panosc
photon and neutron
open science cloud

ExPaNDS
European Open Science Cloud Photon and Neutron Data Services

2nd European Photon & Neutron EOSC Symposium

26 October 2021

DOI, FAIR, an MX COVID-19 Case Study

Author: Frank Von Delft

Affiliation: Diamond Light Source

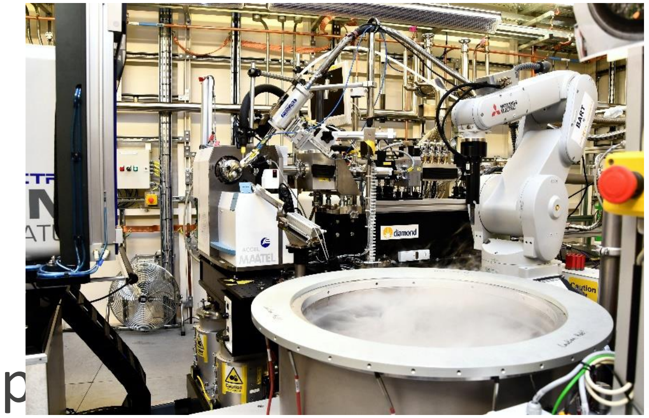
26th October 2021

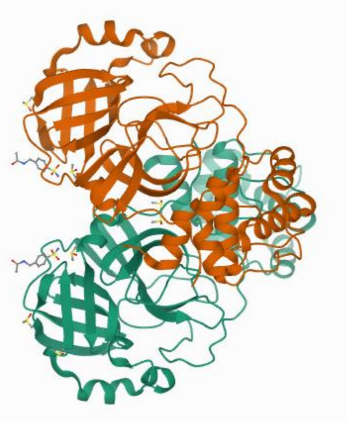
PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Driver

Urgency to contribute to global pandemic response

- XChem experiment on SARS-Cov-2 Main Protease (Mpro): <u>starting data for drug discovery</u>
- Results *(derived data)* made public for non-specialist community: Fragalysis Cloud
- Triggered COVID Moonshot – *crowdsourced open drug discovery effort*





Requests from specialist community to make raw data p

- With necessary metadata
- With links to published structures

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

XChem experiment (Animation)

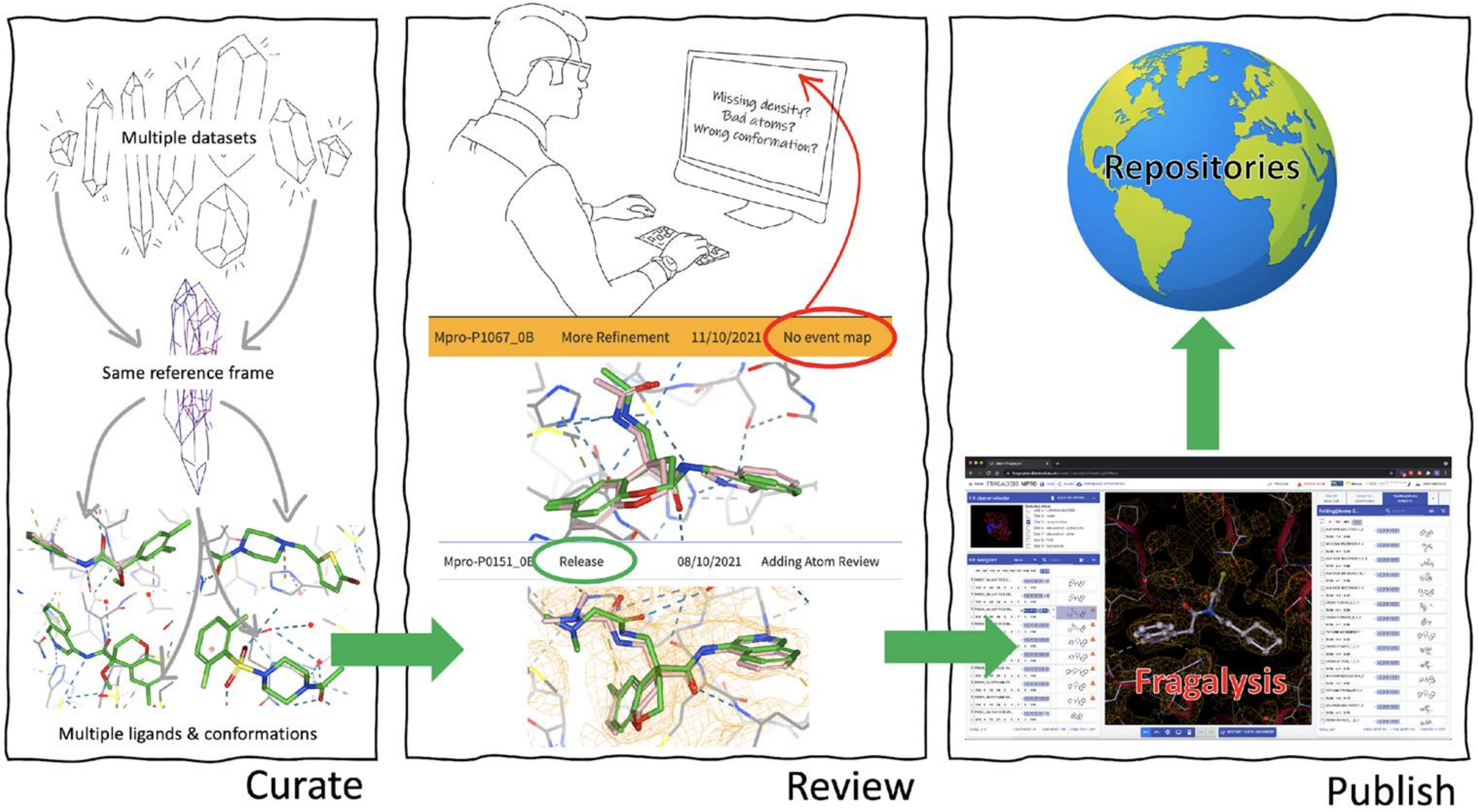
2000 diffraction datasets (8Gb each)

