NOMAD

NOVEL MATERIALS DISCOVERY

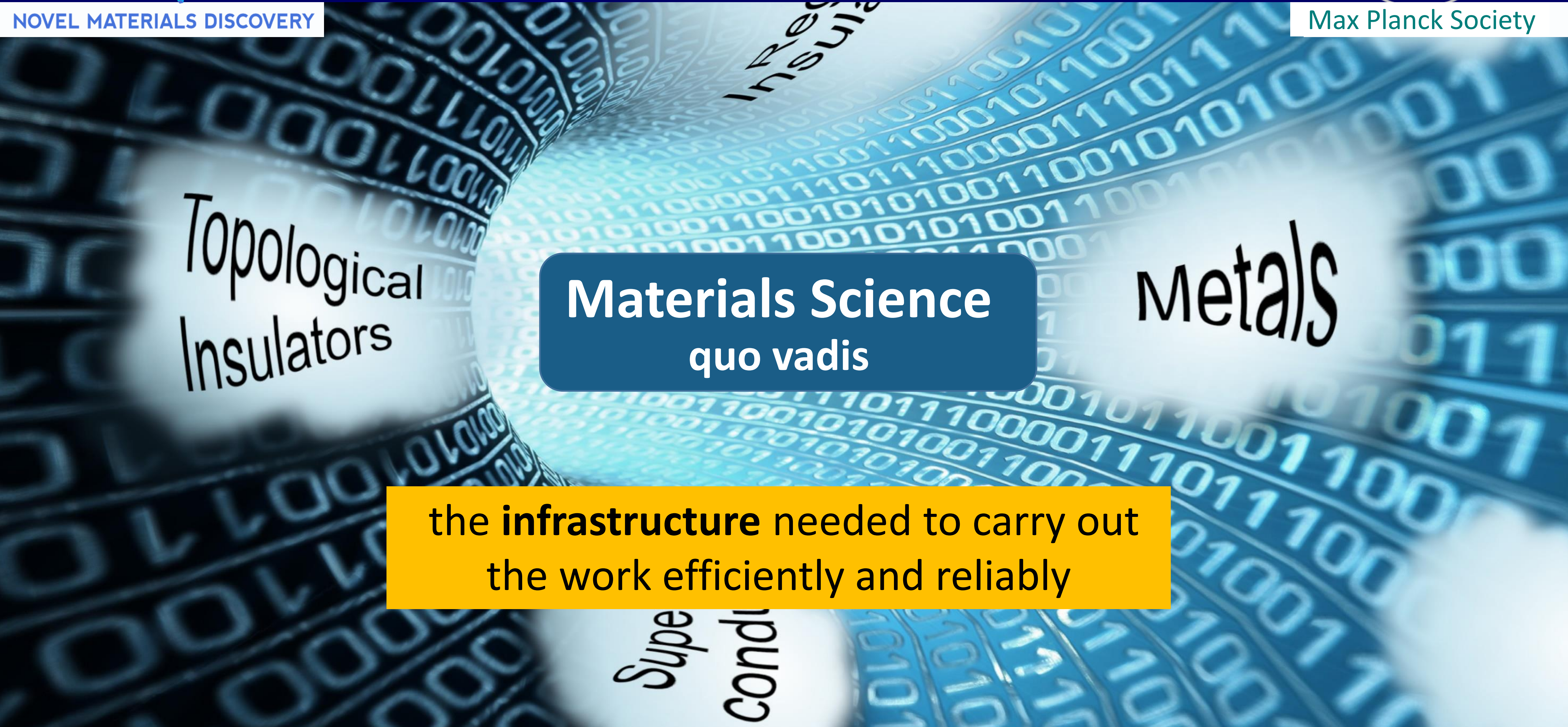
Perspectives in Materials Science

Max Planck Society

Materials Science
quo vadis

Big-Data Driven Materials Science
(The 4th Paradigm of Materials Research)

Perspectives in Materials Science



Materials Science
quo vadis

the **infrastructure** needed to carry out the work efficiently and reliably

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

NOMAD

NOVEL MATERIALS DISCOVERY

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.

