

Field-specific repositories for STEM

Jan Vališ 

Dpt. of Analytical Chemistry, Faculty of Chemical Engineering, UCT Prague
& Academic Services, NTK



Co-funded by
the European Union





Types of repositories – by focus

- Institutional (e.g. CERN, ESO, ASEP)
- General purpose (e.g. Zenodo, National Repository)
- Field-specific/Disciplinary
 - Instrument/analysis/data type (e.g. MS, FT-IR, XRD)
 - Topic (e.g. drug discovery)

Types of repositories – other aspects

- Position in lifecycle
 - raw primary data repository
 - processed data repository
 - repository for supporting data
- By curation
 - open
 - curated





Why field-specific repositories?

- Relevant
 - content
 - metadata models
 - file formats
 - data limits (size/number of files)
- Community
 - standards (e.g. quality control, access control)
 - curation
 - discoverability

What to ask when looking for a repository?

- Retrieving or contributing?
- Motivation? (community, publication or grant compliance?)
- Metadata & ontologies?
- File format compatibility?
- File size/number limitation?
- Access control/embargo?
- Licences?
- Reputation, persistence, sustainability?
- Price?

Ideal characteristics

- Relevant metadata and ontologies
- PIDs (DOI, ORCID, ROR, IGSN ...)
- Supports versioning
- FAIR
- Grant dedications
- Stable funding
- Free to contribute to

The reality

- Relevant metadata, not so much ontologies
- PIDs lacking, often only local IDs
- Versioning often lacking
- Common formats often not preferred and interoperable
- Sometimes access behind login – not open
- Grant dedications often lacking

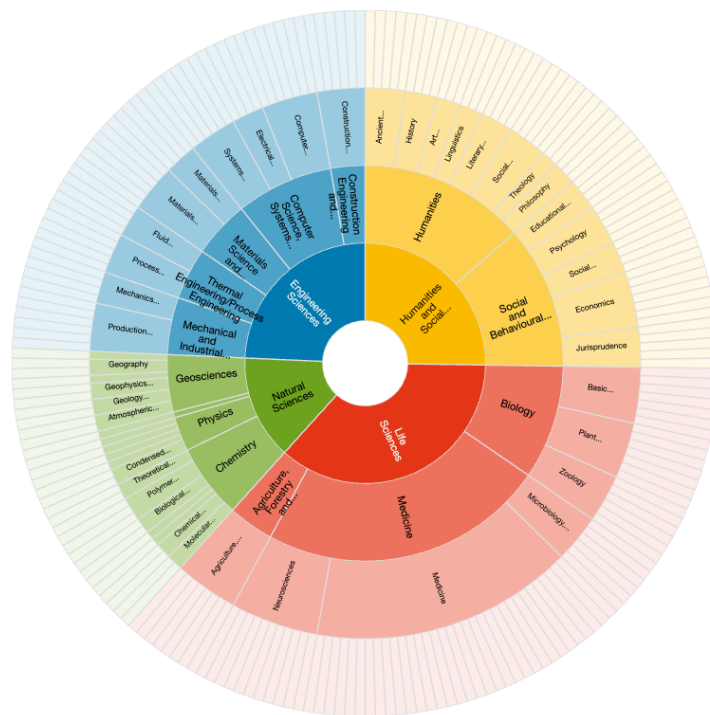
Where to look for repositories?

- All-purpose aggregators (e.g. [re3data](#), [FAIRsharing](#))
- Societies and grant agencies (e.g. [NFDI4Chem](#), [ERC](#))
- Publishers?

Browse by subject

Graphical **Text**

click to zoom into subjects or to select a bottommost subject in the hierarchy as filter for the re3data search page
shift + click on a top subject to select it as filter



re3data.org

Search

Browse

Suggest

Resources

Contact

Filter

Reset all

Subjects

Content Types

Countries

AID systems

API

Certificates

Citation References

Data access

Data access restrictions

Database access

Database licenses

Data licenses

Data upload

Enhanced publication

Institution responsibility type

Institution type

Keywords

Metadata standards

Metrics

PID systems

Provider types

Quality management

Repository languages

Software

Syndications

Repository types

Versioning

Search...

Previous

1

2

3

4

Next

Found 89 result(s)

Illinois Data Bank

Humanities and Social Sciences

Life Sciences

Natural Sciences

Engineering Sciences

Subject(s)

Institutional

Repository type(s)

dataProvider

Provider type(s)

United States

Country

United States

The Illinois Data Bank is a public access data repository that collects, disseminates, and provides persistent and reliable access to the research data of faculty, staff, and students at the University of Illinois at Urbana-Champaign. Faculty, staff, graduate students can deposit their research data directly into the Illinois Data Bank and receive a DOI for citation purposes.

Research Data at Essex

University of Essex Research Data Repository

Humanities and Social Sciences

Life Sciences

Natural Sciences

Engineering Sciences

Subject(s)

Institutional

Repository type(s)

dataProvider

Provider type(s)

United Kingdom

Country

United Kingdom

The Research Data Repository is the University of Essex's online data repository where research data resulting from research taking place within the university can be deposited, published and made accessible to the research community.

RWTH Publications Research Data

RWTH Publications Forschungsdaten

Humanities and Social Sciences

Life Sciences

Natural Sciences

Engineering Sciences

Subject(s)

Institutional

Repository type(s)

dataProvider

Provider type(s)

Germany

Country

Germany

RWTH Publications Research Data offers all RWTH Aachen University affiliates the organizational and technical means to electronically document and publish research data at this institutional repository. Certainly, researchers are encouraged to prefer a subject specific repository whenever appropriate and available. RWTH Aachen University is the largest technical university in Germany and one of nine "German Universities of Excellence" (elite university). The University library Aachen operates the repository as a member of the join community.

Discover

UB Discover

Humanities and Social Sciences

Life Sciences

Natural Sciences

Engineering Sciences

Subject(s)

Institutional

Repository type(s)

serviceProvider

Provider type(s)

Germany

Country

Germany

Discover provides access to research data from projects related to LMU Munich, which are stored and preserved in repositories of the University Library LMU.

PubData

Research Data Repository Leuphana University Lüneburg

Humanities and Social Sciences

Social and Behavioural Sciences

Educational Research

Economics

Life Sciences

Natural Sciences

Engineering Sciences

Subject(s)

Institutional

Repository type(s)

dataProvider

Provider type(s)

Germany

Country

Germany

PubData is Leuphana's institutional research data repository for the long-term preservation, documentation and publication of research data from scientific projects. PubData is maintained by Leuphana's Media and Information Centre (MIZ) and is free of charge. The service is primarily aimed at Leuphana employees and additionally at researchers from cooperation partners contractually associated with Leuphana.

Research Data Repository FDAT

Research Data Repository of the University of Tübingen

Humanities and Social Sciences

Life Sciences

Natural Sciences

Engineering Sciences

Subject(s)

Institutional

Repository type(s)

dataProvider

Provider type(s)

Germany

Country

Germany

FDAT is a research data repository hosted by the University of Tübingen, designed to facilitate long-term archiving and publication of research data. Managed by the Information, Communication and Media Center (IKM), it primarily caters to the humanities and social sciences, while welcoming researchers from all scientific disciplines at the university. Committed to high-quality data management, FDAT emphasizes the importance of adhering to the FAIR Data Principles, promoting findability, accessibility, interoperability, and reusability of the research data it contains.

