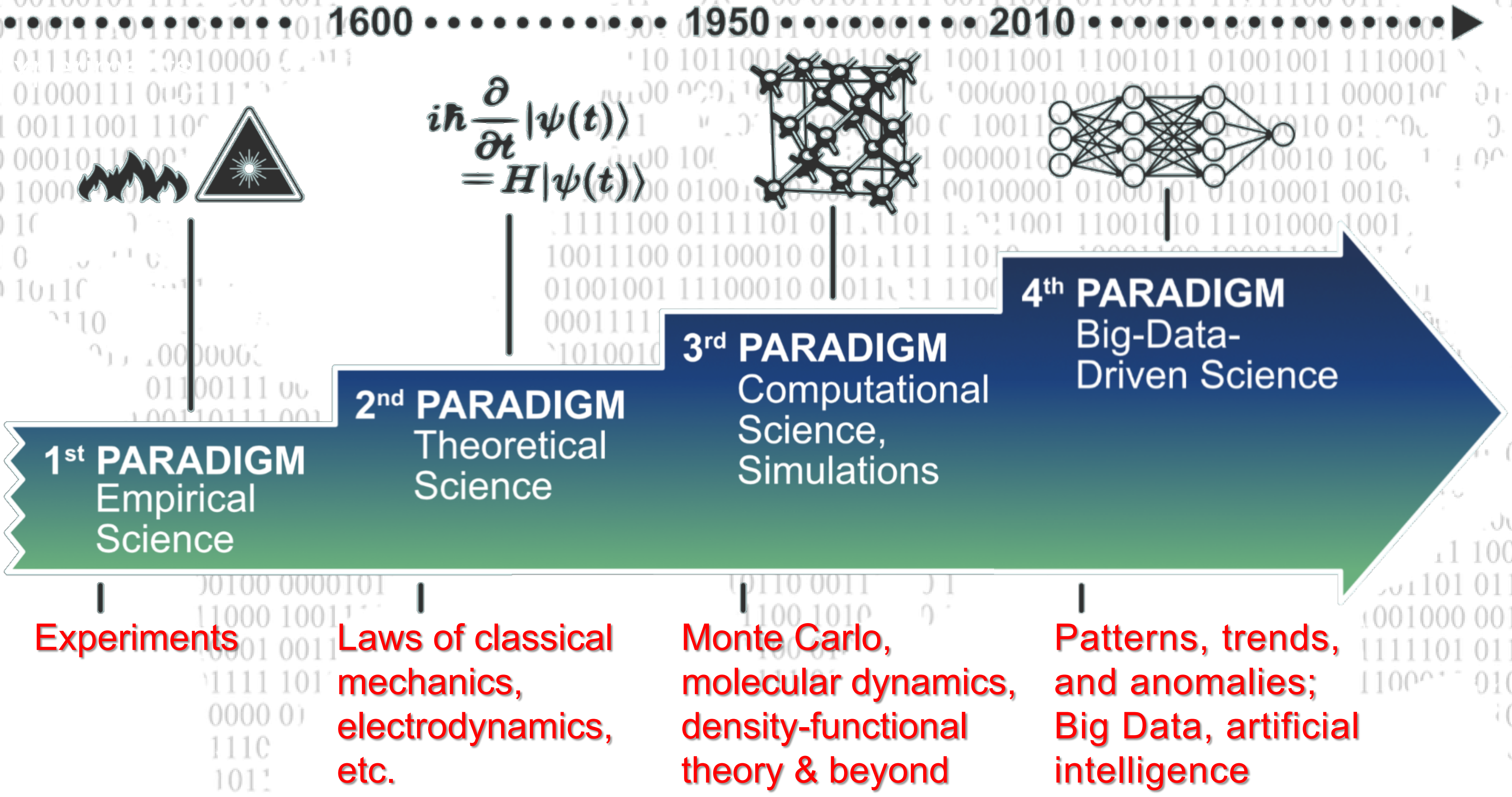
Making Materials Data Findable and AI Ready

Claudia Draxl

Where is materials research going?