$O(10^1)$ compounds selected

Animation by G.-M. Rignanese

NOMAD

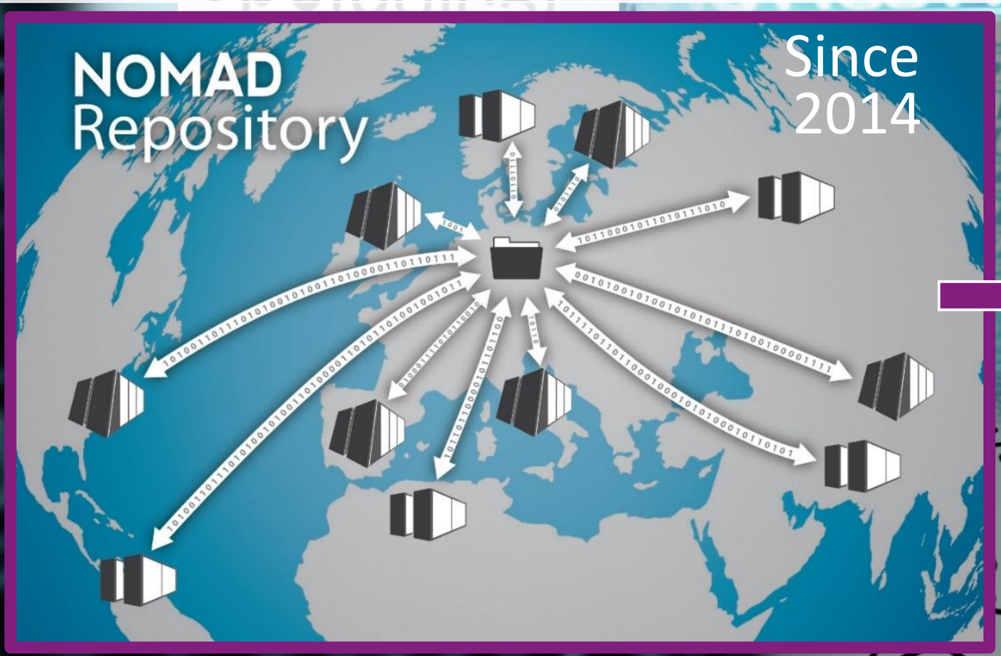
NOVEL MATERIALS DISCOVERY

Findable Accessible Interoperable Reusable

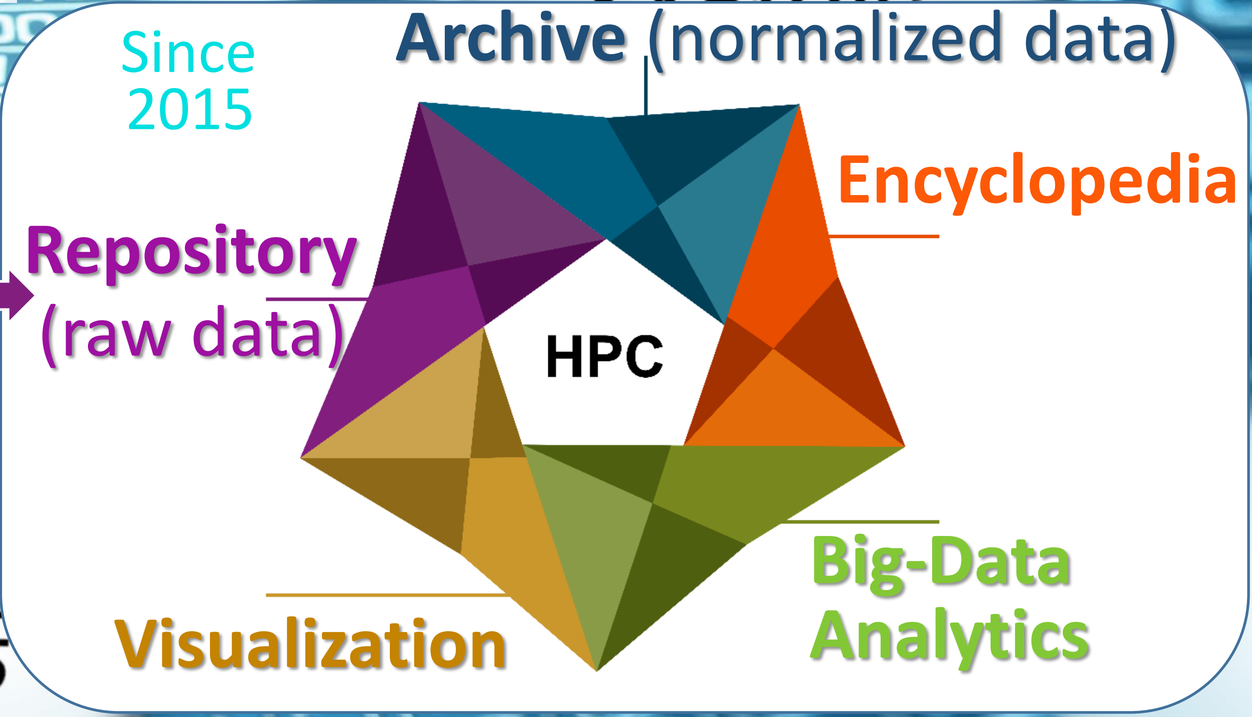
M. D. Wilkinson et al., Scientific Data **3**, 160018 (2016)

