

# The NOMAD Laboratory:



Claudia Draxl

## Vom internationalen Erfolg zur nachhaltigen Infrastruktur?

This project has received funding from the European Union's Horizon 2020 research and innovation programme, grant agreement No 676580.



# Open Science

NOMAD

nature

International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

Archive > Volume 546 > Issue 7658 > Editorial > Article

NATURE | EDITORIAL



## Empty rhetoric over data sharing slows science

Governments, funders and scientific communities must move beyond lip-service and commit to data-sharing practices and platforms.

12 June 2017



PDF



Rights & Permissions



## Correspondence

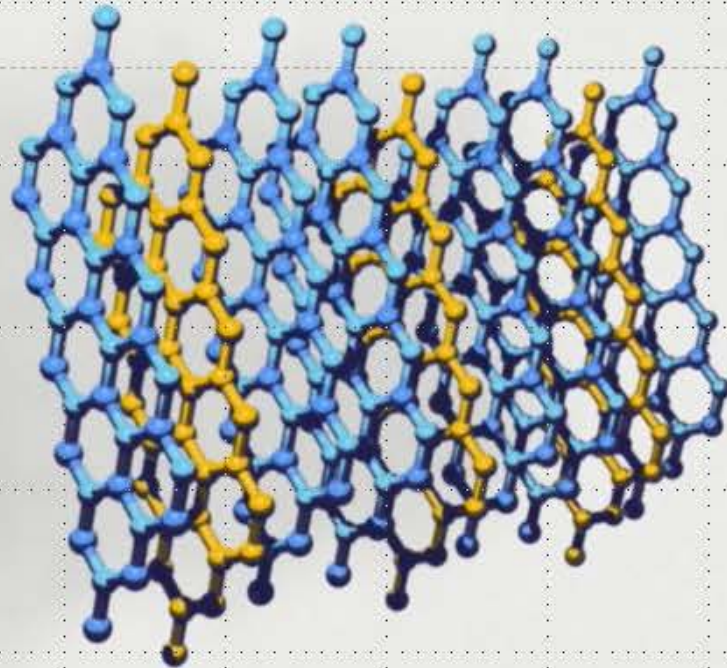
C. Draxl, F. Illas, and M. Scheffler  
Nature 548, 523 (2017).

### Open data settled in materials theory

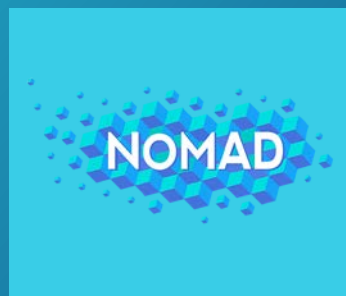
Your concerns over impediments to data sharing (*Nature* 546, 327; 2017) are no longer an issue in computational materials science. This is because the Novel Materials Discovery project (NOMAD; <https://nomad-coe.eu>) has stimulated a cultural shift in attitudes towards open data as a result of the valuable knowledge that has emerged from data mining since early 2014.

<https://www.youtube.com/watch?v=sI2cPuIGNUU&feature=youtu.be>

# COMPUTER SIMULATIONS AND DATA



# Materials data and their structure



Level	Properties	Methods	Size
I	Atomic positions and nuclear charges, properties of free atoms, symmetry, temperature, pressure	Input: definition of material <i>gene</i>	10 kB - 10 MB
II	Total wavefunction, geometry		10 MB - 10 TB
III	Excitation matrix, optical phonon spectra, thermal conductivity, etc.	as little as 10 MB	1 GB - 1 TB
IV	Efficiency of solar cell, thermoelectric figure of merit, turn-over frequency of catalyst, etc. as a function of temperature and pressure	Modeling, output derived from levels I-III <i>phenotype</i>	10 kB - 1 MB

