PSDI
PHYSICAL SCIENCES DATA INFRASTRUCTURE

Elevator pitch for PSDI

Open Databases Integration for Materials Design
September 2025

Tom Underwood[1], Mehmet Giritli[1], Alexander Belozerov[1], Aileen Day[2]

[1] *Science Technology Facilities Research Council, UK*
[2] *University of Southampton, UK*

University of Southampton

UKRI Science and Technology Facilities Council
Scientific Computing

UKRI Engineering and Physical Sciences Research Council

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

What is PSDI?

- ‘An integrated data infrastructure for empowering your physical sciences data and streamlining your processes’
  - Led by University of Southampton, with STFC
  - Collaboration nationally & internationally
- Aim:
  - Make research data in the physical sciences more FAIR
- Interested in:
  - Searching over many existing data sources
  - Making data publicly available and easy to use
- Foundations:
  - Standards
  - APIs
  - Digital infrastructure

What do we offer?


Access to Data Sources


Data Services


Data Tools

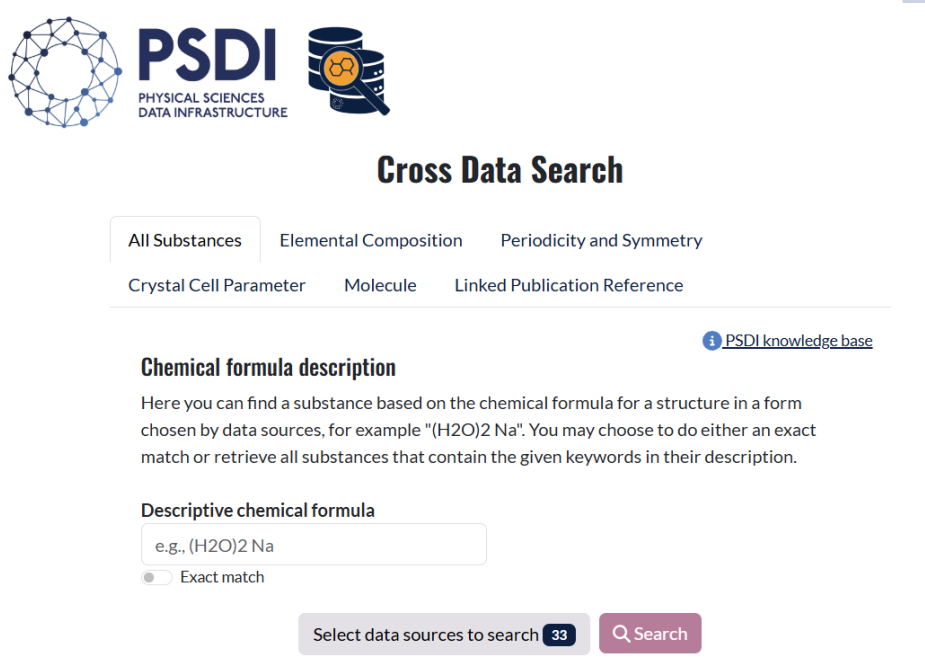

Guidance, Training & Case Studies


Collaboration & Community

See https://www.psdi.ac.uk for more information

PSDI Cross Data Search

- Powered by OPTIMADE (Python tools)!
- Covers many data sources
  - ‘External’ OPTIMADE providers
  - ‘Internal’ PSDI OPTIMADE providers
- Users can search using various approaches:
  - Elemental composition
  - Crystal cell parameters
  - Chemical identifier (SMILES or InChI)
  - Publication reference
  - Advanced search = full OPTIMADE query
- Users can select data sources to query
- Explore custom fields for data sources



