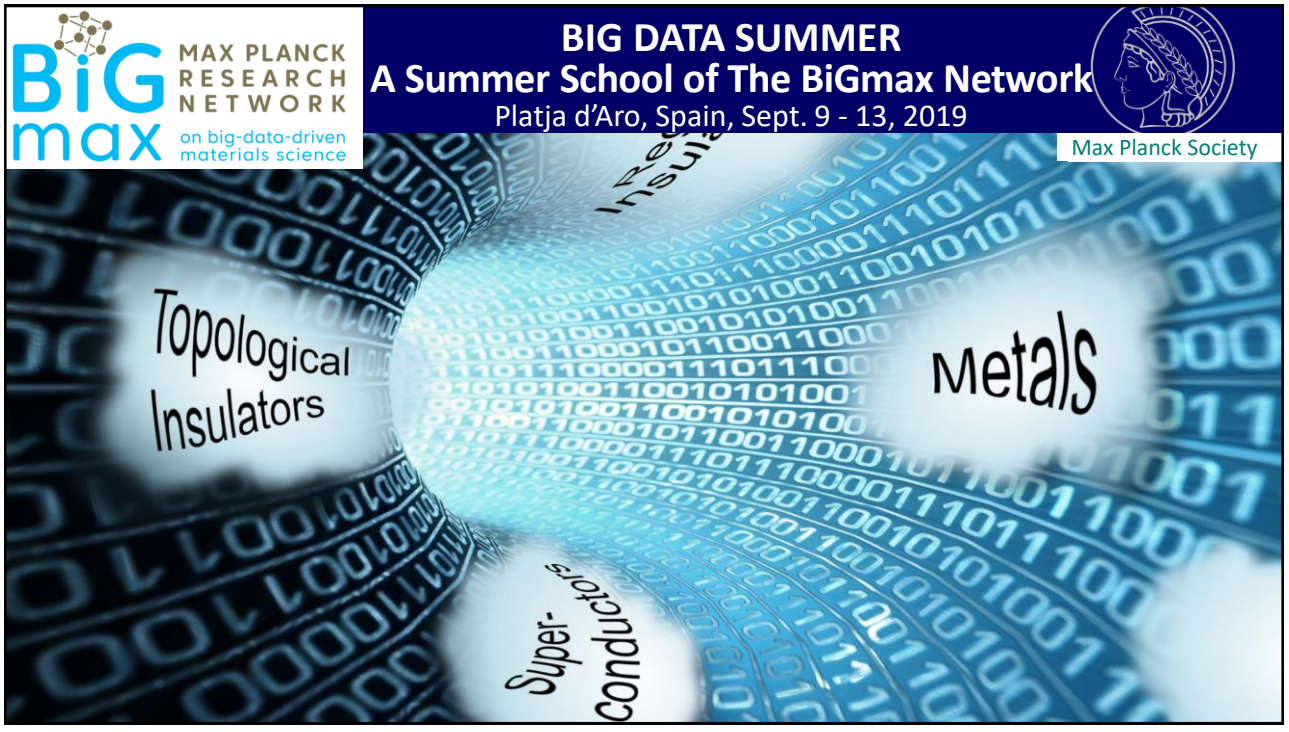
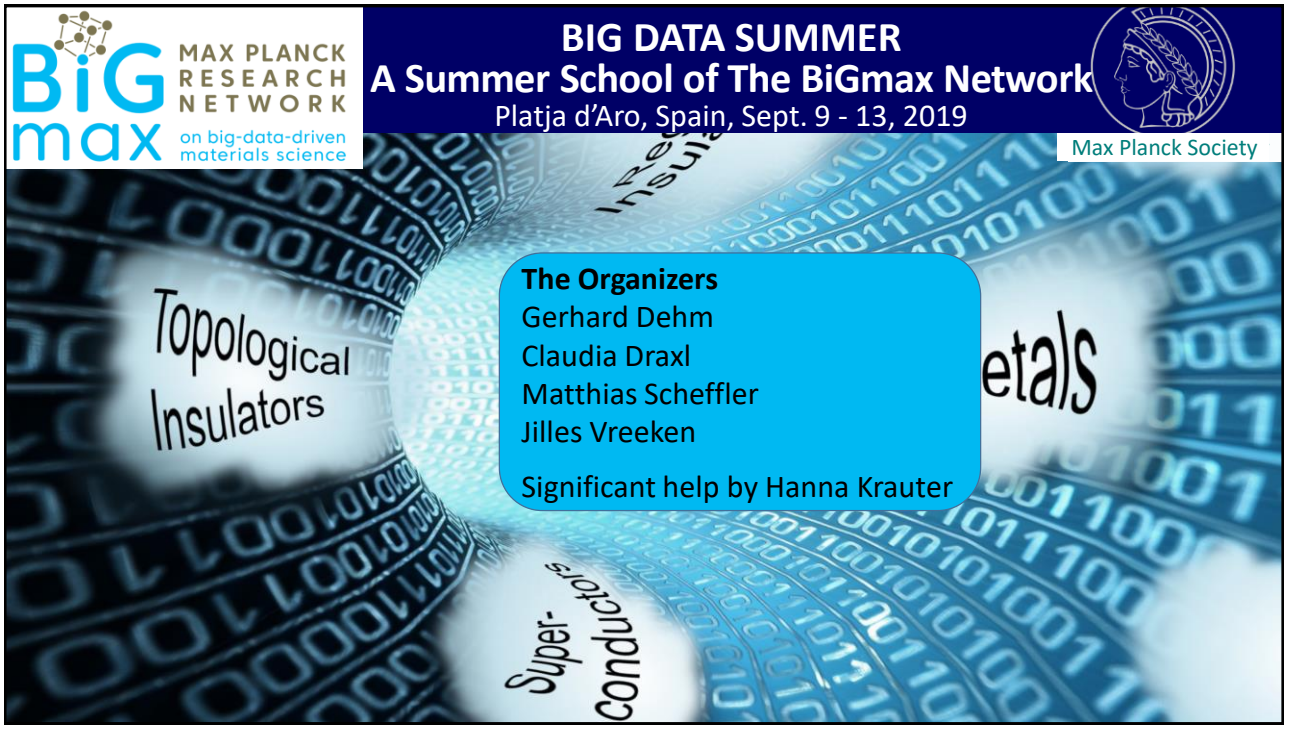

