Status of Data Science in Materials Science

Karsten Albe

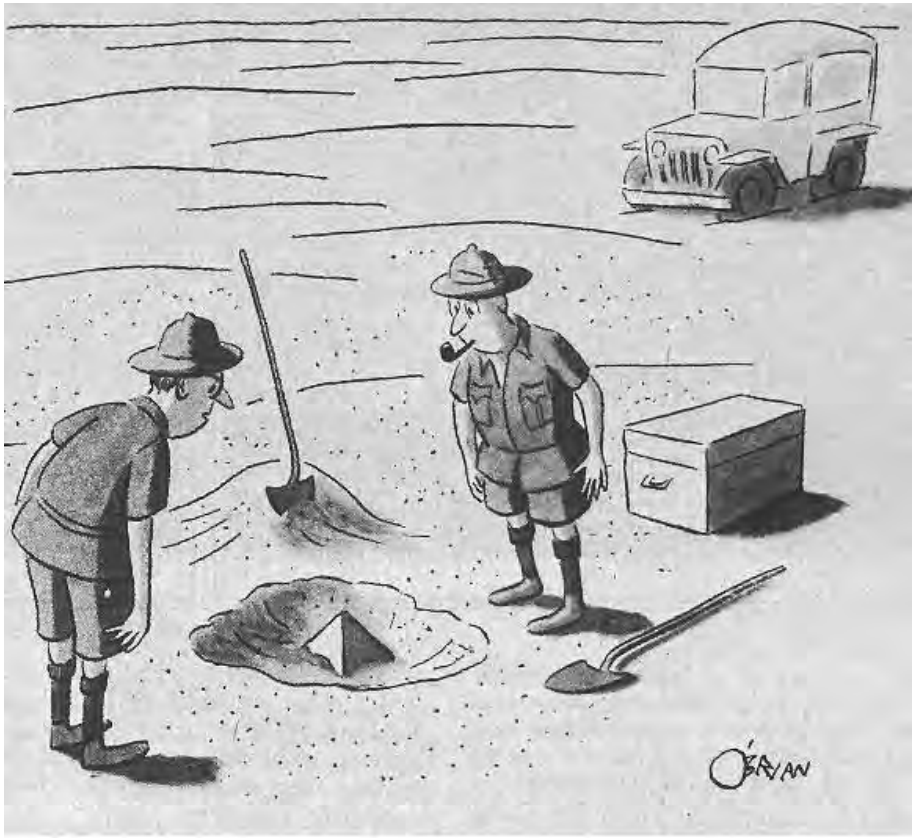
FG Materialmodellierung,
Technische Universität Darmstadt

Content

TECHNISCHE UNIVERSITÄT DARMSTADT

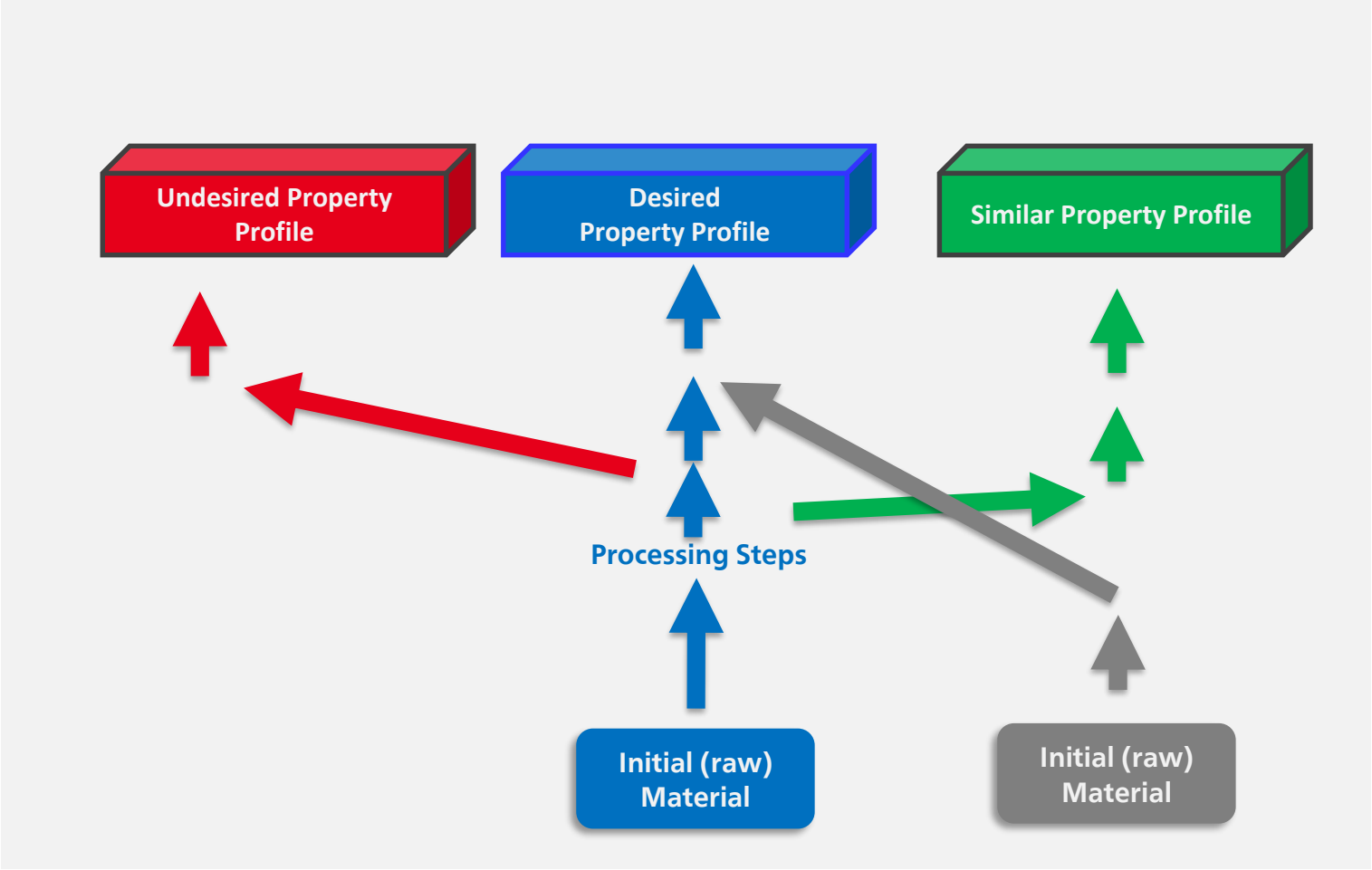
- What is Materials Science ?
- Data from Combinatorial Search
- Data from Large Scale Simulations
- Data from Experiment
- What needs to be done?

