EMDB & EMPIAR Archives

Jack Turner

EMDB

EMBL-EBI

Outline

Archives & EMDB Website

EMDB Deposition

EMPIAR Deposition

EMBL-EBI

Archiving across scales

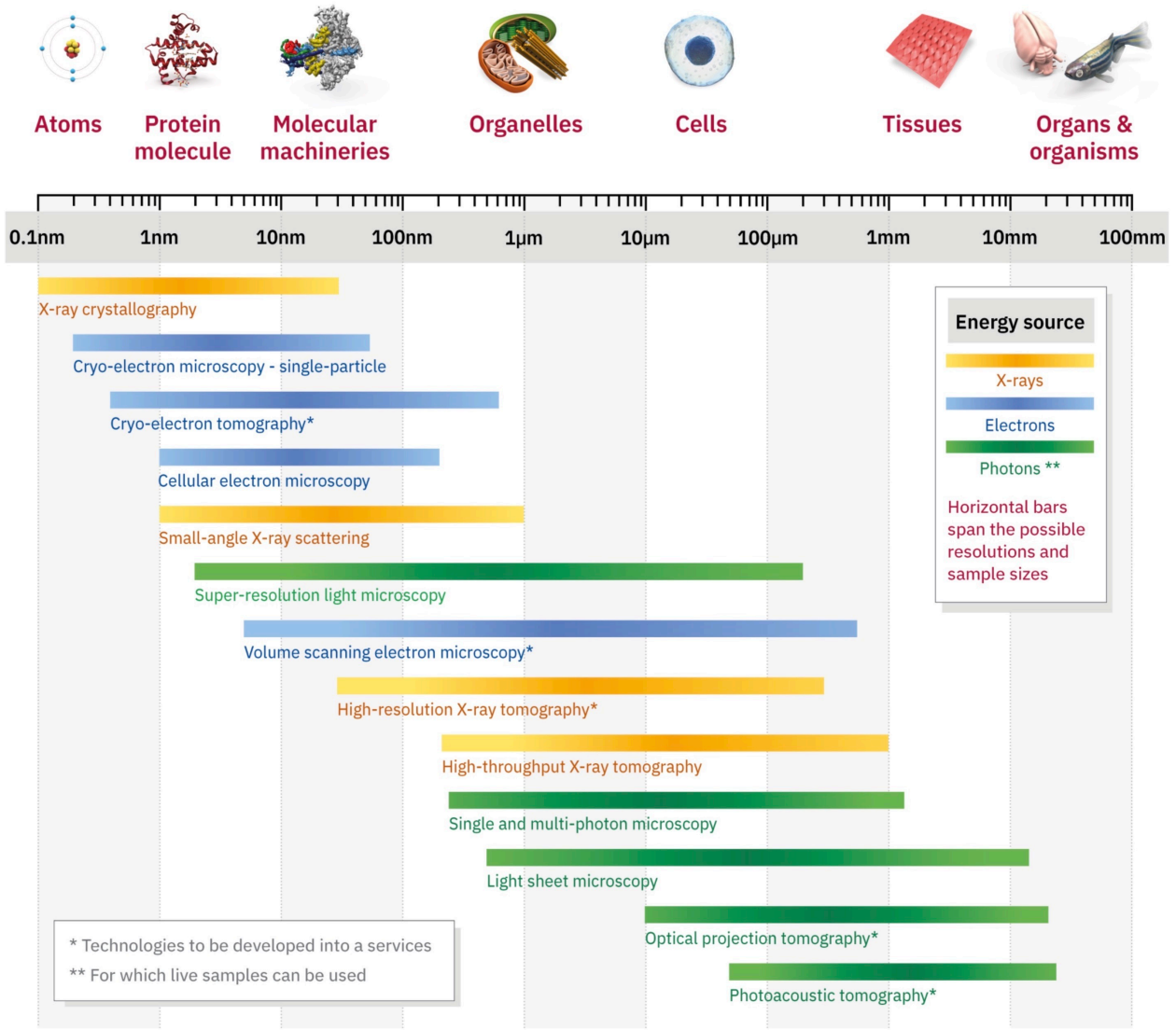


Image credit: EMBL Indicative Scheme pg150 (https://www.embl.org/documents/wp-content/uploads/2022/04/embl-programme-2022-2026.pdf)