UltraViolet

Humanities and Social Sciences

Life Sciences

Natural Sciences

Engineering Sciences

Subject(s)

Institutional

Repository type(s)

dataProvider

Provider type(s)

United States

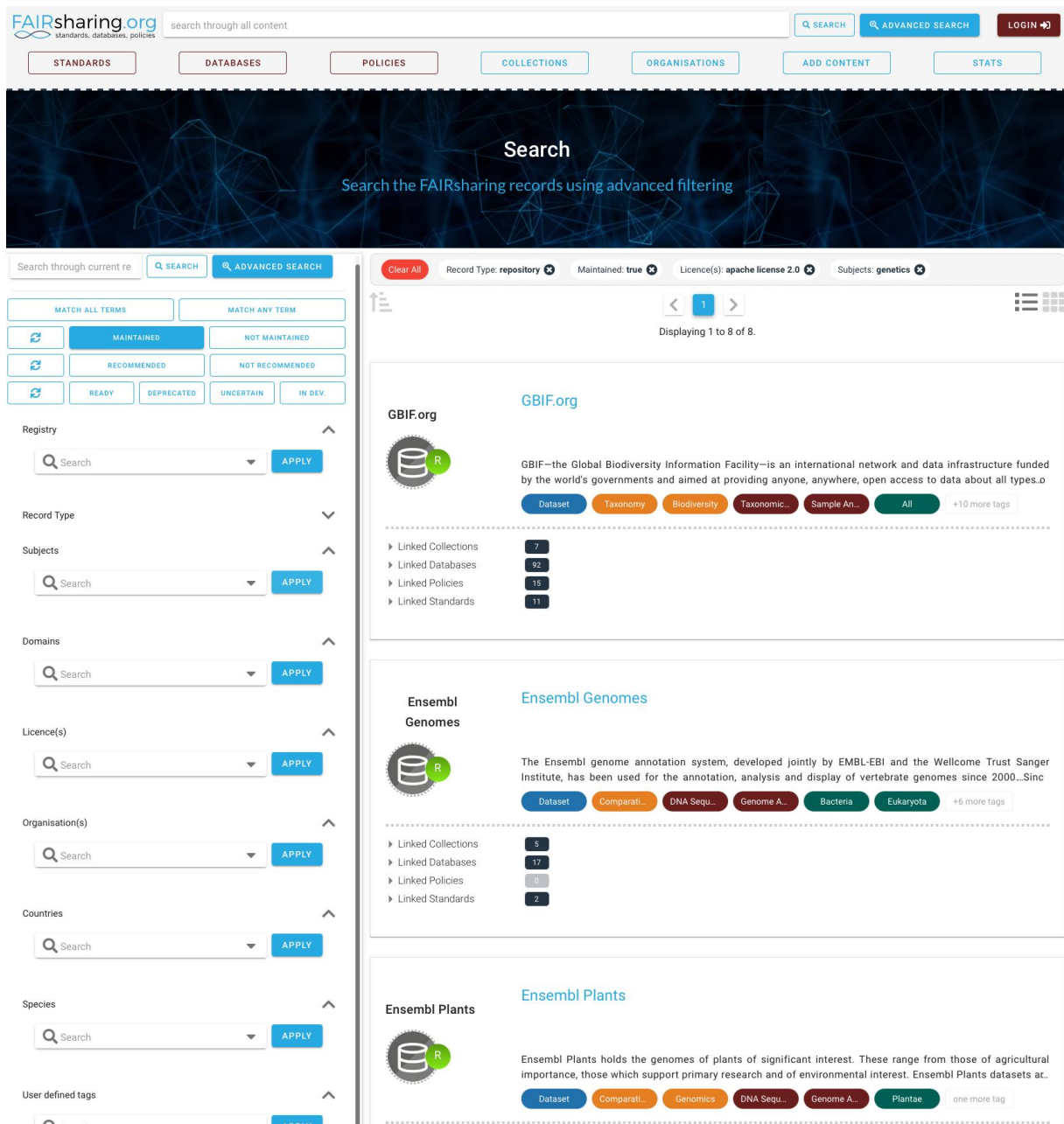
Country

United States

UltraViolet is part of a suite of repositories at New York University that provide a home for research materials, operated as a partnership of the Division of Libraries and NYU IT's Research and Instruction Technology. UltraViolet provides faculty, students, and researchers within our university community with a place to deposit scholarly materials for open access and long-term preservation. UltraViolet also houses some NYU Libraries collections, including proprietary data collections.

10

FAIRsharing



The screenshot displays the FAIRsharing.org search interface. At the top, the header includes the FAIRsharing.org logo and navigation links for STANDARDS, DATABASES, POLICIES, COLLECTIONS, ORGANISATIONS, ADD CONTENT, and STATS. A search bar is located in the top right corner.

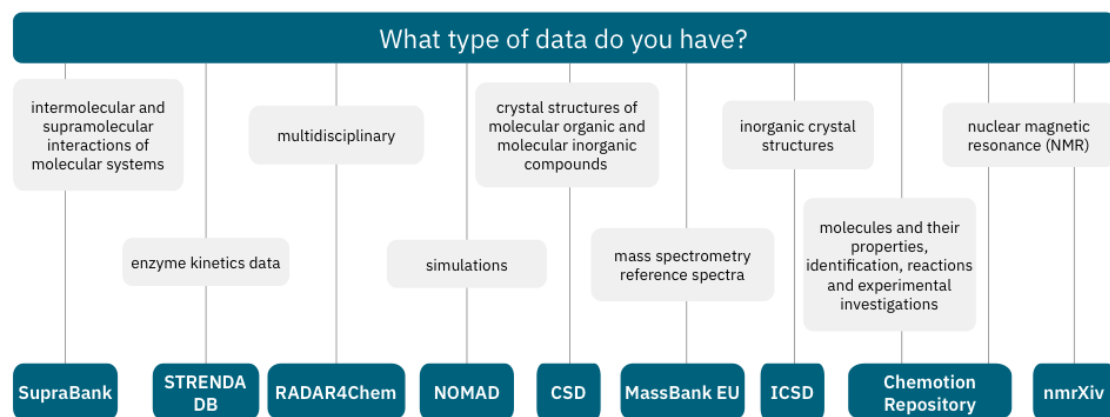
The main search area features a large banner with the text "Search" and "Search the FAIRsharing records using advanced filtering". Below this, a sidebar on the left provides various filters: "Search through current re" (with SEARCH and ADVANCED SEARCH buttons), "MATCH ALL TERMS" and "MATCH ANY TERM" options, and a series of buttons for MAINTAINED, NOT MAINTAINED, RECOMMENDED, NOT RECOMMENDED, READY, DEPRECATED, UNCERTAIN, and IN DEV. Below these are dropdown menus for Registry, Record Type, Subjects, Domains, Licence(s), Organisation(s), Countries, Species, and User defined tags, each with an APPLY button.

The main content area shows search results for "genetics". It includes a "Clear All" button, filters for Record Type (repository), Maintained (true), Licence(s) (apache license 2.0), and Subjects (genetics). A pagination bar indicates "Displaying 1 to 8 of 8".

The results are displayed in a list format, showing three entries:

- GBIF.org**: GBIF—the Global Biodiversity Information Facility—is an international network and data infrastructure funded by the world's governments and aimed at providing anyone, anywhere, open access to data about all types.o. Tags: Dataset, Taxonomy, Biodiversity, Taxonomic..., Sample An..., All, +10 more tags. Linked Collections: 7, Linked Databases: 92, Linked Policies: 15, Linked Standards: 11.
- Ensembl Genomes**: The Ensembl genome annotation system, developed jointly by EMBL-EBI and the Wellcome Trust Sanger Institute, has been used for the annotation, analysis and display of vertebrate genomes since 2000...Sinc. Tags: Dataset, Comparati..., DNA Sequ..., Genome A..., Bacteria, Eukaryota, +6 more tags. Linked Collections: 5, Linked Databases: 17, Linked Policies: 8, Linked Standards: 2.
- Ensembl Plants**: Ensembl Plants holds the genomes of plants of significant interest. These range from those of agricultural importance, those which support primary research and of environmental interest. Ensembl Plants datasets ar... Tags: Dataset, Comparati..., Genomics, DNA Sequ..., Genome A..., Plantae, one more tag.