80 protein crystal structures

rich metadata (kbs)

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

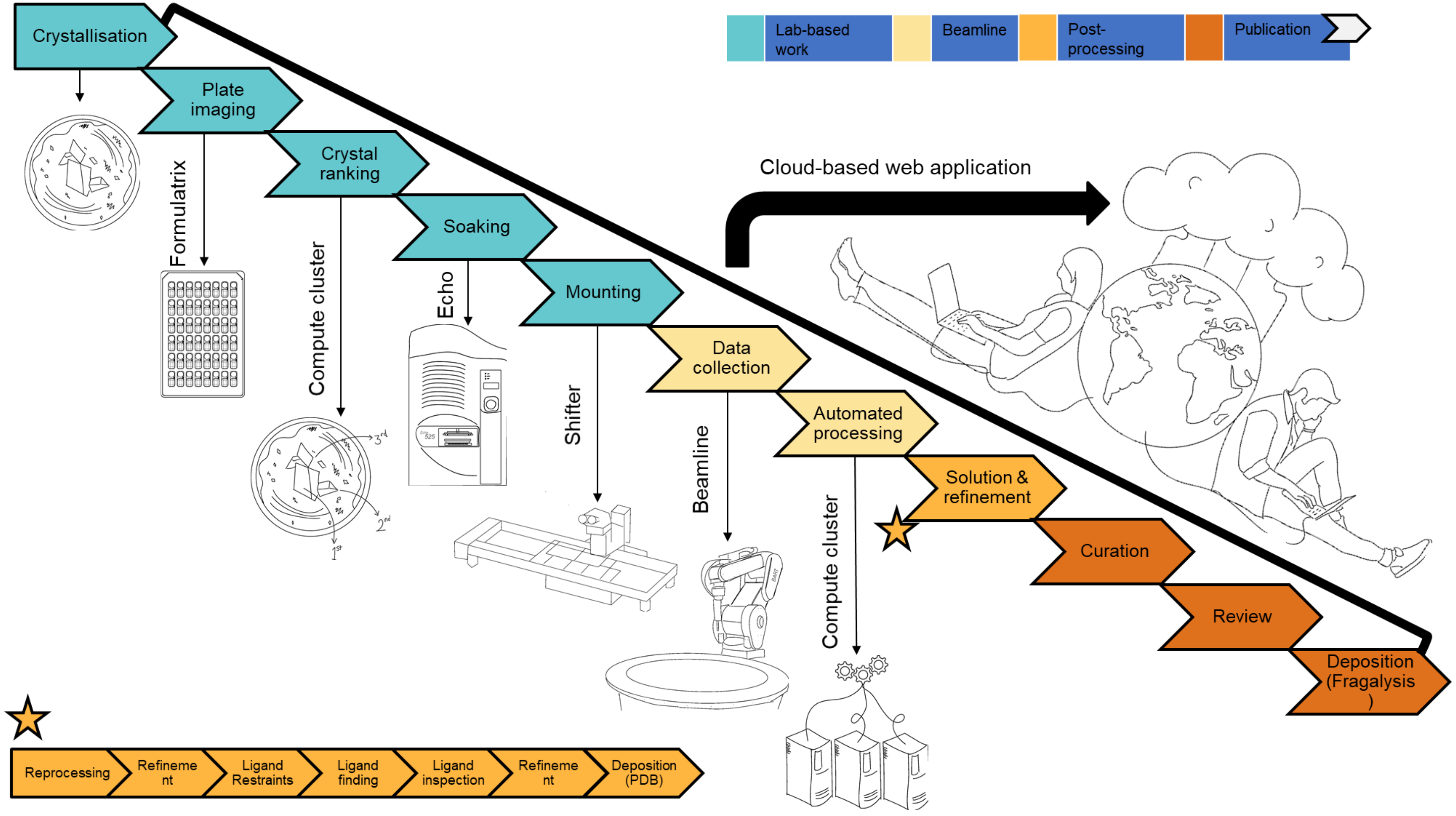
Publishing in Fragalysis



The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.
The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

Overall: lengthy experiment, analysis & release



Lab-based work

Beamline

Post-processing

Publication

Crystallisation

Plate imaging

Crystal ranking