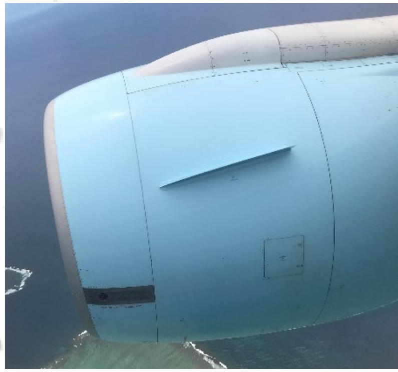
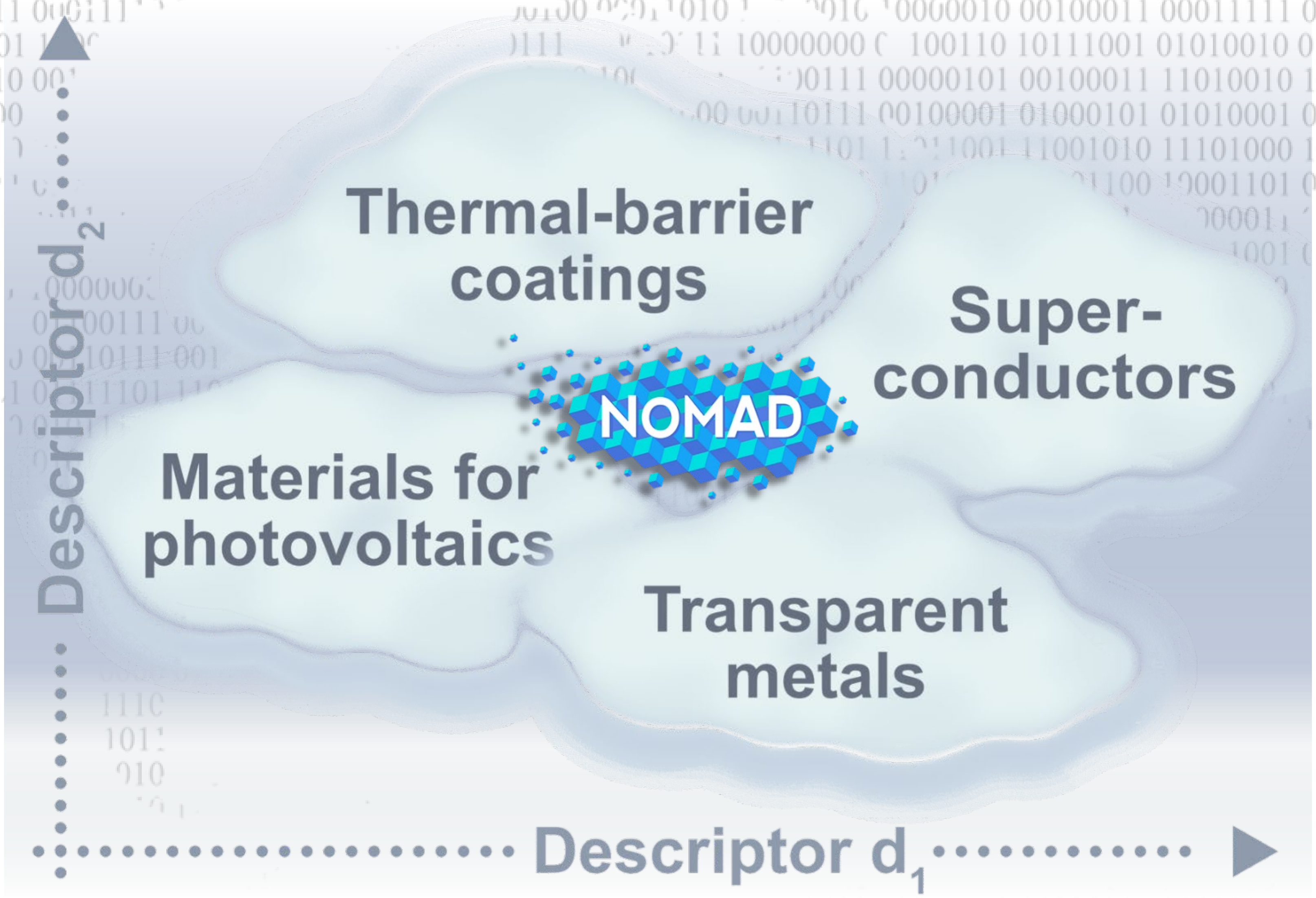
Jim Gray (Jan. 11. 2007): The 4th Paradigm, Data Intensive Discovery, edited by Hey, Tansley, and Tolle