Typical Material Scientist



"This could be the discovery of the century. Depending, of course, on how far down it goes."

Materials Development



Materials Development Continuum

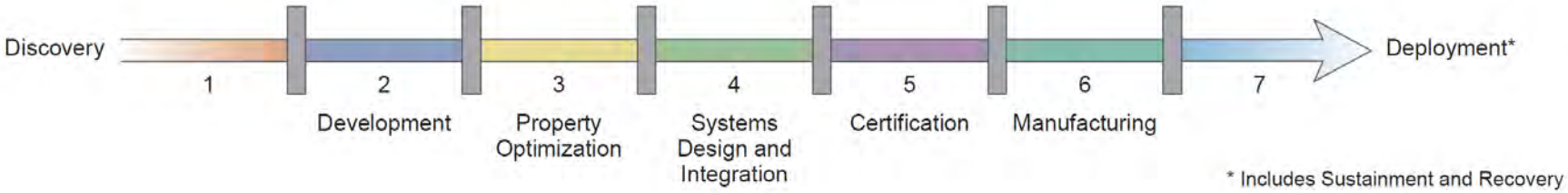


Table 1 | Commercialization time, costs and uncertainties for software, biotechnology and advanced-materials ventures.

	Development time (years)	Research and development costs* (US$ millions)	Commercialization costs† (US$ millions)	Technology uncertainty	Market uncertainty
Software	0–2	0–3	1–10	Low	Medium
Biotechnology	10–15	5–10	300–900	Very high	Medium
Advanced materials	5–15	2–20	50–500	High	High

The reported development times and research and development (R&D) costs are based on interviews and our experience, as well as on published case studies involving advanced materials[2,5,10,34,38], biotechnology[39–42] and software development[43–47]. *R&D costs refer to software development costs up to a first commercial release, to the cost of preclinical trials and to lab-scale R&D costs for a new material and its potential applications. †Commercialization costs refer to the necessary investment to commercialize an invention, which can involve pilot plants, iterative development, process scale-up, clinical trials, regulatory approval, marketing and distribution.

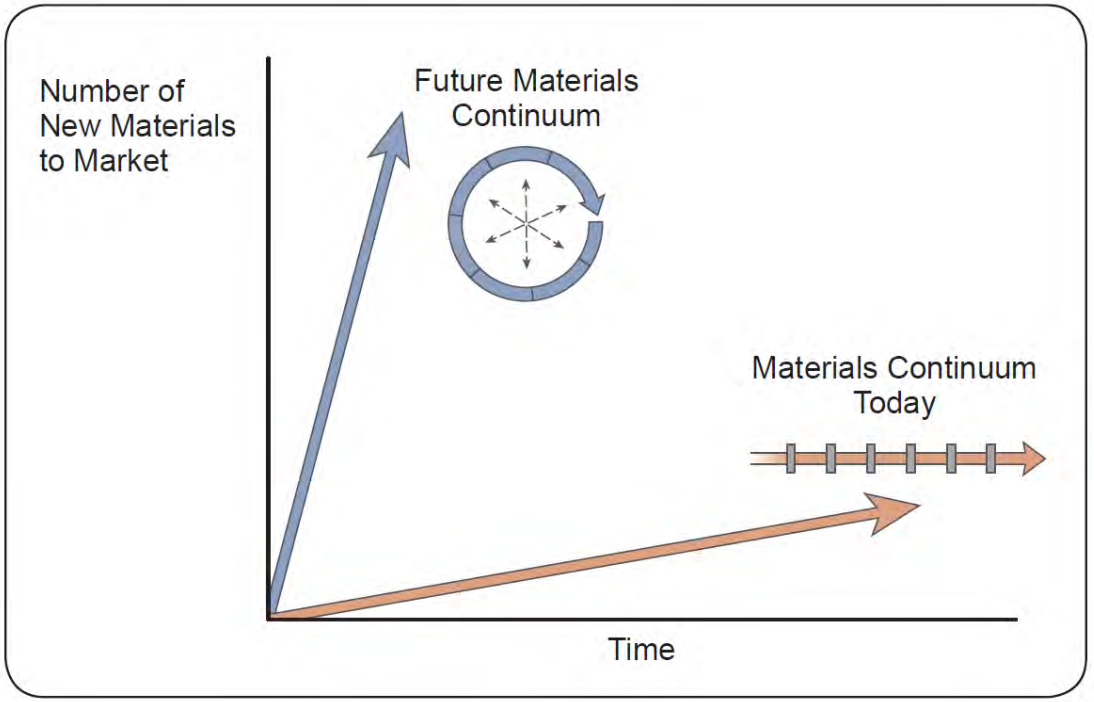
NATURE MATERIALS | VOL 15 | MAY 2016 | www.nature.com/naturematerials

Materials Development Continuum

TECHNISCHE UNIVERSITÄT DARMSTADT

Discovery

1



Deployment*

es Sustainment and Recovery

Table 1 | Commercializati ventures.

