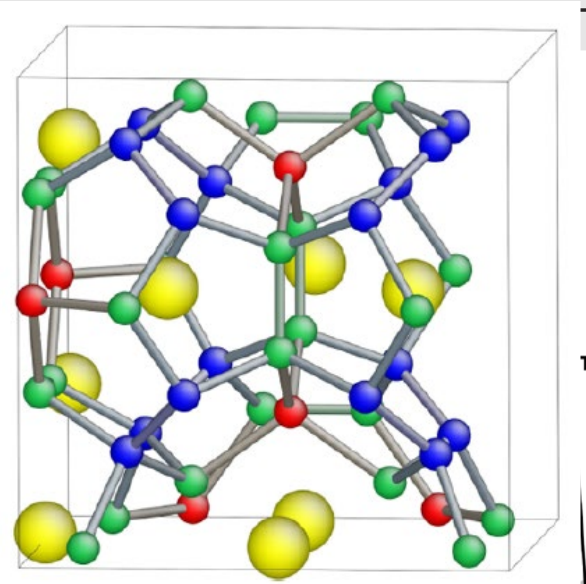
A structure?

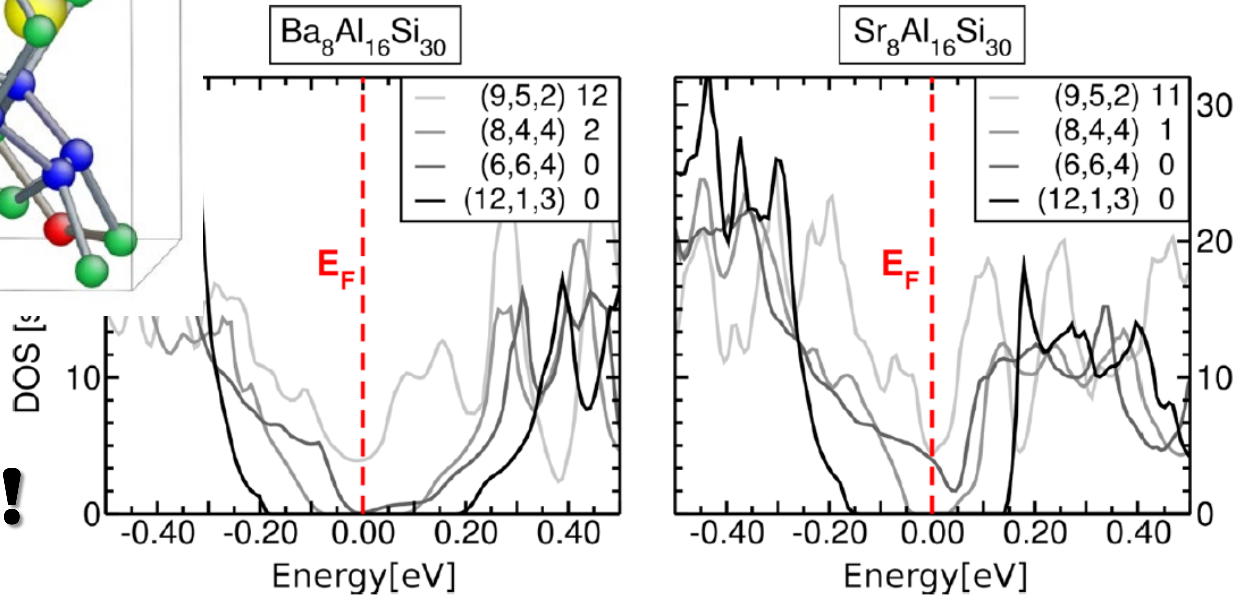
A few numbers?

Table 1. Cluster Expansion for $Ba_8Al_xSi_{46-x}$ Corresponding to Figure 2 (LDA and PBEsol)[a]

	$Ba_8Al_xSi_{46-x}$		
	LDA	PBEsol	
		CE1	CE2
N_t	35	25	17
N_c	8	8	10
J_0	−21	−2133568.7971	−2133571.8115
$J_k - J_c$	239.	262.3	206.4
$J_i - J_c$	316.	342.8	284.3
	1281	1282.9376	1283.2045
J_{kk}	182.		
J_{ii}	456.		
J_{ki}	248.		
J_{kc}	337.		
J_{k-i}	-		
J_{i-i}	-		
J_{k-c}	-		
CV	1.43,		
RMSE	1.04		

[a] N_t indicates the number of ab initio values to perform the CE and N_c … given in eV. CV and RMSE are in meV/atom. In the indices, k, i, ar…





We never provide all information!