NFDI4Chem



Common Data Types and Formats

Based on the results generated by a [survey](#) carried out by the NFDI4Chem consortium in 2020 and the yearly interviews with the repository leaders performed by NFDI4Chem's Task Area 3 repository team, a list of the most common data types and formats within this community was collected and is reported in Table 1. This table displays which data types are most commonly collected within the chemistry community and suggests which repository is most suitable for storing your specific data type.

The information provided in Table 1 will also guide you to efficiently and quickly select the best repository for the storage of your specific type of chemistry data.

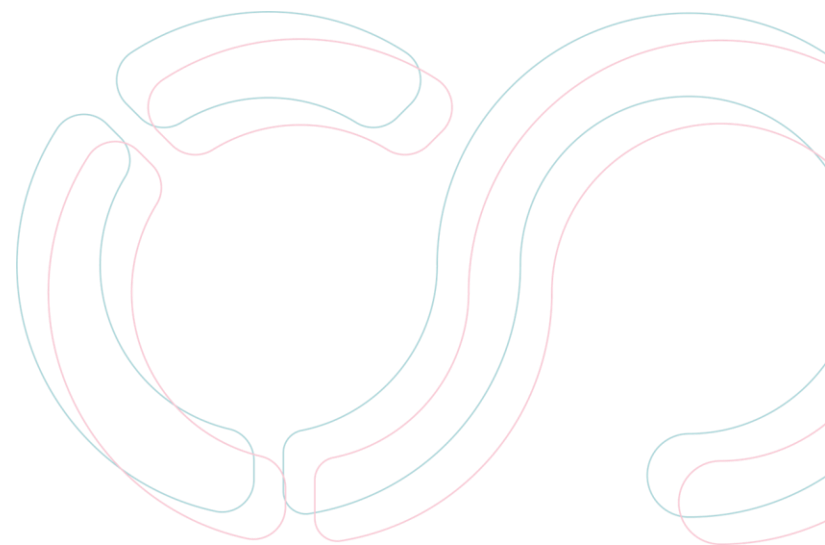
Data type	Data format	Suggested Repository	Criteria for selection
Nuclear Magnetic Resonance	Bruker format (as ZIP), JCAMP-DX	Chemotion	Passing quality checks, curation/review prior publishing
Nuclear Magnetic Resonance	Bruker format (as ZIP), JOEL format, JCAMP-DX, NMReData, nmrML	nmrXiv	Validations / Minimum information reporting standards
Mass Spectrometry	MassBank format see RMassBank	MassBank EU	Passing validation, curation prior publishing
Mass Spectrometry	mzML, JCAMP-DX; Vendor formats such as Thermo RAW are accepted and converted in Chemotion to mzML with Proteowizard's msconvert.	Chemotion	Passing quality checks, curation/review prior publishing
Infrared and Raman	JCAMP-DX; certain vendor formats are accepted and converted to JCAMP-DX by ChemConverter.	Chemotion	Passing quality checks, curation/review prior publishing
X-Ray Diffraction	JCAMP-DX; certain vendor formats are accepted and converted to JCAMP-DX by ChemConverter.	Chemotion	Passing quality checks, curation/review prior publishing
	JCAMP-DX; certain vendor formats are		Passing quality checks,

ERC

Repository name	Repository short name (if available)	Repository URL	Is the repository trusted?	Trusted - Certified repository	Trusted - Community Endorsement	Trusted - Basic Features	Exemplary Readiness Level	Essential Readiness Level	Close to Essential Readiness Level	Data Inconclusive	Coverage	ERC Panels	Content type	Specific Literature type	Access to repository content	Access to data upload	Presence in OpenAIRE	Certification	Community Endorsement	Internationally recognised	Public Policy for Preservation, Curation and Security of the content	Policy URL	
4TU.ResearchData		https://data.4tu.nl/	Yes	Yes	No	Yes	No	No	Yes	No	Disciplinary; PE1 Mathematical Data; Software/Code	No barriers	Country specific	Yes, my repository Core Trust Seal	Yes	No	Yes	Yes	Yes	Yes	https://data.4tu.nl/		
ARCHE - A Research Data Repository	ARCHE	https://arche.eur-ex.eu/	Yes	Yes	Yes	Yes	No	No	No	Yes	Disciplinary; SH4 The Human Data; Software/Code; Other	No barriers	Community/Disciplinary	Yes, my repository Core Trust Seal	Yes	Yes	Yes	Yes	Yes	Yes	https://arche.eur-ex.eu/		
Biological and Chemical Data Repository (BCO-DMO)		https://www.bco-dmo.org/	Yes	Yes	Yes	Yes	No	No	Yes	No	Disciplinary; PE10 Earth System Data; Software/Code; Other	No barriers	Funder specific	No, my repository Core Trust Seal	Yes	Yes	Yes	Yes	Yes	Yes	https://www.bco-dmo.org/		
BioStudies		https://www.ebi.ac.uk/biostudies/	Yes	No	Yes	Yes	No	No	Yes	No	Disciplinary; LS1 Molecules Data	No barriers	Community/Disciplinary	Not yet/Not sure	No certification	Yes	Yes	Yes	Yes	Yes	https://www.ebi.ac.uk/biostudies/		
DANS Data Station Social Sciences		https://ssh.dans.knaw.nl/	Yes	No	Yes	Yes	No	No	Yes	No	Disciplinary; SH1 Individual Data	No barriers	Country specific	Yes, my repository Core Trust Seal	Yes	Yes	Yes	Yes	Yes	Yes	https://dans.knaw.nl/		
Finnish Social Science Data	FSD	https://www.fsd.fi/	Yes	Yes	Yes	No	No	No	Yes	No	Disciplinary; SH1 Individual Data	No barriers	Other	Yes, my repository Core Trust Seal	Yes	Yes	Yes	Yes	Yes	Yes	Not Applicable		
GFZ Data Services		https://data.gfz-potsdam.de/	Yes	No	Yes	Yes	No	No	Yes	No	Disciplinary; PE10 Earth System Data; Software/Code	No barriers	Community/Disciplinary	Not yet/Not sure	No certification	Yes	Yes	Yes	Yes	Yes	https://data.gfz-potsdam.de/		
Lithuanian Data Repository	LIDA	https://lida.dia.lt/	Yes	No	Yes	No	No	No	Yes	No	Disciplinary; SH1 Individual Data; Software/Code	Authentication	Other	Yes, my repository Core Trust Seal	Yes	Yes	Yes	Yes	Yes	Yes	Not Applicable		
PANGAEA - Data in Earth and Environmental Sciences	PANGAEA	https://www.pangaea.de/	Yes	Yes	Yes	Yes	No	No	Yes	No	Disciplinary; PE10 Earth System Data	No barriers	Community/Disciplinary	Yes, my repository Core Trust Seal	Yes	Yes	Yes	Yes	Yes	Yes	https://www.pangaea.de/		
Protein Data Bank	PDB	https://pdbe.org/	Yes	Yes	Yes	Yes	No	No	No	Yes	Disciplinary; LS1 Molecules Data	No barriers	Community/Disciplinary	Yes, my repository Core Trust Seal	Yes	Yes	Yes	Yes	Yes	Yes	https://www.pdb.org/		
Social Data Network	SoDaNet	https://sodanet.nl/	Yes	No	Yes	Yes	No	No	Yes	No	Disciplinary; SH1 Individual Data	No barriers	Community/Disciplinary	No, my repository Core Trust Seal	Yes	Yes	Yes	Yes	Yes	Yes	https://sodanet.nl/		
The Austrian Science Data Repository	AUSSDA	https://data.austrian.ac.at/	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Disciplinary; SH1 Individual Data; Software/Code	No barriers	Community/Disciplinary	Yes, my repository Core Trust Seal	Yes	Yes	Yes	Yes	Yes	Yes	https://aussda.austrian.ac.at/		
The Tromsø Research Data Repository	TROLLING	https://trolling.no/	Yes	No	Yes	Yes	No	No	Yes	No	Disciplinary; PE6 Computer Data; Software/Code	No barriers	Community/Disciplinary	Yes, my repository Core Trust Seal	Yes	Yes	Yes	Yes	Yes	Yes	https://site.uio.no/trolling/		
UK Polar Data Repository	UK PDC	https://www.ukpdc.org/	Yes	Yes	Yes	No	No	No	No	Yes	Disciplinary; PE2 Fundamental Data	No barriers	Community/Disciplinary	Not yet/Not sure	No certification	Yes	Yes	Yes	Yes	Yes	Not Applicable		