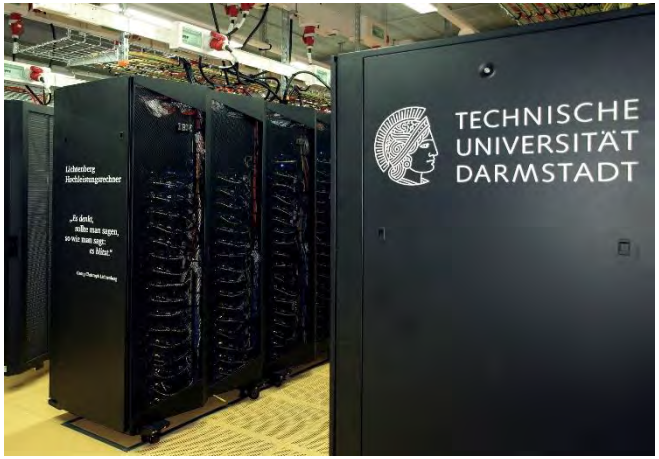
	Dev	tainty	Market uncertainty
Software	0–2		Medium
Biotechnology	10–1		Medium
Advanced materials	5–15		High

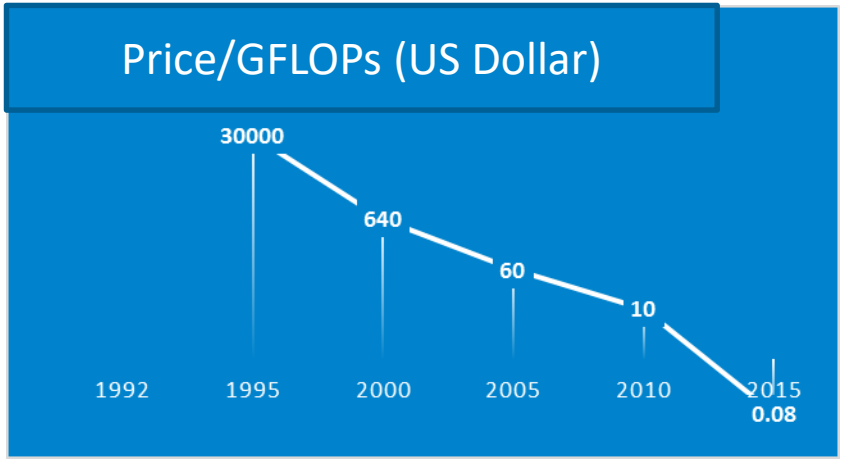
The reported development times and re… d materials[2,5,10,34,38], biotechnology[39–42] and software development[43–47]. *R&D costs refer to software development costs up to a first commercial release, to the cost of preclinical trials and to lab-scale R&D costs for a new material and its potential applications. †Commercialization costs refer to the necessary investment to commercialize an invention, which can involve pilot plants, iterative development, process scale-up, clinical trials, regulatory approval, marketing and distribution.

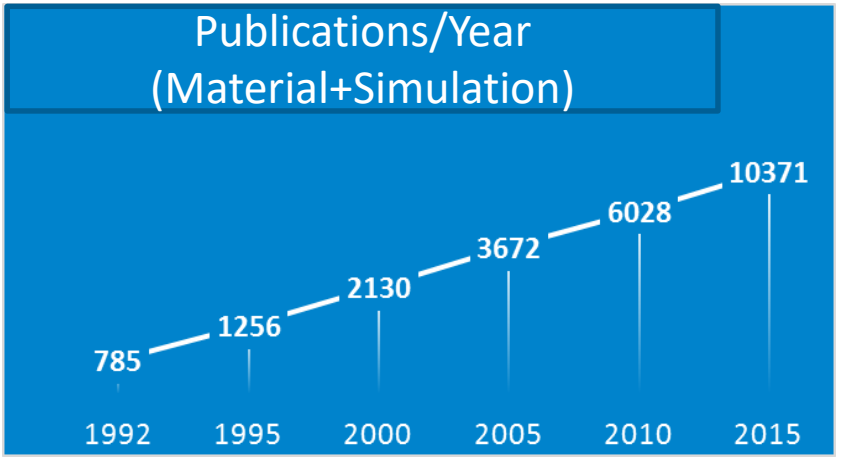
NATURE MATERIALS | VOL 15 | MAY 2016 | www.nature.com/naturematerials