MAX PLANCK RESEARCH NETWORK
on big-data-driven materials science
BIG DATA SUMMER
A Summer School of The BiGmax Network
Platja d'Aro, Spain, Sept. 9 - 13, 2019
Max Planck Society
Topological Insulators
Metals
Super-conductors


MAX PLANCK RESEARCH NETWORK
on big-data-driven materials science
BIG DATA SUMMER
A Summer School of The BiGmax Network
Platja d'Aro, Spain, Sept. 9 - 13, 2019
Max Planck Society
Topological Insulators
The Organizers
Gerhard Dehm
Claudia Draxl
Matthias Scheffler
Jilles Vreeken
Significant help by Hanna Krauter
Super-conductors

What Is The BiGmax Network?

https://www.bigmax.mpg.de

BiG max MAX PLANCK RESEARCH NETWORK on big-data-driven materials science

In BiGmax, Max Planck researchers from 10 Max Planck Institutions and the Humboldt University Berlin address the big-data challenges of modern materials science. This concerns the 4V issues (*Volume*, *Variety*, *Velocity*, and *Veracity)* and, in particular, the development of artificial-intelligence tools to analyze the data. This new research direction aims to identify better or even novel materials for specific purposes for basic sciences and applications.

The first period of the network is for five years. It started in April 2017.

BiG max MAX PLANCK RESEARCH NETWORK on big-data-driven materials science

Member Institutions:

- Fritz Haber Institute of the MPG, Berlin
- MPI for Dynamics of Complex Technical Systems, Magdeburg
- MPI for Iron Research, Düsseldorf
- MPI for the Physics of Complex Systems, Dresden
- MPI for Informatics, Saarbrücken
- MPI for Structure and Dynamics of Matter, Hamburg
- MPI for Intelligent Systems, Tübingen/Stuttgart
- MPI for Polymer Research, Mainz
- MPI of Colloids and Interfaces, Potsdam
- MP Computing and Data Facility, Garching
- Humboldt University, Berlin


MAX PLANCK
RESEARCH
NETWORK
on big-data-driven
materials science
BIG DATA SUMMER
A Summer School of The BiGmax Network
Platja d'Aro, Spain, Sept. 9 - 13, 2019
Max Planck Society
CAP ROIG
HOTEL