Our daily research is great, but overall a slow process.


Imagine …
All data from
theory
… brought together

FAIR data - Findable

since 2014

NOMAD

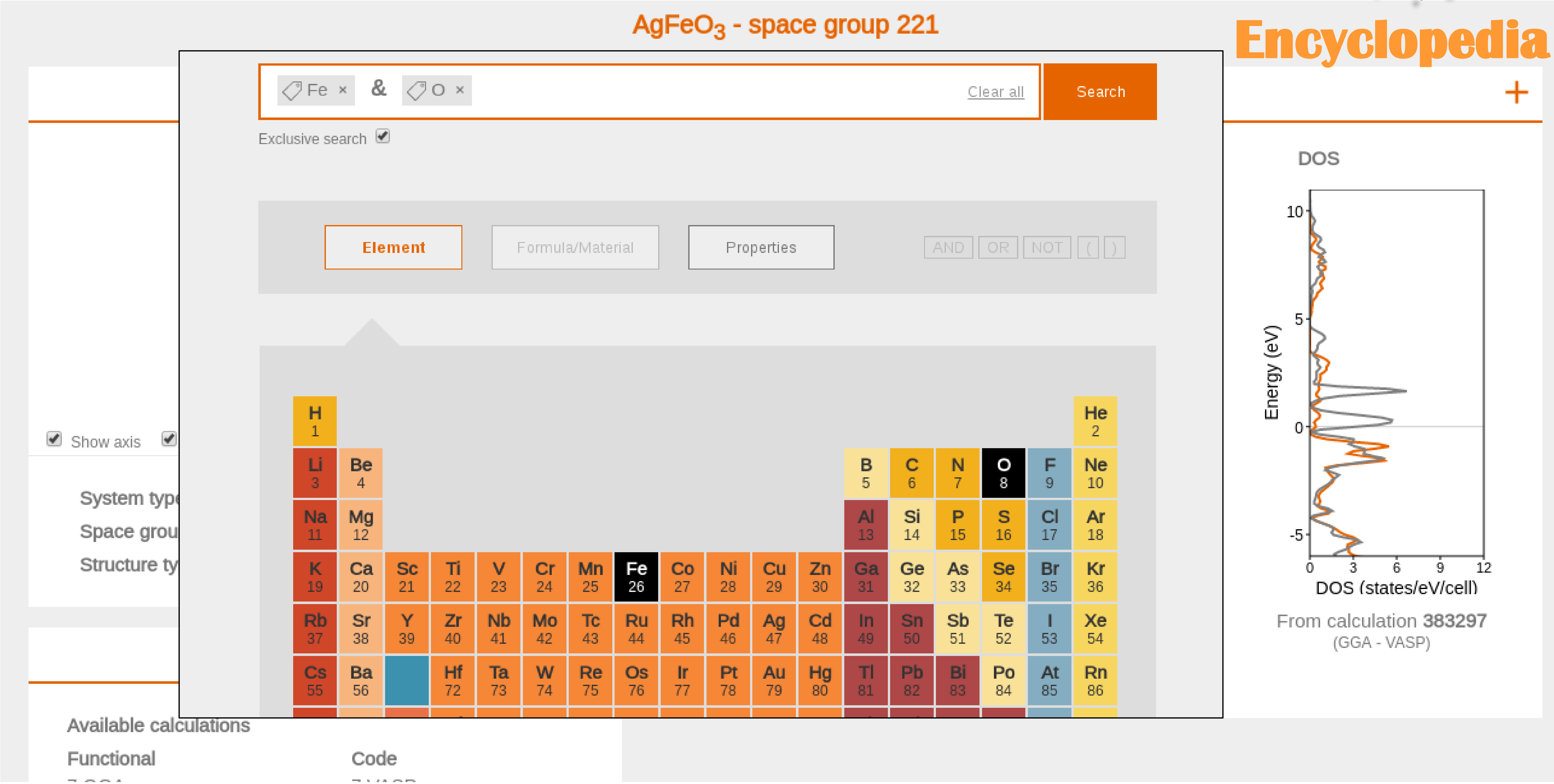
The Novel
Materials
Discovery
Laboratory

more than 100,000,000 calculations


NOMAD

FAIR data - Accsessible

Encyclopedia



FAIR data - Interoperable

More than 100 million calculations coming from ...