Scientific vision

Accelerated materials discovery





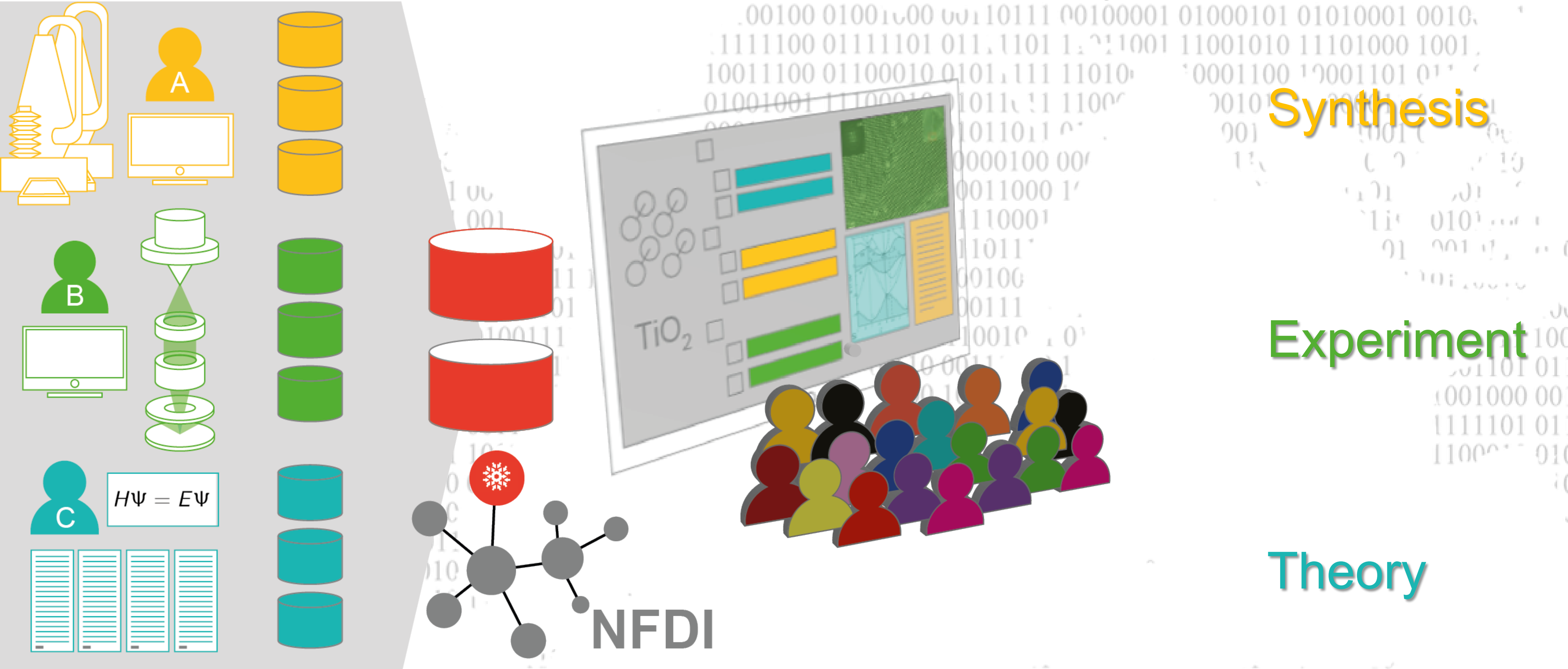
FAIRmat

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

FAIRmat‘s goal

Federated data infrastructure

Science ▸ Federated data ▸ Centralized metadata ▸ Aggregated information ▸ Global community



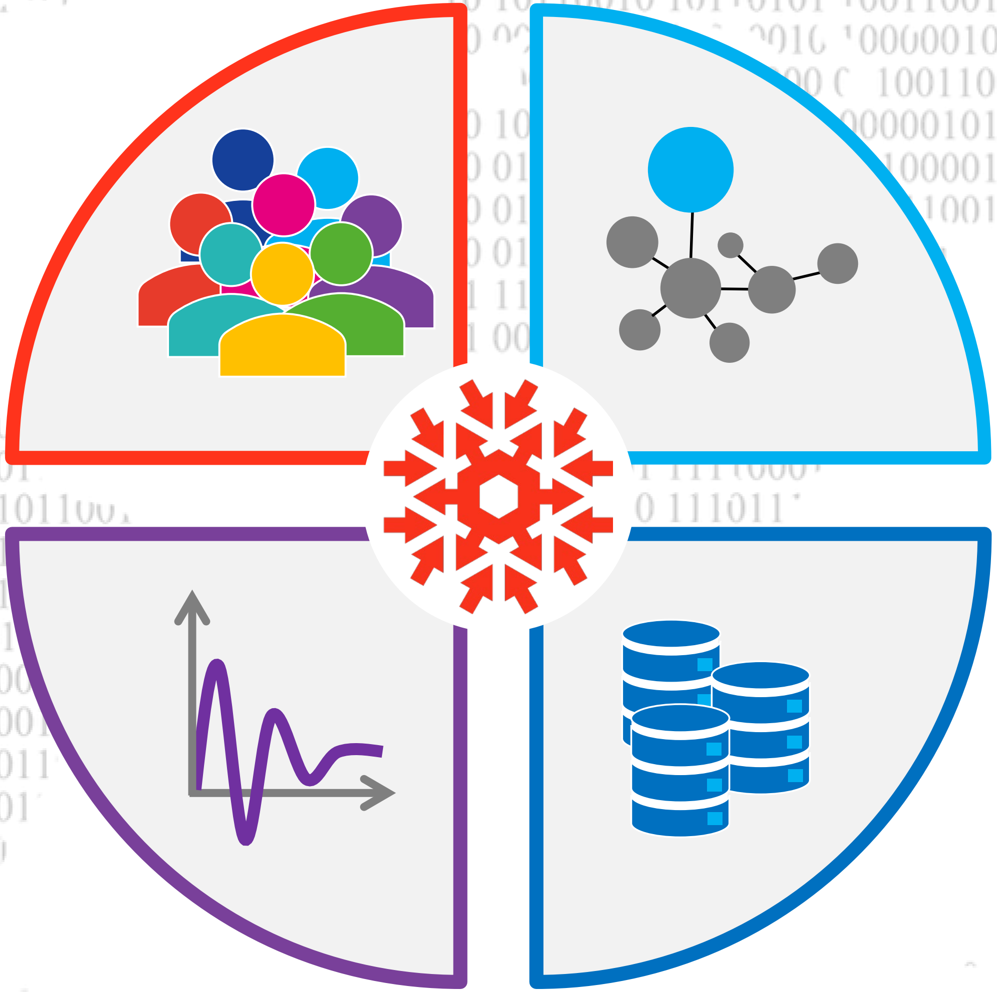
Main challenges

PEOPLE

Convince people that data sharing will advance science and engineering, also their own scientific work.

METADATA AND ONTOLOGIES

Develop metadata schemas, parsers, converters, and ontologies.



DATA PROCESSING AND ANALYSIS

Develop ontology-derived workflows, a materials encyclopedia, AI tools.

INFRASTRUCTURE

Develop software for storing, processing, and retrieving exponentially growing data volumes in a federated data infrastructure.