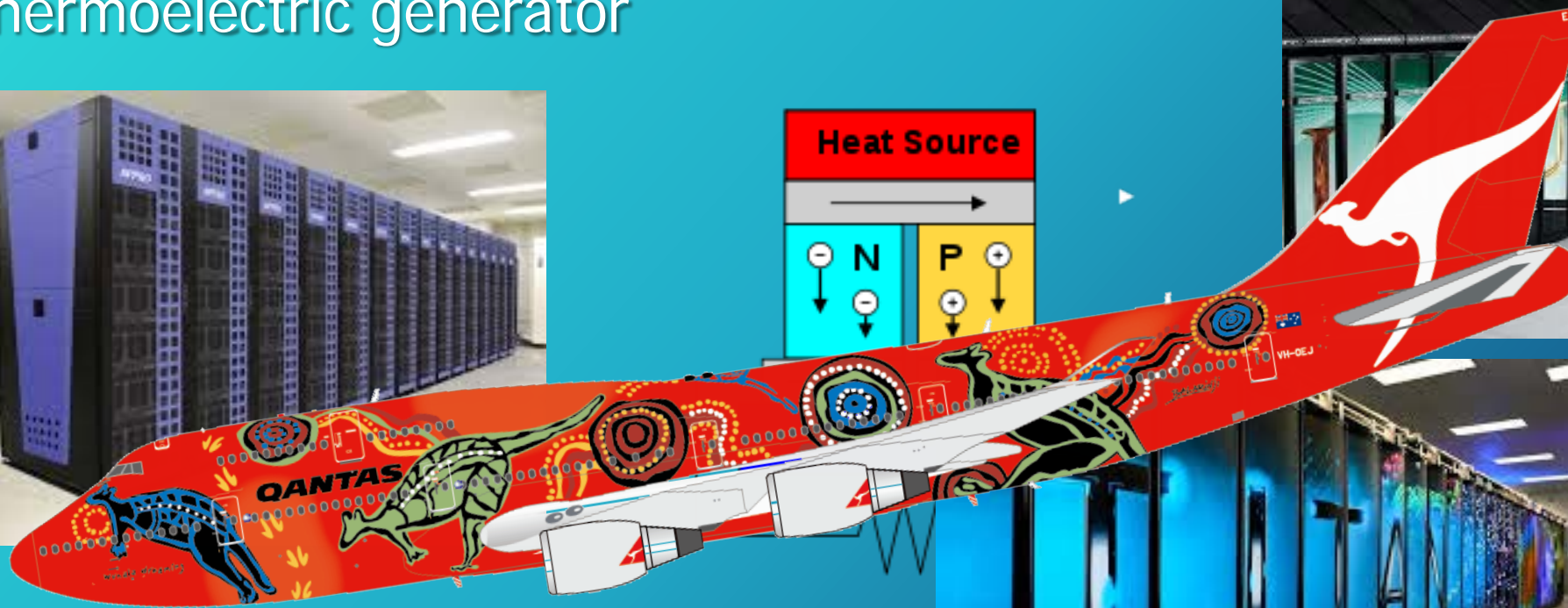
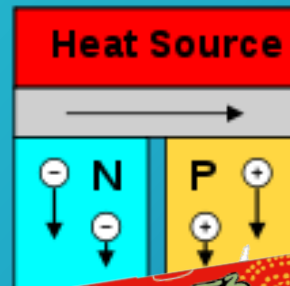
The amount of materials data produced on workstations, compute clusters, and supercomputers is growing exponentially. Most of it is thrown away ....

figure of merit

# Waste-heat recovery

NOMAD

## Thermoelectric generator



100 MW

# Waste-heat recovery



## What makes a good thermoelectric?

Figure of merit ZT with

$$Z = \frac{\sigma S^2}{\kappa_{el}^0 + \kappa_{ph}}$$

S Seebeck coefficient  
 $\sigma$  electronic conductivity  
 $\kappa$  thermal conductivity

Problem:

High electrical conductivity  $\sigma$  **and** low thermal conductivity  $\kappa$  is required