40 different codes

Every output fully parsed

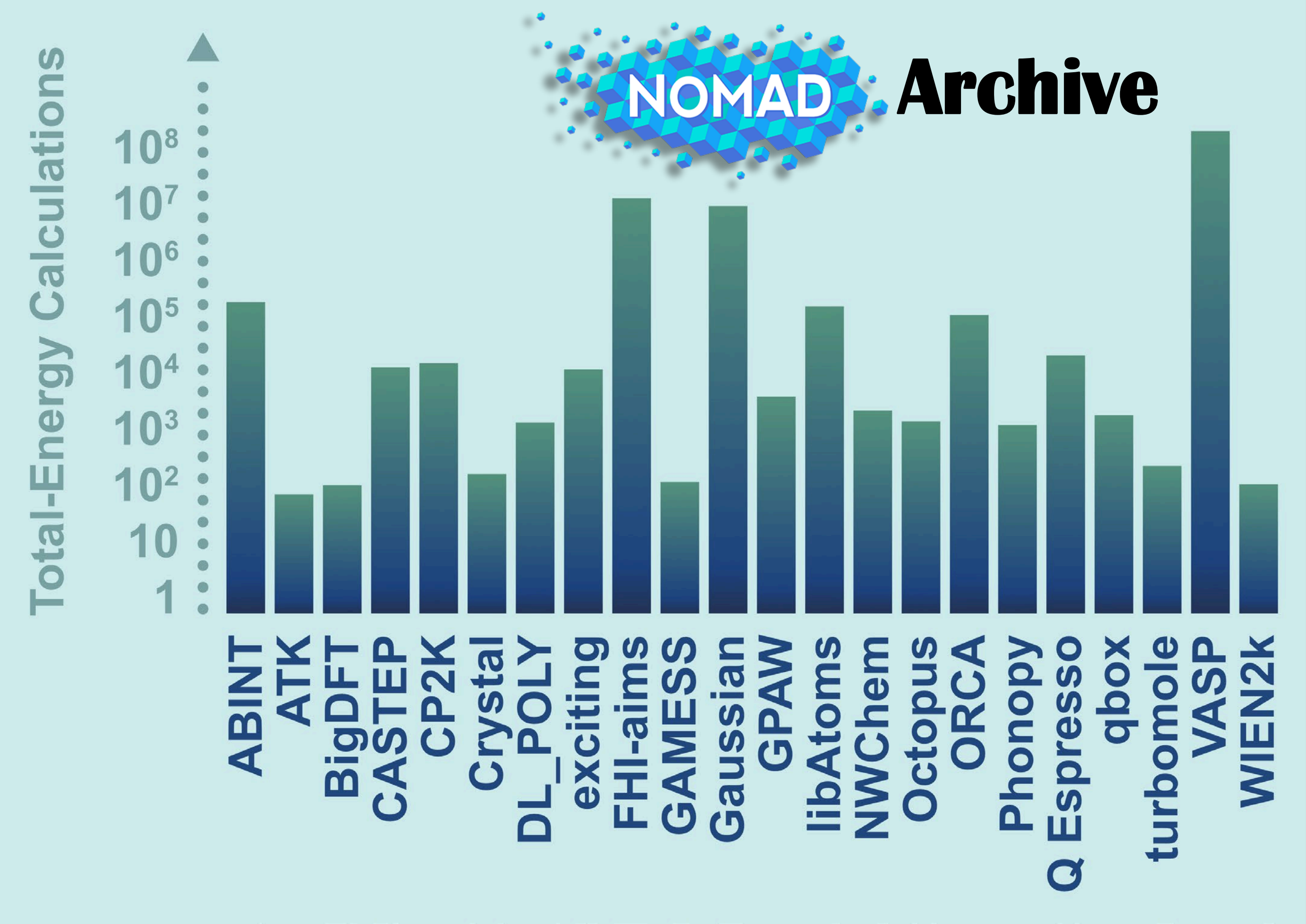
Metadata !!!!!!!!!!!

Unique description of data

Normalized data

Unified format, units, ...