Soaking

Mounting

Data collection

Automated processing

Solution & refinement

Curation

Review

Deposition (Fragalysis)

Formulatrix

Compute cluster

Echo

Shifter

Beamline

Compute cluster

Cloud-based web application

Reprocessing

Refinement

Ligand Restraints

Ligand finding

Ligand inspection

Refinement

Deposition (PDB)

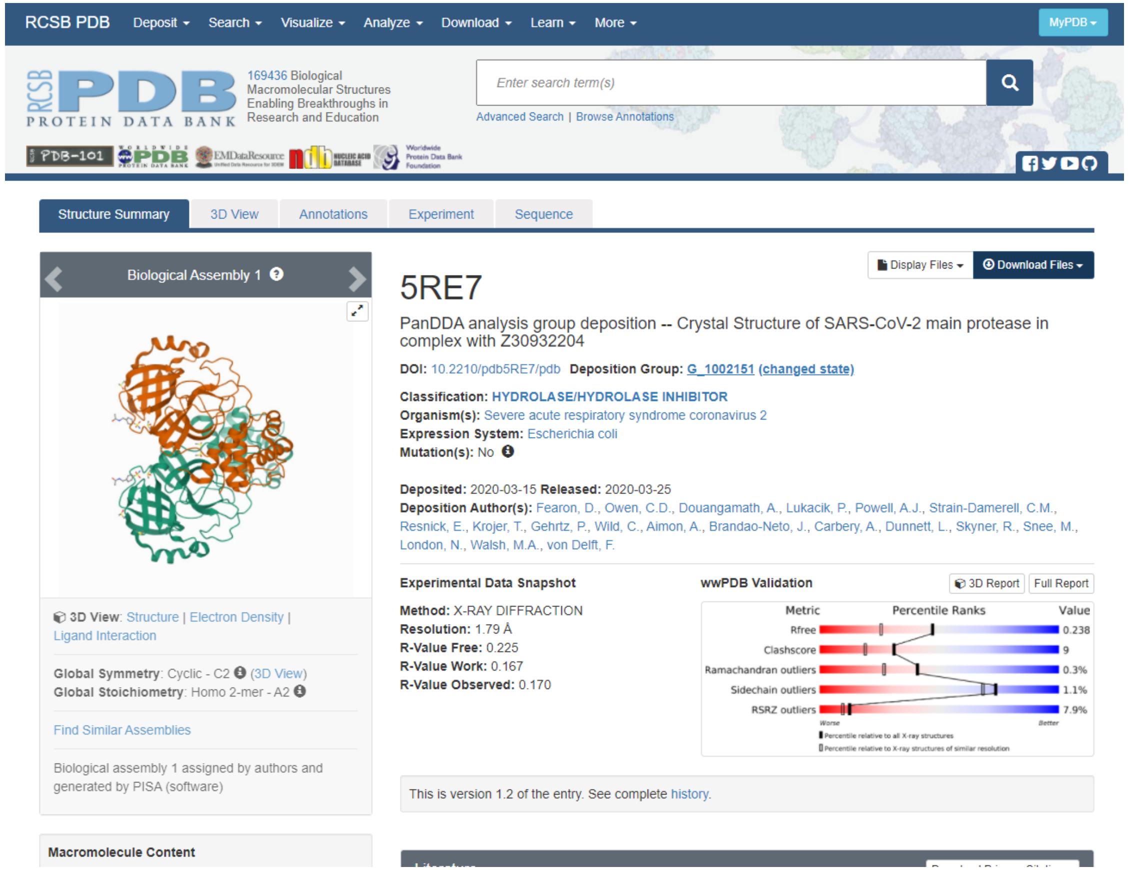
PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