# Materials data and their structure



Level	Properties	Methods	Size
I	Atomic positions and nuclear charges, properties of free atoms, symmetry, temperature, pressure	Input: definition of material <i>gene</i>	10 kB - 10 MB
II	Total wavefunction, geometry		10 MB - 10 TB
III	Excitation matrix, optical spectra, phonon spectra, electrical conductivity, thermal conductivity, etc.	DF perturbation theory, <i>ab initio</i> MD	1 GB - 1 TB
IV	Efficiency of solar cell, thermoelectric figure of merit, turn-over frequency of catalyst, etc. as a function of temperature and pressure	Modeling, output derived from levels I-III <i>phenotype</i>	10 kB - 1 MB

Much of the value of high-throughput calculations is wasted without deeper Big-Data driven analysis of the results

# The NOMAD Repository



Data is the raw material of the 21<sup>st</sup> century

The NOMAD (Novel Materials Discovery) Repository was established to host, organize, and share materials data.

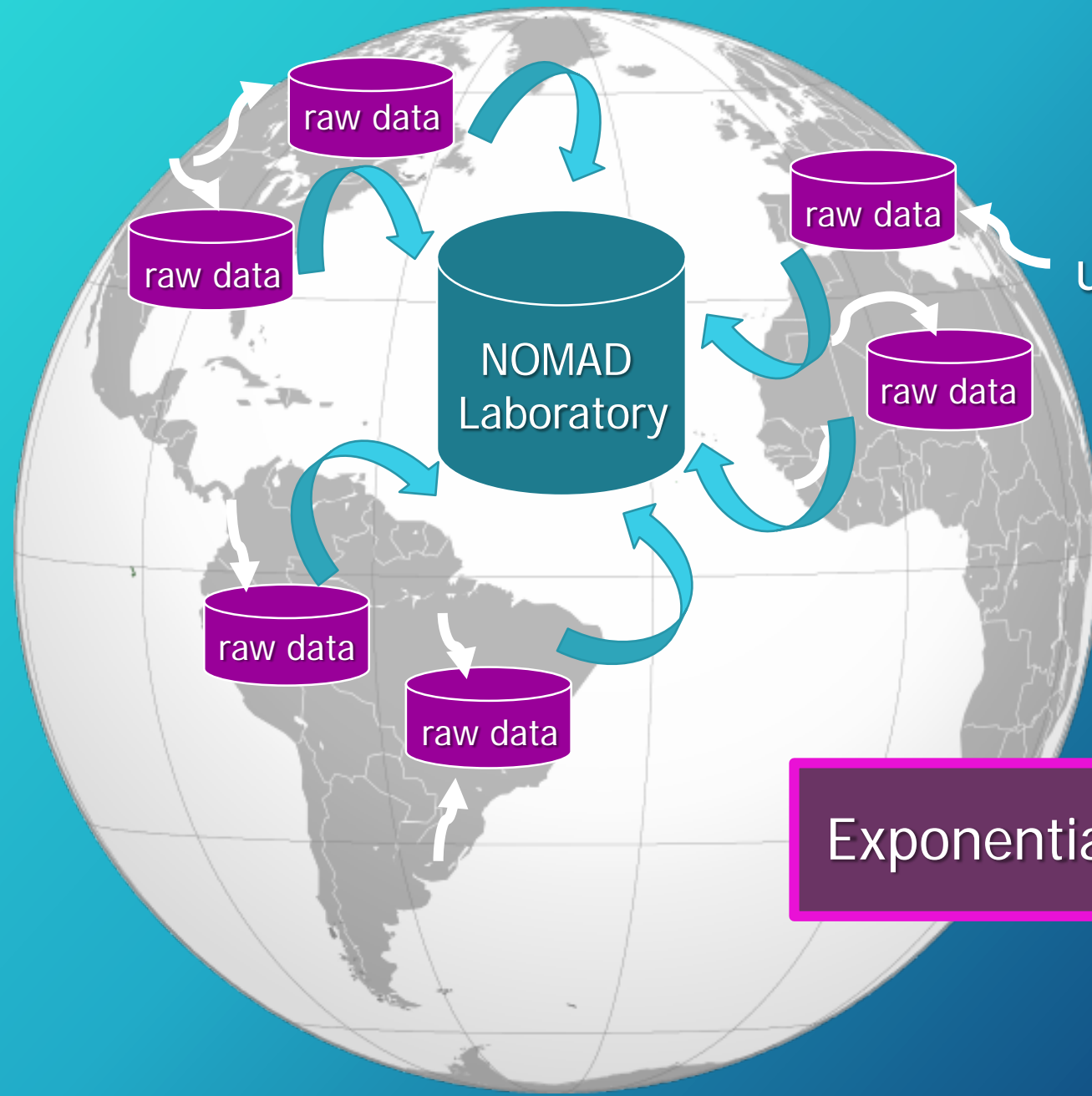
The NOMAD Repository accepts input and output files of all major codes.