Commonalities

Recurrent themes – maximizing synergies

Metadata, ontologies, and workflows

Foundation of FAIRness

The 4V challenges

Concerns all areas and methods, but differently

Parsers, normalizers, and converters

Inevitable for getting / keeping community on board

Data curation and quality assessment

Crucial aspect of interoperability

Materials Encyclopedia and AI tools

Important components of reusability and exploitation

FAIRmat

AREA A: Synthesis

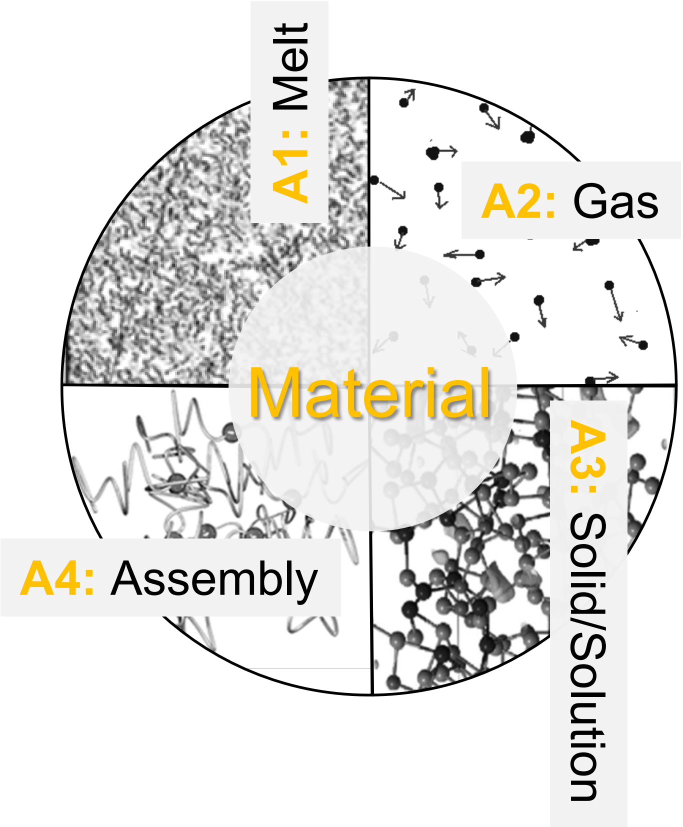
Area A

Challenges & goals

M. Albrecht

C. Felser

Reproducible growth of materials from various synthesis routes



- Four different synthesis routes
- Reference database for crystal growth
- Harmonize metadata schemas of synthesis and experimental characterization
- Towards computer-aided development of synthesis recipes

FAIRmat

FAIRmat

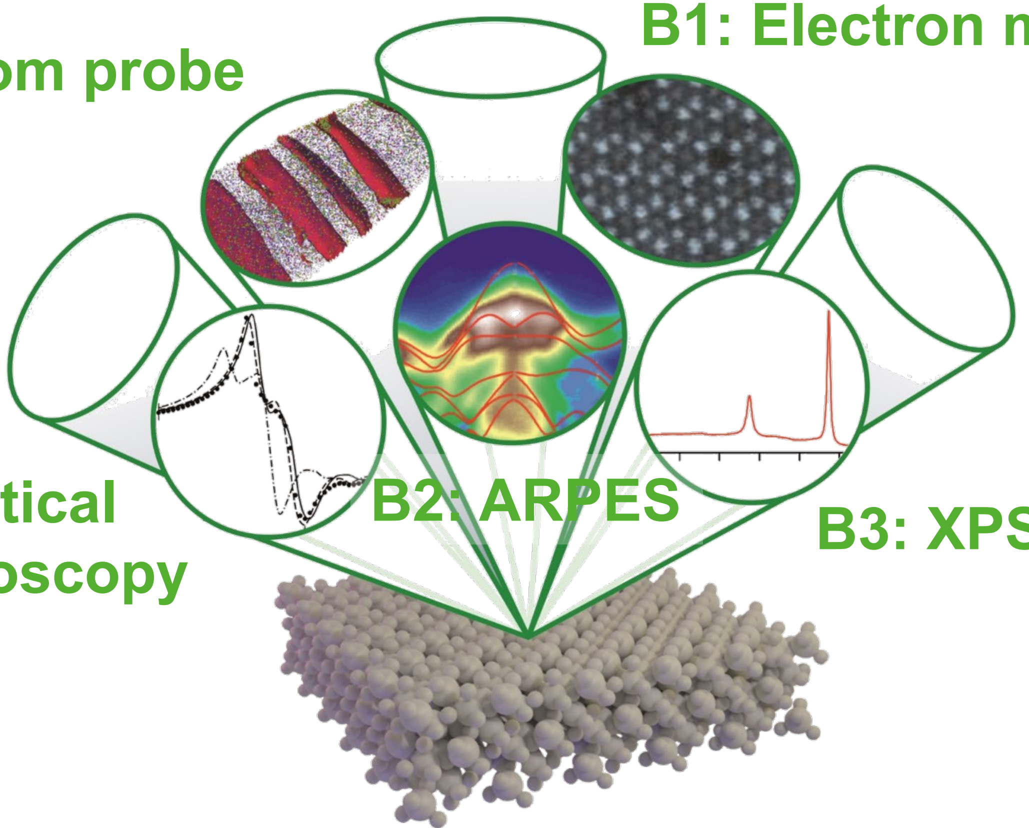
AREA B: Experiment

Area B

Challenges & goals


M. Greiner C. Koch

- Metadata and workflows for the ***extremely diverse characterization methods*** used by the community
- Focus on Electronic Lab Notebooks and Laboratory Information Management Systems



Each method has its specific challenges regarding processing, curation, and storage.

