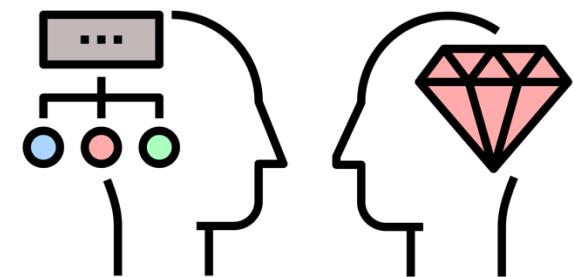


# Machine Learning for Materials

## 3. Materials Data

Aron Walsh

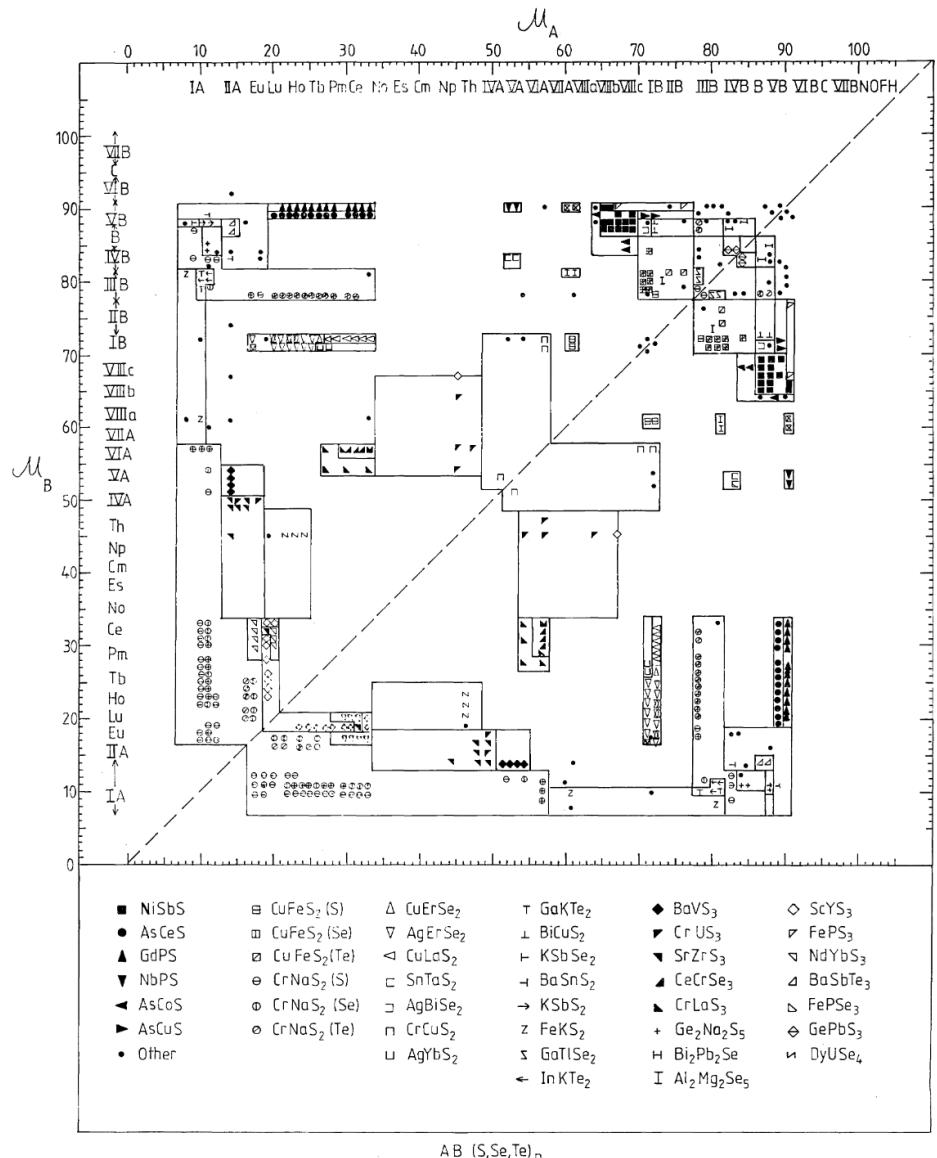
Department of Materials  
Centre for Processable Electronics



# Module Contents

1. Introduction
2. Machine Learning Basics
- 3. Materials Data**
4. Crystal Representations
5. Classical Learning
6. Artificial Neural Networks
7. Building a Model from Scratch
8. Accelerated Discovery
9. Generative Artificial Intelligence
10. Recent Advances

# Data-Driven Materials Research

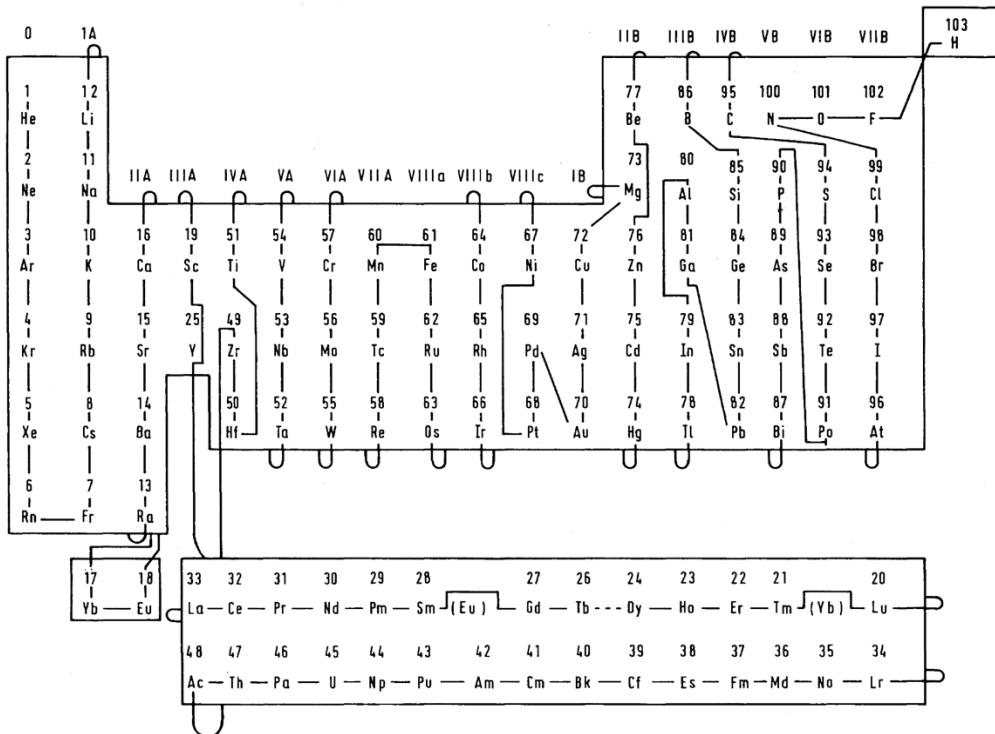


Pettifor maps

A series of work  
on structural classification  
of compounds and alloys

Quickly predict the structure  
types of new compositions

# Data-Driven Materials Research



Hand-built features  
Mendeleev number is used  
for efficient grouping of  
structure types  
(to capture periodic trends)

# Data-Driven Materials Research

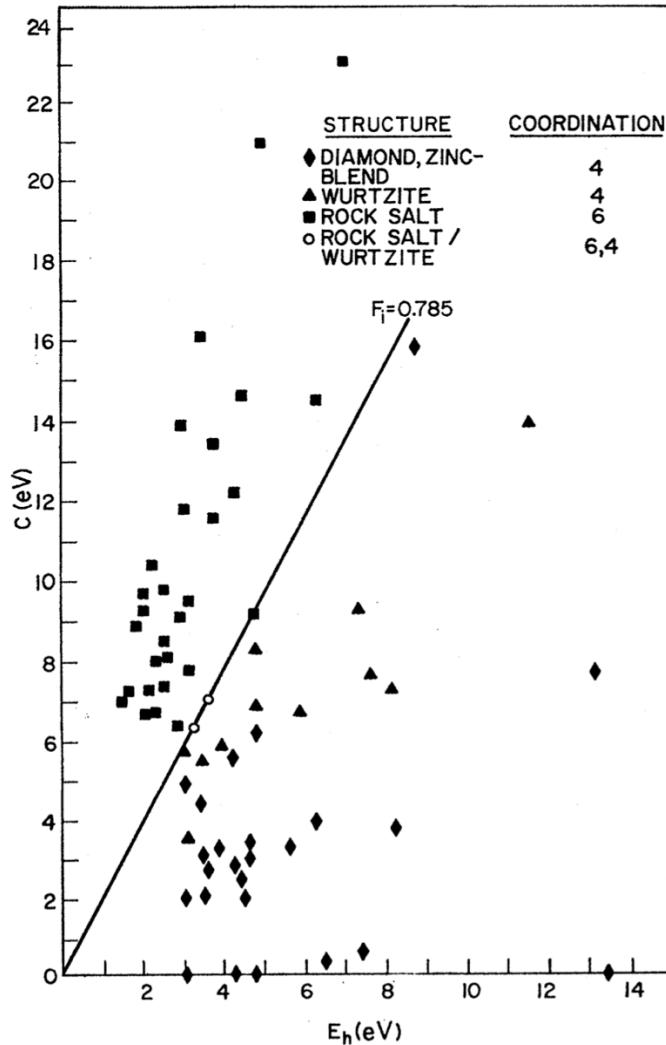


FIG. 10. The separation of fourfold and sixfold coordinates structures, using the spectroscopically defined covalent and ionic energy gaps  $E_h$  and  $C$ , respectively.