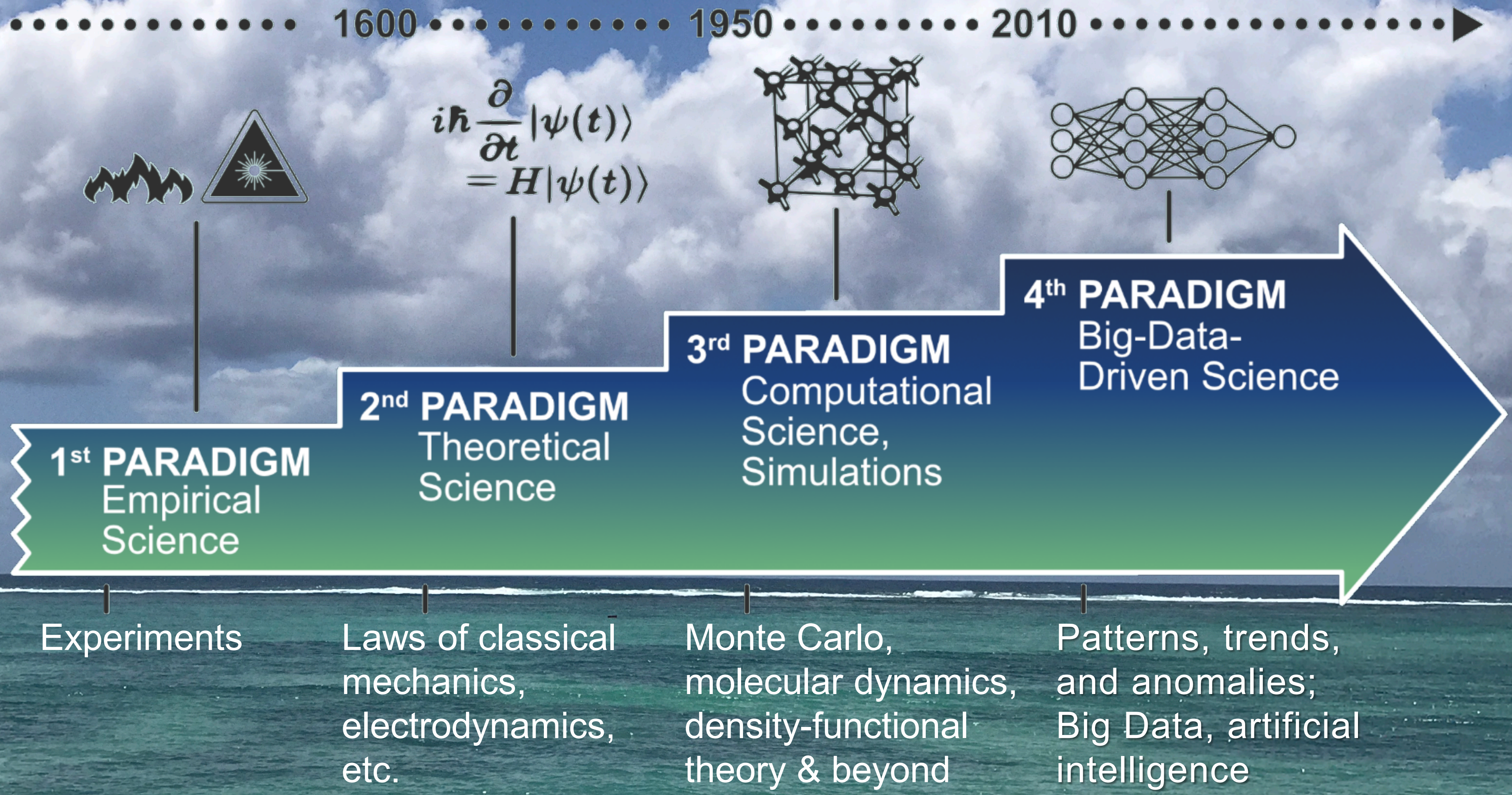
Only the first step!!




Imagine …
All data from
synthesis
experiment
theory
… brought together


superconductors
transparent metals
solar absorbers
thermoelectrics
thermal barriers
dielectrics
Our scientific vision ...