FAIRmat

Simple example: Veracity / Variety

Heterogeneity

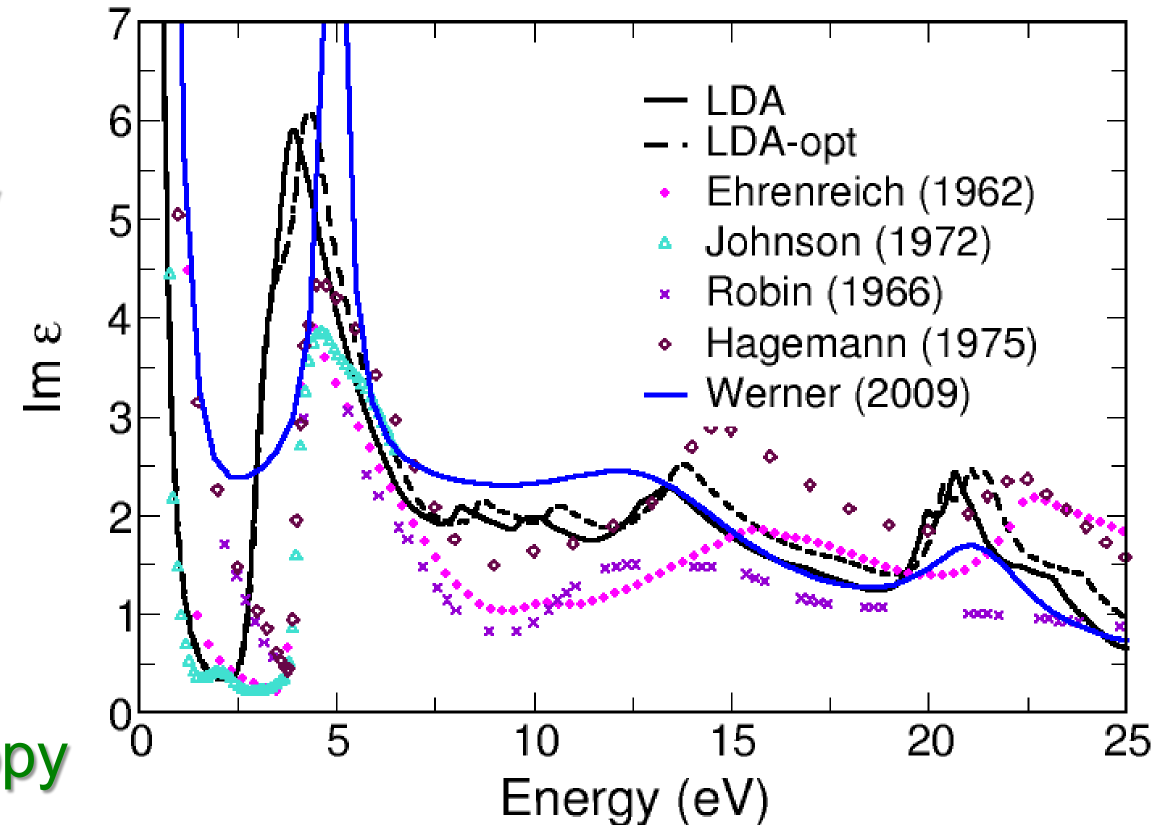
Optical spectra of *silver*

Decisive role of sample quality

Many ways of measuring one property

- Ellipsometry
- Absorption spectroscopy
- Reflectance spectrocopy
- Electron microscopy / spectroscopy