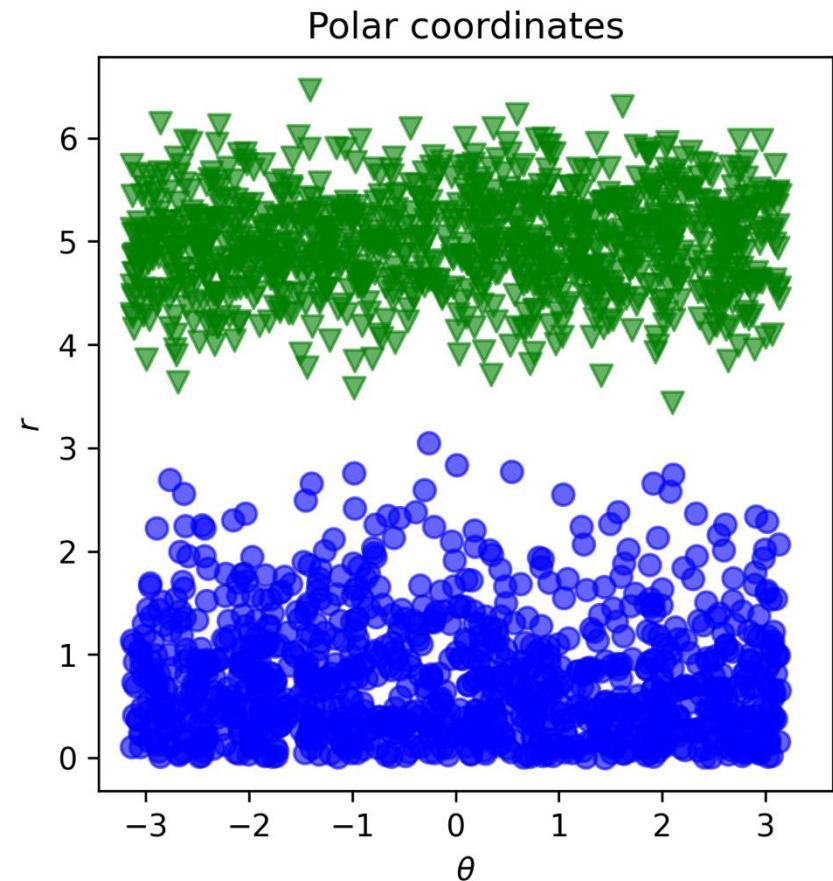
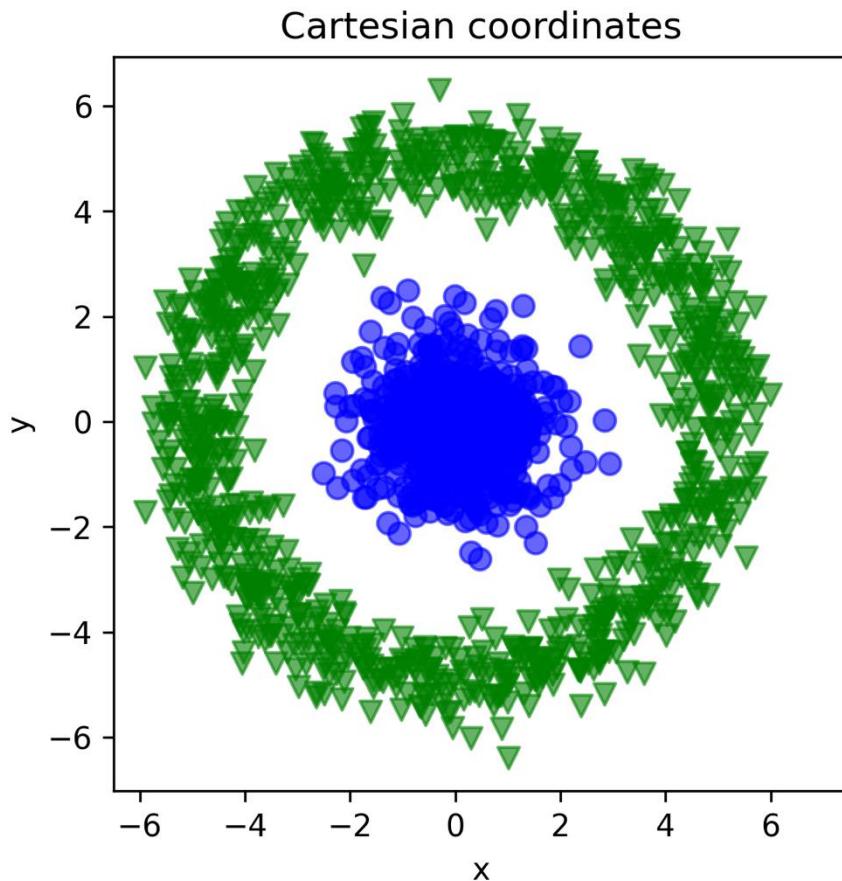
## Structure-property correlations

Connect crystal structure with measurable properties (mechanical, electronic, etc.)

Early analysis was manual and often focused on linear relations with physics-informed features

# Data Representation

Choice of units or coordinate system  
can greatly impact model performance



More on this in the next class

# Class Outline

## Materials Data

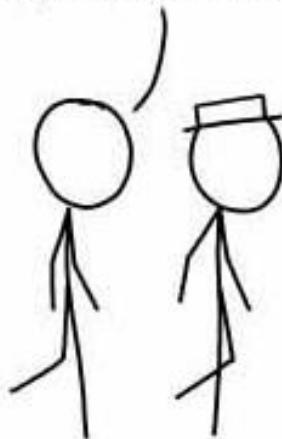
A. *Data sources and formats*

B. *API queries*

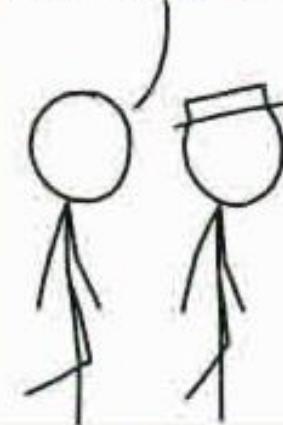
THE GREAT THING ABOUT  
DIGITAL DATA IS THAT  
IT NEVER DEGRADES.



HARD DRIVES FAIL,  
OF COURSE, BUT THEIR  
BITS CAN BE COPIED  
FOREVER WITHOUT LOSS.



FILM DEGRADES, PAINT  
CRACKS, BUT A COPY OF A  
CENTURY-OLD DATA FILE IS  
IDENTICAL TO THE ORIGINAL.



• Verizon LTE  
4:45 PM  
IF HUMANITY HAS A  
PERMANENT RECORD,  
WE ARE THE FIRST  
GENERATION IN IT.