‘Internal’ PSDI OPTIMADE sources

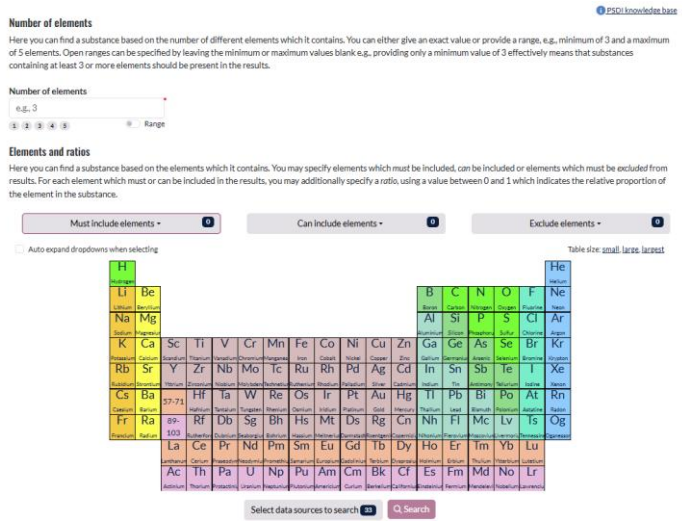
- Propersea
- ChASe
- CCDC
- UK Catalysis Hub Data
- UK Catalysis Hub Publications
- STFC Open Data
- Chemotion

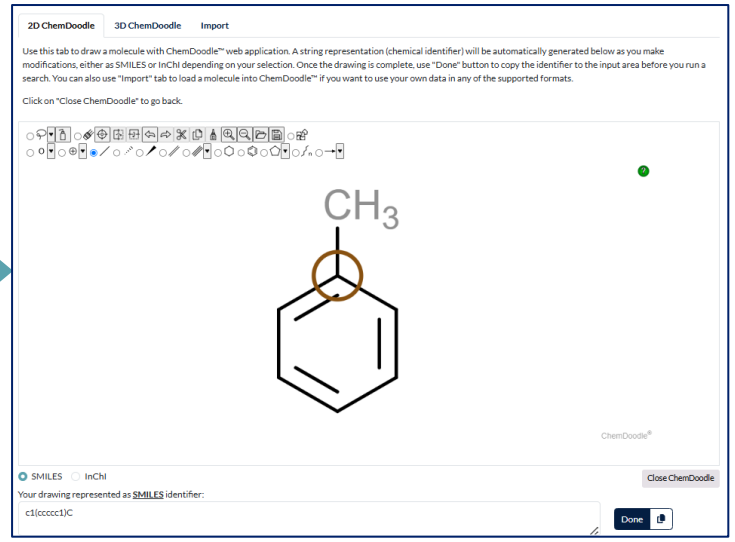
Try it yourself: https://data-search.psdi.ac.uk

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

PSDI Cross Data Search

- Powered by OPTIMADE (Python tools)!
- Covers many data sources
  - ‘External’ OPTIMADE sources
  - ‘Internal’ PSDI OPTIMADE sources
- Users can search using various approaches:
  - Elemental composition
  - Crystal cell parameters
  - Chemical identifier (SMILES or InChI)
  - Publication reference
  - Advanced search = full OPTIMADE query
- Users can select data sources to query
- Explore custom fields for data sources