Many instruments, data formats, …

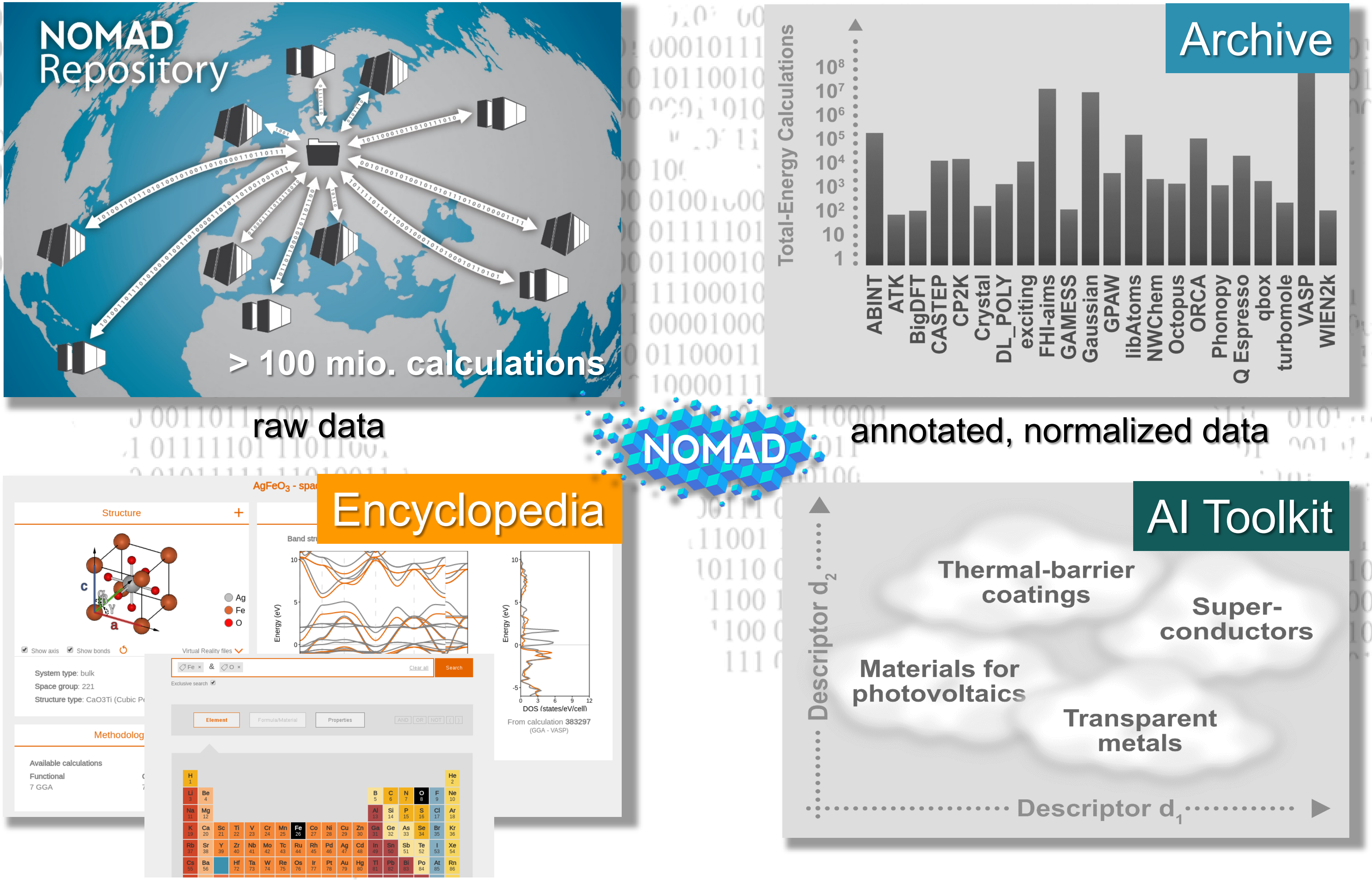


Many theoretical methods to compute the same

FAIRmat

AREA C:
Theory and Computations

Forerunner of a FAIR data infrastructure



Area C

Challenges & goals

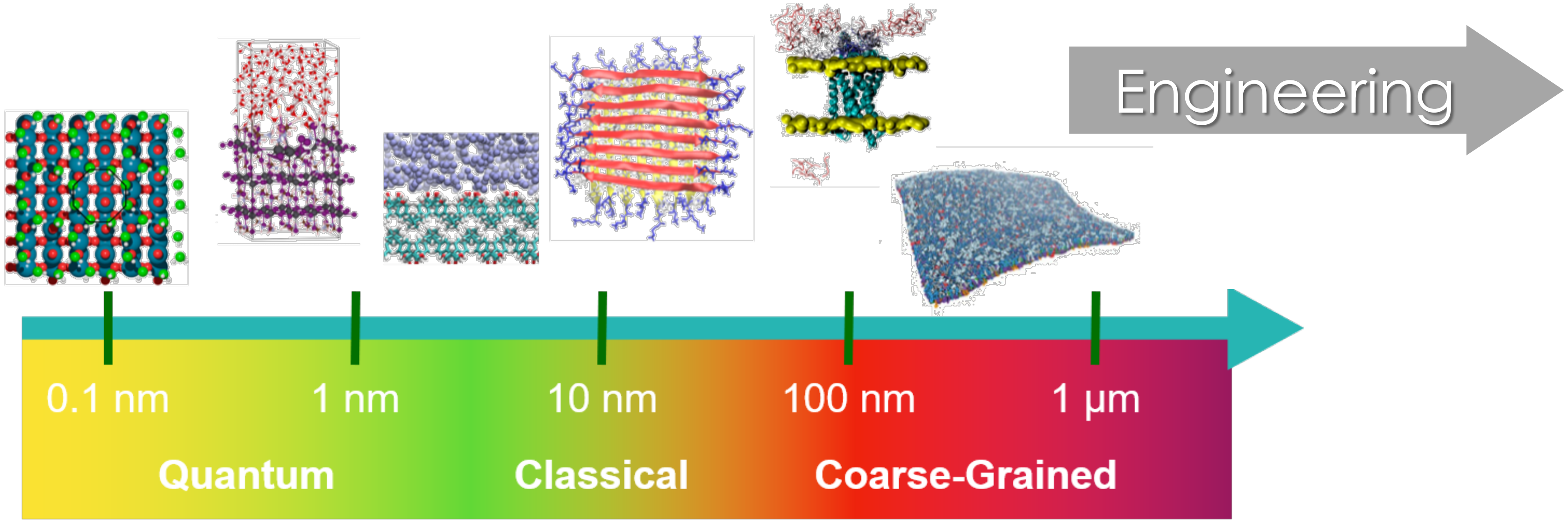

M. Scheffler


K. Kremer


T. Bereau

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

- Integration of the NOMAD Laboratory into FAIRmat
- Significant enhancement of its services
- DFT and higher-level methods, molecular dynamics, Monte-Carlo