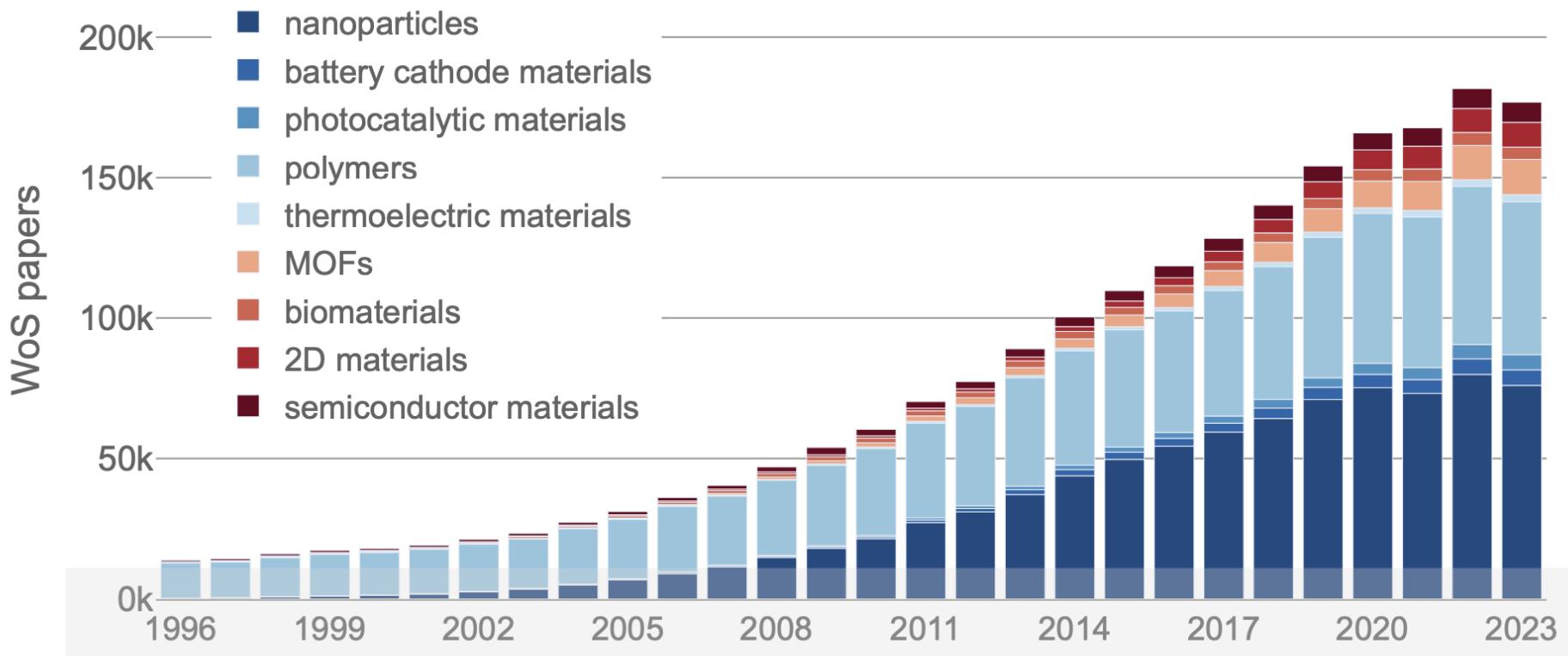


# Where to Find Data?

- **Manual collection** – go through papers, extract data and tabulate (*takes time*)
- **Accelerated collection** – use of natural language processing (*requires model and workflow*)
- **Pre-built databases** – excellent when they exist in your area (*may require access fees*)
- **Automated experiments** – generate your own data over a given parameter space (*expensive*)

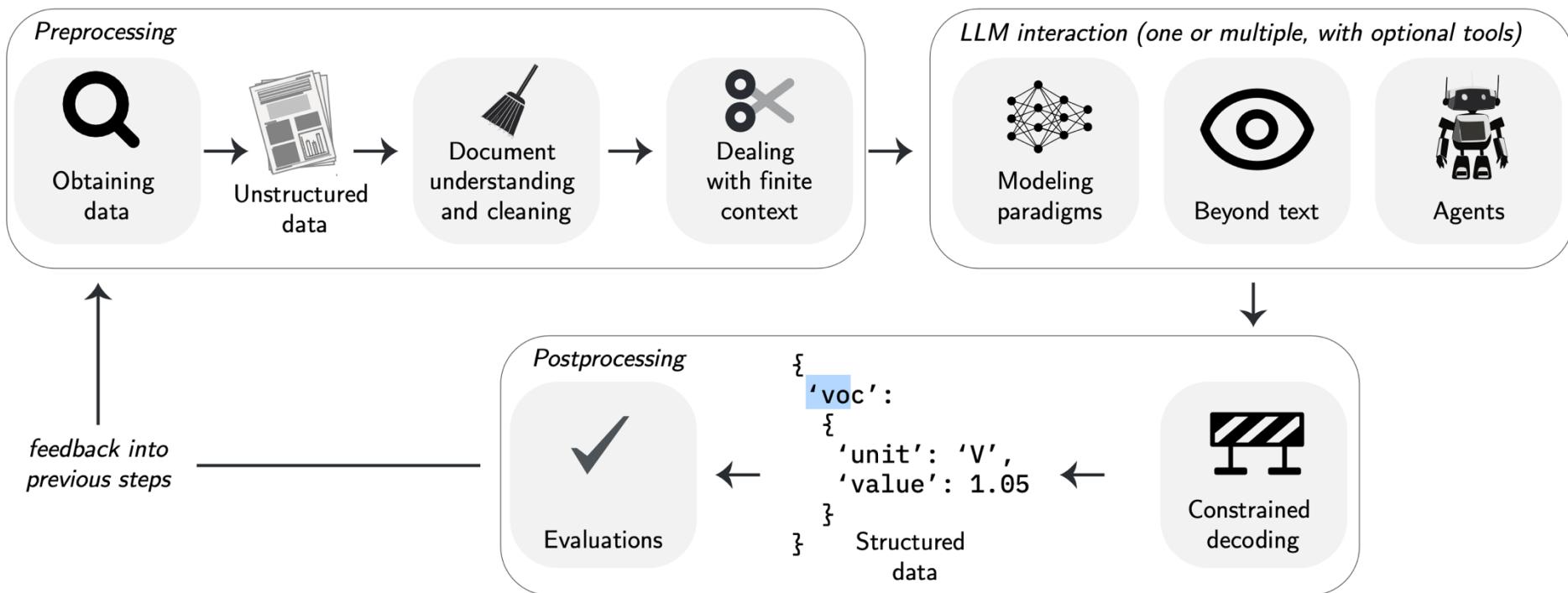
# Data Extraction from the Literature

Leverage the vast literature of published papers



# Data Extraction from the Literature

Many tailored workflows are available based on regular expressions and/or statistical models



Examples include <https://github.com/mcs07/ChemDataExtractor> and [https://github.com/CederGroupHub/text-mined-synthesis\\_public](https://github.com/CederGroupHub/text-mined-synthesis_public)

# Why Share Data?

- **Reproducibility** – allow direct comparison with published literature beyond static tables and figures, e.g. raw spectra and diffraction patterns
- **Reuse** – facilitate meta-studies comparing results from multiple experiments, e.g. variation in UV-vis spectra for different samples
- **Statistical models** – power of machine learning depends on the quantity, quality, and diversity of training data

# Common Forms of Data Sharing

- **Supporting information with publications** – often in the form of static pdf files (*increasingly obsolete*)
- **Data repositories** – most institutions offer data upload portals, but often lack guidelines and metadata, e.g. zip or tar files
- **Community-specific repositories** – best option if available, usually in a common format and searchable, with error detection

# Common Forms of Data Sharing

Many file types that differ in how data is structured, stored, and compressed, but all easy to read in

```
import pandas as pd

# read in many structured data formats

df_csv = pd.read_csv('file.csv')

df_excel = pd.read_excel('file.xlsx')

df_hdf = pd.read_hdf('file.h5')

df_json = pd.read_json('file.json')

df_stata = pd.read_stata('file.dta')

df_clipboard = pd.read_clipboard()

df_pickle = pd.read_pickle('file.pkl')
```

JSON is common as an open, flexible, and human-readable format

# FAIR Data Standards

- **Findable:** discoverable by humans & machines with metadata & persistent identifiers (e.g. DOI)
- **Accessible:** archived in long-term storage with clear access terms (e.g. CC open license)
- **Interoperable:** exchangeable between different applications and systems using open file formats
- **Reusable:** well documented and curated with clear terms and conditions on usage

# Data Security

Not all databases are public, e.g. companies  
and academic-industrial collaborations

- **Privacy:** protection of personal data  
e.g. General Data Protection Regulation (GDPR)
- **Encryption:** protocols for storage and transfer  
e.g. public key encryption, hashing
- **Access control:** limiting users or computers  
e.g. passwords, firewalls
- **Data integrity:** avoid corruption or modification  
e.g. data provenance tracking, regular versioning

# Crystallography in the Lead

## Community Databases

Cambridge  
Structural Database  
(from 1960)

