PSDI

PHYSICAL SCIENCES
DATA INFRASTRUCTURE

The Physical Sciences Data Infrastructure (PSDI): What is it and what does it do?

Workshop: NMR Data Analysis of Paramagnetic Metal Complexes
30th March 2026

Tom Underwood
Science & Technology Facilities Council (STFC)

https://www.psdi.ac.uk/

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

Presentation Outline

- Motivation for PSDI
- What is PSDI?
- What does PSDI provide?
- Getting involved with PSDI

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

Motivation

- Challenges faced by scientists using other people's data…

The data is not…	Possible sources of problem. The data is…	Possible solutions
Findable	• not included in search engines • lost in too many search results • no longer located where it once was	• Search engines with more data sources • Detailed metadata for precise searches • Persistent identifiers
Accessible	• not published: 'Available on request' • behind a paywall • too large to download/share	• Open data • Infrastructure for archiving and sharing data
Interoperable / Reusable	• in the wrong format for my purposes	• Data standards • Tools to convert between data formats
Reproducible	• not accompanied with enough information for me to recreate it	• Software for capturing data creation process - ELNs, workflow engines • Automation of workflows

This describes a 'data infrastructure'.
PSDI is this for the physical sciences*

+ community engagement + training

PSDI
PHYSICAL SCIENCES DATA INFRASTRUCTURE

What is PSDI?

A platform for data collection, sharing, aggregation, integration and curation

Supporting analysis across experimental, simulation and reference data

Connecting, combining and enhancing existing data infrastructures

Sustaining data resources beyond lifespan of individual research projects

Driven by community needs

What it isn't:

A data dump for any kind of data

Replacing existing facilities

An HPC facility

Led by:

UKRI Science and Technology Facilities Council
Scientific Computing

University of Southampton

Funded by:

UKRI UK Research and Innovation

UKRI Engineering and Physical Sciences Research Council

Digital Research Infrastructure Programme

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

PSDI Principles



Connecting existing infrastructures
Workflows to connect between systems
National and international collaboration



Best use of data
Find Data across multiple data sources (reference, experimental, simulation)
Better structured data & metadata
Reproducible data processing workflows



Best use of people
Community co-ordination
Build skills and providing training
Support and sharing guidance



Best use of technology
Deploying extensible tools and services
Use of standards and ontologies

PSDI: filling a Gap in Provision

Other domains have established provision **in the UK**, e.g.

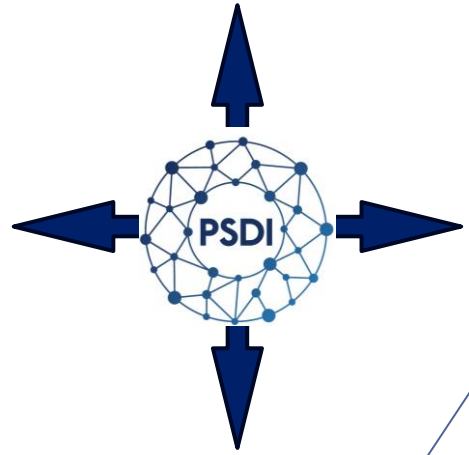
- EBI in Life Sciences
- NERC Data centres in Environmental Science
- UK Data Archive in Social Science

Other countries have initiatives underway **in this domain**, e.g.

- USA: Materials Genome Initiative
- Japan: NIMS
- European data infrastructures, e.g. E-CAM, MaX and NOMAD
- German National Research Data Infrastructure (NFDI)

We are building a *UK*, *Physical Science*, Data Infrastructure

- Supporting UK Chemistry, Materials and related disciplines
- Interfacing to other domains:
  - e.g. Life, Medical, Engineering and Environmental Sciences
- Interfacing with international initiatives



The growth in partners

PSDI
PHYSICAL SCIENCES DATA INFRASTRUCTURE

UKRI Science and Technology Facilities Council
Scientific Computing

University of Southampton

CatalysisHub

QUEEN'S UNIVERSITY BELFAST

UCL

CCPBioSim

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

PSDI

UNIVERSITY OF BATH

WARWICK
THE UNIVERSITY OF WARWICK

University of Sheffield | HENRY ROYCE INSTITUTE

HERIOT WATT UNIVERSITY

DataRevival

UNIVERSITY OF CAMBRIDGE

CCDC
advancing structural science

CCP5
Collaborative Computational Project
Computer Simulation of Condensed Phases

Loughborough University

University of Strathclyde Glasgow

CCP-NC

DCC

HEC MCC

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

What does PSDI provide?