Max Planck Society

Requests the full input and output files



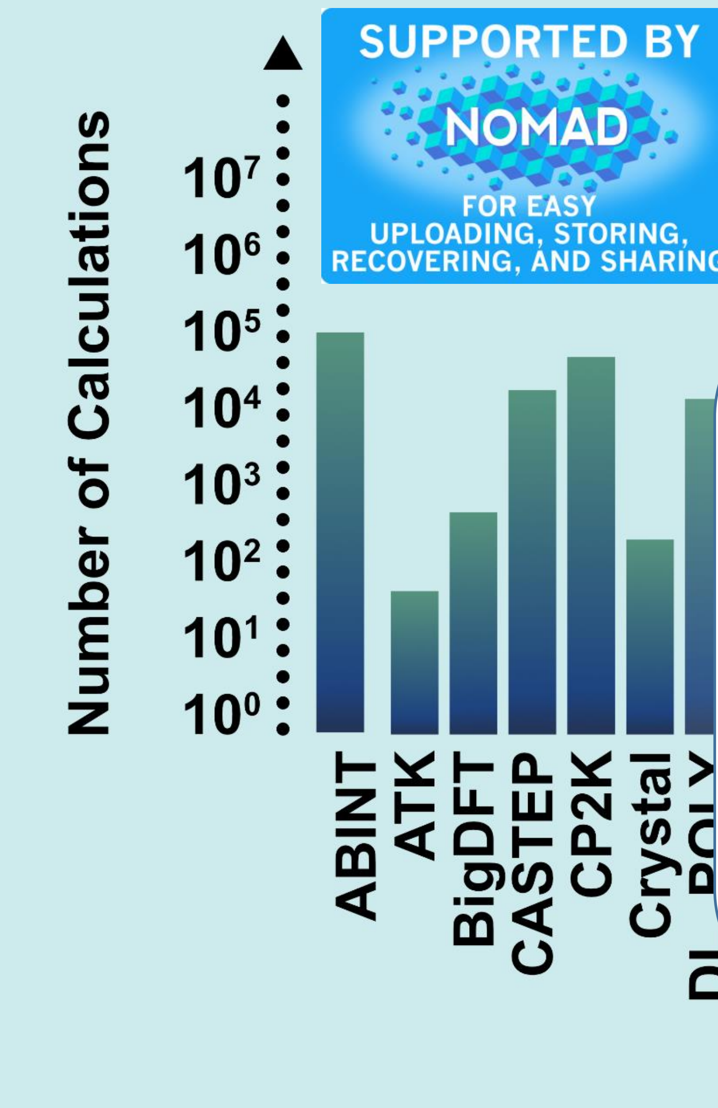
The NOMAD Center of Excellence

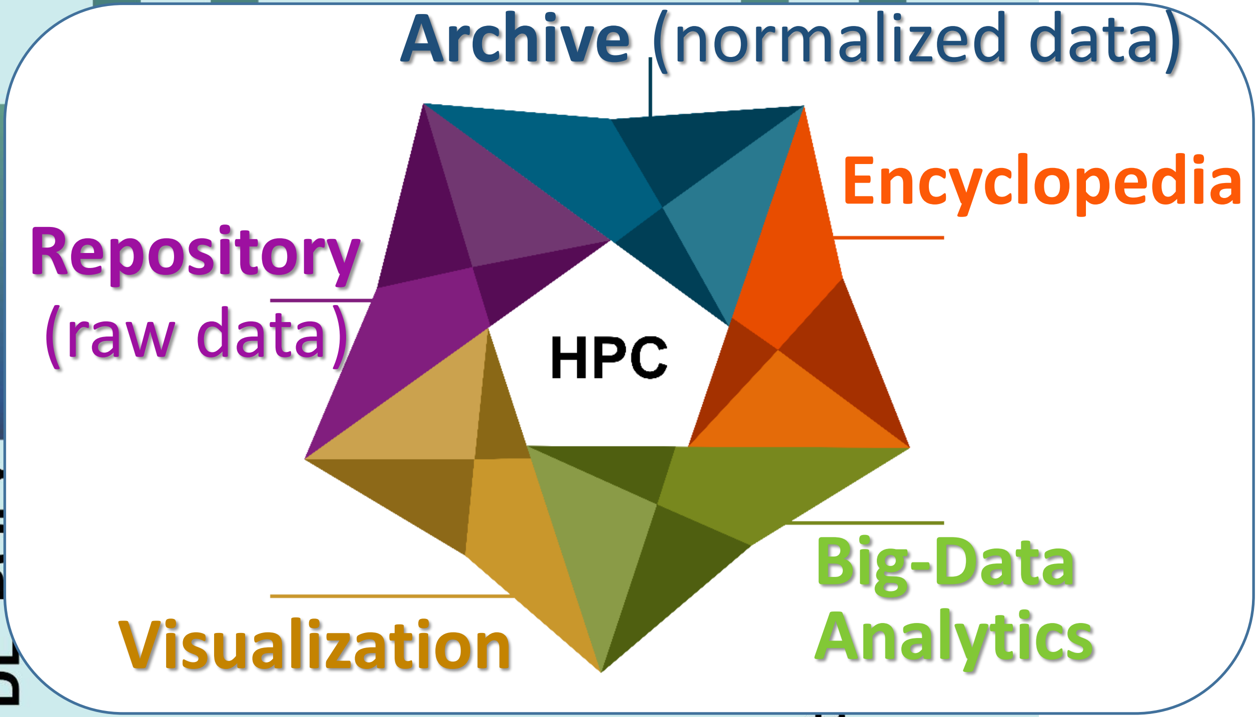


The NOMAD Repository