Metadata contains "License" field	PID Assignment	Machine-actionable	Standardized Metadata	Metadata Standard	Linked resources	Linked resources PID(s)	CC0 or PD or equivalent Metadata	Author(s)	Description	Title	Date of publication/deposit	Venue of publication/deposit	Embargo	Metadata allow reference to "Horizon Europe"	Grant Project name	Grant project acronym	Grant project number	Record PID	Author(s) PID	Organisation PID	Grant PID
Yes	Yes	Yes	Yes	Dublin Core; DC	Yes, through a	Yes, through a	Yes	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	No	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a
Yes	Yes	Yes	Yes	We have a dedicated	Yes, through a	Yes, through a	Yes	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a
Yes	Yes	Yes	Yes	ISO19115-2, DC	Yes, through a	Yes, through a	Yes	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a
Yes	Yes	Yes	Yes	schema.org/JSON-LD	Yes, through a	Yes, through a	Yes	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a
Yes	Yes	Yes	Yes	DataCite	Yes, through a	Yes, through a	Yes	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	No	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a
Yes	Yes	Yes	Yes	DDI	Yes, through a	Yes, through a	Yes	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	No	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a
Yes	Yes	Yes	Yes	DataCite, ISO15924	Yes, through a	Yes, through a	Yes	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a
Yes	Yes	Yes	Yes	Dublin Core, DC	Yes, through a	Yes, through a	Yes	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a
Yes	Yes	Yes	Yes	Dublin Core, DC	Yes, through a	Yes, through a	Yes	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a
Yes	Yes	Yes	Yes	PDBx/mmCIF	Yes, through a	Yes, through a	Yes	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	No	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a
Yes	Yes	Yes	Yes	DDI	Yes, through a	Yes, through a	Yes	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a
Yes	Yes	Yes	Yes	CESSDA Metadata	Yes, through a	Yes, through a	Yes	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a
Yes	Yes	Yes	Yes	See the section on data	Yes, through a	Yes, through a	Yes	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a
Yes	Yes	Yes	Yes	ISO 19115	Yes, through a	Yes, through a	Yes	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a	Yes, through a

Examples - Chemistry



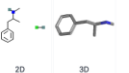
PubChem

NIH National Library of Medicine
National Center for Biotechnology Information

PubChem About Docs Submit Contact Search PubChem

COMPOUND SUMMARY

Benzeneethanamine, N,alpha-dimethyl-, hydrochloride (1:1)

PubChem CID	9306
Structure	
Primary Hazards	 Acute Toxic Health Hazard Laboratory Chemical Safety Summary (LCSS) Datasheet
Molecular Formula	$C_{10}H_{15}N.ClH$ $C_{10}H_{15}ClN$
Synonyms	di-Methamphetamine hydrochloride 300-42-5 Amphedroxyn Desfedrin Doxephlin View More...
Molecular Weight	185.69 g/mol Computed by PubChem 2.2 (PubChem release 2025.09.15)
Parent Compound	 CID 1206 (DL-methamphetamine)
Component Compounds	 CID 1206 (DL-methamphetamine) - CID 313 (Hydrochloric Acid)
Dates	Create: 2005-08-08 Modify: 2025-09-27
Description	DL-methamphetamine hydrochloride appears as odorless white crystals or fluffy crystalline powder. Bitter taste, pH (aqueous solutions) about 6. (NTP, 1992) National Toxicology Program, Institute of Environmental Health Sciences, National Institutes of Health (NIH). 1992. National Toxicology Program Chemical Repository Database. Research Triangle Park, North Carolina. ► CAMEO Chemicals

1 Structures  

✓ Comprehensive record
(structure, physical data,
characterisation ...)

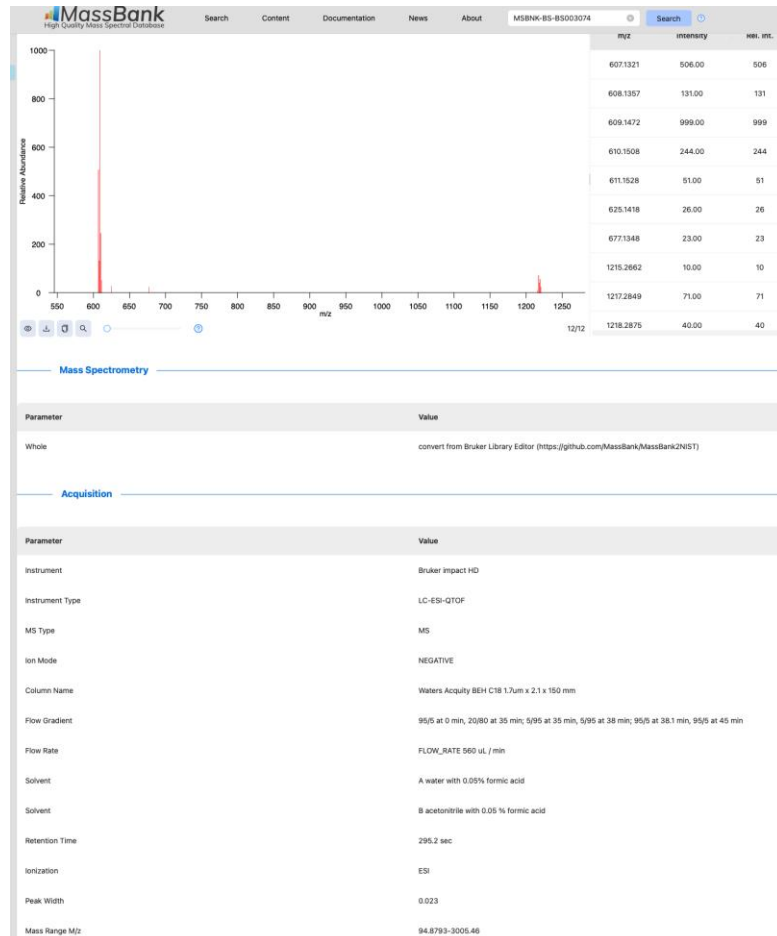
✓ Relevant metadata

• Stable? (US govt. funding)

✗ No PIDs (no DOI, ORCiD ... just
PubChem CID)