Computational Materials Science

TECHNISCHE UNIVERSITÄT DARMSTADT

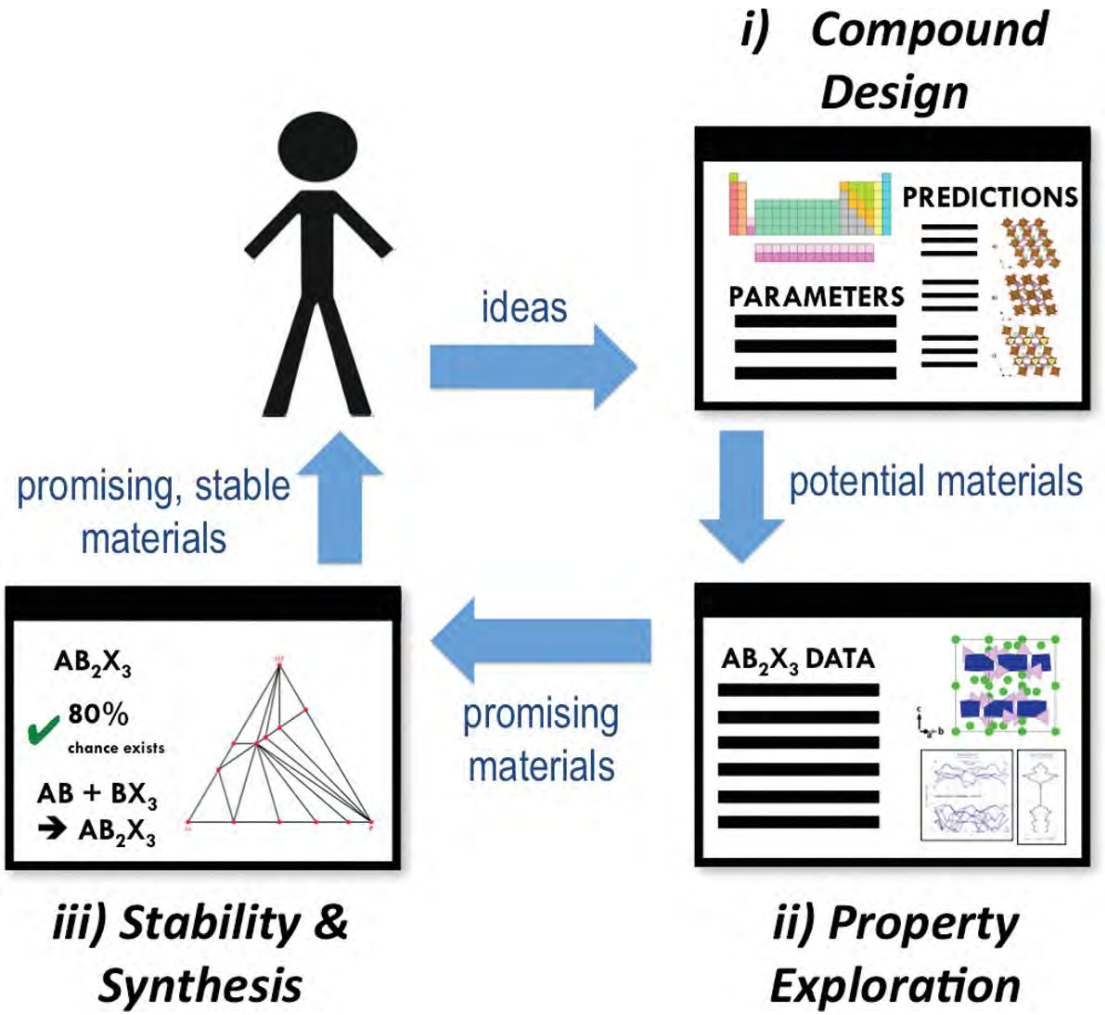






Data Driven Materials Design

TECHNISCHE UNIVERSITÄT DARMSTADT



APL Materials **2013,** 1, DOI: 10.1063/1.4812323

Materials Genome Project: Number of predictied compounds

TECHNISCHE UNIVERSITÄT DARMSTADT

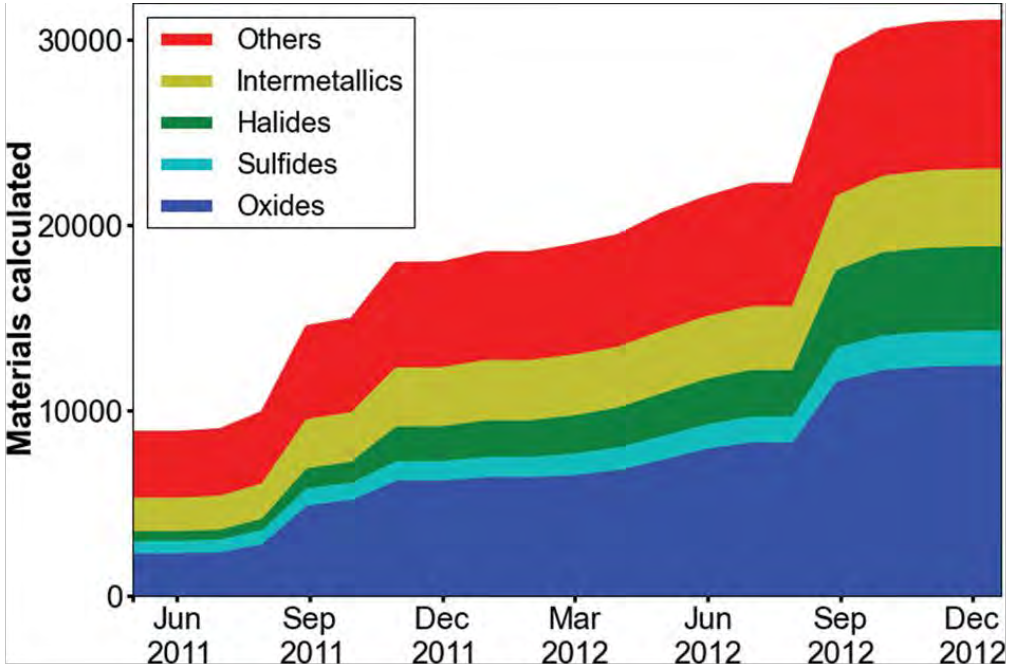
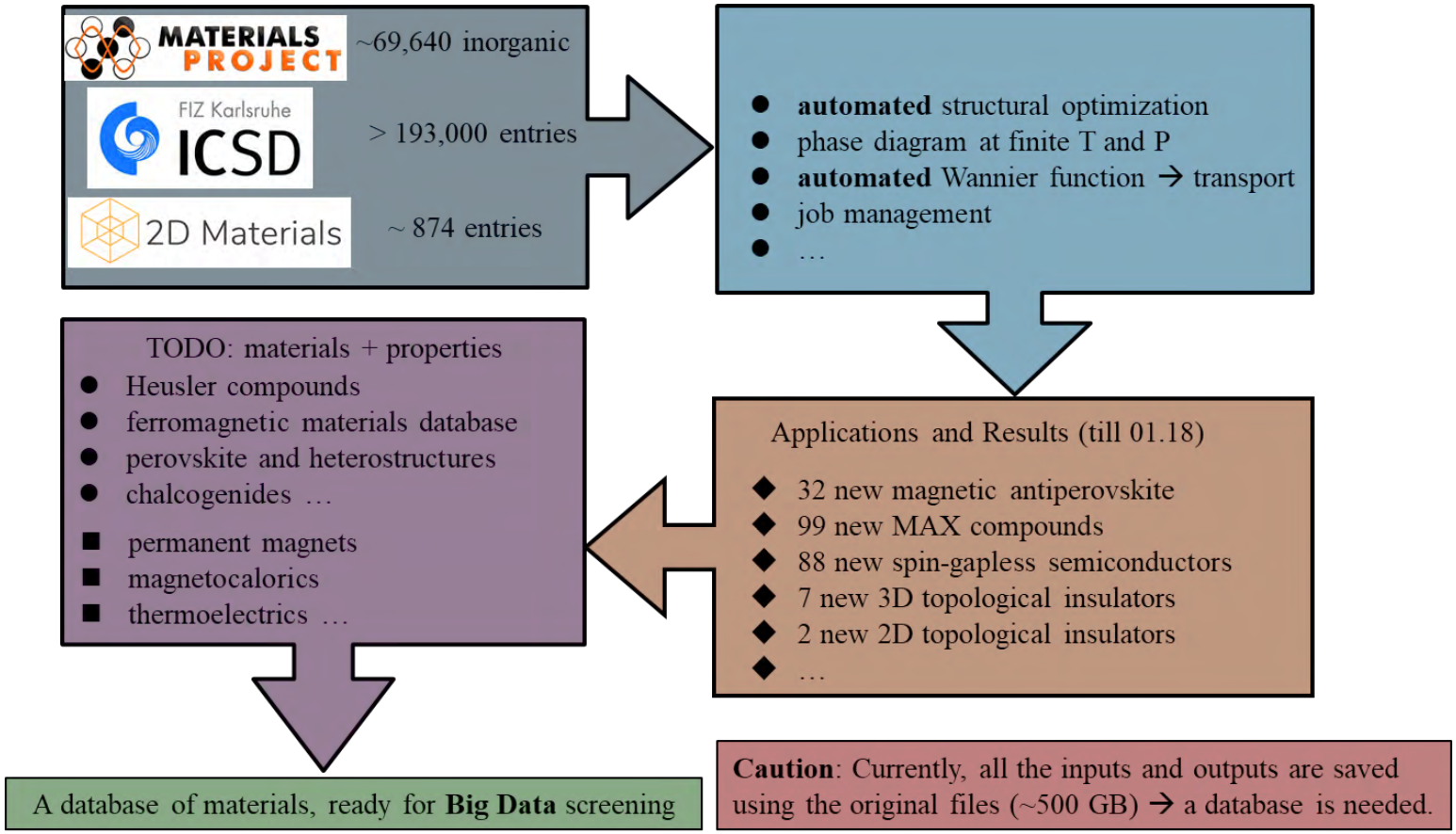