>100 Mio. DFT Calculations

NOVEL MATERIALS DISCOVERY

Max Planck Society

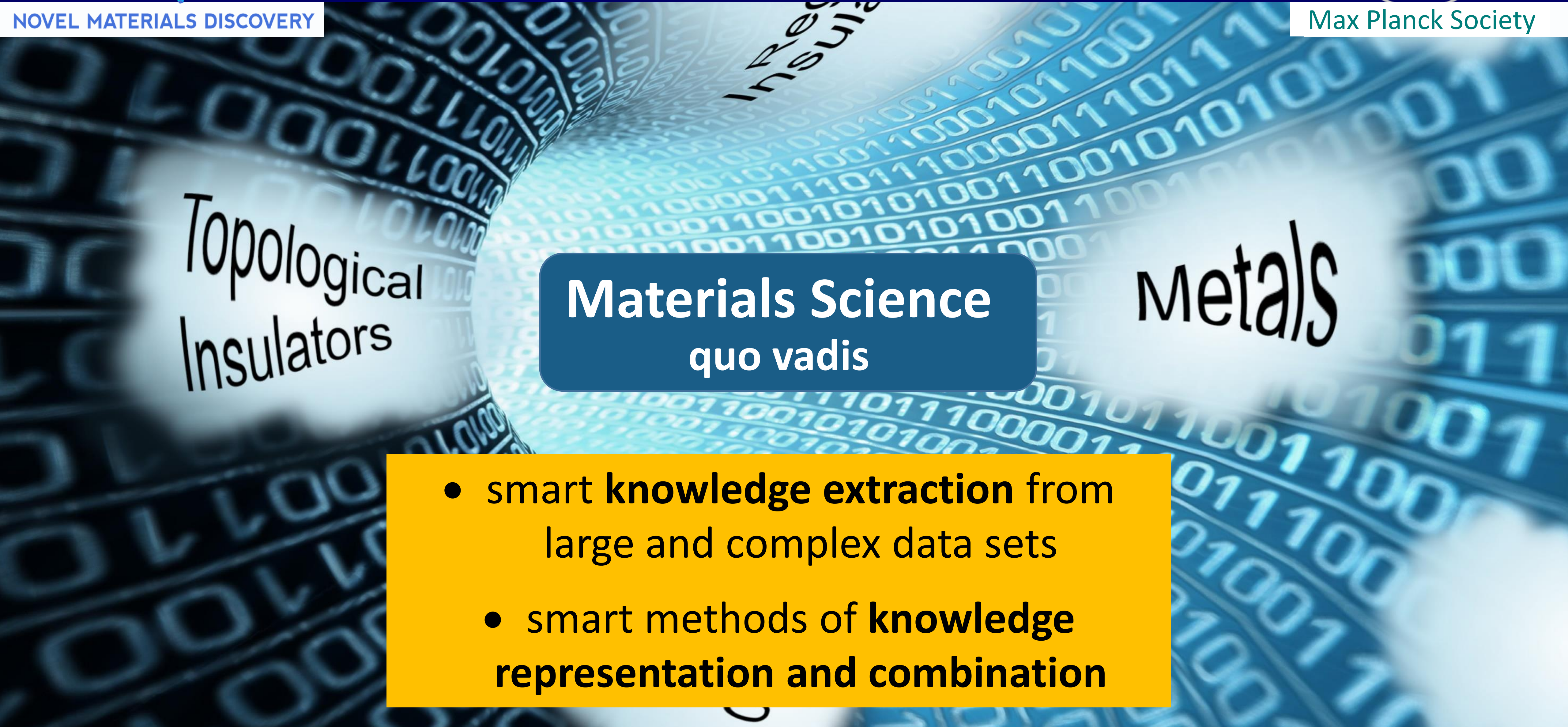




NOMAD
NOVEL MATERIALS DISCOVERY

Perspectives in Materials Science

Max Planck Society