Currently, the NOMAD Repository contains 45 Mio. calculations.

<https://repository.nomad-coe.eu>







Exponential increase ...

# The NOMAD Archive



Code-independent representation of the computed properties

Nomenclature, data representation, and file formats of the input and output files of the community codes are very heterogenous

Normalization requires definition of **metadata**

# The NOMAD Encyclopedia



A materials-oriented view on the Archive data that allows for seeing, comparing, exploring, and understanding.

Whatever property of a given material has been computed, is easily accessible through a user-friendly graphical user interface.

This spans from structural features, to mechanical behavior, thermal properties, electronic structure, transport characteristics, and the response to light and other excitations.

A screenshot of a periodic table interface. The element Niobium (Nb) is highlighted with a white box and a red border. The table is color-coded by groups. The interface includes dropdown arrows at the top and bottom for navigation. The highlighted element is Niobium (Nb), atomic number 41, located in the 5th period and 5th group.

H 1																	He 2	
Li 3	Be 4											B 5	C 6	N 7	O 8	F 9	Ne 10	
Na 11	Mg 12											Al 13	Si 14	P 15	S 16	Cl 17	Ar 18	
K 19	Ca 20	Sc 21	Ti 22	V 23	Cr 24	Mn 25	Fe 26	Co 27	Ni 28	Cu 29	Zn 30	Ga 31	Ge 32	As 33	Se 34	Br 35	Kr 36	
Rb 37	Sr 38	Y 39	Zr 40	<b>Nb 41</b>	Mo 42	Tc 43	Ru 44	Rh 45	Pd 46	Ag 47	Cd 48	In 49	Sn 50	Sb 51	Te 52	I 53	Xe 54	
Cs 55	Ba 56			Hf 72	Ta 73	W 74	Re 75	Os 76	Ir 77	Pt 78	Au 79	Hg 80	Tl 81	Pb 82	Bi 83	Po 84	At 85	Rn 86
Fr 87	Ra 88			Rf 104	Ha 105	Sg 106	Ns 107	Hs 108	Mt 109	Ds 110	Rg 111	Cn 112	Nh 113	Fl 114	Mc 115	Lv 116	Ts 117	Og 118
		Lanthanides and Actinides																
		La 57	Ce 58	Pr 59	Nd 60	Pm 61	Sm 62	Eu 63	Gd 64	Tb 65	Dy 66	Ho 67	Er 68	Tm 69	Yb 70	Lu 71		
		Ac 89	Th 90	Pa 91	U 92	Np 93	Pu 94	Am 95	Cm 96	Bk 97	Cf 98	Es 99	Fm 100	Md 101	No 102	Lr 103		