FIG. 2. Number of compounds available on the Materials Project web site since the initial release in October 2011, broken down by type of compound.

Published in: Anubhav Jain; Shyue Ping Ong; Geoffroy Hautier; Wei Chen; William Davidson Richards; Stephen Dacek; Shreyas Cholia; Dan Gunter; David Skinner; Gerbrand Ceder; Kristin A. Persson; *APL Materials* **2013,** 1,
DOI: 10.1063/1.4812323

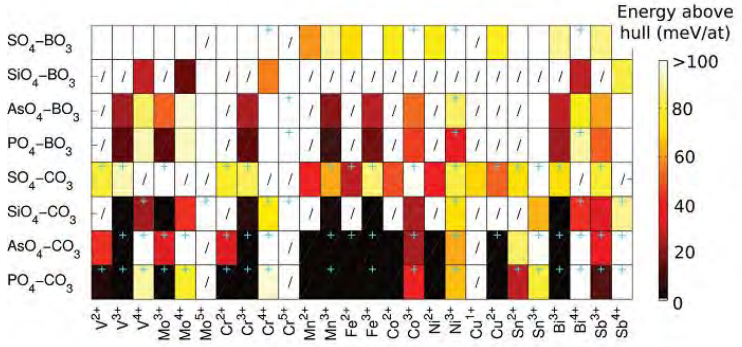
High Troughput@TUD /Prof. Zhang

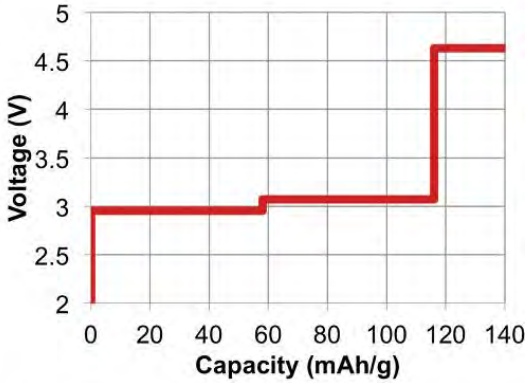


Example: Battery Materials

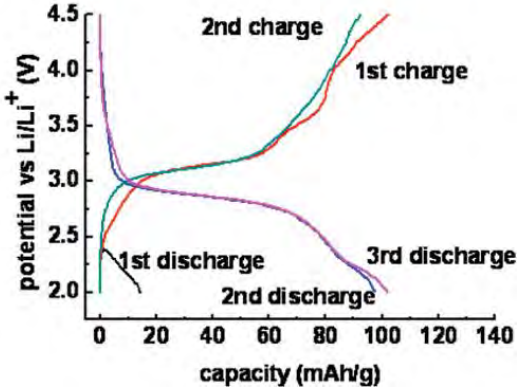
TECHNISCHE UNIVERSITÄT DARMSTADT