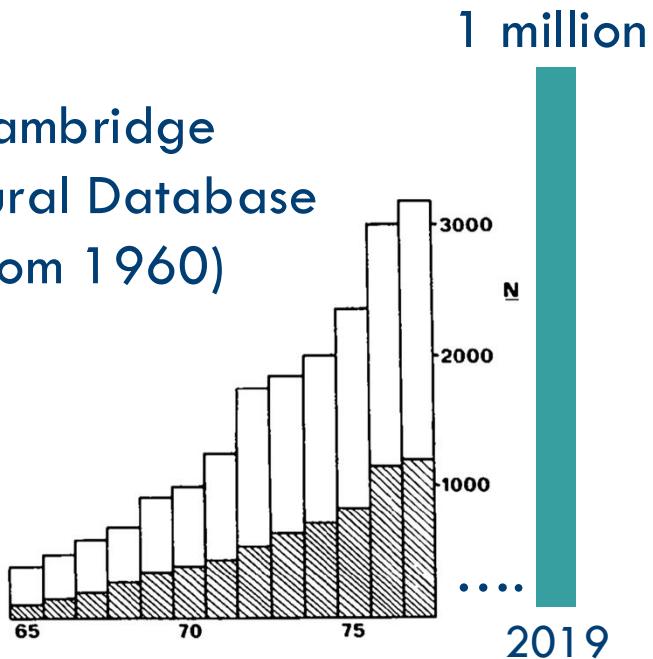


Fig. 1. Growth of CCDC database, expressed as number of entries ( $N$ ) per publication year, for the period 1965–1977. The shaded area represents organometallics and metal complexes, the unshaded area represents organics.

## Standard Format

Human and Machine Readable

```
1 #####  
2 # Crystallographic Information Format file  
3 # Produced by PyCifRW module  
4 #  
5 # This is a CIF file. CIF has been adopted by the International  
6 # Union of Crystallography as the standard for data archiving and  
7 # transmission.  
8 #  
9 # For information on this file format, follow the CIF links at  
10 # http://www.iucr.org  
11 #####  
12  
13 data_SnS2  
14 _symmetry_space_group_name_H-M      'P 1'  
15 _cell_length_a                      3.6999199  
16 _cell_length_b                      3.6999199  
17 _cell_length_c                      6.97795855  
18 _cell_angle_alpha                  90.0  
19 _cell_angle_beta                  90.0  
20 _cell_angle_gamma                  119.999999964  
21 _chemical_name_systematic          'Generated by pymatgen'  
22 _symmetry_Int_Tables_number        1  
23 _chemical_formula_structural       SnS2  
24 _chemical_formula_sum              'Sn1 S2'  
25 _cell_volume                      82.7263116646  
26 _cell_formula_units_Z              1  
27 loop_  
28   _symmetry_equiv_pos_site_id  
29   _symmetry_equiv_pos_as_xyz  
30     1 'x, y, z'
```

# Crystallography in the Lead

Compound composition ,  
unit cell dimensions  
and symmetry

```
data_2010jaa0012_p21n
chemical_formula_sum
'F3 H4 Mn N O3 P'
chemical_formula_weight      208.94
symmetry_cell_setting       monoclinic
symmetry_space_group_name_H -M   P21/n
cell_length_a                5.3999(2)
cell_length_b                13.1644(6)
cell_length_c                7.4117(3)
cell_angle_alpha              90.00
cell_angle_beta               107.501( 3)
cell_angle_gamma              90.00
cell_volume                   502.48(4)
cell_formula_units_Z          4
```

```
cell_measurement_temperature  293(2)
diffrn_ambient_temperature   120(2)
diffrn_radiation_wavelength  0.71073
diffrn_radiation_type        MoK \a
diffrn_radiation_source      'Bruker   -Nonius FR591 rotating anode'
diffrn_detector_area_resol_mean '4096x4096pixels / 62x62mm'
diffrn_reflns_number          6305
diffrn_reflns_av_R_equivalents 0.0561
diffrn_reflns_av_sigma/netl   0.0356
refine_ls_R_factor_all        0.0344
refine_ls_R_factor_gt          0.0294
refine_ls_wR_factor_ref        0.0782
refine_ls_wR_factor_gt          0.0737
refine_ls_goodness_of_fit_ref  0.928
refine_ls_restrained_S_all     0.928
```

Data collection and  
refinement information

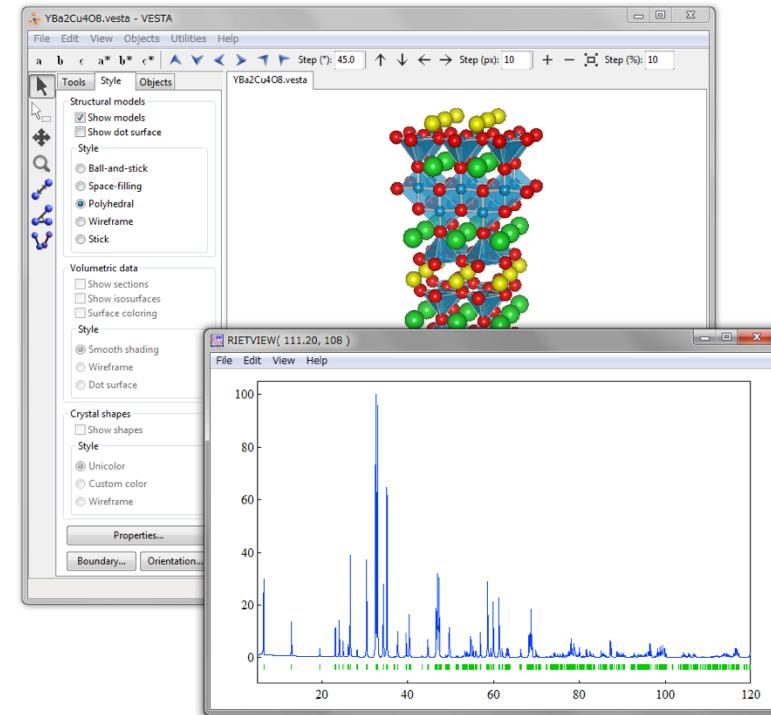
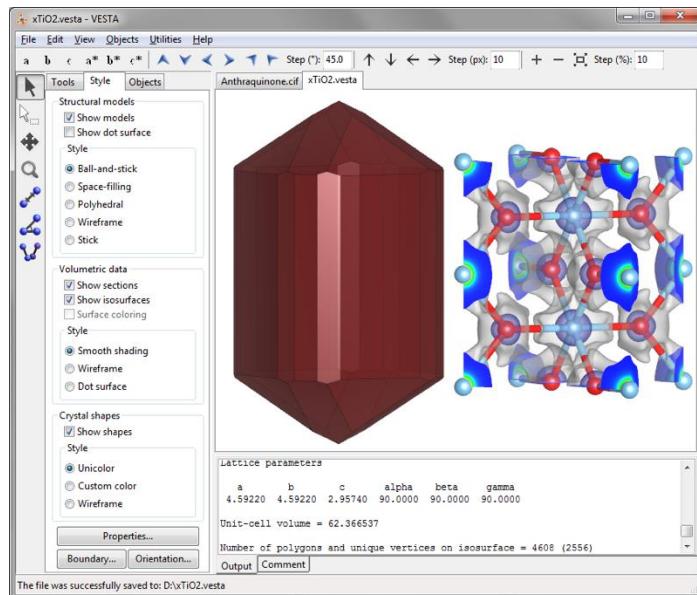
Atom coordinates