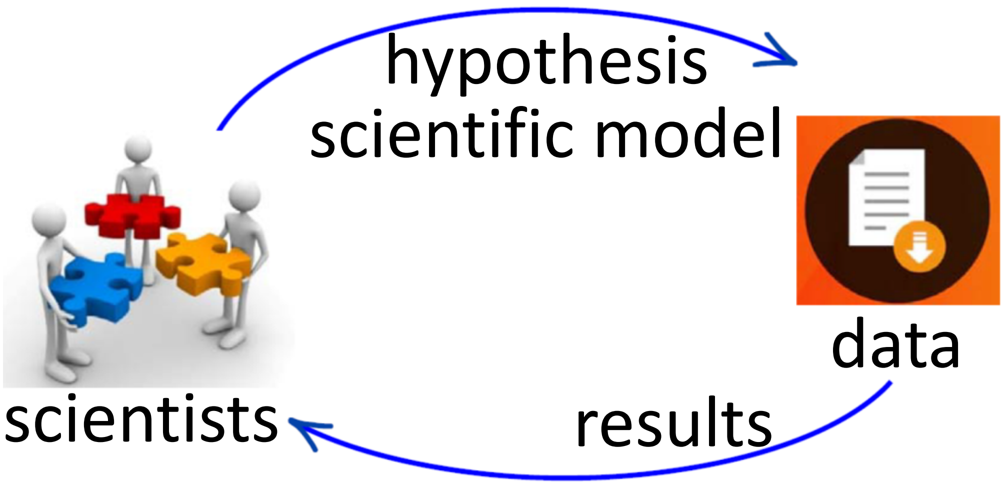
Materials Science
quo vadis

- smart **knowledge extraction** from large and complex data sets
- smart methods of **knowledge representation and combination**

“Maps of Materials Properties”