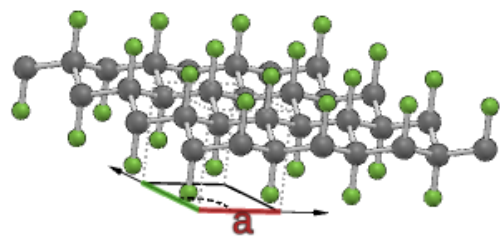
<https://encyclopedia-gui.nomad-coe.eu/>

# The NOMAD Encyclopedia



CF

Structure



● C  
● F

Show axis  Show bonds



Virtual Reality files 

System type: 2D

Methodology



Available calculations

Functional

15 GGA

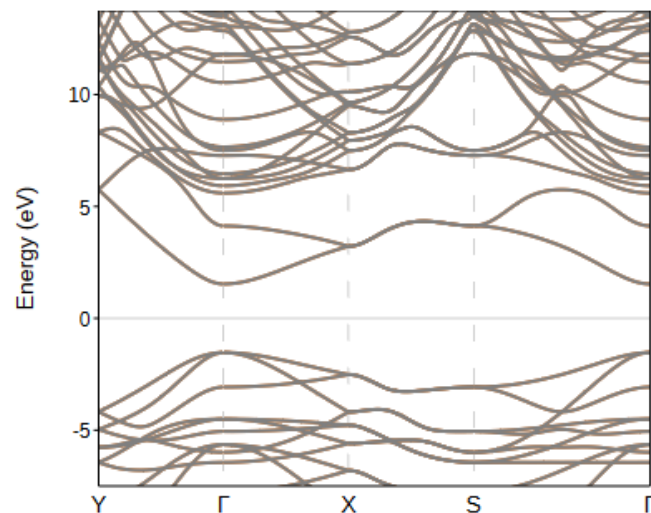
Code

15 exciting

Electronic structure



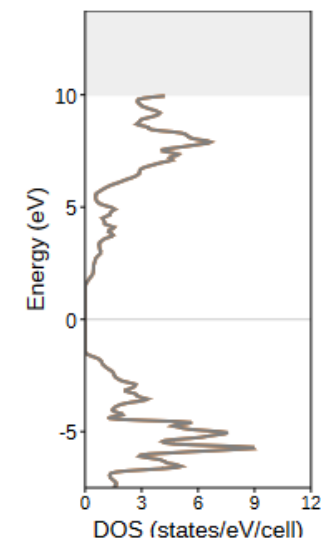
Band structure



From calculation **3492148**  
(GGA - exciting)

— Spin ↑ — Spin ↓

DOS



From calculation **3492199**  
(GGA - exciting)

# Seeing helps understanding!



## Remote visualization

Enables users to interactively perform comprehensive data visualization tasks on their computers without having to deal with hardware or software installations

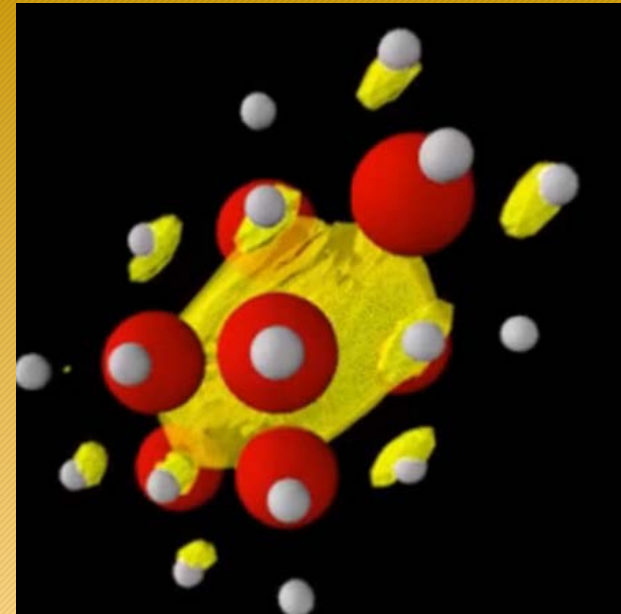
## Graphics implemented in Encyclopedia

## Virtual-reality tools

Working on very different devices

From a few to order mio €

Can we see electron-hole pairs?