H4	H	0.563(8)	0	.763(4)	1.186(6)	0.036(12)	Uiso	1	1	d ...
Mn1	Mn	0.0000	0.5000	0.5000	0.00741(16)	Uani	1	2	d S ..	
P1	P	0.44334(12)	0.39660(5)	1.31194(9)	0.00852(17)	Uani	1	1	d ...	
Mn2	Mn	0.0000	0.5000	1.0000	0.00813(17)	Uani	1	2	d S ..	
F1	F	0.4112(3)	0.28867(12)	1.2	175(2)	0.0134(3)	Uani	1	1	d ...
F2	F	0.0669(3)	0.44659(11)	0.7416(2)	0.0115(3)	Uani	1	1	d ...	
O3	O	0.2656(3)	0.59984(13)	0.5925(2)	0.0105(4)	Uani	1	1	d ...	
F3	F	0.0951(3)	0.62898(12)	0.9601(2)	0.0130(3)	Uani	1	1	d ...	
O1	O	0.0097(9)	0.0128(9)	0.011	4(8)	0.0016(6)	0.0056(7)	0.0019(7)		
O2	O	0.0088(9)	0.0152(9)	0.0098(8)	0.0022(7)	0.0025(7)	0.0001(7)			
N1	N	0.0122(12)	0.0121(12)	0.0133(12)	-0.0005(9)	0.0021(9)	-0.0009(9)			
Mn1	Mn	0.0068(3)	0.0091(3)	0.0062(3)	0.00054(17)	0.0017(2)	-0.00003(18)			
P1	P	0.0075(3)	0.0103(3)	0.0077(3)	-0.0006(2)	0.0023(2)	0.0004(2)			
Mn2	Mn	0.0075(3)	0.0096(3)	0.0067(3)	0.00087(18)	0.0012(2)	-0.00011(19)			
F1	F	0.0134(8)	0.0112(7)	0.0157(8)	-0.0032(6)	0.0042(6)	-0.0006(6)			
F2	F	0.0124(7)	0.0134(7)	0.0085(7)	0.0020(6)	0.0030(6)	0.0024(6)			
O3	O	0.0079(9)	0.0113(9)	0.0120(9)	0.0002(6)	0.0025(7)	0.0009(7)			
F3	F	0.0132(8)	0.0128(7)	0.0115(7)	0.0019(6)	0.0014(6)	-0.0014(6)			

Atomic displacement parameters

# Crystallography in the Lead

Many open-source programs for cif visualisation  
(including Miller indices, diffraction patterns...)



# Example: General Repository

The screenshot shows the Zenodo website interface. At the top, there is a blue header bar with the Zenodo logo, a search bar, an upload button, and a communities link. Below the header, the page title is "Materials Design Group". Underneath, the section title "Recent uploads" is displayed. There are two entries listed:

**July 25, 2023 (v1) Journal article Open Access**  
**Environmentally-friendly mixed titanium/tin (IV) vacancy-ordered double perovskites (Cs<sub>2</sub>Ti<sub>1-x</sub>S<sub>x</sub>X<sub>6</sub>; X = I, Br, Cl)**  
View  
Kavanagh, Seán R.; Liga, Shanti; Walsh, Aron; Scanlon, David O.; Konstantatos, Gerasimos;  
Lead toxicity and poor stability under operating conditions are major drawbacks delaying the commercialization of perovskite solar cells. Titanium(IV) has been considered as an alternative species to Pb(II) because it is non-toxic, abundant and its perovskites have demonstrated promising performance  
Uploaded on July 25, 2023

**June 2, 2023 (v1) Dataset Open Access**  
**Molecular Dynamics Trajectory of CsPbI<sub>3</sub> produced using ACE MLFF**  
View  
William J. Baldwin; Xia Liang; Aron Walsh; G'abor Cs'anyi;  
Molecular dynamics trajectories of CsPbI<sub>3</sub>, with 2 fs timestep between computed frames, frames are recorded every 200 fs. Under each folder, there are three npz files (each contains 100 ps of trajectory) and one initial configuration LAMMPS data file. The npz files can be read with 'num'  
Uploaded on June 2, 2023

# Example: Community Repository

The screenshot shows the NOMAD community repository interface. The top navigation bar includes links for PUBLISH, EXPLORE, ANALYZE, and ABOUT, along with LOGIN / REGISTER and UNITS options. On the left, there's a sidebar for ENTRIES SEARCH and FILTERS, which includes dropdown menus for Material, Elements / Formula, Symmetry, Method, Simulation, DFT, GW, Experiment, and EELS. Under Elements, 'Fe' and 'O' are selected. The main content area displays a table of 20/21,504 search results. The columns are: a checkbox column, Formula, Method name, Upload time (sorted by descending time), and Authors. All results listed are for 'Bi19Fe20La060' and are attributed to 'The Perovskite Database Project'. Each row has a three-dot menu icon on the far right.

	Formula	Method name	Upload time ↓	Authors	
<input type="checkbox"/>	Bi19Fe20La060		07/09/2022, 09:27:02	The Perovskite Database Project	...
<input type="checkbox"/>	BiFeO3		07/09/2022, 09:27:02	The Perovskite Database Project	...
<input type="checkbox"/>	BiFeO3		07/09/2022, 09:27:02	The Perovskite Database Project	...
<input type="checkbox"/>	BiFeO3		07/09/2022, 09:27:02	The Perovskite Database Project	...
<input type="checkbox"/>	BiFeO3		07/09/2022, 09:27:02	The Perovskite Database Project	...
<input type="checkbox"/>	Bi19Fe20La060		07/09/2022, 09:27:02	The Perovskite Database Project	...
<input type="checkbox"/>	Bi19Fe20La060		07/09/2022, 09:27:02	The Perovskite Database Project	...
<input type="checkbox"/>	Bi19Fe20La060		07/09/2022, 09:27:02	The Perovskite Database Project	...

# Example: Curated Repository

Welcome to ICSD Web. IP authenticated (130.246.140.4). Royal Society Chemistry

FIZ Karlsruhe | Contact  
Close session

Detailed View

Entry 1 of 1

Back to Query Back to List CCDC Export Cif Print Feedback to Editor

Collection Code 26170

**Summary**

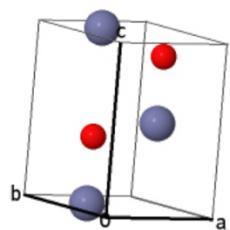
Struct. formula	Zn O	Structure type	Wurtzite#ZnS(2H)
Cell parameter	3.24986(1) 3.24986(1) 5.20662(1) 90. 90. 120.	Space group	P 63 m c (186)
Cell volume	47.62 [Å <sup>3</sup> ]	Z	2
Temperature	room temperature	Pressure	atmospheric
Data quality	High quality	R-value	0.0378
Author	Abrahams, S.C.; Bernstein, J.L.	Title	Remeasurement of the structure of hexagonal Zn O
Reference	Acta Crystallographica, Section B: Structural Crystallography and Crystal Chemistry (1969) 25, p. 1233-1236	DOI	10.1107/S0567740869003876

**Details**

▼ Visualization

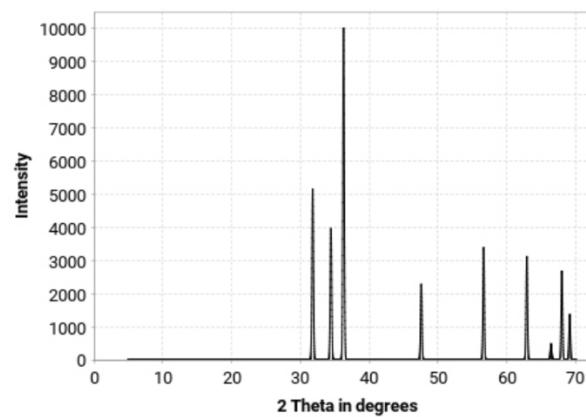
**Published Crystal Structure**

HM:P 63 m c #186  
a=3.250 Å  
b=3.250 Å  
c=5.207 Å  
 $\alpha=90.000^\circ$   
 $\beta=90.000^\circ$   
 $\gamma=120.000^\circ$



Interactive Visualization

**Powder Pattern**



Intensity

2 Theta in degrees

Interactive Visualization

# Example: Materials Modelling

 **Materials Explorer**  
App by Materials Project

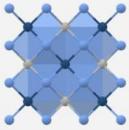
 **Ac<sub>2</sub>AgIr**  
mp-861724

TABLE OF CONTENTS

- Summary
- Crystal Structure
- Properties
  - Thermodynamic Stability
  - Electronic Structure
  - Phonon Dispersion
  - Diffraction Patterns
  - Aqueous Stability
  - Magnetic Properties

Energy Above Hull: 0.000 eV/atom

Space Group: Fm $\bar{3}$ m

Band Gap: 0.00 eV

Predicted Formation Energy: -0.372 eV/atom

Magnetic Ordering: Non-magnetic

Total Magnetization: 0.00  $\mu$ B/f.u.