THEORY





EXPERIMENT

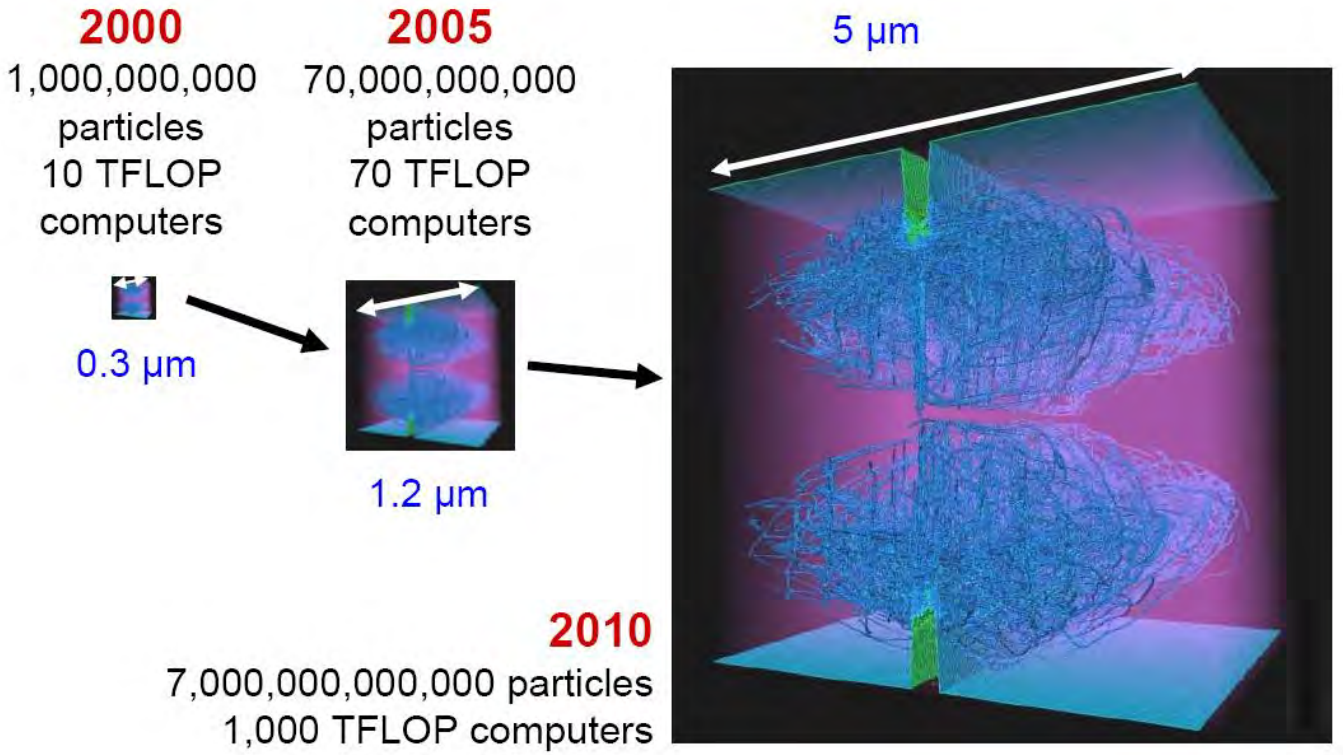


APL Materials **2013,** 1, DOI: 10.1063/1.4812323

What would Boltzmann have done with a computer?

Increase in Computational Power by Parallelization

TECHNISCHE UNIVERSITÄT DARMSTADT



(Abraham et a., 2002, Buehler, Hartmaier *et al.*, 2004)

Data filtering and compression are key

TECHNISCHE UNIVERSITÄT DARMSTADT

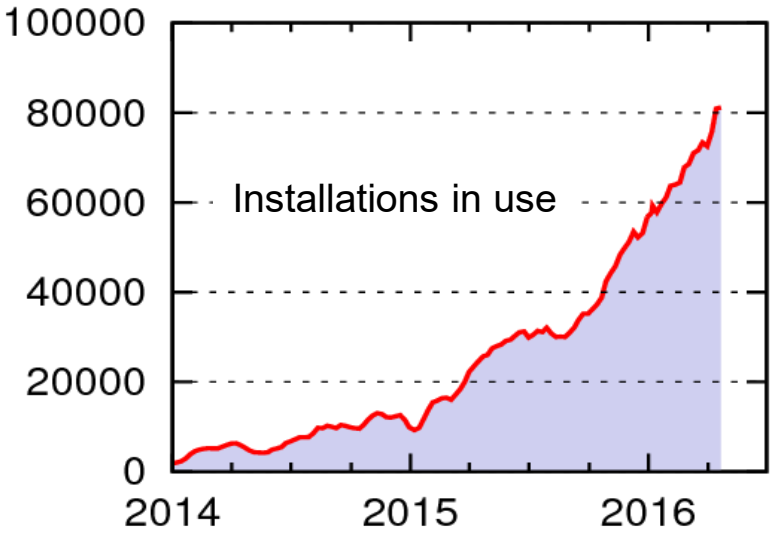
One day of metal strength MD simulation on the Sequoia supercomputer would produce trajectory data worth:

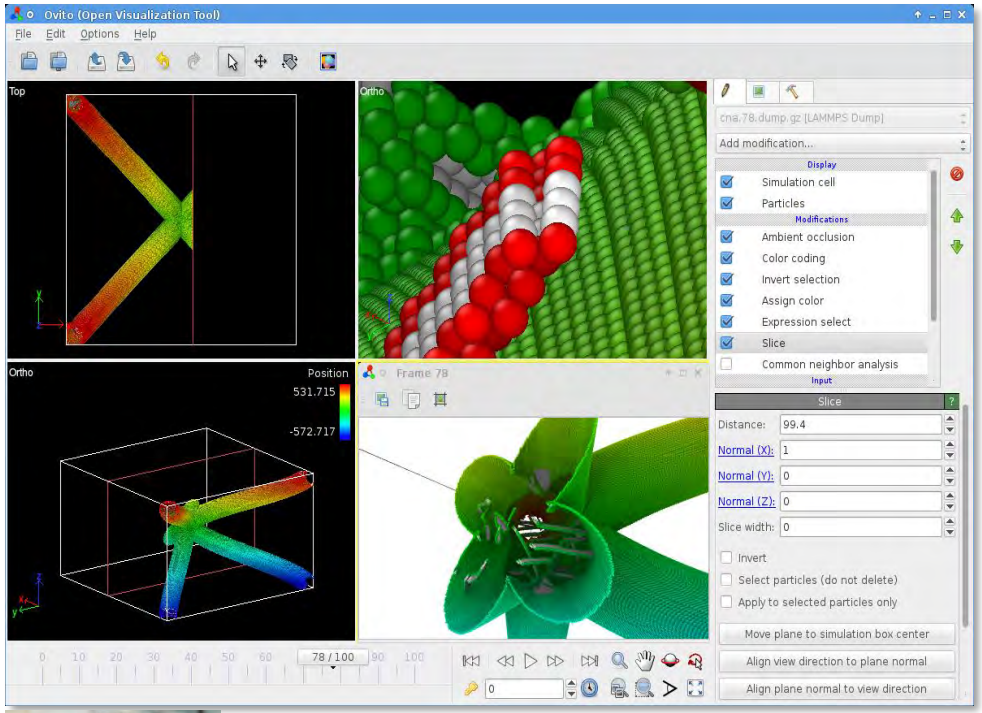
10^{18} bytes = 1000 petabytes = 1 exabyte