FAIR data – Repurposable



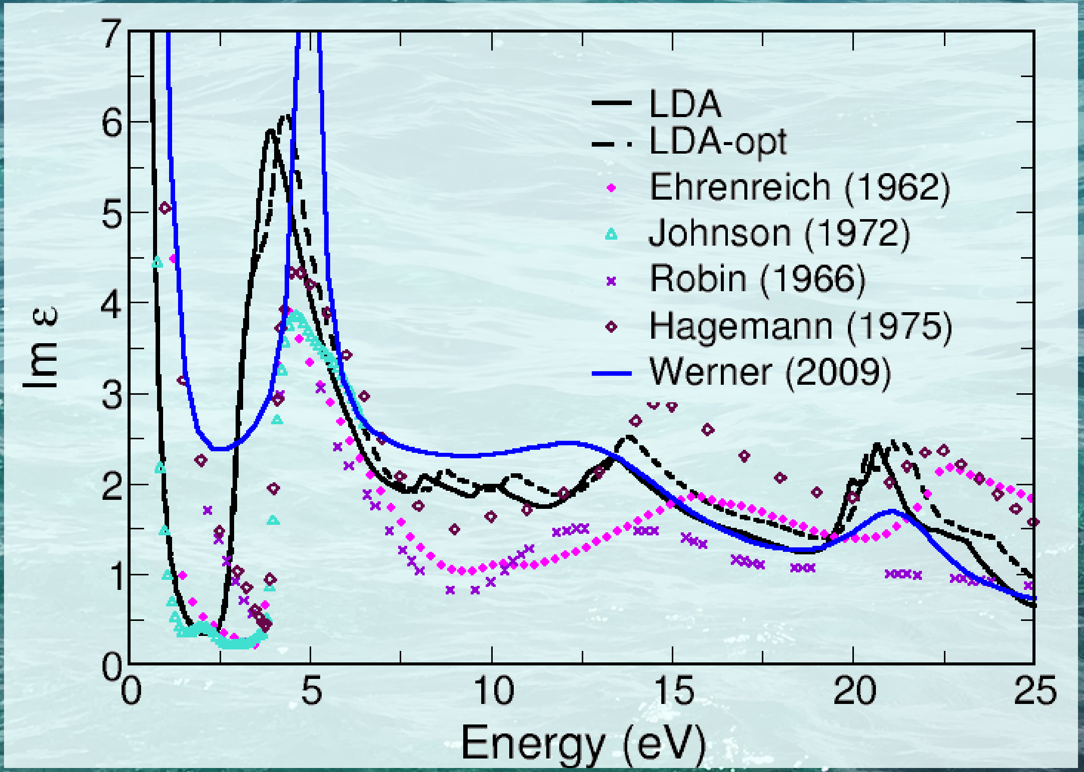
Jim Gray (Jan. 11. 2007): The 4th Paradigm, Data Intensive Discovery, edited by Hey, Tansley, and Tolle


Variety / Veracity
Data quality
Hidden parameters
Trust
levels
Benchmarks
Reproducibilty
What impacts the I?

Examples

Veracity

Optical spectra of samples called silver



Variety

Many ways of obtaining the same property – here: dielectric funtion

Ellipsometry
Absortion spectroscopy
Reflectance spectrocopy
Electron-loss spectroscopy

Theory on various levels

Interoperability?

Experimental data in a nutshell …

Data at instrument

Sample data

- Type
- Synthesis
- Cha
- Pre
- …

Instru

- Typ
- Sof
- Mai
- …

Measu

- Experimental conditions
- Instrument conditions
- Calibration
- Reference to sample & instrument