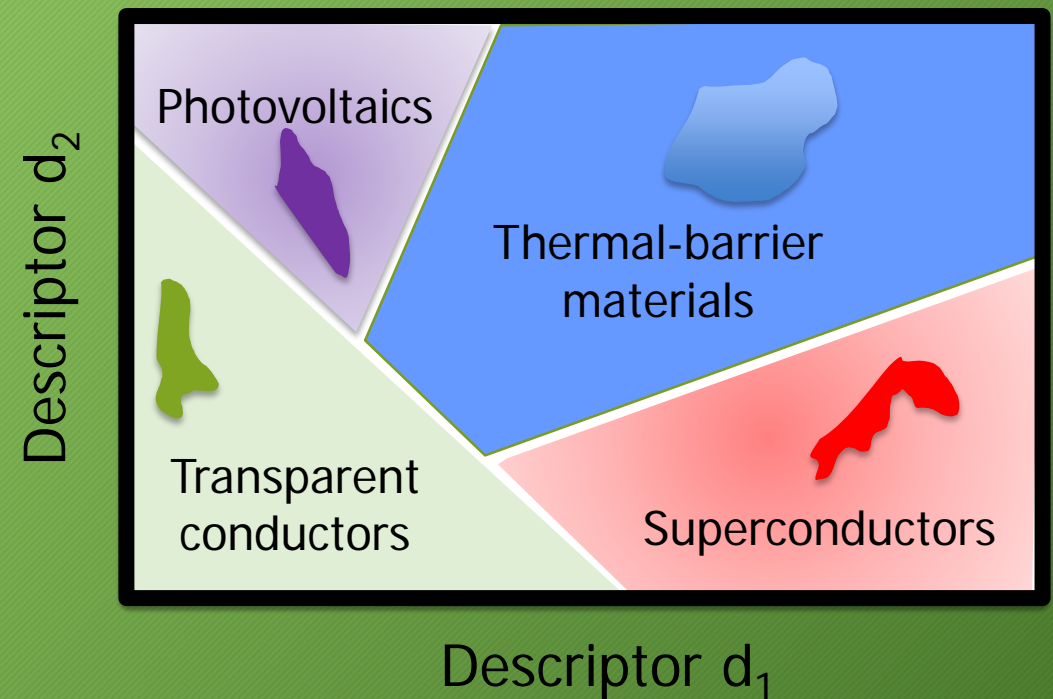


# Big-Data Analytics

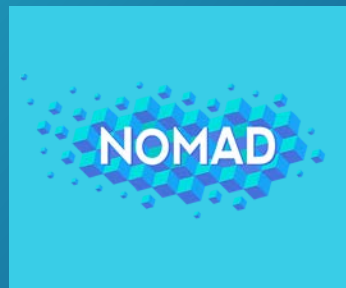


Identify correlations and structure in the data

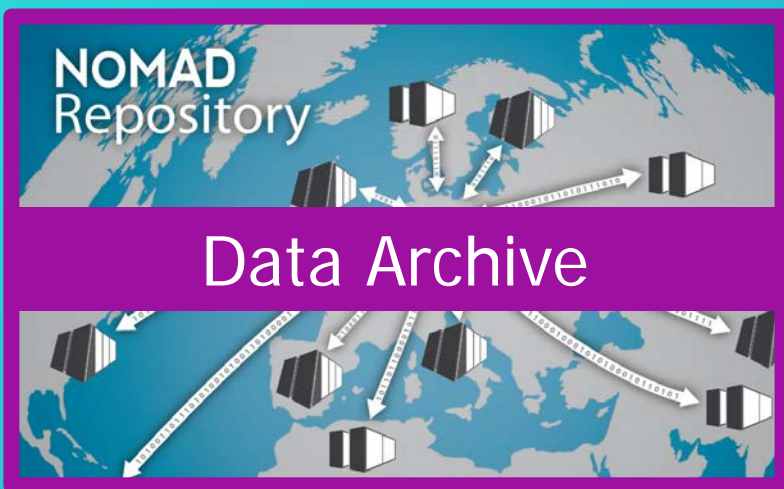
Enable scientists and engineers to identify materials for potential use in novel products and to decide which materials should be the focus of future studies.