MAX PLANCK
RESEARCH
NETWORK
on big-data-driven
materials science
BIG DATA SUMMER
A Summer School of The BiGmax Network
Platja d'Aro, Spain, Sept. 9 - 13, 2019
Max Planck Society
CAP ROIG
HOTEL

Tomorrow afternoon:

16:20 - 16:50	Break
16:50 - 20:00	**Poster Parade and Poster Session**
20:00	Dinner

Lucas Foppa

Thursday afternoon 14:30: Excursion

By chartered bus to Figueres
(70 min ride)

Dalí Theatre-Museum
(https://www.salvador-dali.org/en/)
with a guided tour

NOMAD
NOVEL MATERIALS DISCOVERY

If You Need Help, Please Contact:

Max Planck Society

Lucas Foppa

Marcin Krynski

Andreas Leitherer

Thomas Purcell

Daniel Speckhard

Maria Troppenz

BiG max MAX PLANCK RESEARCH NETWORK on big-data-driven materials science

BIG DATA SUMMER
A Summer School of The BiGmax Network
Platja d'Aro, Spain, Sept. 9 - 13, 2019

Max Planck Society

Program

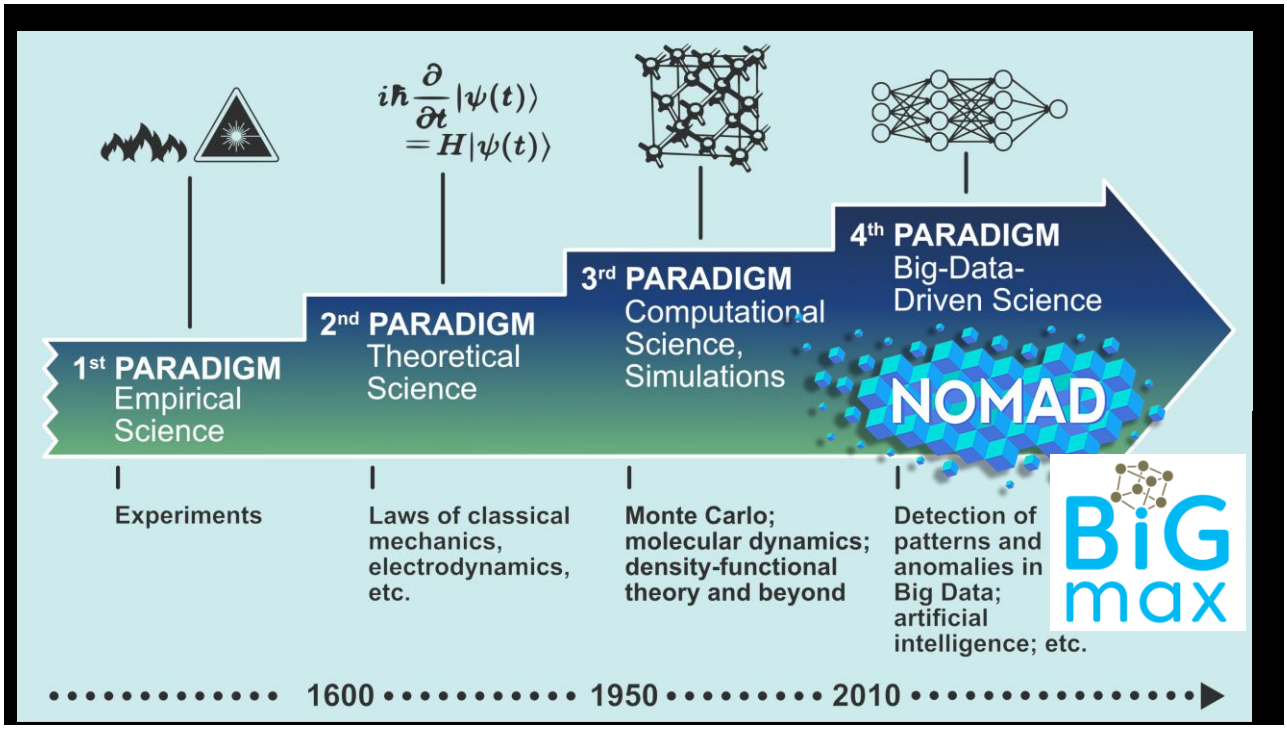
Monday, September 9, 2019

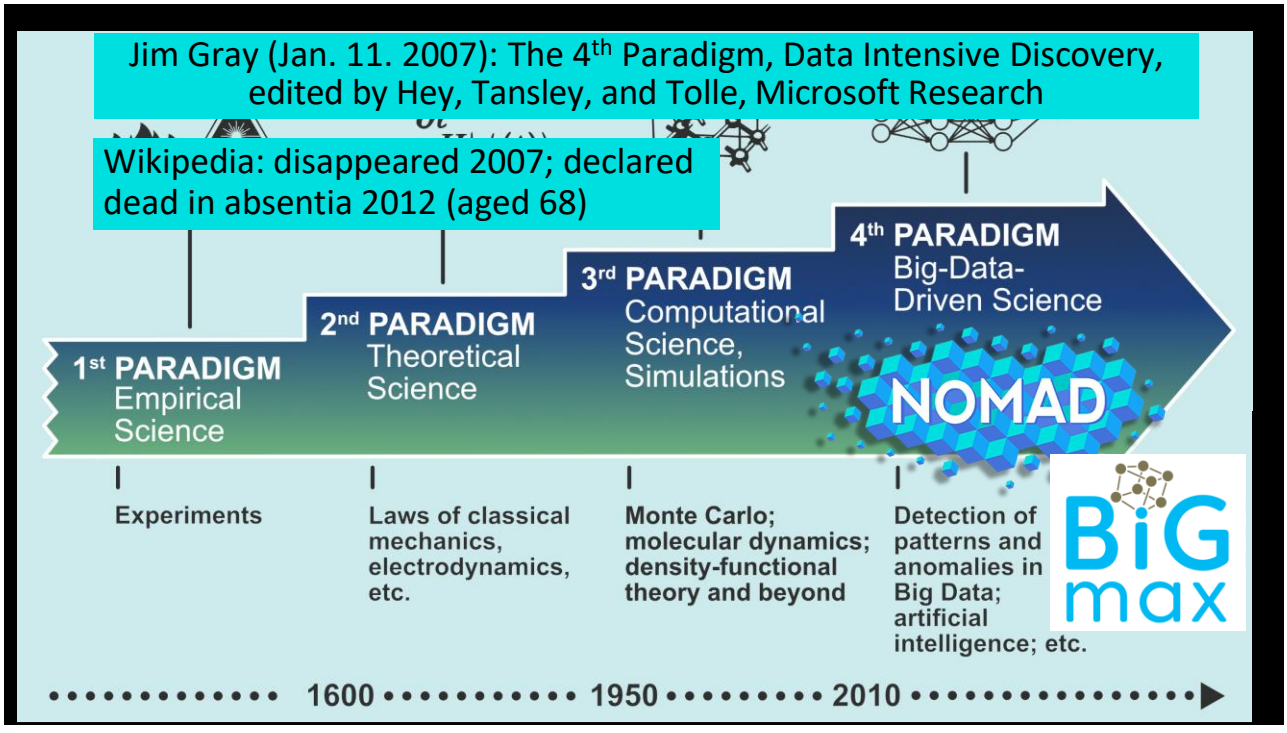
15:00	*Arrival – Coffee break*	
Session chair, morning session: Claudia Draxl		
15:30 - 16:30	Matthias Scheffler	*Welcome and Introduction*
16:30 - 17:30	Jilles Vreeken	*Material Subgroups*
17:30 - 18:00	Break	
18:00 - 19:00	Hans-Joachim Bungartz	*Research Data Infrastructures – How generic can & should they be?*
19:30	Welcome Cocktail - Dinner	

BiG max MAX PLANCK RESEARCH NETWORK on big-data-driven materials science

BIG DATA SUMMER
A Summer School of The BiGmax Network
Platja d'Aro, Spain, Sept. 9 - 13, 2019

Max Planck Society


$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = H|\psi(t)\rangle$
1st PARADIGM Empirical Science
2nd PARADIGM Theoretical Science
3rd PARADIGM Computational Science, Simulations
4th PARADIGM Big-Data-Driven Science
NOMAD
BiG max
Experiments
Laws of classical mechanics, electrodynamics, etc.
Monte Carlo; molecular dynamics; density-functional theory and beyond
Detection of patterns and anomalies in Big Data; artificial intelligence; etc.
1600
1950
2010