FAIRmat

FAIRmat

AREA D:
Digital Infrastructure

Area D

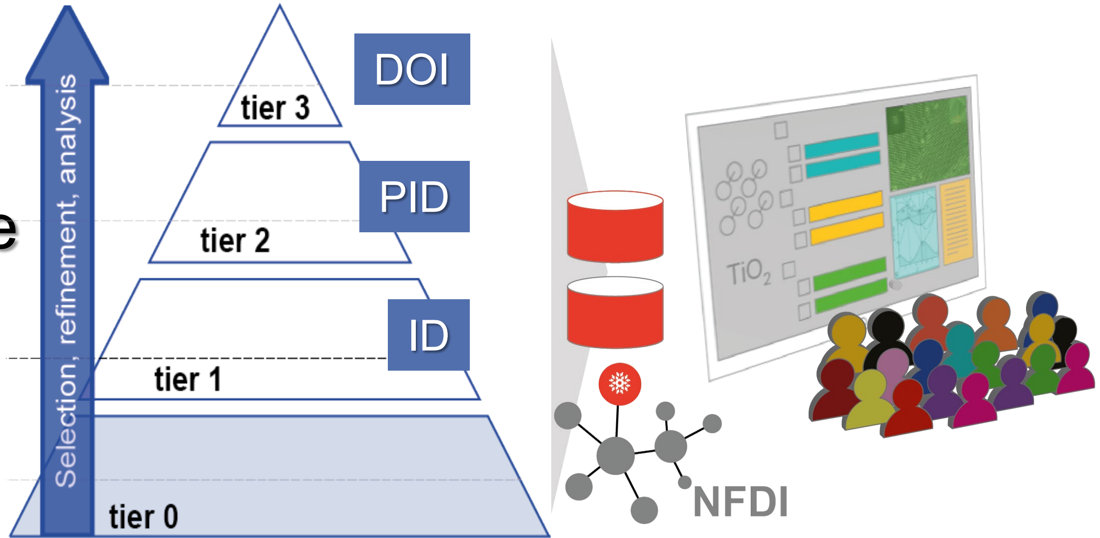
Challenges & goals

H. Bungartz

W. Nagel

Federated data infrastructure for the community

- ▲ Architecture: federated data - central metadata
- ▲ Authentication: single sign-on (AAI) infrastructure; ORCID
- ▲ Security: compute-center and network standards
- ▲ Curation: metadata / AI based
- ▲ Access: FAIRmat Portal
- ▲ “Oasis” – stand-alone infrastructure for managing data of individual groups



FAIRmat

AREA E:
Use-case Demonstrators

Area E

Challenges & goals

C. Wöll

A. Groß

Our tools should not only get us organized but enable researchers to enhance their daily scientific life.

- Test and demonstrate the functionality of the FAIRmat data infrastructure.
- Make sure that the developed DI tools will support the research of the ***various research fields and sub-communities***.
- Exemplify interfaces and hand-shakes with other NFDI consortia.

FAIRmat

FAIRmat

AREA F:
User Support, Training, and Outreach

Area F

Challenges & goals

Inform, involve, and embrace the community

M. Scheffler

M. Aeschlimann

- Reach students, postdocs, and professors and explain why and how a *Findable and AI Ready Data Infrastructure* will open ***new horizons*** for our research and science in general.
- Tutorials, schools, hackathons, workshops, international conferences, and university lectures.
- ***White Paper*** *on the establishment of modern research data management in the physics curricula*.
- Collaborations with DPG and other consortia.

FAIRmat

DPG

Area F

First FAIRmat Colloquium

Barend Mons:

How to materialise FAIR

HU Berlin, Adlershof, October 7, 2021

https://www.fair-di.eu/fairdi-colloqium-home

Registration for on-site participation

FAIRmat

FAIRmat

AREA G:

Administration and Coordination

Team

In total 58 PIs

FAIRmat

Participating institutions	Location	Task
Deutsche Physikalische Gesellschaft (DPG)	Bad Honnef	F3
TIB Leibniz Information Centre for Science and Technology (TIB)	Hannover	G2

Participating individuals	Institution, location	Task
Martin Aeschlimann*	Technical University of Kaiserslautern, Kaiserslautern	B2, F2
Sören Auer	TIB Leibniz Information Centre for Science and Technology (TIB), Hannover	G2
Carsten Baldauf	Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin	E6, F2
Tristan Bereau*	Max Planck Institute for Polymer Research (MPIP), Mainz	C2
Stefan Blügel	Research Center Jülich (FZJ), Jülich	E4
Silvana Botti	Friedrich Schiller University Jena, Jena	C1
Christoph J. Brabec	University Erlangen-Nürnberg (FAU), Erlangen	E3
Malte Dreyer	Humboldt-Universität zu Berlin, Berlin	F1
Natasha Dropka	Leibniz-Institut für Kristallzüchtung (IKZ), Berlin	A1
Ralph Ernstorfer	Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin	B2
Norbert Esser	Leibniz Institute for Analytical Sciences (isas), Berlin	B4
Claudia Felser*	MPI for Chemical Physics of Solids (MPI-CPfS), Dresden	A3
Luca Ghiringhelli	Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin	D1, E7
Roger Gläser	Leipzig University, Leipzig	E2
Axel Groß*	Ulm University, Ulm	E1
Marius Grundmann	Leipzig University, Leipzig	B4
Aleksander Gurlo	Technical University Berlin (TU Berlin), Berlin	A3
Thomas Hammerschmidt	Ruhr-University Bochum (RUB), Bochum	E7
Tamas Haraszti	DWI - Leibniz Institute for Interactive Materials, Aachen	A4
Kerstin Helbig	Humboldt-Universität zu Berlin, Berlin	F1
Stefan Hecht	DWI - Leibniz Institute for Interactive Materials Aachen	A4
Thomas Heine	Technical University Dresden (TU Dresden), Dresden	E5
Jürgen Janek	Justus Liebig University Giessen, Giessen	E1
Heinz Junkes	Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin	D5
Josef A. Käs	Leipzig University, Leipzig	E6
Christoph Koch*	Humboldt-Universität zu Berlin, Berlin	B1, G1
Sarah Köster	Georg-August-University Göttingen, Göttingen	E6
Kurt Kremer*	Max Planck Institute for Polymer Research (MPIP), Mainz	C
Michael Krieger	University Erlangen-Nürnberg, Erlangen	D5
Markus Kühbach	Max-Planck-Institut für Eisenforschung GmbH (MPIE), Düsseldorf	B5
Hermann Lederer	MP Computing and Data Facility (MPCDF), Garching	D2
Ingrid Mertig	Martin-Luther-University Halle-Wittenberg, Halle	E4
Wolfgang Nagel*	TU Dresden, Dresden	D2
Rossitza Pentcheva	University Duisburg-Essen, Duisburg	F1
Dierk Raabe	Max-Planck-Institut für Eisenforschung GmbH (MPIE), Düsseldorf	B5
Alexander Reinefeld	Zuse Institute Berlin, Berlin	D3
Karsten Reuter	Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin	C2, E2
Raphael Ritz	MP Computing and Data Facility, Garching	D1
Cesar Rodriguez-Emmenegger	DWI - Leibniz Institute for Interactive Materials, Aachen	A4
Erich Runge	*ex officio*, German Physical Society (DPG), Bonn	F3
Markus Scheidgen	Humboldt-Universität zu Berlin, Berlin	D3
Robert Schlögl	Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin	E2
Thomas Schröder	Leibniz Institute for Crystal Growth (IKZ), Berlin	A1
Godehard Sutmann	Jülich Supercomputing Centre, Jülich	D2
Annette Trunschke	Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin	E2
Thomas Unold	Helmholtz Zentrum Berlin (HZB), Berlin	E3
Roser Valenti	Goethe University Frankfurt, Frankfurt/Main	C3
Holger von Wenckstern	Leipzig University, Leipzig	A2
Dieter Weber	Research Center Jülich, Jülich	B1
Heiko Weber	University Erlangen-Nürnberg, Erlangen	D5
Stefan Wesner	Ulm University, Ulm	D4
Joachim Wosnitza	Helmholtz Zentrum Dresden Rossendorf, Dresden	E4