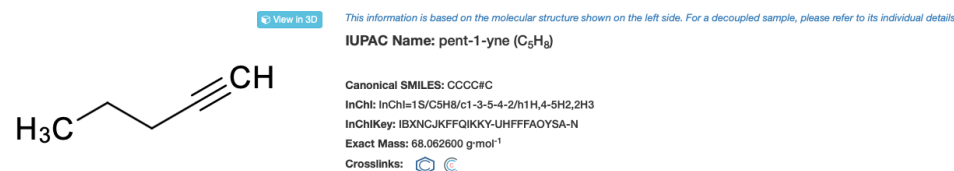
✗ No clear licencing (often ©)

Massbank



- ✓ Relevant metadata
- ✓ Open and free to use
- ✓ Supported by NFDI4Chem
- ✗ No DOI for each spectrum
(but database versioned on Zenodo
from GitHub → some DOI)

Chemotion



Click to enable zooming

 Sample Published on 2023-12-01 

Contributor:  Simone Gräßle^{1,2}

Author:  Simone Gräßle^{1,2}

1. Institute of Organic Chemistry, Karlsruhe Institute of Technology, Germany

2. Institute of Biological and Chemical Systems - Functional Molecular Systems, Karlsruhe Institute of Technology, Germany

Sample type: Consists of molecule with defined structure

Sample DOI: [10.14272/IBXNCJKFFQIKKY-UHFFFAOYSA-N.1](#)  [PDF Data](#)

Sample ID: [CRS-37169](#) 

Relations of this sample:

Reference in the Literature:

Analyses  1H NMR, 13C NMR, DEPT, COSY, HSQC, HMBC

1H nuclear magnetic resonance spectroscopy (1H NMR)

Analysis DOI: [10.14272/IBXNCJKFFQIKKY-UHFFFAOYSA-N/CHM00000593](#)  [PDF Data](#)

Analysis ID: [CRD-37164](#) 

¹H NMR (400 MHz, Chloroform-d [7.26 ppm], ppm) δ = 2.15 (td, J = 7.0 Hz, J = 2.6 Hz, 2H), 1.93 (t, J = 2.6 Hz, 1H), 1.54 (n, J = 7.3 Hz, 2H), 0.99 (t, J = 7.3 Hz, 3H).

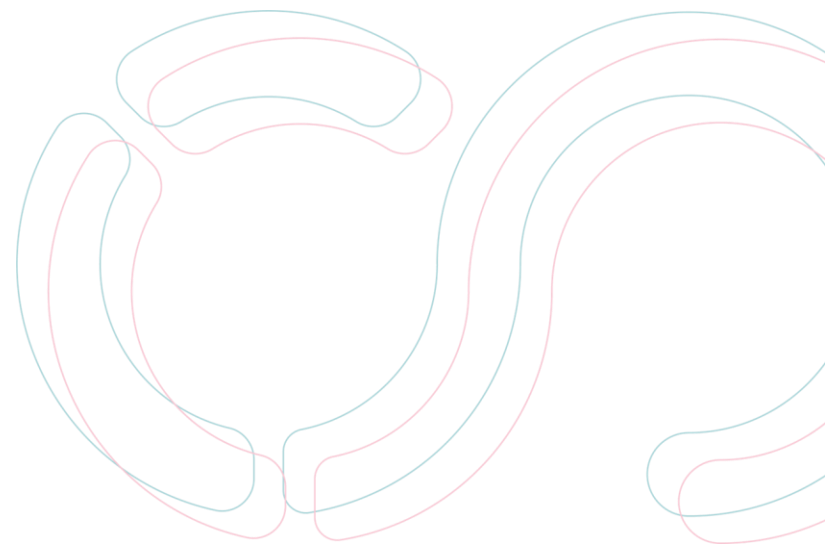
Datasets

[1H NMR in DMSO-d6](#)

[1H NMR in CDCl3](#)

- ✓ Relevant data types (structures, reactions, spectra)
- ✓ Relevant metadata & ontologies
- ✓ PIDs (on individual analysis level)
- ✓ Embargo & CC licences
- ✓ Peer-reviewed
- ✓ Stable (German govt. funding)
- ✓ Integration with ELN
- 10 GB limit per user

Examples - Biology



RCSB PDB

9OG7 | pdb_00009og7
APO SARS-COV-2-6P-MUT7 S PROTEIN 1 RBD UP CONFORMATION
PDB DOI: <https://doi.org/10.2210/pdb9OG7/pdb> EM Map EMD-70455: EMD EMDDataResource

Classification: VIRAL PROTEIN
Organism(s): Severe acute respiratory syndrome coronavirus 2
Expression System: Homo sapiens
Mutation(s): Yes

Deposited: 2025-04-30 Released: 2025-10-01
Deposition Author(s): Niu, L., Chandrasekhar, M., Tolbert, W.D., Pazgier, M.
Funding Organization(s): Canadian Institutes of Health

Experimental Data Snapshot
Method: ELECTRON MICROSCOPY
Resolution: 3.09 Å
Aggregation State: PARTICLE
Reconstruction Method: SINGLE PARTICLE
Starting Model: experimental
View more details

wwPDB Validation
Metric: Clusters, Ramachandran outliers, Sidechain outliers, Q-scores, Model-Map Fit Percentile Ranks, Value: 6, 0.4%, 4.1%, 0.306

Macromolecule Content
• Total Structure Weight: 438.28 kDa
• Atom Count: 25,396
• Modeled Residue Count: 3,130
• Deposited Residue Count: 3,840
• Unique protein chains: 1

Find similar proteins by: Sequence (by identity cutoff) | 3D Structure

Entity ID: 1	Chains	Sequence Length	Organism	Details	Image
Molecule	A, B [auth C], C [auth B]	1,280	Severe acute respiratory syndrome coronavirus 2	Mutation(s): 11 Gene Name(s): S, 2	

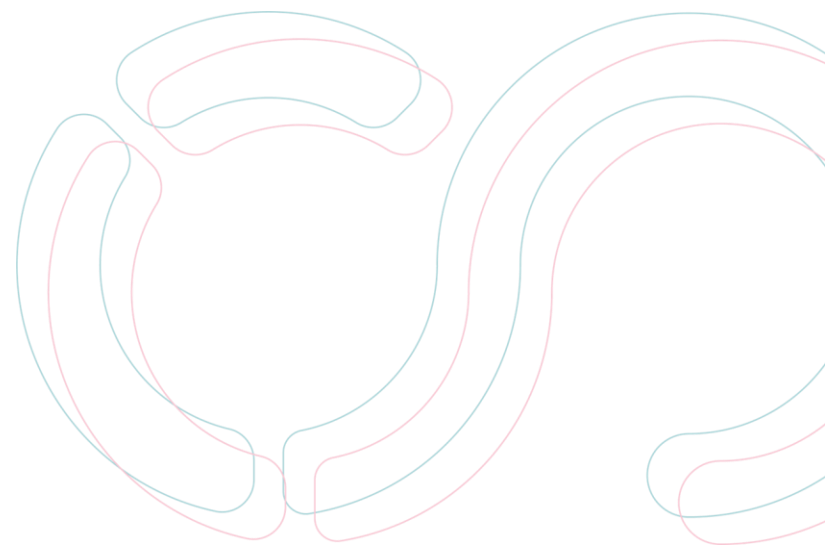
UniProt
Find proteins for **P0DTG2** (Severe acute respiratory syndrome coronavirus 2)
Explore **P0DTG2** | Go to UniProtKB: **P0DTG2**

Entity Groups
Sequence Clusters: 80% Identity, 80% Identity, 70% Identity, 60% Identity, 50% Identity, 40% Identity, 30% Identity
UniProt Group: **P0DTG2**

Glycosylation
Glycosylation Sites: 14
Go to GlyGen: **P0DTG2-1**

- ✓ Relevant data types
- ✓ Relevant metadata & ontologies
- ✓ PID
- ✓ CC 0
- ✓ automatic quality assurance
 - not for –omics, computations ...
 - limited versioning

Examples - Physics



ESO Science Archive



EsoArch

SCIENCE ARCHIVE FACILITY

Observational Raw Data Query Form

Use our tool?
Instrument-specific Interfaces
ESO Archive Overview
Archive FAQ
ESO Facility Home
ESO HOME

This query interface allows us to search and to request raw observational data taken by telescopes of the La Silla Paranal Observatory.