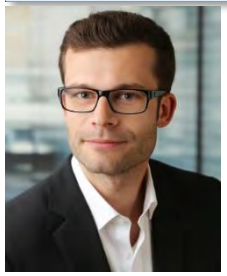
OVITO – *The Open Visualization Tool*

TECHNISCHE UNIVERSITÄT DARMSTADT

The leading visualization and analysis software for MD simulation data





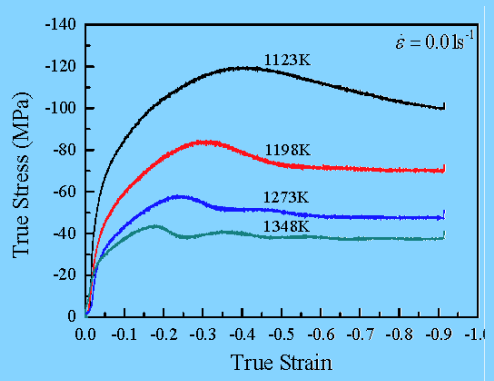


Modelling Simul. Mater. Sci. Eng. 18 (2010), 015012

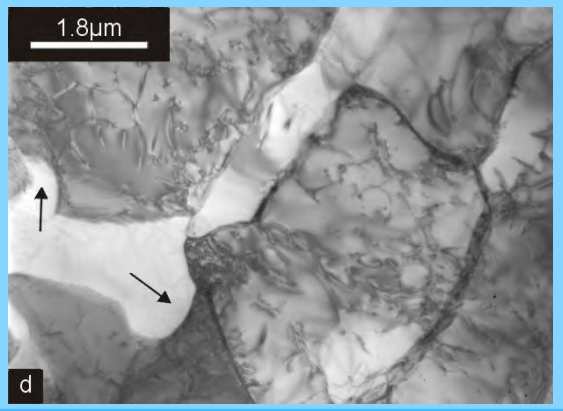
Visualization and data analysis – from atomistic simulations to insights