The ExPaNDS project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 857641.

The PaNOSC project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 823852.

FAIR data in Macromolecular Crystallography (MX)

- Protein Data Bank (PDB) since 1971
- Structures made public
  - more recently (20 years) with “reduced data” *(=not raw)*
  - allows re-analysis
- Most relevant data and metadata included
- Free to all users
- PDB uniquely defined by 4-character name e.g. 5RE7



169436 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

5RE7

PanDDA analysis group deposition -- Crystal Structure of SARS-CoV-2 main protease in complex with Z30932204

DOI: 10.2210/pdb5RE7/pdb **Deposition Group:** G_1002151 (changed state)

Classification: HYDROLASE/HYDROLASE INHIBITOR
Organism(s): Severe acute respiratory syndrome coronavirus 2
Expression System: Escherichia coli
Mutation(s): No

Deposited: 2020-03-15 **Released:** 2020-03-25
Deposition Author(s): Fearon, D., Owen, C.D., Douangamath, A., Lukacik, P., Powell, A.J., Strain-Damerell, C.M., Resnick, E., Krojer, T., Gehrtz, P., Wild, C., Aimon, A., Brandao-Neto, J., Carbery, A., Dunnett, L., Skyner, R., Snee, M., London, N., Walsh, M.A., von Delft, F.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 1.79 Å
R-Value Free: 0.225
R-Value Work: 0.167
R-Value Observed: 0.170

wwPDB Validation

3D View: Structure | Electron Density | Ligand Interaction

Global Symmetry: Cyclic - C2 (3D View)
Global Stoichiometry: Homo 2-mer - A2

Find Similar Assemblies

Biological assembly 1 assigned by authors and generated by PISA (software)

This is version 1.2 of the entry. See complete history.

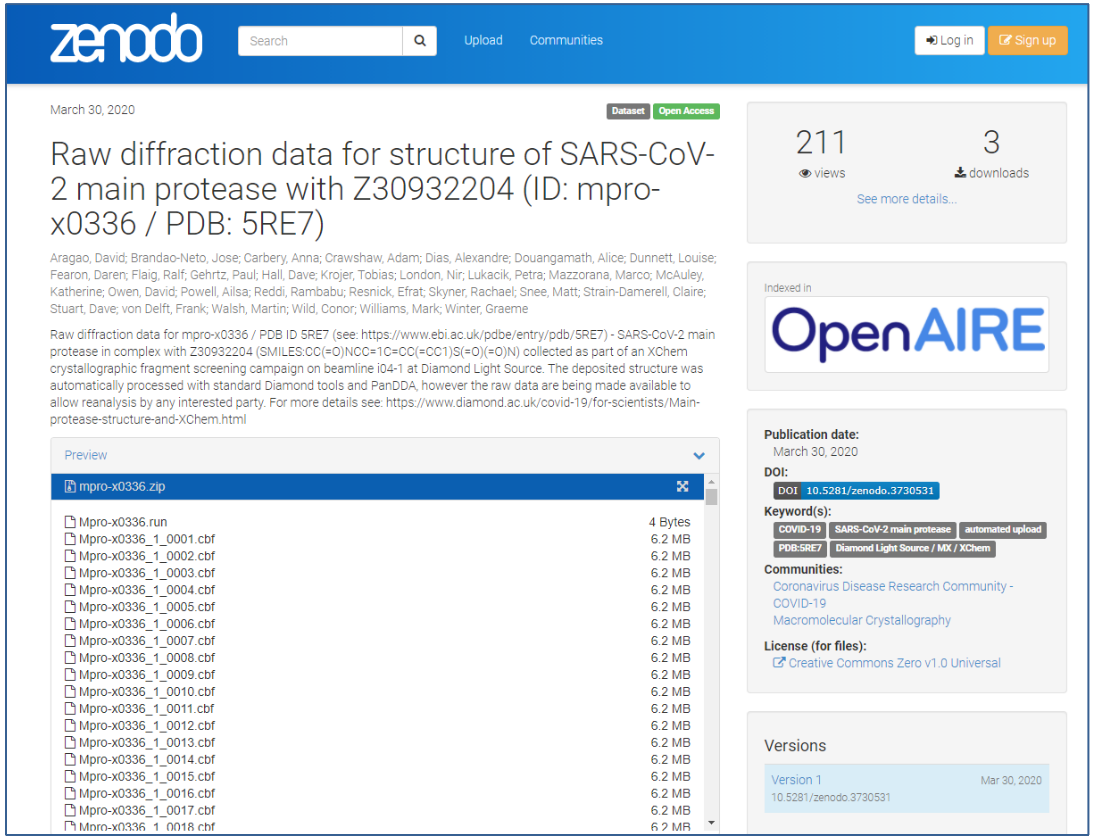
panosc photon and neutron open science cloud