Jim Gray (Jan. 11. 2007): The 4th Paradigm, Data Intensive Discovery, edited by Hey, Tansley, and Tolle, Microsoft Research
Wikipedia: disappeared 2007; declared dead in absentia 2012 (aged 68)
1st PARADIGM Empirical Science
2nd PARADIGM Theoretical Science
3rd PARADIGM Computational Science, Simulations
4th PARADIGM Big-Data-Driven Science
NOMAD
BiG max
Experiments
Laws of classical mechanics, electrodynamics, etc.
Monte Carlo; molecular dynamics; density-functional theory and beyond
Detection of patterns and anomalies in Big Data; artificial intelligence; etc.
1600
1950
2010

NOMAD

NOVEL MATERIALS DISCOVERY

High-Throughput Screening in Computational (and Experimental) Materials Science

Max Planck Society

Consider as many compounds a possible, typically $O(10^3)$ – $O(10^5)$

Sharing Advances Science

Needs for a FAIR, Efficient Research-Data Infrastructure

Recycle the "waste"! Enable re-purposing.

$O(10^1)$ – $O(10^2)$ compounds selected

Antoine Lavoisier

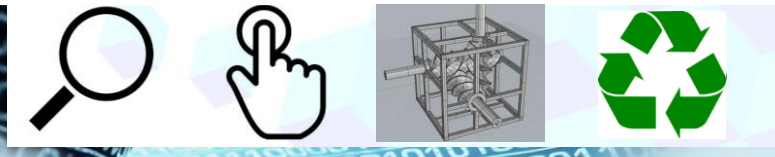
NOMAD

NOVEL MATERIALS DISCOVERY

Findable Accessible Interoperable Reusable

M. D. Wilkinson et al., Scientific Data **3**, 160018 (2016)

Max Planck Society



Topological

Metals

The materials science challenge:

- The number of possible materials is practically infinite
- No way to cover this huge structural and chemical compound space by high-throughput screening
- The interesting materials are very few (they are statistically exceptional)

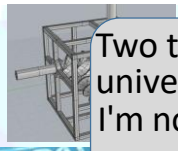
NOMAD

Findable Accessible Interoperable Reusable

M. D. Wilkinson et al., Scientific Data **3**, 160018 (2016)

NOVEL MATERIALS DISCOVERY

Max Planck Society



Two things are infinite: the universe and human stupidity; I'm not sure about the universe.

Albert Einstein

The materials science challenge:

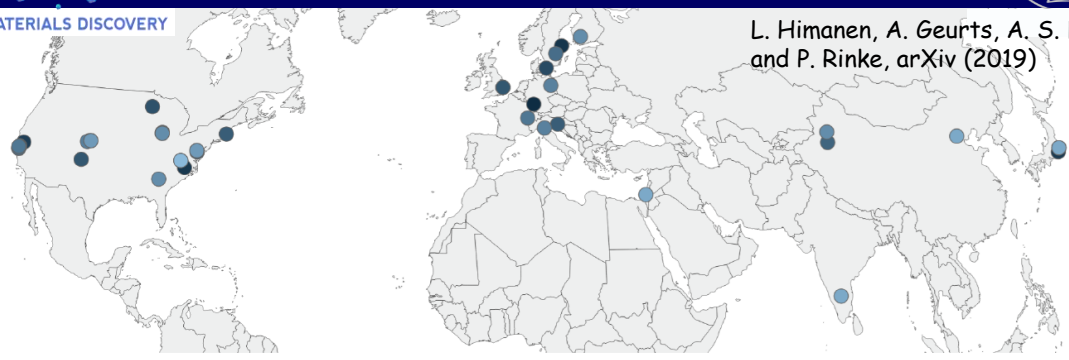
- The number of possible materials is practically infinite
- **No way to cover this huge structural and chemical compound space by high-throughput screening**
- The interesting materials are very few (they are statistically exceptional)

NOMAD

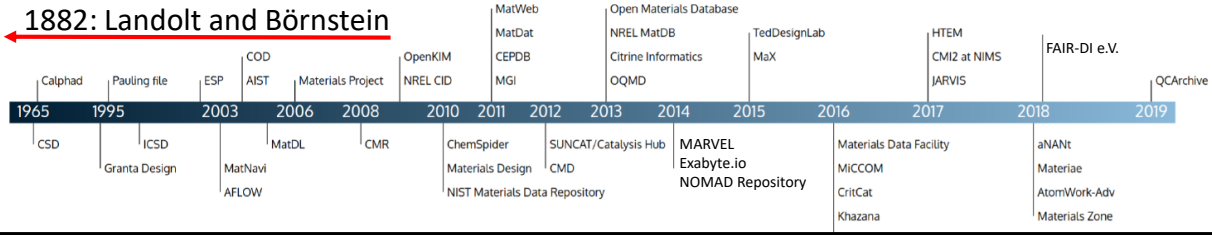
World-Wide Large-Scale Data Projects and Facilities