TECHNISCHE UNIVERSITÄT DARMSTADT

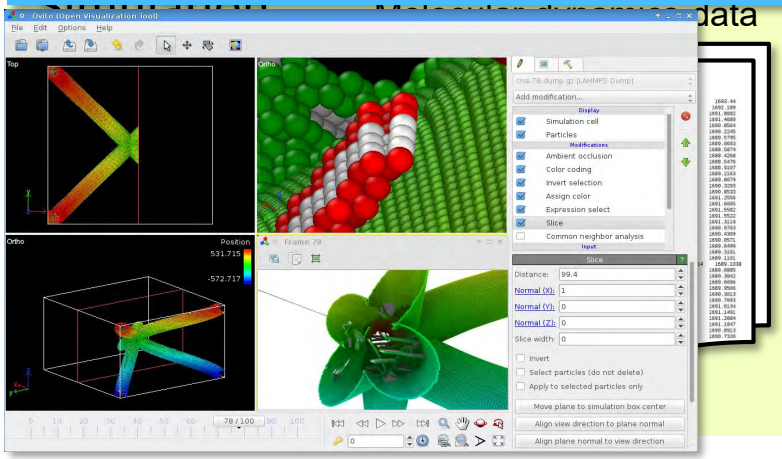
Experiment



Microscopy & Imaging



OVITO – The Open Visualization Tool



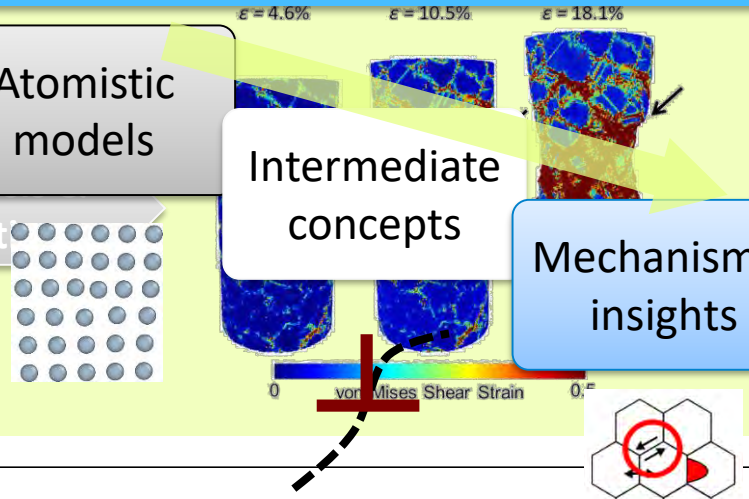
Data a… Visualizat…

Computational analysis methods

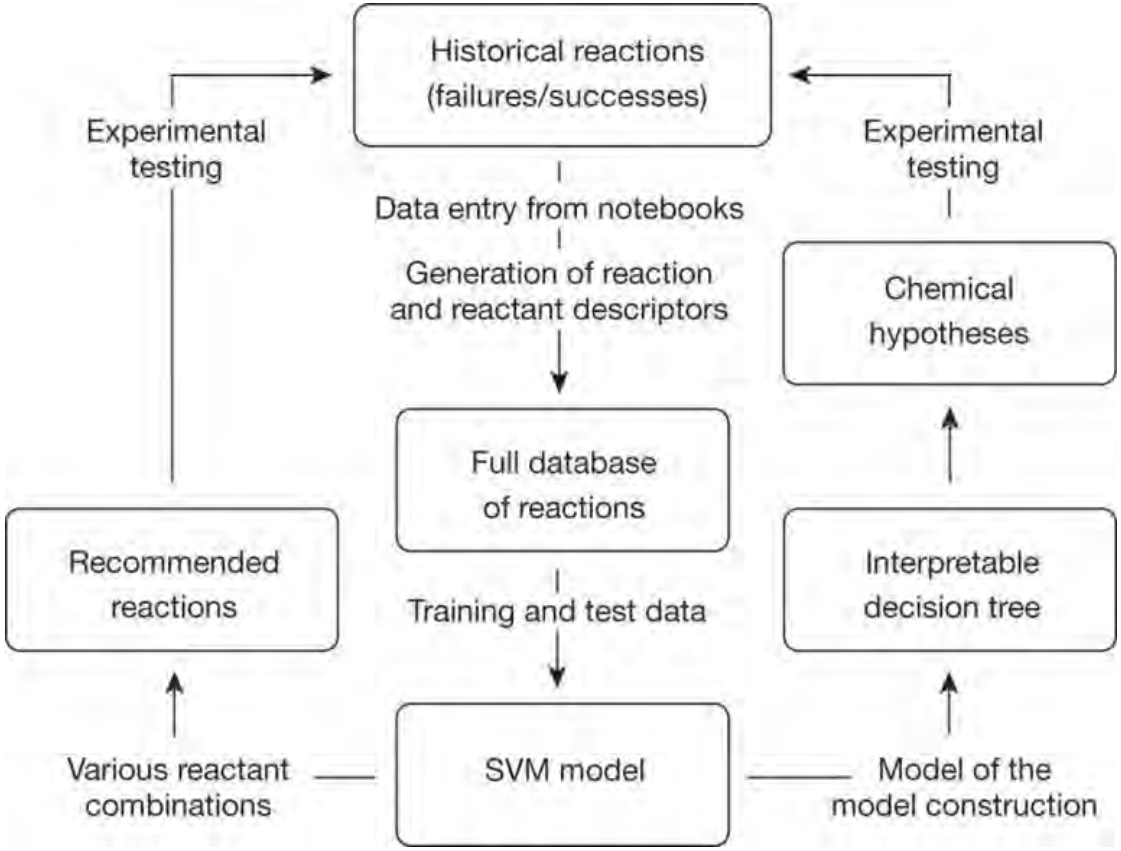
Atomistic models

Intermediate concepts

Mechanisms/ insights



Machine-learning-assisted materials discovery



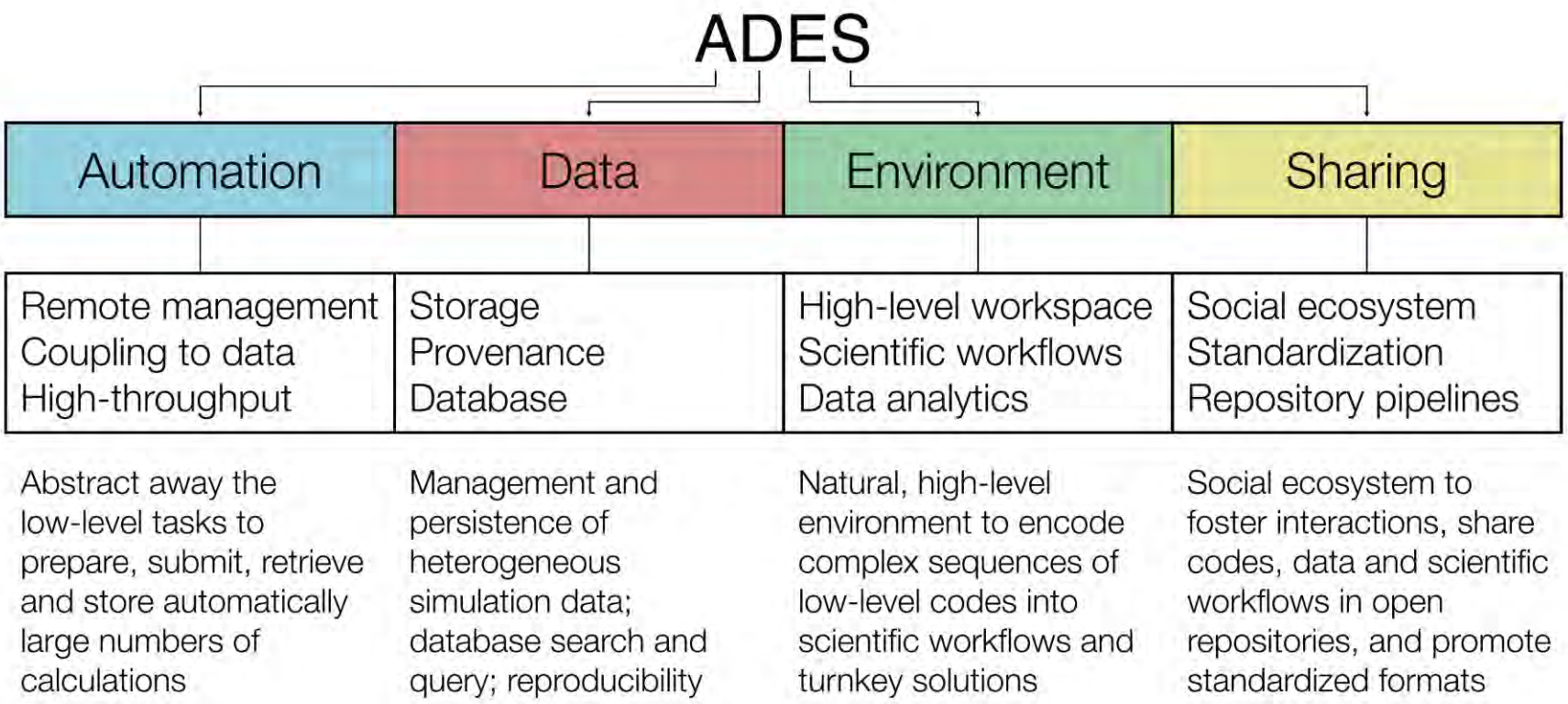
P Raccuglia *et al. Nature* **533,** *73*–76 (2016) doi:10.1038/nature17439

TECHNISCHE UNIVERSITÄT DARMSTADT

We are not lacking data, but profound strategies how to link information from experiment, theory and simulation in a concise way.

Automation, Data, Environment, Sharing

TECHNISCHE UNIVERSITÄT DARMSTADT



G. Pizzi, et al., Comp. Mat. Sci 111, 218-230 (2016)