# Sustainability



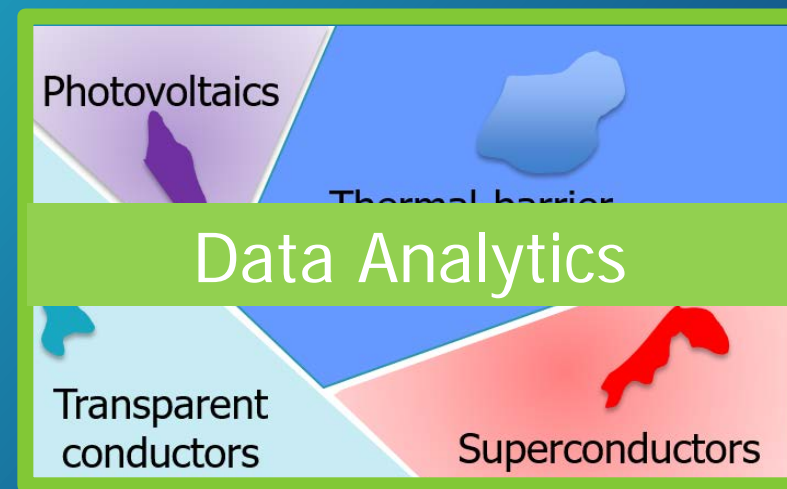
10 years from last upload  
guaranteed by MPG



Infrastructure

Further developments?

A diagram of an encyclopedia interface. At the top, there are several dropdown menus. Below them is a periodic table of elements. The element Niobium (Nb, atomic number 41) is highlighted with a white border. Below the periodic table, the word "Encyclopedia" is written in large white letters on an orange background.



Industrial exploitation

# Sustainability



A screenshot of a Nature journal article page. The top navigation bar is dark red with white text for "nature" and "International weekly journal of science". Below it is a secondary navigation bar with links for Home, News &amp; Comment, Research, Careers &amp; Jobs, Current Issue, Archive, Audio &amp; Video, and For Authors. A breadcrumb trail shows Archive &gt; Volume 546 &gt; Issue 7658 &gt; Editorial &gt; Article. The article title is "Empty rhetoric over data sharing slows science" and the sub-headline is "Governments, funders and scientific communities must move beyond lip-service and commit to data-sharing practices and platforms." The date "12 June" is visible on the left. The main title "Not-so-open data" is overlaid on the screenshot in a large, bold, black font. A teal box highlights the phrase "funding agencies" in the article's lead sentence.

12 June **Not-so-open data**

*Too many **funding agencies** resist supporting the sharing of data and too many research communities struggle with its practicalities. The result is empty rhetoric and slow science.*





**Kristian Thygesen**  
DTU Lyngby



**Ciaran Clissman**  
Pintail Dublin



**Arndt Bode**  
LRZ Munich



**Jose Maria Cela**  
BSC Barcelona



**Alessandro De Vita**  
Kings College London



nomad-coe.eu

**Matthias Scheffler**  
FHI Berlin



# NOMAD



**Claudia Draxl**  
HU Berlin



**Angel Rubio**  
MPSD Hamburg



**Risto Nieminen**  
Aalto Univ. Helsinki



**Kimmo Koski**  
CSC Helsinki



**Francesc Illas**  
Univ. Barcelona



**Stefan Heinzl**  
MPSCD Garching

**Thanks !**