NOVEL MATERIALS DISCOVERY

L. Himanen, A. Geurts, A. S. Foster, and P. Rinke, arXiv (2019)



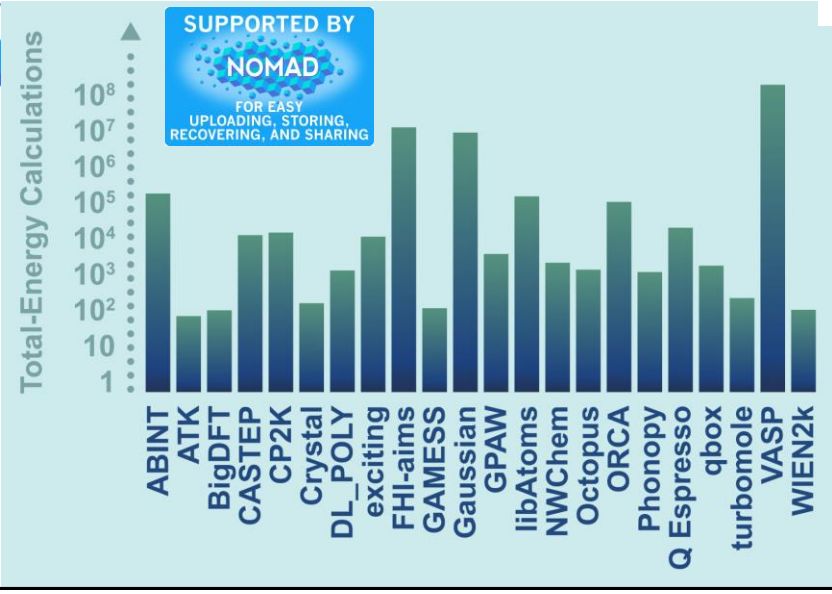
1882: Landolt and Börnstein



The NOMAD Repository
>50 Mio. Total-Energy Calculations

NOMAD
NOVEL MATERIALS DISCOV

Max Planck Society



The NOMAD Repository
>50 Mio. Total-Energy Calculations

NOMAD
NOVEL MATERIALS DISCOVERY

Max Planck Society

90% of the VASP files are from

AFLOW

S. Curtarolo

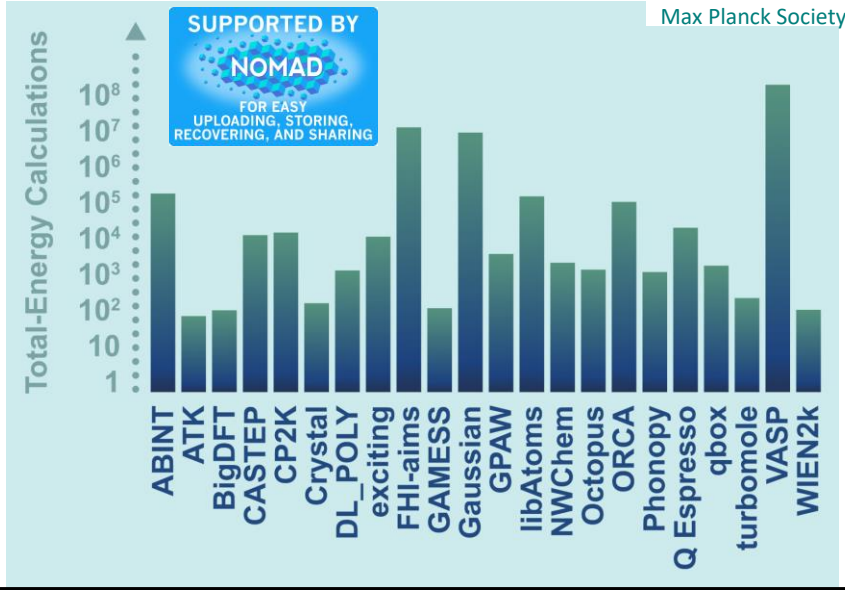
OQMD

C. Wolverton

Materials Project

G. Ceder

K. Persson



NOMAD
NOVEL MATERIALS DISCOVERY

What Is Needed for A FAIR Data Infrastructure?