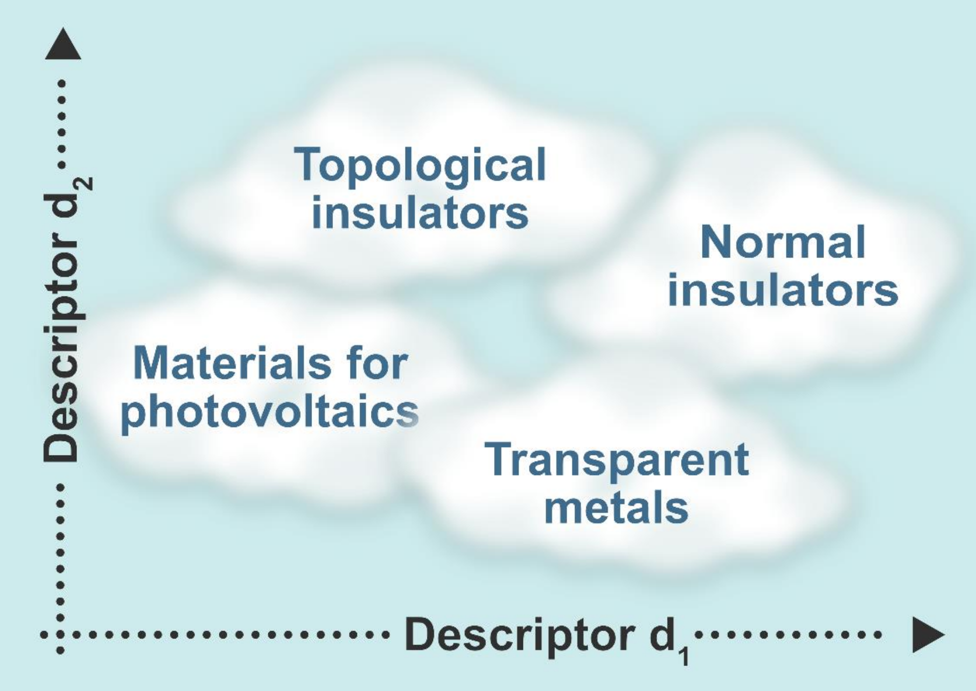
Adaptivity in science and data analysis

question



Feedback loop defeats
all available statistics

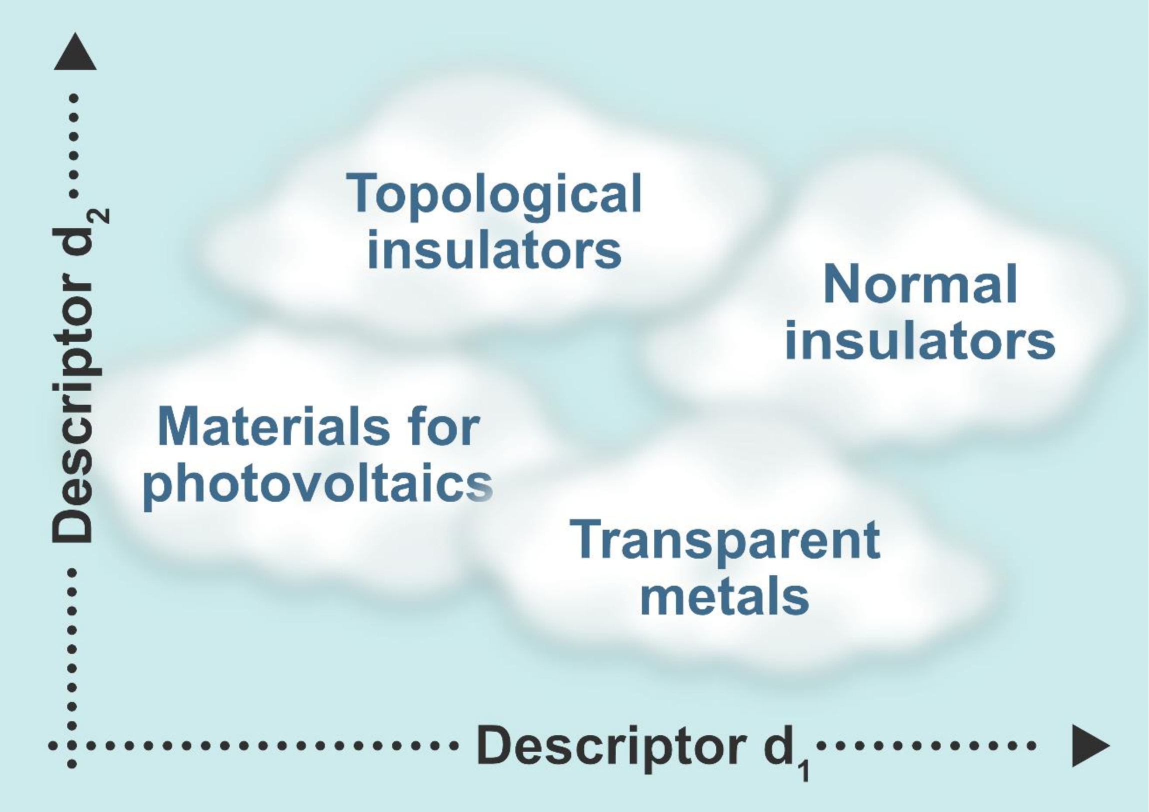
Building a “map of materials properties” from data, without a scientific model