- You can limit your queries to either optical or infrared instruments using the MarkOptical and MarkInfrared buttons in the instrument blue panel
- When looking for images in a certain bandpass, constraints can now be provided using wavelengths in nm, or standard bandpass name (see [help page](#))
- All requested time slots can decide whether to download any of the following:

Output preferences: Hind table **Return max:** 200 **FOWS:** All Fields Syntax Help

Target, Program, and Scheduling Information

Target Name RA DEC Search Box DO ID Input RA(h) DEC(deg) Output Sexagesimal (%) deg List of Targets Vixtrain source next vixtrain 5dmg source	Resolved by SIMBAD ☐ J2000	Night (YYYY MM(M) DD) Otherwise give a query range using the following start/end dates: Start: 12 hrs [UTC] End: 12 hrs [UTC] Program ID Program Type Any ☐ PI Col. SV Any ☐ Title
---	---	---

Observing Information

Imaging	MarkOptical	MarkInfrared	Spectroscopy	Interferometry	Other	Data Product Info
☒ ALL ☐ NONE ☒ EFOSC2/LaSilla ☒ EMIR/LaSilla ☒ ERS/VLT ☒ FORSL/LaSilla ☒ FORSS/VLT ☒ HARPS/LaSilla ☒ GROND/LaSilla ☒ IMACS/VLT ☒ NACO/VLT ☒ OMEGACAM/MST ☒ SOF/LaSilla ☒ SPHERE/VLT ☒ SUBARU/LaSilla ☒ TMM2/LaSilla ☒ VMOS/VLT ☒ VIRCAM/VISTA ☒ VISIR/VLT ☒ WFI/LaSilla ☒ XSHOOTER/VLT	☐	☐	☒ ALL ☐ NONE ☒ CES/LaSilla ☒ CRISSE/VLT ☒ ESOSSC/VLT ☒ EMMI/LaSilla ☒ ERIS/VLT ☒ EBREXESOV/VLT ☒ FEROCALaSilla ☒ FORS/VLT ☒ FORS2/VLT ☒ GIRAFFE/VLT ☒ HARPS/LaSilla ☒ IBAC/VLT ☒ IRAC/VLT ☒ MUSE/VLT ☒ NACO/VLT ☒ NRPS/LaSilla ☒ SINION/VLT ☒ SPHERA/VLT ☒ TMM2/LaSilla ☒ ULVISE/VLT ☒ VMOS/VLT ☒ VIRSVL/VLT ☒ XSHOOTER/VLT	☒ ALL ☐ NONE ☒ AMBER/VLT ☒ GRAVITY/VLT ☒ MATISSE/VLT ☒ MID/VLT ☒ PIONEER/VLT ☒ VINCI/VLT ☒ ALL ☐ NONE ☒ EFOSC2/LaSilla ☒ FORS/VLT ☒ FORS2/VLT ☒ IMACS/VLT ☒ NACO/VLT ☒ SOF/LaSilla ☒ SPHERE/VLT ☒ ALL ☐ NONE ☒ EFOSC2/LaSilla ☒ FORS/VLT ☒ FORS2/VLT ☒ IMACS/VLT ☒ NACO/VLT ☒ SOF/LaSilla ☒ SPHERE/VLT ☒ VIRSVL/VLT	☒ ALL ☐ NONE ☒ ALPACA/Panorama ☒ APICAM/Panorama ☒ COLAPRAX ☒ CHETAPEX ☒ FAMM/OASIS ☒ FAMT/OASIS ☒ GBRPS/OASIS ☒ CLGSPAT ☒ MADVLT ☒ MASCOPT/Panorama ☒ SPECULOOS/Panorama ☒ WECAIM/JUKIT	Type Any ☐ User defined input: Mode Any ☐ User defined input: Dataset ID Orig Name Release Date OB Name OB ID TPL START Instrumental Setup TPL ID Exptime (imaging only) Filter bandpass (imaging only) Bandpass FWHM Griam Grating Slit

Instrument & Mode

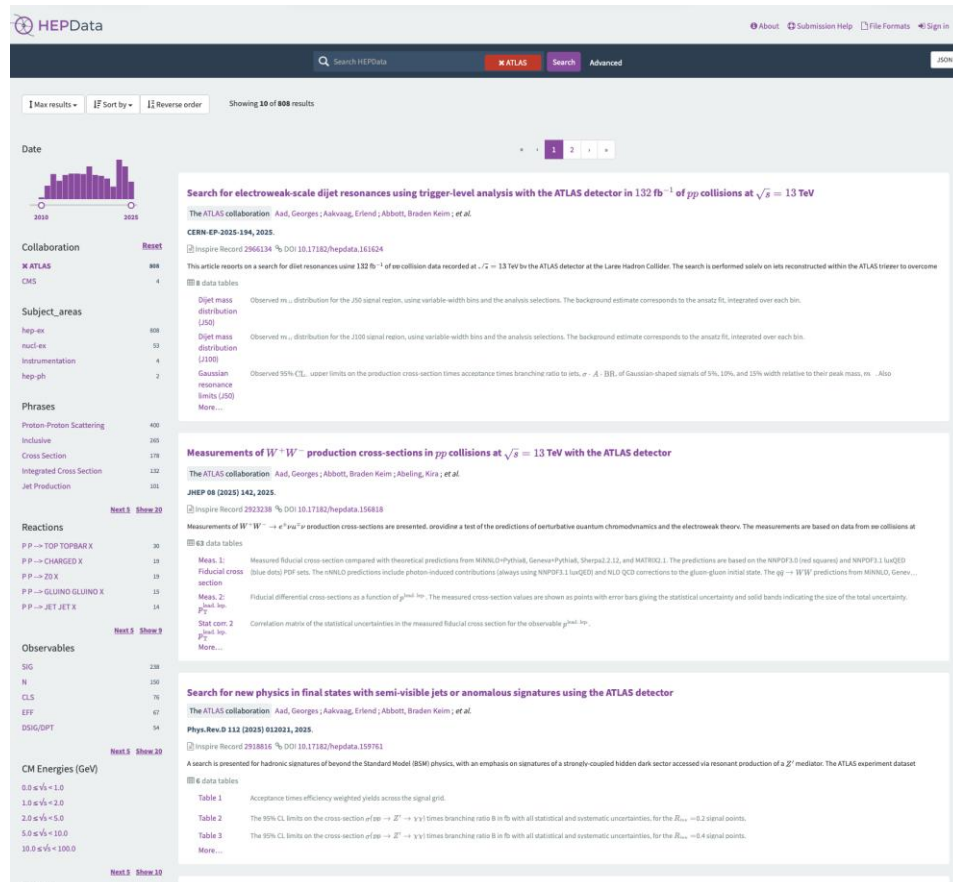
Additional Tools and Parameters

Airmass	Distance	Number of Frames / Instrument
DIMM Seeing	Position Angle	Cumulative Exptime / Filter
FITS Header Display	MJD-OBS	Sky Map