Max Planck Society

- Scientific results are only meaningful and worth keeping if they are fully characterized and all individual steps are fully documented.
- Computed data are only meaningful when method, approximations, code, code version, and all computational parameters are known.
- For experimental data, we need a full characterization of the sample, the description of the apparatus, the measurement conditions, and the measured quantity.

FAIR-DI

This requires metadata, ontologies, and workflows.

We also need good search engines, an "encyclopedia" GUI, and appropriate hardware.

NOMAD
NOVEL MATERIALS DISCOVERY

What Do We Know?

Max Planck Society

About **250,000** inorganic compounds have been synthesized so far. ***Many more are possible.***

And what do we know about them?

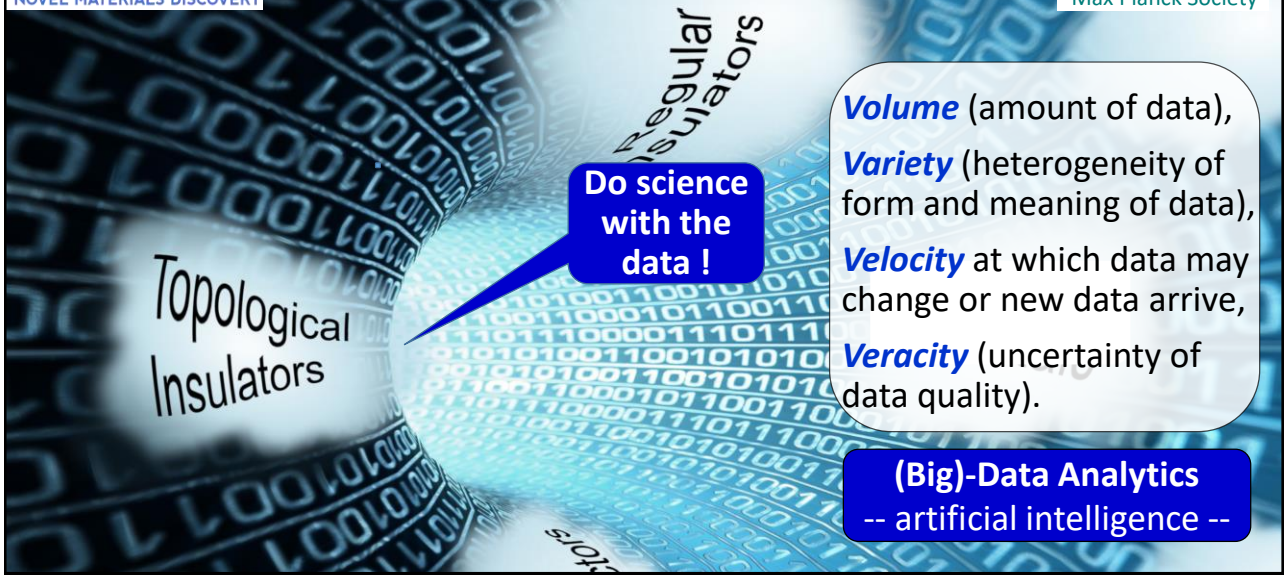
Elastic constants	for ~ 200 compounds
Dielectric constants	for 300-400 compounds
Heat conductance	for ~ 200 compounds
Superconductivity, T_C	for ~ 4,500 compounds
Topological insulators	in 3D: 42, in 2D: 7 compounds
Catalysis, TOF(T, p)	for ??? compounds

For almost every property we are below 1% in coverage ….

NOMAD
NOVEL MATERIALS DISCOVERY

The Big-Data Challenge

Max Planck Society



Volume (amount of data),

Variety (heterogeneity of form and meaning of data),

Velocity at which data may change or new data arrive,

Veracity (uncertainty of data quality).