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

NOMAD
NOVEL MATERIALS DISCOVERY

Max Planck Society



The science is in the determination (and understanding) of the descriptors. How to find them by statistical learning?

"Invert" the big-data problem:

Construct billions of *possible* descriptors and evaluate their values for N materials.

Find the strongest correlations of these N billion data points with computed (or experimental) results (properties) of the N studied materials. – *Sure Independent Screening and Sparsifying Operator (SISSO) by Ouyang, Ghiringhelli et al. (2018).*

Building Maps of Materials

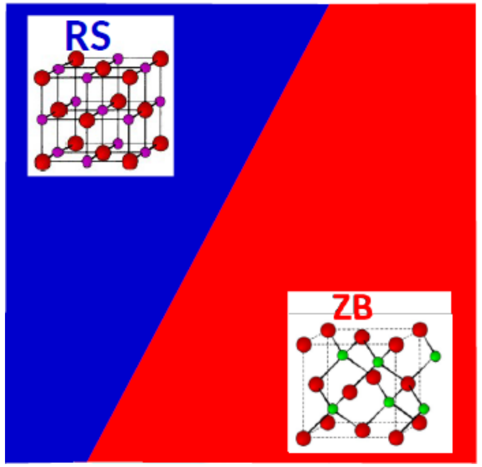
(Role Models: Periodic Table, Ashby Plots)

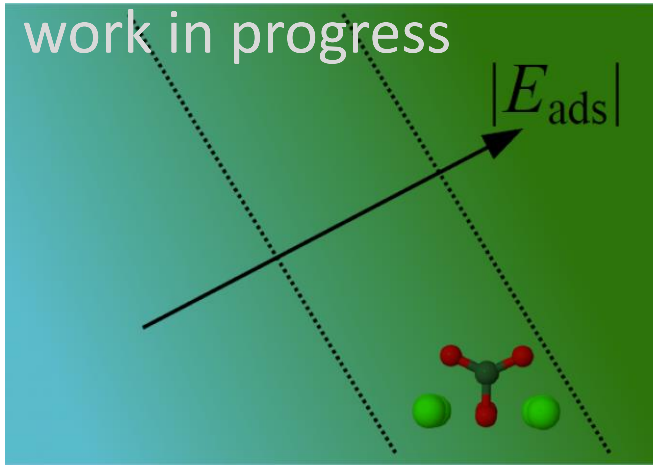
NOVEL MATERIALS DISCOVERY

Max Planck Society

Crystal-structure prediction

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

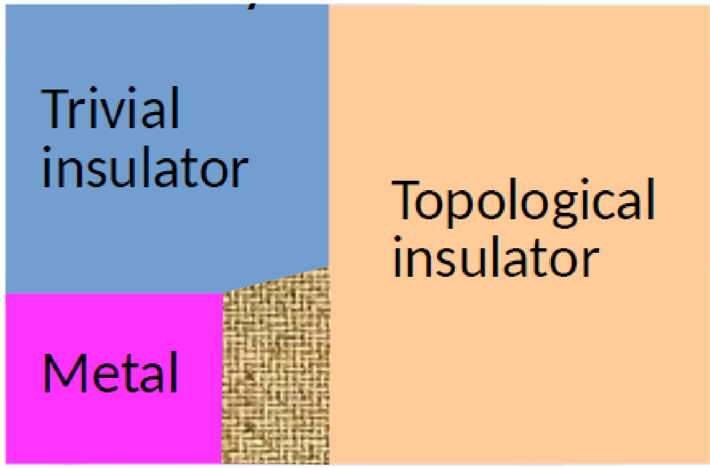


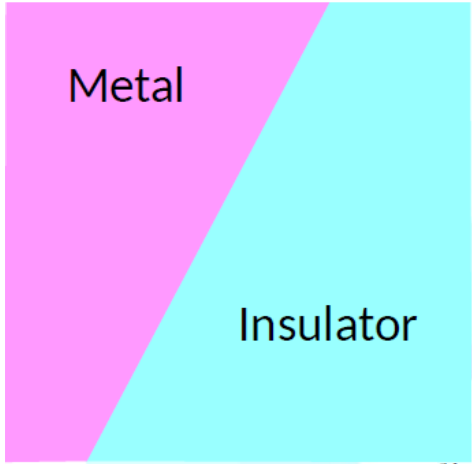


Activation of CO_2 at metal oxides and carbides

Property classification:

- Topological insulators



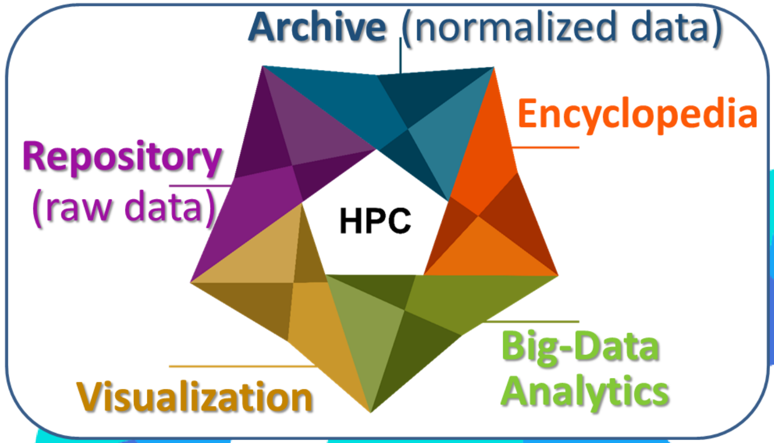


Property classification:

- Metal vs. insulator

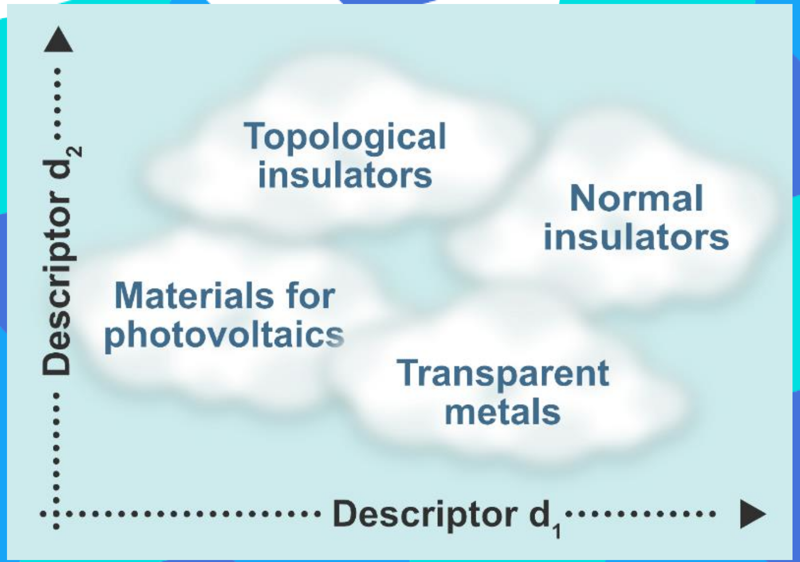
The Materials-Science Challenge Is Different to That of Standard Machine Learning

The **infrastructure** needed to carry out the work efficiently and reliably



Smart **knowledge extraction** from large and complex data sets

Smart methods of **knowledge representation and combination**



NOMAD

NOVEL MATERIALS DISCOVERY

Next Steps

Max Planck Society

FAIR-DI

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

https://www.fair-di.eu

For a **F**ederated, **AI**–**R**eady DI

Video Conference on a
FAIR Data Infrastructure for Materials Genomics
(3-5 June, 2020)
A fantastic line-up of speakers

We need

- a federated system for data, as data are getting too large for been transported
- new (different) AI tools: domain specific AI; outlier detection, local vs. global models; bring codes to (federated/ distributed) data
- "complete" characterization of the measurement conditions by metadata (basis for reproducibility)
- to avoid (much) more load on the researcher (make metadata assignment and data standards automatic)
- IT and practical AI specialists