Experimentally Observed: No

The screenshot shows the Materials Explorer interface for the material Ac<sub>2</sub>AgIr (mp-861724). The left sidebar contains a search icon, the title "Materials Explorer" and "App by Materials Project", a crystal structure diagram, the material name "Ac<sub>2</sub>AgIr", and its ID "mp-861724". Below these are sections for "TABLE OF CONTENTS" and "Properties". The "Properties" section includes links for "Thermodynamic Stability", "Electronic Structure", "Phonon Dispersion", "Diffraction Patterns", "Aqueous Stability", and "Magnetic Properties". The main content area displays a 3D ball-and-stick model of the crystal structure. To the right of the model is a vertical toolbar with icons for zoom, settings, refresh, camera, and export. A dropdown menu next to the toolbar lists file formats: CIF (Symmetrized), CIF, POSCAR, JSON, Prismatic, and VASP Input Set (MPRelaxSet). At the bottom of the main area are two arrows pointing up and right, likely indicating navigation or selection options. The right side of the screen features a card with various material properties: Energy Above Hull (0.000 eV/atom), Space Group (Fm $\bar{3}$ m), Band Gap (0.00 eV), Predicted Formation Energy (-0.372 eV/atom), Magnetic Ordering (Non-magnetic), Total Magnetization (0.00  $\mu$ B/f.u.), and Experimentally Observed (No).

# Example: Microscopy

The screenshot shows the homepage of the Electron Microscopy Data Bank (EMDB). At the top, there is a navigation bar with links to EMBL-EBI, Services, Research, Training, About us, and a search bar labeled "EMBL-EBI". Below the navigation bar is a large banner featuring the EMDB logo and the text "EMDB Electron Microscopy Data Bank". To the right of the banner is a search interface with a search bar containing "Search EMDB..." and a magnifying glass icon. Below the search bar are examples of search terms: "1001, Apoferritin, Tomography, Rossmann MG, 5A1A". The main content area has a dark background with a green textured pattern. It features a section titled "About EMDB" with a brief description of the repository's purpose and history. Below this, there is a list of three collaborating archives: EMDB, PDB, and EMPIAR. A "Quick links" sidebar on the right provides links to various resources like EMDB Policies, Talks & Tutorials, Validation Analysis, Volume Browser, EMDB Citations, EMPIAR, PDBe, BiolImage Archive, EMDataResource, and FM Navigator. There are also social media icons for email, Twitter, and YouTube.

## About EMDB

The Electron Microscopy Data Bank (**EMDB**) is a public repository for electron cryo-microscopy volume maps and tomograms of macromolecular complexes and subcellular structures. It covers a variety of techniques, including single-particle analysis, electron tomography, and electron crystallography (for more information, see the [EMDB Policies](#)).

EMDB was founded at **EMBL-EBI** in 2002 under the leadership of Kim Henrick. From 2007, the archive has been operated jointly by EMBL-EBI and the **Research Collaboratory for Structural Bioinformatics (RCSB)**. In 2013, the **Protein Data Bank Japan (PDBj)** also became involved in EMDB. As of the first of January, 2021, EMDB is an archive operated under the aegis of the **Worldwide Protein Data Bank (wwPDB)**, and has also joined that organisation as a full member.

Results of cryo-EM studies are archived in three collaborating archives:

- EMDB stores the processed 3D volumes and tomograms;
- PDB stores any atomic models that have been constructed based on EMDB data;
- EMPIAR stores the raw EM data underpinning the data in EMDB/PDB.

Data for EMDB and PDB can be deposited through the wwPDB deposition and annotation system [OneDep](#).

**Quick links**

- ✉ EMDB Policies
- 🎥 Talks & Tutorials
- 📊 Validation Analysis
- 🔎 Volume Browser
- 🔗 EMDB Citations
- 🔗 EMPIAR
- 🔗 PDBe
- 🔗 BiolImage Archive
- 🔗 EMDataResource
- 🔗 FM Navigator

# Example: NMR

**NMRShiftDB**

**Current usage is:**  
Registered Users: 2032  
Structures: 236869  
Spectra: Measured 399439, calculated 554

**Login**

Use cookies for persistant login

**Create New Account**  
(Only necessary for contributing data)

**Forgot password?**

**Impressum**

Problems using nmrshiftdb2? See our [tips on browsers to use!](#)

[Home](#) [Search](#) [Results](#) [Quick Check](#) [Predict](#) [Assignment](#) [Submit](#) [Review](#) [Help](#)

**nmrshiftdb2 Links**

[nmrshiftdb2 Documentation](#) [Developers' page](#) [Media coverage](#) [Links](#) [FAQ](#) [Guestbook](#) [Contact](#)

**About nmrshiftdb2**

nmrshiftdb2 is a NMR database (web database) for organic structures and their nuclear magnetic resonance (nmr) spectra. It allows for spectrum prediction ( $^{13}\text{C}$ ,  $^1\text{H}$  and other nuclei) as well as for searching spectra, structures and other properties. The nmrshiftdb2 software is open source, the data is published under an open content license. The core of nmrshiftdb2 are fully assigned spectra with raw data and peak lists (we have pure peak lists as well). Those datasets are peer reviewed by a [board of reviewers](#). The project is supported by a [scientific advisory board](#). nmrshiftdb2 is part of the [NFDI4Chem initiative](#) and will provide a component for a curated repository there. Please consult the [documentation](#) for more detailed information.

**Problem of the Month - May 2021**

The sample in this problem is an oligopeptide. Note that one of the amino acid residues is not one of the 21 naturally occurring eukaryotic amino acids (it appears twice), and one of the naturally occurring ones is methylated.  
Deduce the constitution and assign all nuclei (proton, carbon and nitrogen)  
Continue to slide 3 for an additional challenge.

$\text{C}_{13}\text{H}_{22}\text{N}_4\text{O}_4$

$^1\text{H}$  NMR spectrum recorded at 500.13 MHz

View Problem of the Month archive from 2020

View old Problem of the Month archive

**Latest Additions**

**Quick Check**

Which molecule is this? See our [Problem of the Month presentation](#) to find out.

Check your assignment here

# Class Outline

## Materials Data

A. *Data sources and formats*

B. *API queries*

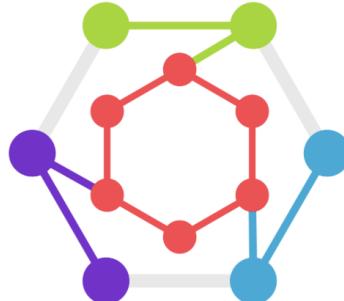
# Database Access

Mode	Advantage	Disadvantage
Web browser	No knowledge of database software is required	Often one material at a time – slow for large datasets
Data file	All data is downloaded as one (e.g. zip or tar) file	Specialist software often needed; data is not up-to-date
API* (e.g. Python)	Access latest data with advanced queries	Some programming knowledge required

\*API = Application Programming Interface

Tip: Keep a record of the database version you are using; data can change

# Materials Database Access: Python API



**OPTIMADE**  
Open Databases Integration  
for Materials Design

About

Resources

Specification

Contributors

Github

Forum

Try It!

## About us

The **Open Databases Integration for Materials Design** (OPTIMADE) consortium aims to make materials databases interoperable by developing a specification for a common REST API.

We have released version 1 of the OPTIMADE specification, with several databases already providing implementations. A full list is available on the [OPTIMADE providers dashboard](#).

24

29

27,730,600

PROVIDERS

DATABASES

STRUCTURES

# Query – Optimade



```
# Import OptimadeClient
from optimade.client import OptimadeClient

# Initialise client for OQMD provider, synchronous
client = OptimadeClient(include_providers=[ "oqmd" ], use_async=False)

# Get structures with reduced chemical formula "OPb"
results = client.get(filter='chemical_formula_reduced="OPb"')
```