Data evaluation

Transformation of raw data

- New file format

Documentation data

- Descriptive text
- Summaries of findings
- Reference to analysis

Publication

Data published in article

- Results combined with prose text

igures
rting

e

ge

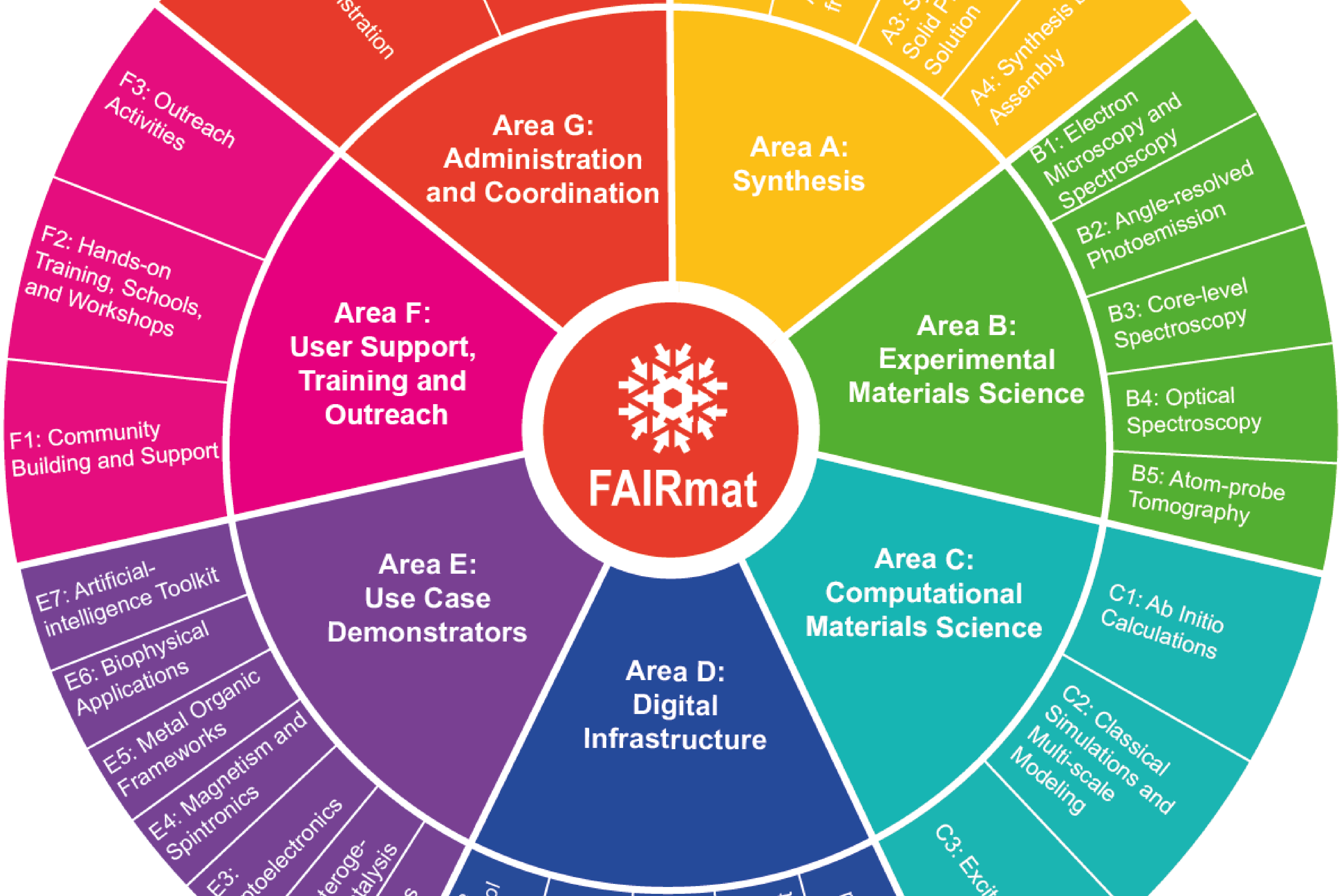
- Reference to sample, instruments
- Hosted in repository
- PID

Making materials-science data meaningful requires an in-depth description of how they have been obtained.

It starts with the synthesis and treatment of the sample, captures the instrument and the entire measurement process, …


FAIRmat
Synthesis
Use Cases
Theory
Experiment

Task Areas



FAIRmat

A: Synthesis

Challenges & goals

M. Albrecht

C. Felser

Worldwide, synthesis recipes are collected for personal use of the scientists, often documented in handwritten lab notebooks.

Log files created by the synthesis instruments, often not kept.

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

Goal 2: Harmonize metadata schemes of synthesis and experimental characterization