FAIRmat Headquarter – HUB

Center for Materials Science Data currently established at Humboldt-Universität zu Berlin.

Postdocs and research engineers not assigned to individual research group, specific Task or Area, but shared in a pool of creative minds with complementary skill-sets.

This maximally exploits synergies and provides the required flexibility and efficiency.

https://nomad-lab.eu/career

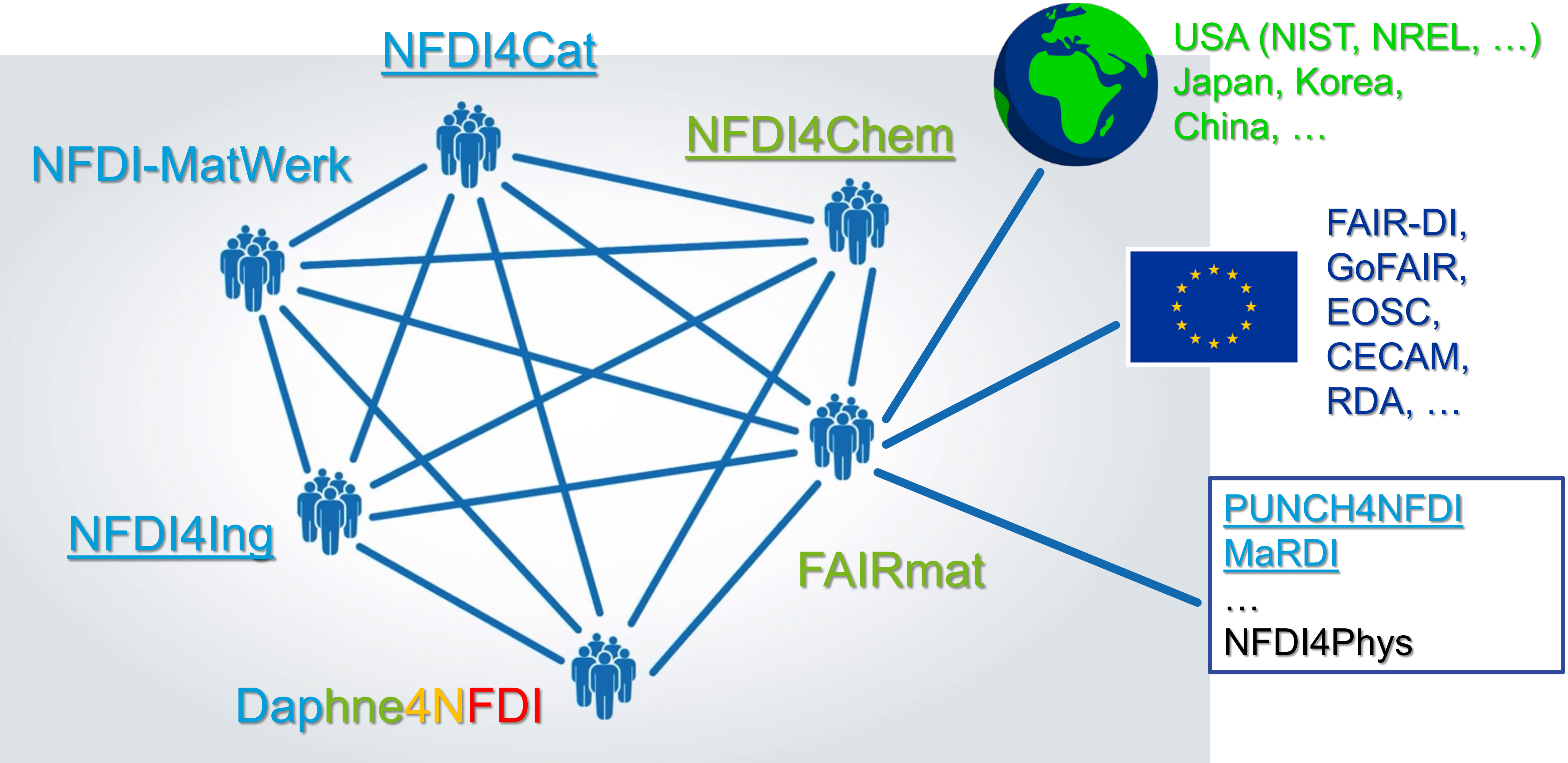
We are hiring!

Area G

FAIRmat

Networking, collaborations, interactions, hand-shakes

NFDI as a whole



FAIRmat

Thanks!

https://fair-di.eu/fairmat/

DFG