# Query – Materials Project (MPRester)

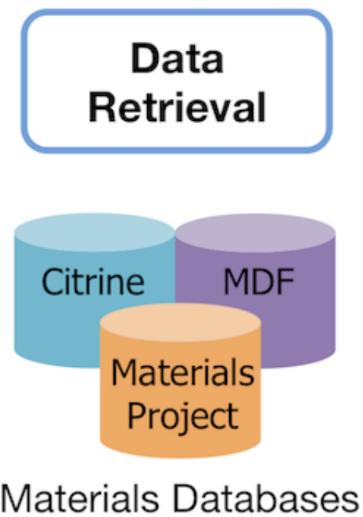


```
# Import MPRester from the Materials Project API
from mp_api.client import MPRester
from emmet.core.summary import HasProps

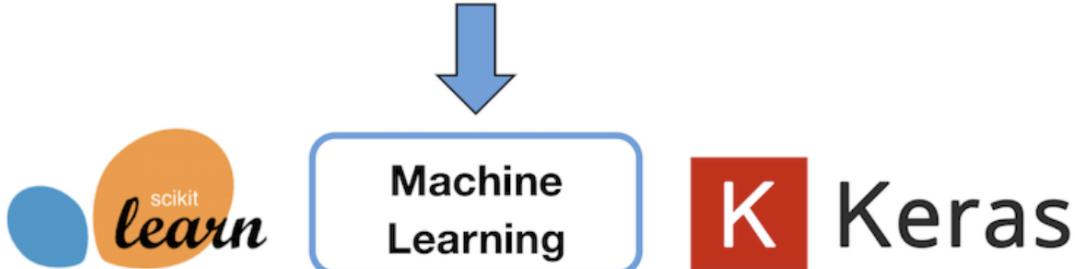
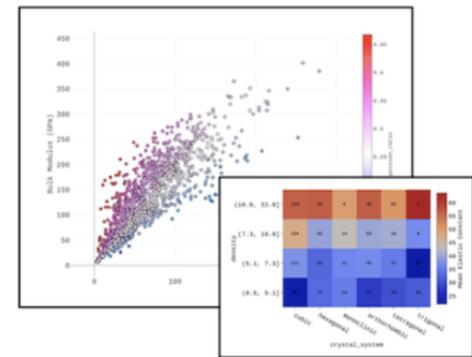
# Access the Materials Project
with MPRester("your_api_key_here") as mpr:
    # Search for materials with dielectric properties
    docs = mpr.materials.summary.search(
        has_props=[HasProps.dielectric],
        fields=["material_id"] # Only retrieve the material_id field
    )

    # Extract material IDs from the search results
    mpids = [doc.material_id for doc in docs]
```

# Load a Dataset



MATERIAL	FEATURES				PROPERTY
TiO <sub>2</sub> rutile	F <sub>11</sub>	F <sub>12</sub>	...	F <sub>1N</sub>	gap = 3.0 eV
C diamond	F <sub>21</sub>	F <sub>22</sub>	...	F <sub>2N</sub>	gap = 5.5 eV
...	...	...	...	...	...
PbTe rocksalt	F <sub>M1</sub>	F <sub>M2</sub>	...	F <sub>MN</sub>	gap = 0.3 eV



# Load a Dataset

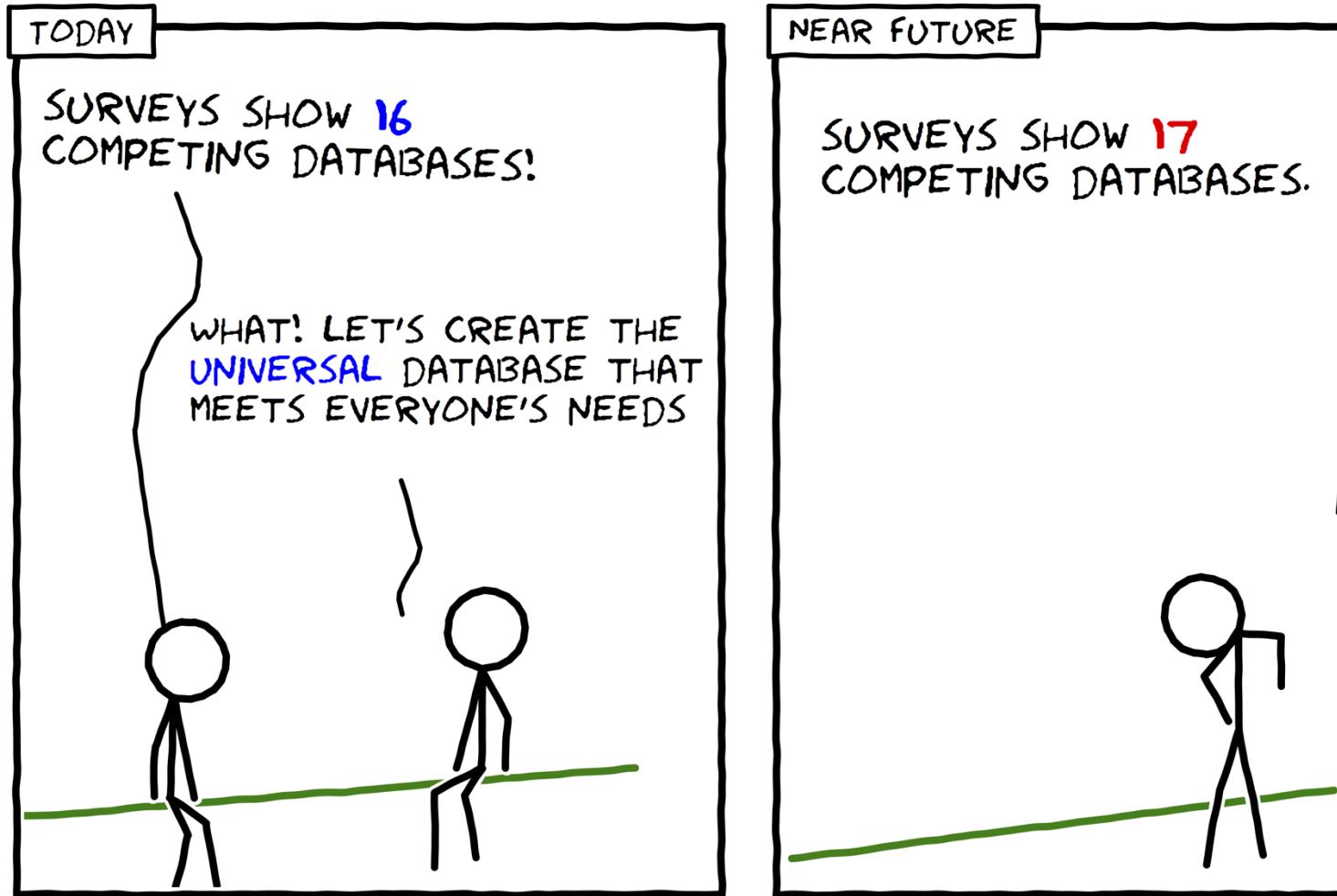


```
# Load modules
from matminer.datasets import get_available_datasets
from matminer.datasets import load_dataset

# See what's available
get_available_datasets()

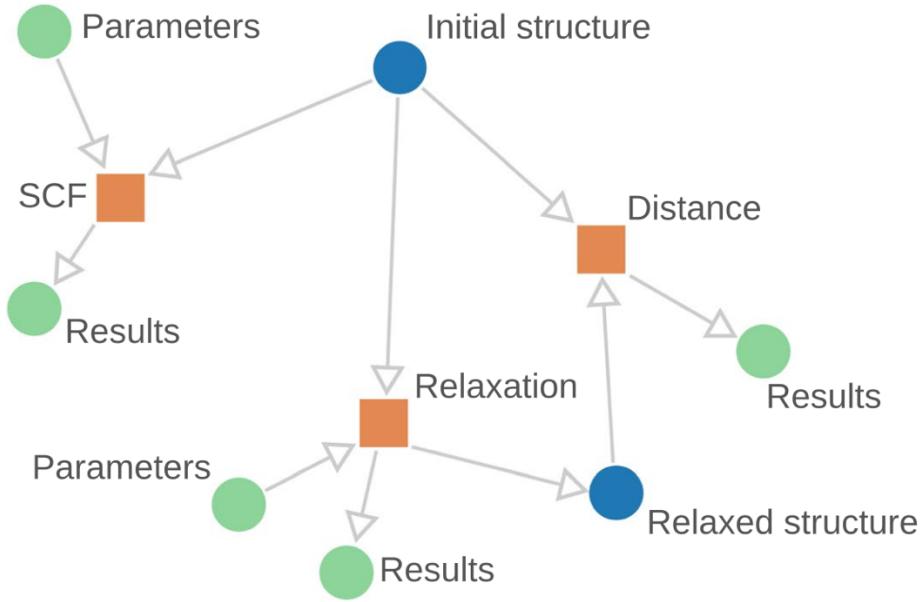
# Load dataset
df = load_dataset("citrine_thermal_conductivity")
```

# Structure and Property Databases

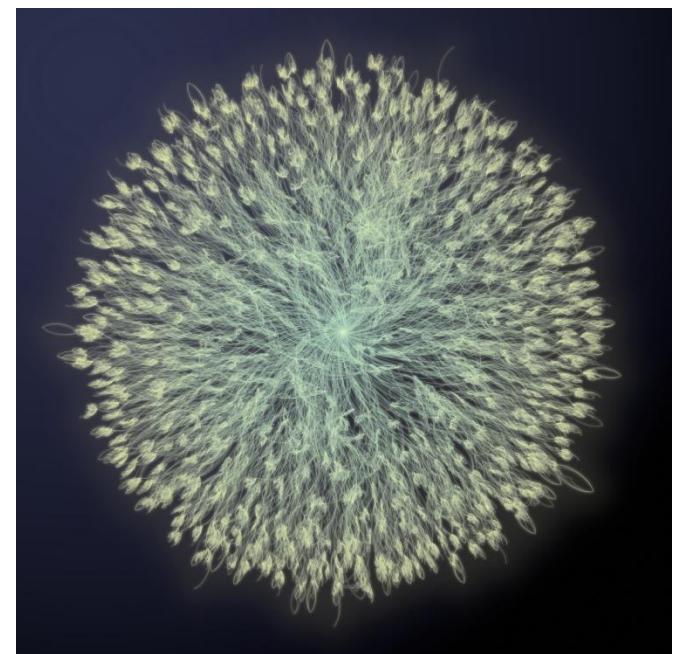


# Data Provenance

Projects can combine data from many sources.  
Provenance graphs are one way to link them