ExPaNDS European Open Science Cloud Photon and Neutron Data Services

Action for *raw data*

- Use Zenodo: well-defined API, and familiarity
- “Mail merge” spreadsheet with general data
  - *(authors, text descriptions etc.)*
- Zip data for convenience for end user
  - (one file / deposition)
- Programmatic uploaded Python)
  - (unpublished) API
  - good help from developers

- Main work: *organising the metadata*



March 30, 2020

Raw diffraction data for structure of SARS-CoV-2 main protease with Z30932204 (ID: mpro-x0336 / PDB: 5RE7)

Aragao, David; Brandao-Neto, Jose; Carbery, Anna; Crawshaw, Adam; Dias, Alexandre; Douangamath, Alice; Dunnett, Louise; Fearon, Daren; Flaig, Ralf; Gehrtz, Paul; Hall, Dave; Krojer, Tobias; London, Nir; Lukacik, Petra; Mazzorana, Marco; McAuley, Katherine; Owen, David; Powell, Ailsa; Reddi, Rambabu; Resnick, Efrat; Skyner, Rachael; Snee, Matt; Strain-Damerell, Claire; Stuart, Dave; von Delft, Frank; Walsh, Martin; Wild, Conor; Williams, Mark; Winter, Graeme

Raw diffraction data for mpro-x0336 / PDB ID 5RE7 (see: https://www.ebi.ac.uk/pdbe/entry/pdb/5RE7) - SARS-CoV-2 main protease in complex with Z30932204 (SMILES:CC(=O)NCC=1C=CC(=CC1)S(=O)(=O)N) collected as part of an XChem crystallographic fragment screening campaign on beamline i04-1 at Diamond Light Source. The deposited structure was automatically processed with standard Diamond tools and PanDDA, however the raw data are being made available to allow reanalysis by any interested party. For more details see: https://www.diamond.ac.uk/covid-19/for-scientists/Main-protease-structure-and-XChem.html

mpro-x0336.zip

Mpro-x0336.run	4 Bytes
Mpro-x0336_1_0001.cbf	6.2 MB
Mpro-x0336_1_0002.cbf	6.2 MB
Mpro-x0336_1_0003.cbf	6.2 MB
Mpro-x0336_1_0004.cbf	6.2 MB
Mpro-x0336_1_0005.cbf	6.2 MB
Mpro-x0336_1_0006.cbf	6.2 MB
Mpro-x0336_1_0007.cbf	6.2 MB
Mpro-x0336_1_0008.cbf	6.2 MB
Mpro-x0336_1_0009.cbf	6.2 MB
Mpro-x0336_1_0010.cbf	6.2 MB
Mpro-x0336_1_0011.cbf	6.2 MB
Mpro-x0336_1_0012.cbf	6.2 MB
Mpro-x0336_1_0013.cbf	6.2 MB
Mpro-x0336_1_0014.cbf	6.2 MB
Mpro-x0336_1_0015.cbf	6.2 MB
Mpro-x0336_1_0016.cbf	6.2 MB
Mpro-x0336_1_0017.cbf	6.2 MB
Mpro-x0336_1_0018.cbf	6.2 MB

211 views

3 downloads

See more details...