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

FAIRmat

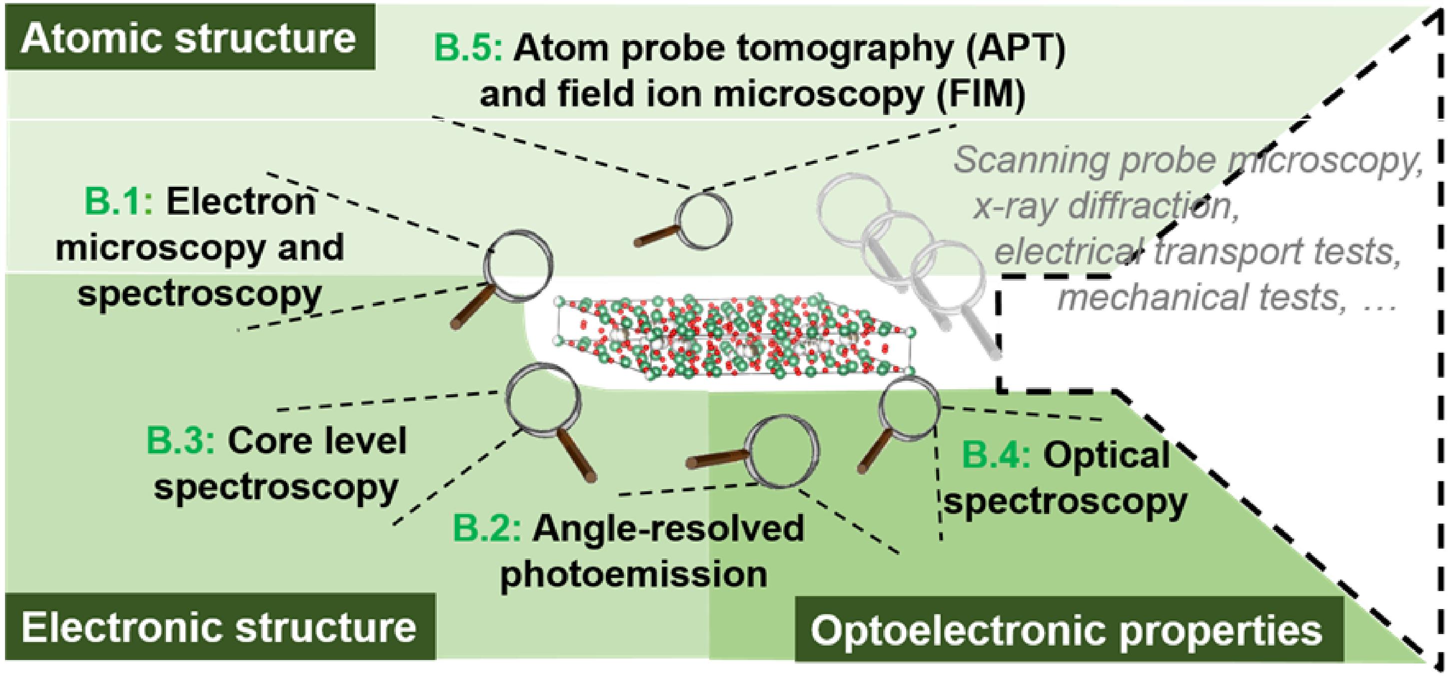
Area B: Experiment

FAIRmat

Challenges & goals

Goal 1: Metadata and workflows for the ***extremely diverse*** characterization methods used by the experimental condensed-matter community

Goal 2: Efficient and persistent linkage of data types to be implemented by means of LIMS and ELN solutions.



Each experimental probe has its specific challenges concerning processing, curation, and storage, owing to differences in volume, velocity, data formats, etc.