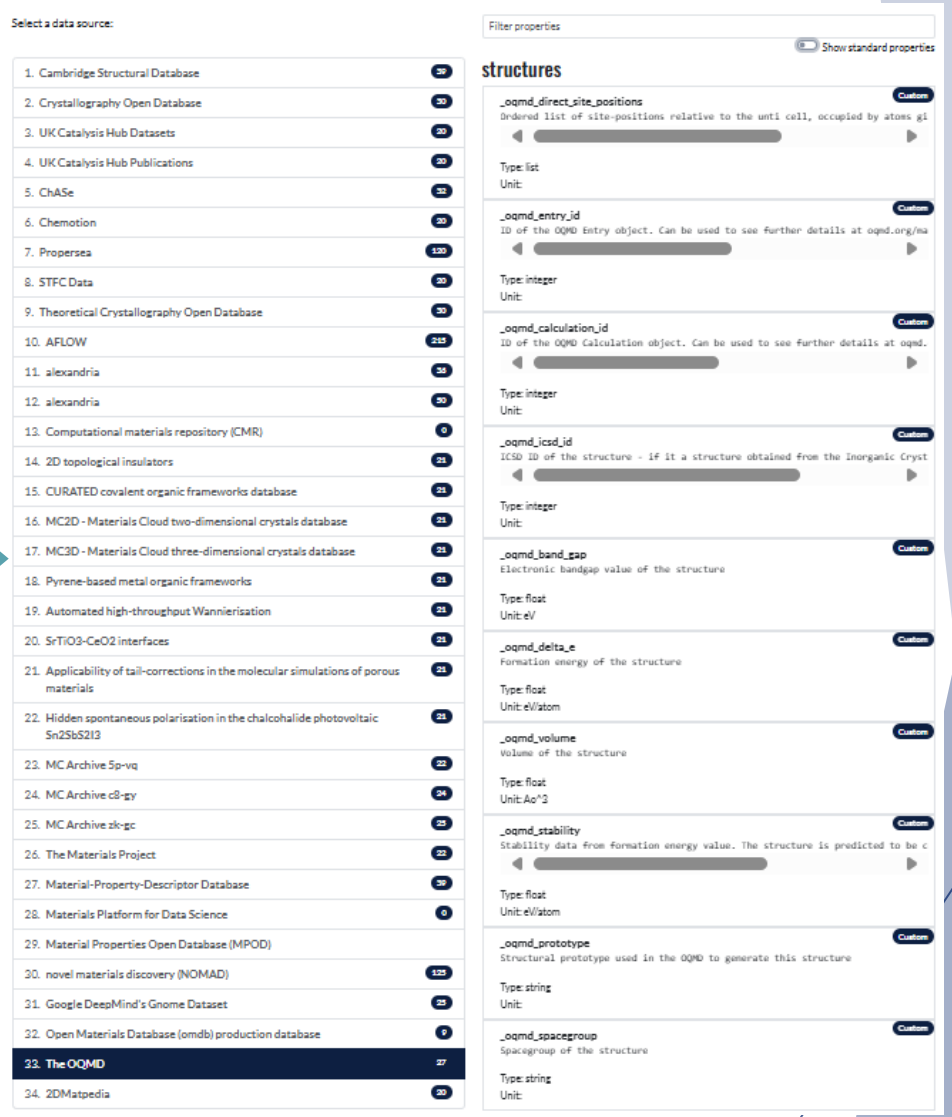
Try it yourself: https://data-search.psdi.ac.uk

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

PSDI Cross Data Search

https://data-search.psdi.ac.uk/fields-explorer

- Powered by OPTIMADE (Python tools)!
- Covers many data sources
  - ‘External’ OPTIMADE sources
  - ‘Internal’ PSDI OPTIMADE sources
- Users can search using various approaches:
  - Elemental composition
  - Crystal cell parameters
  - Chemical identifier (SMILES or InChI)
  - Publication reference
  - Advanced search = full OPTIMADE query
- Users can select data sources to query
- Explore custom fields for data sources



Try it yourself: https://data-search.psdi.ac.uk

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

Future work

- Custom property standardisation for PSDI Cross Search
  - The internal PSDI OPTIMADE sources have different custom property schemas
  - Set up a property namespace for PSDI
    - To provide definitions for custom properties used in PSDI Cross search
  - Link to cheminformatics namespace work?
    - To improve current text-string-based SMILES and InChI searches from
    - To enable molecule substructure searches: SMILES & SMARTS
- Make more OPTIMADE sources
  - CCP-NC – see Sathya's talk later!
  - 'OPTIMADEise' more PSDI data sources
    - BioSim trajectories
    - MLIP Invenio
- Expose 'internal' OPTIMADE source to the world

Thank you for your attention!