Indexed in

OpenAIRE

Publication date: March 30, 2020

DOI: DOI 10.5281/zenodo.3730531

Keyword(s): COVID-19 · SARS-CoV-2 main protease · automated upload · PDB:5RE7 · Diamond Light Source / MX / XChem

Communities:
Coronavirus Disease Research Community - COVID-19
Macromolecular Crystallography

License (for files): Creative Commons Zero v1.0 Universal

Versions

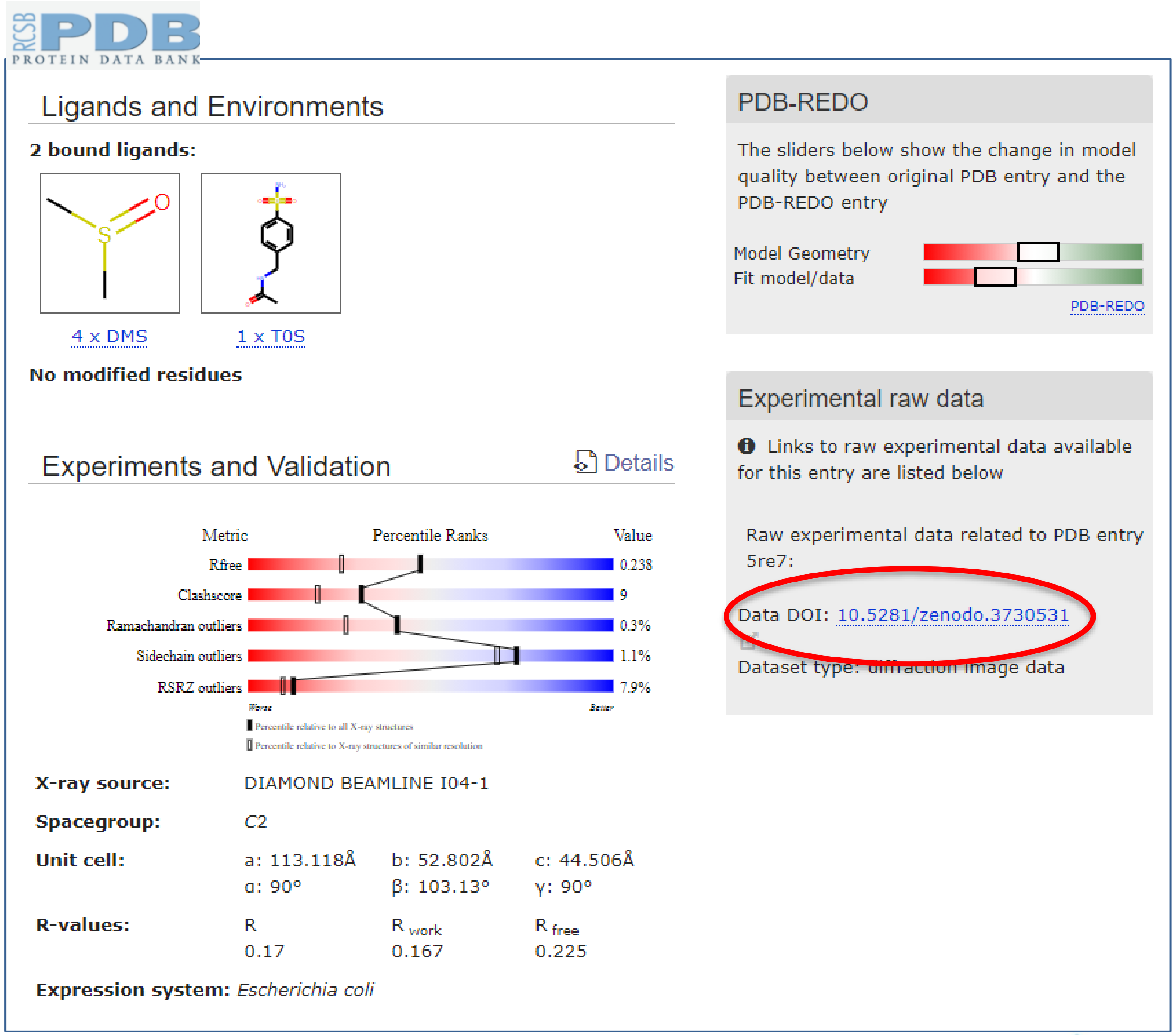
Version 1 Mar 30, 2020
10.5281/zenodo.3730531