Data Services



Data Tools



Community

Underpinning services

Authentication & Authorisation

Data Indexing

Containerisation & deployment

Resource Catalogue



Access to Data Sources



Guidance, Training & Case Studies



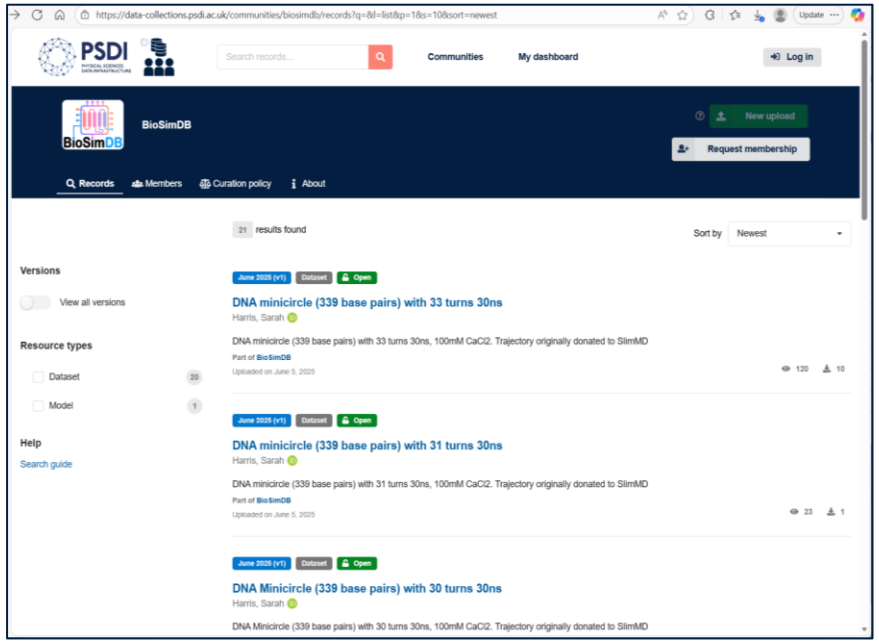
Collaboration

Find our resources on www.psdi.ac.uk!

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

Data hosting

Reference quality data collections



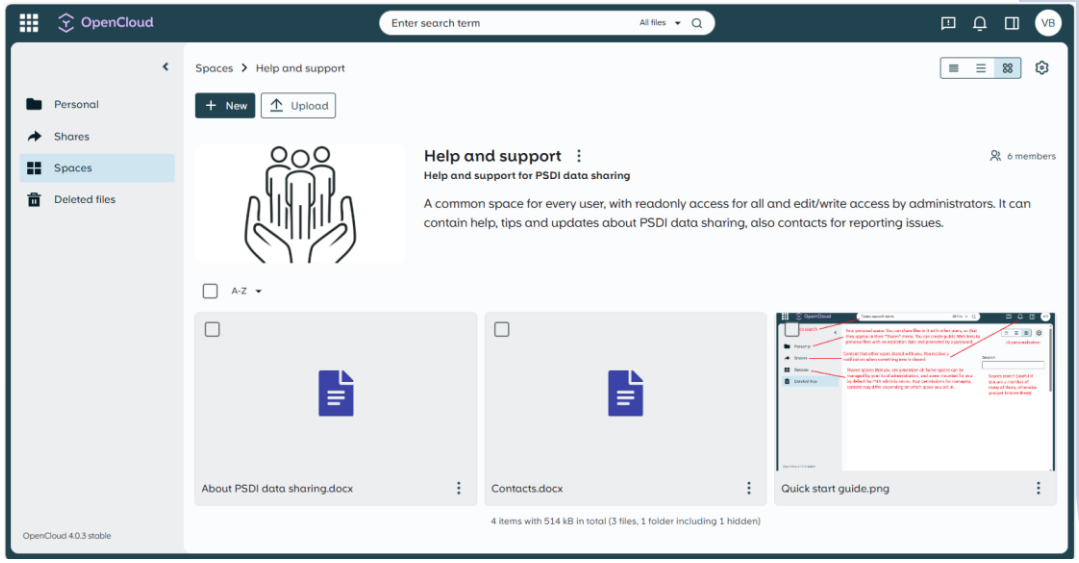
Collections include:

- BioSimDB
- Data to Knowledge
- SimpNMR_DB
- AFMDB

data-collections.psdi.ac.uk

Sharing 'active' research data

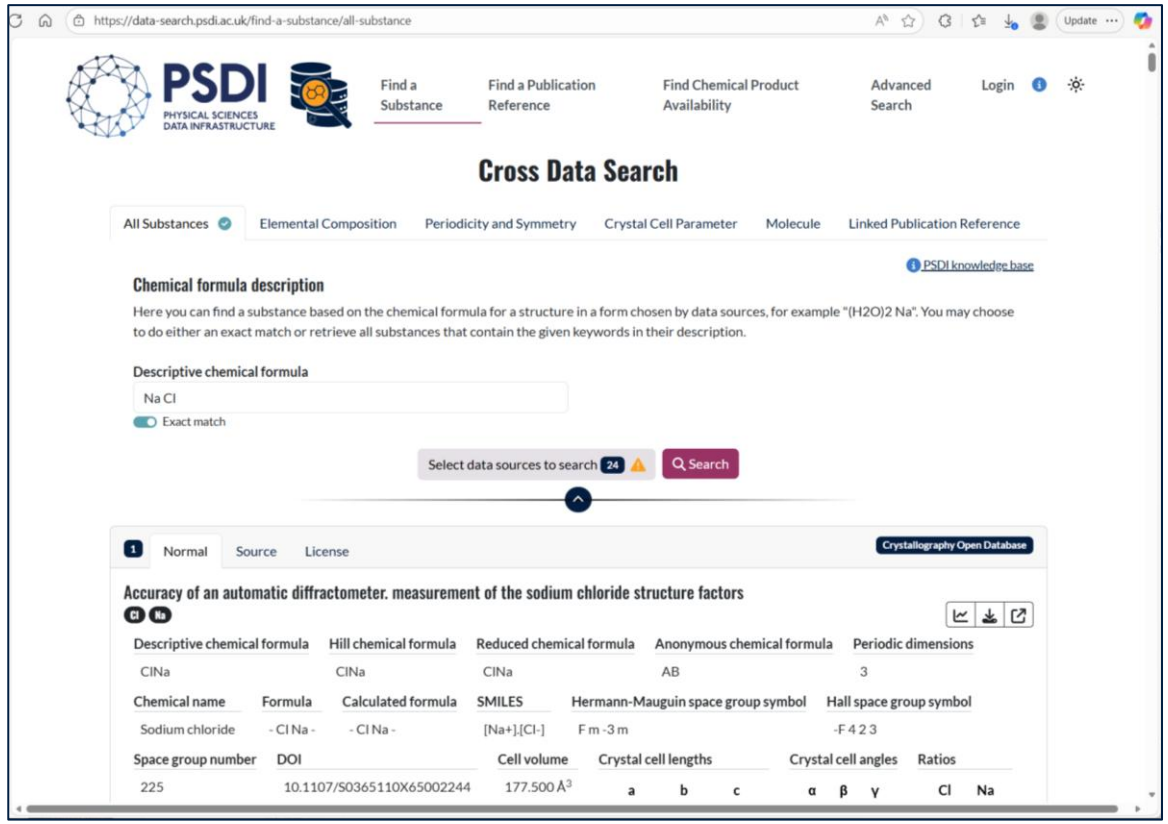
Coming soon



PSDI instance of OpenCloud:

- Personal/team/community spaces for data files
- Share big files easily across institutions
- Integrate data with other PSDI services

PSDI Cross Data Search



Find data for a substance, searching over many sources (PSDI, external & licensed)

Over 30 different data sources

Data sources include:

- Cambridge Structural Database (CSD)
- Propersea (Property Prediction)
- Chemical Availability Search (ChASe)
- Chemotion
- STFC Open Data
- OPTIMADE data sources

PSDS data sets currently being migrated:

- ICSD
- DETHERM

Coming soon

Data sources coming soon:

- PSDI community data collections

data-search.psdi.ac.uk

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

Tools & Services

PSDI Data Conversion data-conversion.psdi.ac.uk	PSDI Data Conversion Python App	Interactive learning environment learn.psdi.ac.uk
MagresView	AiiDA MLIP	PSDI Knowledge Base guidance.psdi.ac.uk
µaterials Galaxy materialsgalaxy.stfc.ac.uk	Metadise container	CatHub Thematic Portal
CrystaLLM-pi Coming soon crystallm-pi.psdi.ac.uk	CrystaLLM-pi container Coming soon	

See full list of tools & services at **resources.psdi.ac.uk**

PSDI
PHYSICAL SCIENCES
DATA INFRASTRUCTURE

Got an idea to solve a data problem?

Speak to us!



You can reach the team through psdi@soton.ac.uk and we will connect you

- Access our services and resources! Use these in your research www.psdi.ac.uk
- Co-run Events on targeted areas e.g. domain focused workshops, metadata etc. [speak to us now]
- Present Webinars to kickstart collaboration [~1 hour webinar]
- Extend, adapt and build tools with PSDI
- Attend our events to meet likeminded people and discuss your data problems and solutions
- Join us in developing guidance, best practice and toolkits