Connections between structures,  
calculations, and data



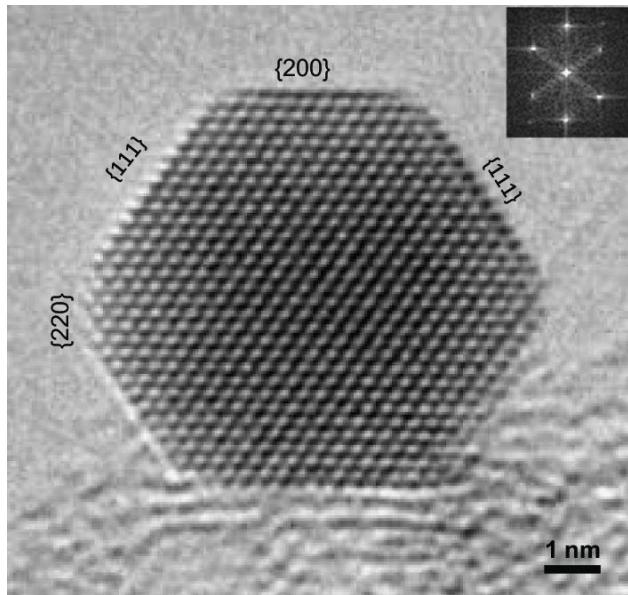
Graph for a project on 324  
covalent organic frameworks

# Image Data

Images are widely used in materials science.  
The building blocks are pixels (e.g.  $128 \times 128$ )

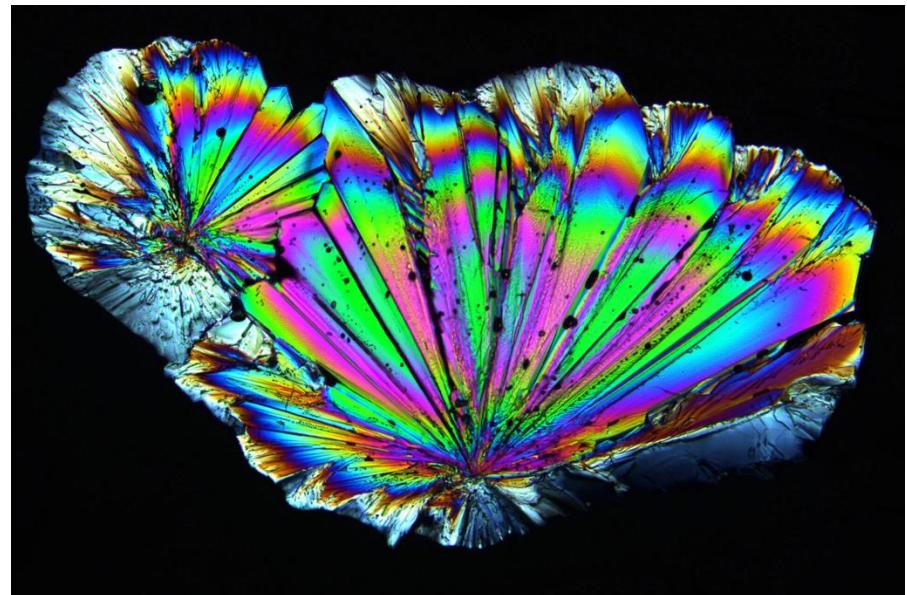
**Greyscale**

Pixel  $\in [1, 255]$



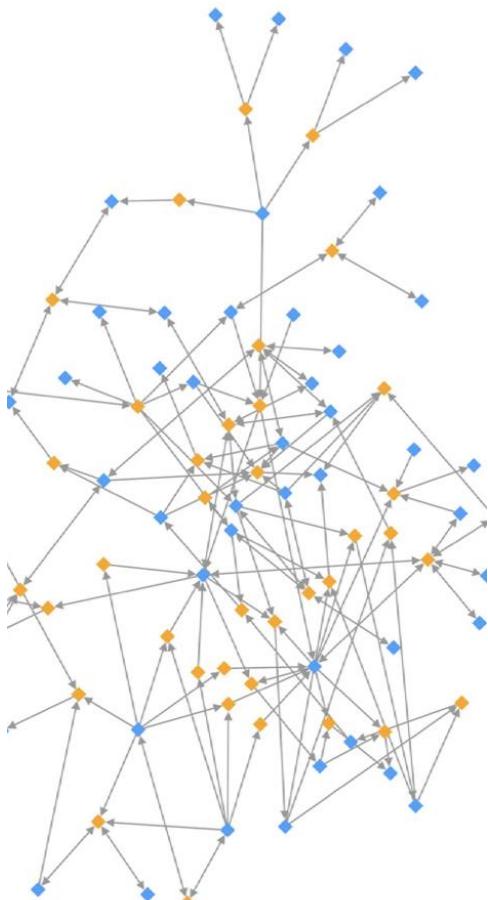
**Colour**

$P_R, P_G, P_B \in [1, 255]$

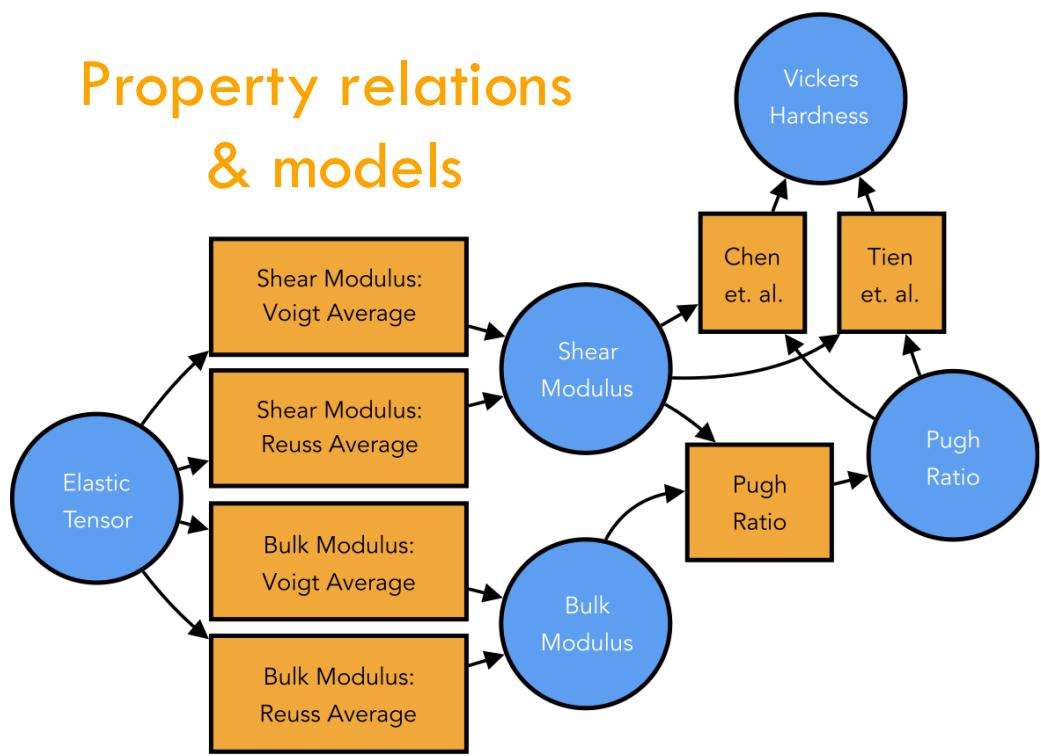


# Knowledge Graphs

Structured representation of knowledge to model properties and their interrelations in a graph format



Property relations  
& models



# Class Outcomes

1. Describe the importance of materials data for research and development
2. Demonstrate an understanding of the types of data that are shared in the materials community
3. Perform simple queries using an application programming interface

*Activity:*

Chemical space