panosc
photon and neutron open science cloud

ExPaNDS
European Open Science Cloud Photon and Neutron Data Services

Links

- List of Protein Data Bank (PDB) ID / Zenodo DOI sent to the PDB team who set up the data DOI links
- Links are bidirectional: **Zenodo ←→ PDB**
- **Easier to find for users.**



Ligands and Environments

2 bound ligands:

4 x DMS

1 x TOS

No modified residues

Experiments and Validation

Details

X-ray source: DIAMOND BEAMLINE I04-1

Spacegroup: *C2*

Unit cell: a: 113.118Å b: 52.802Å c: 44.506Å
α: 90° β: 103.13° γ: 90°

R-values: R 0.17 R $_{work}$ 0.167 R $_{free}$ 0.225

Expression system: *Escherichia coli*

PDB-REDO

The sliders below show the change in model quality between original PDB entry and the PDB-REDO entry

Model Geometry

Fit model/data

PDB-REDO

Experimental raw data

Links to raw experimental data available for this entry are listed below

Raw experimental data related to PDB entry 5re7:

Data DOI: 10.5281/zenodo.3730531

Dataset type: diffraction image data

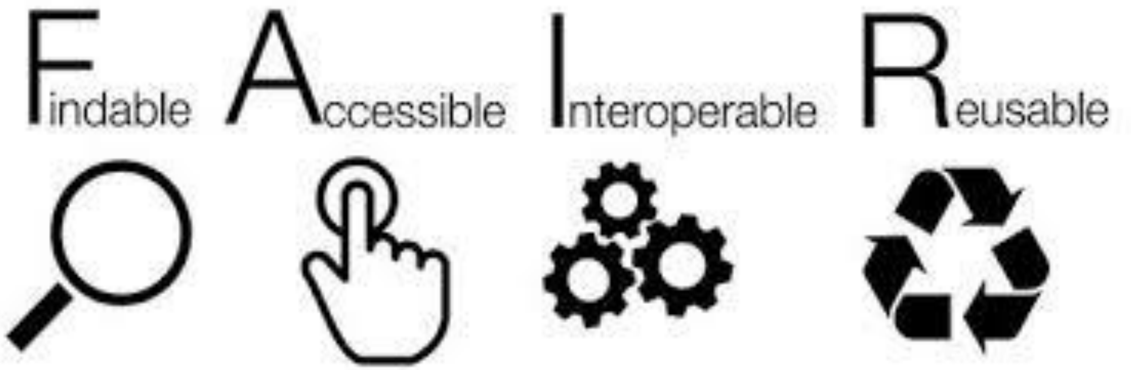
panosc
photon and neutron open science cloud