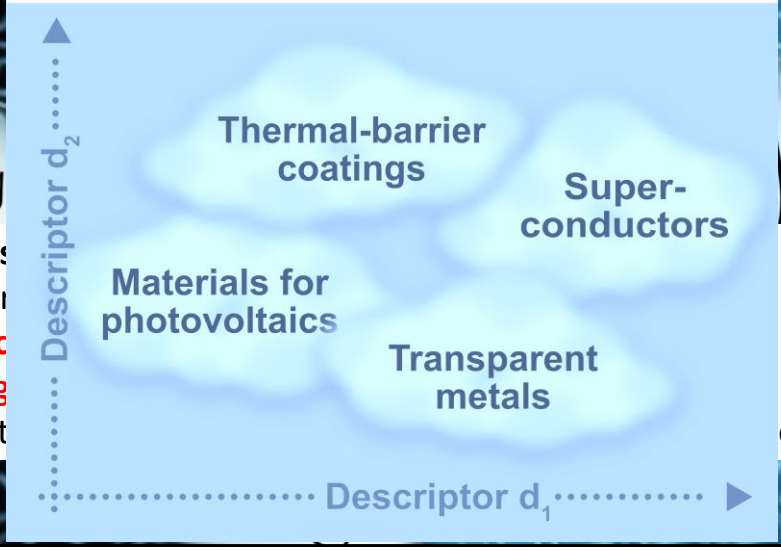
(Big)-Data Analytics
-- artificial intelligence --

NOMAD
NOVEL MATERIALS DISCOVERY

Findable **A**ccessible **I**nteroperable **R**eusable

Max Planck Society

Create maps of materials properties using AI methods/concepts



The materials s

- The numbe
- **No way to c … space by high-throug**
- The interest … eptional)

Developments in many areas – many tools

Cheminformatics
Computer Science
Knowledge Graphs
Clustering
Compressed Sensing
Linguistics
Neural Networks
Signal Processing
Subgroup Discovery
Social media
Kernel Ridge Regression
Robotics
Bioinformatics
Support Vector Machines
Random Forest
Intelligent Systems
and more …

NOMAD
NOVEL MATERIALS DISCOVERY

Learning from "Big" Data: Very Many Methods and Concepts, Very Interdisciplinary

Max Planck Society

Artificial Intelligence (AI)
Any technique that enables computers to mimic human intelligence, using logical if-then rules, compressed sensing, decision trees, machine learning

Machine Learning
A subset of AI that includes statistical techniques that enable machines to improve at tasks with more data

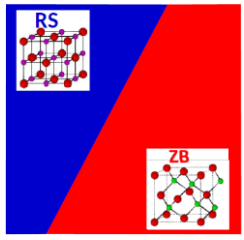
Deep Learning
The subset of machine learning composed of algorithms that permit software to train itself to perform tasks, like speech and image recognition, by exposing multilayered neural networks to vast amounts of data

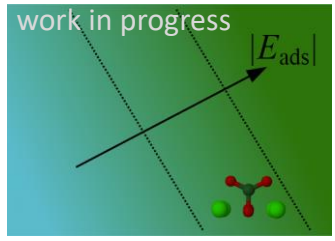
Building Maps of Materials
(Role Models: Periodic Table, Ashby Plots)

NOMAD
NOVEL MATERIALS DISCOVERY

Max Planck Society

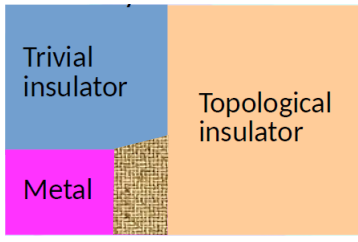
Crystal-structure prediction
- Octet binaries (ZB vs. RS)
- $Al_xGa_yIn_zO_3$ ($x+y+z=2$)
- Perovskites (Goldschmidt tolerance factor)

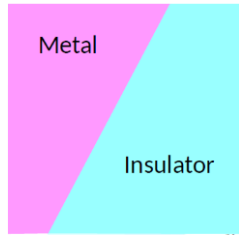




Activation of CO_2 at metal oxides and carbides

Property classification:
- Topological insulators





Property classification:
- Metal vs. insulator

The Materials-Science Challenge Is Different to That of Standard Machine Learning

NOMAD
NOVEL MATERIALS DISCOVERY

Max Planck Society

We are looking for statistically exceptional data groups. This may be needles, or nuts, or bolts, or coins, or … Often, we don't know exactly what we are searching for, except that the data should be statistically exceptional.

FAIR-DI

FAIR Data Infrastructure for Physics, Chemistry, Materials Science, and Astronomy e.V.