ELN: Electronic Lab Notebook; LIMS: Laboratory Information Management System

Area C: Computations

Challenges & goals

M. Scheffler

K. Kremer

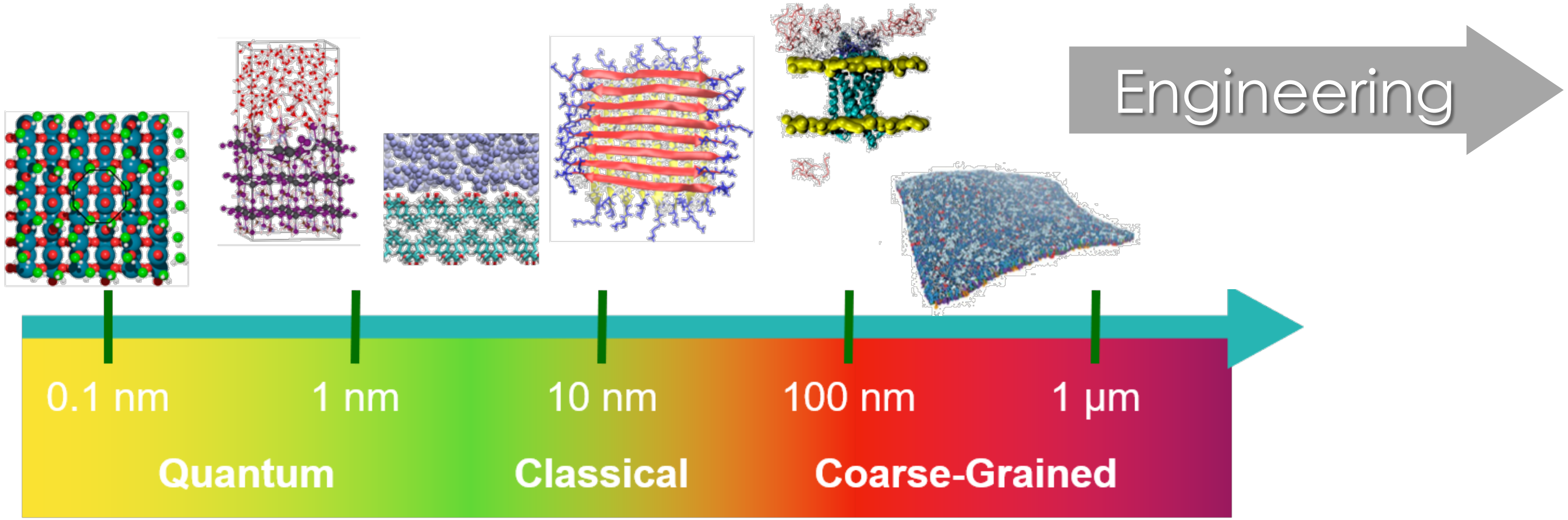
T. Bereau

Huge variety of methodology – from volumenous classical simulations to highly sophisticated quantum-mechanical many-body techniques, all with intricate subtleties

Goal 1: Integration of the NOMAD Laboratory into FAIRmat

Goal 2: Significant enhancement of its services

Goal 3: Much wider scope of methodologies



FAIRmat

Area D: Infrastructure

Challenges & goals


H. Bungartz


W. Nagel

Different scientific methods require vastly different data handling (4V).

Large amounts of very heterogeneous data of various sources need to be integrated.

Long-term availability and data security.

Goal 1: Enabling individual scientists and research institutes to manage data following common principles, with compatible technologies and a shared interface

Goal 2: Creation of a FAIR data exploration and sharing platform

Goal 3: Become role model of data security

FAIRmat

Area E: Use Cases

Challenges & goals

Can we have tools that not only get us organized but really enable us to enhance science in daily life?

C.Wöll

A. Groß

Goal 1: Test and demonstrate the functionality of the FAIRmat data infrastructure and identify weaknesses to be improved.

Goal 2: Show how the developed DI tools will significantly support the research of the various sub-communities.

Goal 3: Demonstrate the interfaces to and hand-shakes with other NFDI consortia.

FAIRmat

Area F: Outreach

Challenges & goals

How to get the entire community on board?

M. Scheffler

M. Aeschlimann

Goal 1: Advise research groups and institutions how to connect to the FAIRmat infrastructure and make use of it.

Goal 2: Inform about and integrate the community in its developments and achievements by dedicated conferences, workshops, schools, hands-on training, and hackathons

Goal 3: Train a new generation of interdisciplinary (data-expert) researchers, offer classes and lab courses, and work on future physics curricula

FAIRmat

FAIRmat

An ***inclusive***, ***user-driven*** approach to develop easy-to-use tools and an infrastructure towards FAIR data processing, storage, curation, sharing, and AI readiness for future use of materials data

Thanks !!


FAIRmat
Area A: Synthesis
A1: Synthesis from Melt
A2: Synthesis from Gas Phase
A3: Synthesis from Solid Phase and Solution
A4: Synthesis by Assembly
Area B: Experimental Materials Science
B1: Electron Microscopy and Spectroscopy
B2: Angle-resolved Photoemission
B3: Core-level Spectroscopy
B4: Optical Spectroscopy
B5: Atom-probe Tomography
Area C: Computational Materials Science
C1: Ab Initio Calculations
C2: Classical Simulations and Multi-scale Modeling
C3: Excitations
Area D: Digital Infrastructure
D1: Metadata, Ontologies, and Workflows
D2: Processing, Storage, and Compute Concept
D3: Central Meta-data Repository and Portal
D4: Connectivity
D5: Configurable Experiment Control System with ELNs
Area E: Use Case Demonstrators
E1: Batteries
E2: Heteroge-neous Catalysis
E3: Optoelectronics
E4: Magnetism and Spintronics
E5: Metal Organic Frameworks
E6: Biophysical Applications
E7: Artificial-intelligence Toolkit
Area F: User Support, Training and Outreach
F1: Community Building and Support
F2: Hands-on Training, Schools, and Workshops
F3: Outreach Activities
Area G: Administration and Coordination
G1: Administration
G2: Cross-cutting Topics