ExPaNDS
European Open Science Cloud Photon and Neutron Data Services

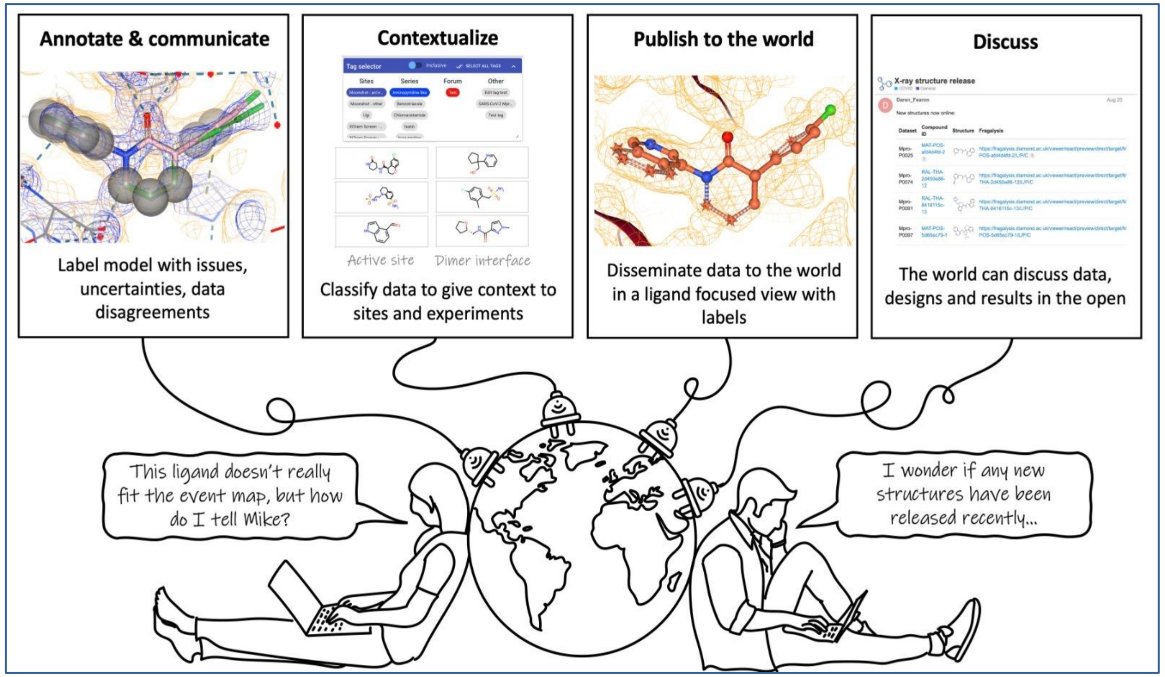
(Specialist) User perspective

- The data can be **downloaded** and used with **standard software**
- No “site” file is needed, all metadata needed are present in the dataset
- The results can be reproduced using the model in PDB
- MX users think in terms of PDB IDs ➔ We are making it the main persistent identifier
- There are no entry barriers (registration etc.)
- Open format – users are familiar with Zip files

The data are

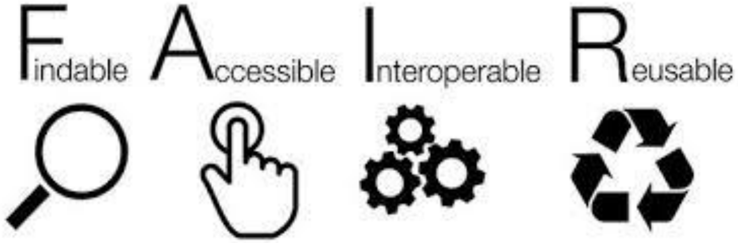


For non-specialist user


Annotate & communicate
Label model with issues, uncertainties, data disagreements
Contextualize
Active site
Dimer interface
Classify data to give context to sites and experiments
Publish to the world
Disseminate data to the world in a ligand focused view with labels
Discuss
X-ray structure release
The world can discuss data, designs and results in the open
This ligand doesn't really fit the event map, but how do I tell Mike?
I wonder if any new structures have been released recently...

- The data can be **viewed, downloaded** and annotated
- No pre-knowledge is needed…
- Trivial to share: click "share", copy URL

The data can *also* be



Implementation: HARD

panosc
photon and neutron open science cloud

ExPaNDS
European Open Science Cloud Photon and Neutron Data Services

Outcome

- COVID Moonshot consortium receives £8m funding from Wellcome on behalf of the COVID-19 Therapeutics Accelerator

- Funds pre-clinical phase: develop 5 candidates to clinical trials

- A growing collaboration:

DNDi
PostEra
Memorial Sloan Kettering Cancer Center
WEIZMANN INSTITUTE OF SCIENCE
diamond
MedChemica
wellcome
UNIVERSITY OF OXFORD
NIHR

COVID Moonshot collaborators (left to right): Ben Perry, Discovery Open Innovation Leader at DND*i*; Frank von Delft, Professor of Structural Chemical Biology at the University of Oxford and Principal Beamline Scientist at Diamond Light Source; John Chodera, Associate Member at the Memorial Sloan Kettering Cancer Center and founding member of the Folding@home Consortium; Annette von Delft, Translational Scientist at the University of Oxford; Ed Griffen, Technical Director and Co-founder of MedChemica; Alpha Lee, Chief Scientific Officer at PostEra and Faculty Member at the University of Cambridge.

Thank you

Contact details: ***frank.von-delft@diamond.ac.uk***

PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.