Display Options

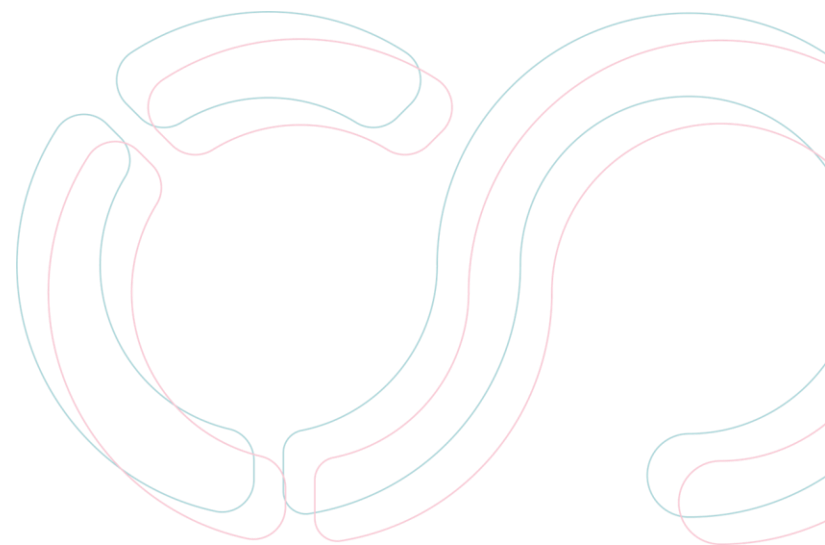
- ✓ Well documented datasets
- ✓ Relevant metadata
- Build around individual projects
 - open to read, not to submit

HEPData



- ✓ Relevant metadata
- ✓ PID
- ✓ Curated
- ✓ Trusted & sustainable (CERN)
- Only for data connected to publication
- Not only for CERN data

Earth & Environmental



PANGAEA

PANGAEA.
Data Publisher for Earth & Environmental Science

Not logged in

SEARCH SUBMIT HELP ABOUT CONTACT

Citation: Doval, María Dolores; Hansell, Dennis A (2000): Part of the global DOC versus AOU (dissolved organic carbon/apparent oxygen utilization) data compilation, OACESWOCEP15S170W [dataset]. PANGAEA, <https://doi.org/10.1594/PANGAEA.807664>,
In supplement to: Doval, MD; Hansell, DA (2000): Organic carbon and apparent oxygen utilisation in the western South Pacific and the central Indian Oceans. *Marine Chemistry*, **68**(3), 249-264, [https://doi.org/10.1016/S0304-4203\(99\)00081-X](https://doi.org/10.1016/S0304-4203(99)00081-X)

Always quote citation above when using data! You can download the citation in several formats below.
Published: 2000 (exact date unknown) • DOI registered: 2013-03-20

RIS Citation BibTeX Citation & Copy Citation Share Show Map Google Earth

Project(s): Joint Global Ocean Flux Study (JGOFS) [Q](#)

Coverage: Median Latitude: -35.055891 * Median Longitude: -170.341860 * South-bound Latitude: -66.990000 * West-bound Longitude: -175.000000 * North-bound Latitude: 0.000000 * East-bound Longitude: -168.600000
Minimum DEPTH, water: 2 m * Maximum DEPTH, water: 5097 m

Event(s): OACESWOCEP15S170W-33 [Q](#) * Latitude: -66.990000 * Longitude: -170.000000 * Location: Pacific Ocean [Q](#) * Campaign: JGOFS_not_specified [Q](#)
OACESWOCEP15S170W-34 [Q](#) * Latitude: -66.340000 * Longitude: -170.000000 * Location: Pacific Ocean [Q](#) * Campaign: JGOFS_not_specified [Q](#)
OACESWOCEP15S170W-35 [Q](#) * Latitude: -65.660000 * Longitude: -170.000000 * Location: Pacific Ocean [Q](#) * Campaign: JGOFS_not_specified [Q](#)
[Show more...](#)

Comment: compilation by Xose Anton Alvarez-Salgado (xsalgado@iim.csic.es)

Parameter(s):

#	Name	Short Name	Unit	Principal Investigator	Method/Device	Comment
1	Event label Q	Event				
2	Date/Time of event Q	Date/Time				
3	Latitude of event Q	Latitude				
4	Longitude of event Q	Longitude				
5	DEPTH, water Q	Depth water	m			Geocode
6	Temperature, water Q	Temp	°C	Doval, María Dolores Q		
7	Oxygen, apparent utilization Q	AOU	µmol/l	Doval, María Dolores Q		
8	Carbon, organic, dissolved Q	DOC	µmol/l	Doval, María Dolores Q		

License: [CC BY](#) Creative Commons Attribution 3.0 Unported (CC-BY-3.0)

Size: 3192 data points

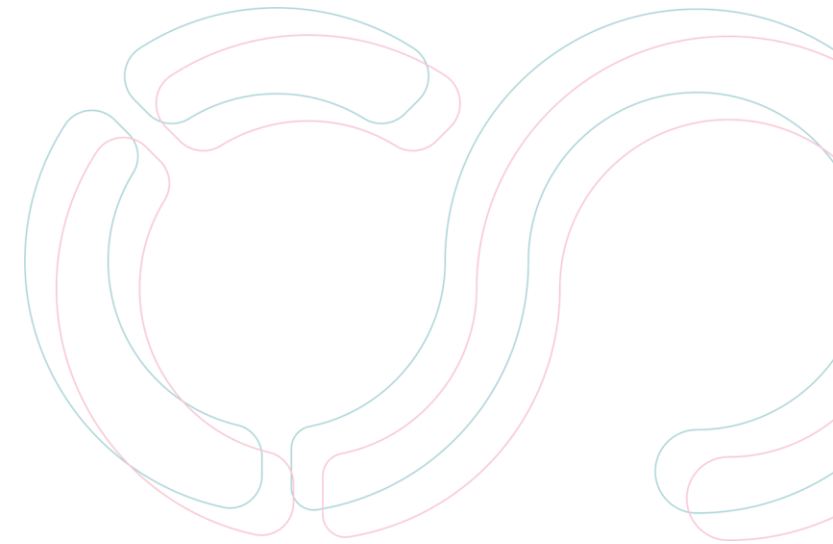
Download Data
Download dataset as tab-delimited text — use the following character encoding: UTF-8; Unicode (PANGAEA default)

View dataset as HTML (shows only first 2000 rows)

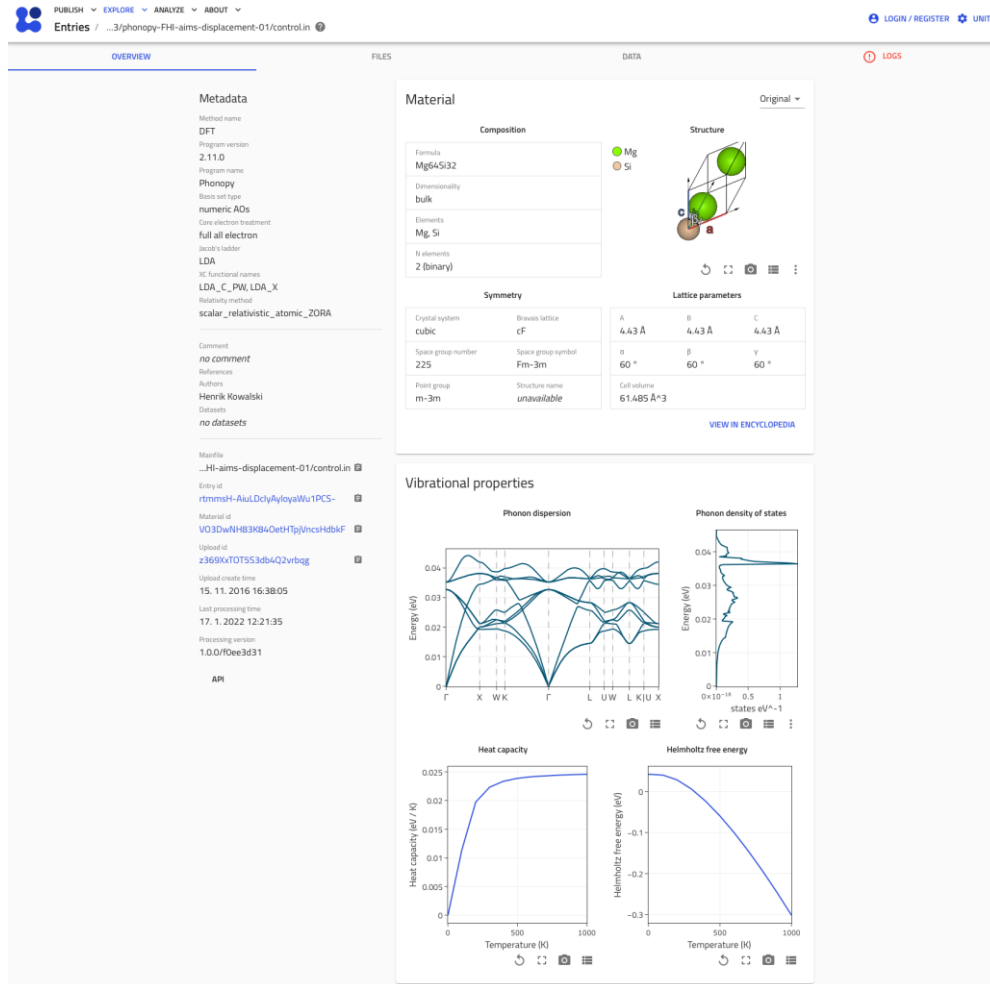
Datasets with similar metadata
Aristegui, J; Duarte, CM; Agustí, S et al. (2002): Part of the global DOC versus AOU (dissolved organic carbon/apparent oxygen utilization) data compilation, FAO9908, <https://doi.org/10.1594/PANGAEA.807626>
Alvarez-Salgado, XA; Pérez, FF; Ríos, AF et al. (2001): Part of the global DOC versus AOU (dissolved organic carbon/apparent oxygen utilization) data compilation, WOCE14, <https://doi.org/10.1594/PANGAEA.807697>
Doval, MD; Alvarez-Salgado, XA; Castro, CG et al. (2000): Part of the global DOC versus AOU (dissolved organic carbon/apparent oxygen utilization) data compilation, Fruela95, <https://doi.org/10.1594/PANGAEA.807628>

- ✓ Certified and supported
- ✓ Relevant metadata
- ✓ PIDs
- ✓ CC licences
- ✓ Community standards (e.g. common CSV headers)
- mainly for small datasets – only metadata for large datasets
- ✗ Versioning

Engineering & material sciences

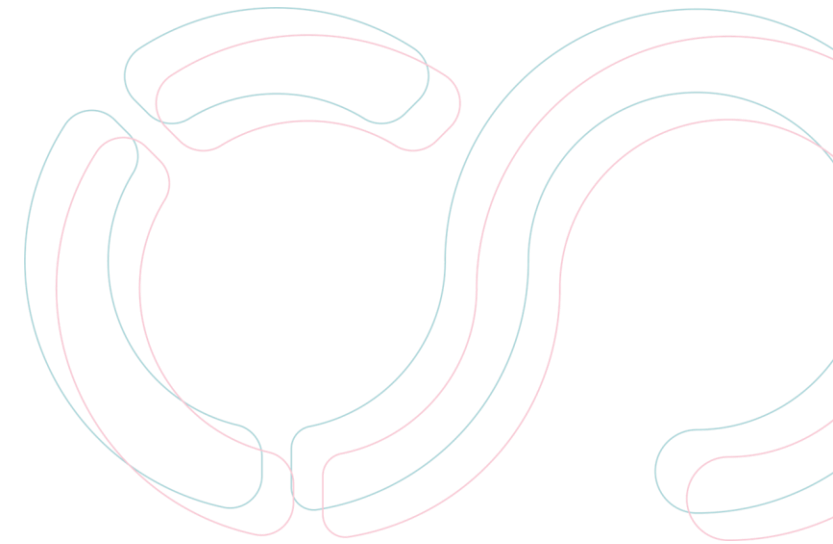


NOMAD



- ✓ Relevant metadata
- ✓ Relevant file formats
- ✓ Rich support
- ✓ DOIs
- ✓ Even for larger datasets

National Repository Platform



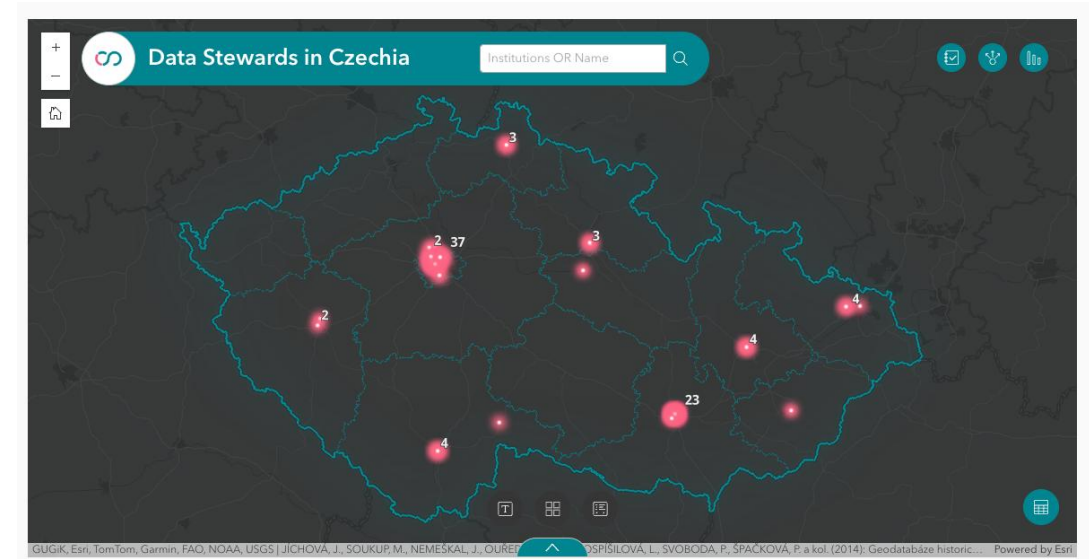
Domain specific repositories under NRP

- HW & SW common and secured for all repositories under [NRP](#)
- High security, legal compliance
- Community-build repositories

- DANTEc – material sciences, incl. chemistry (Q1 2026)
- Repository for laser physics data (Q2 2027)
- Repository for solar data (Q2 2027)
- ...

Where to get support?

- Your institutional
 - Data Stewards
 - Librarians
- Community
 - [Map of Data Stewards](#)
 - [Discord](#)



What to take home?

- Domain-specific repositories have key benefits over generic ones
- Motivation → requirements → repository selection
- Consider places, where you already retrieve data from
- Use the aggregators or your funders for overview
- Look out for the upcoming NRP repositories
- If in doubt, reach out for support – the community is here for this

Questions?

jan.valis@techlib.cz



Co-funded by
the European Union