EMBL-EBI

Archiving across scales

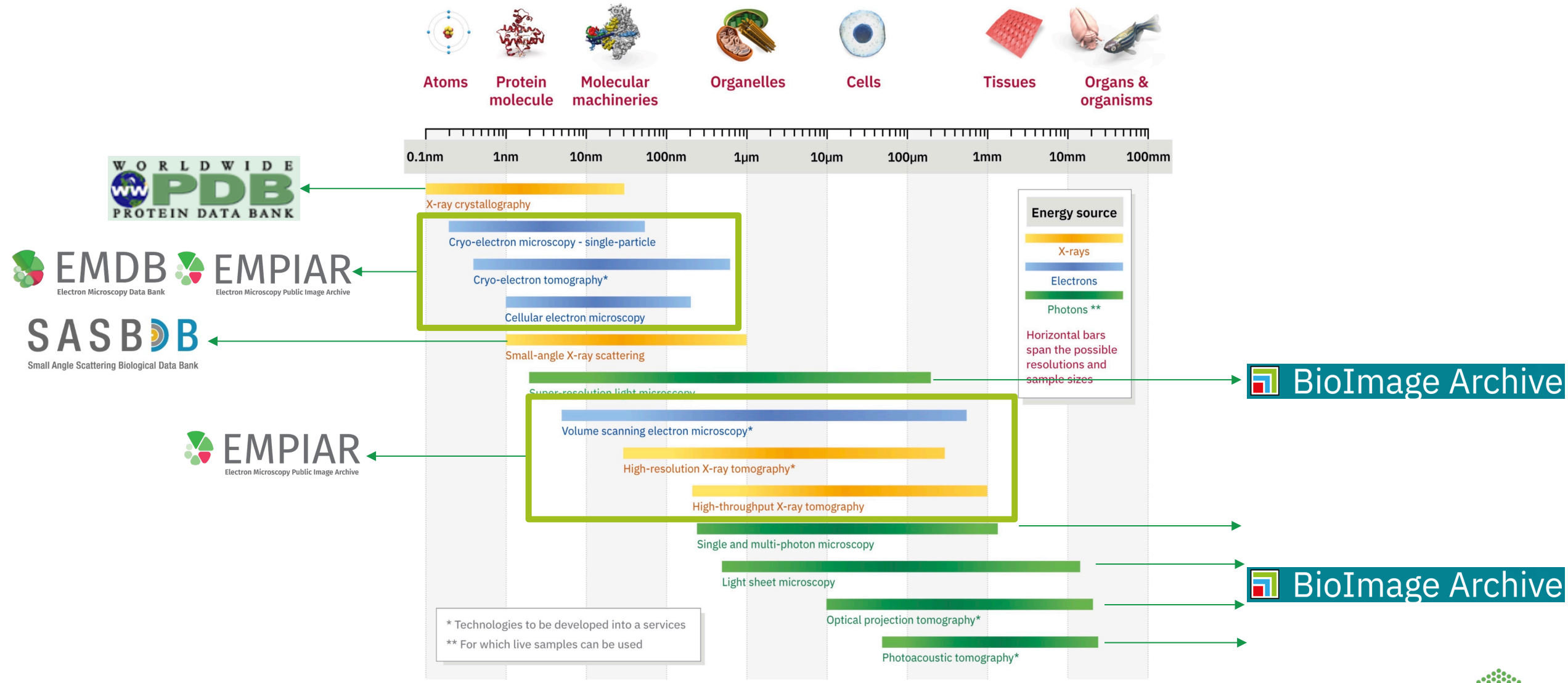


Image credit: EMBL Indicative Scheme pg150 (https://www.embl.org/documents/wp-content/uploads/2022/04/embl-programme-2022-2026.pdf)

EMBL-EBI

Archiving across scales

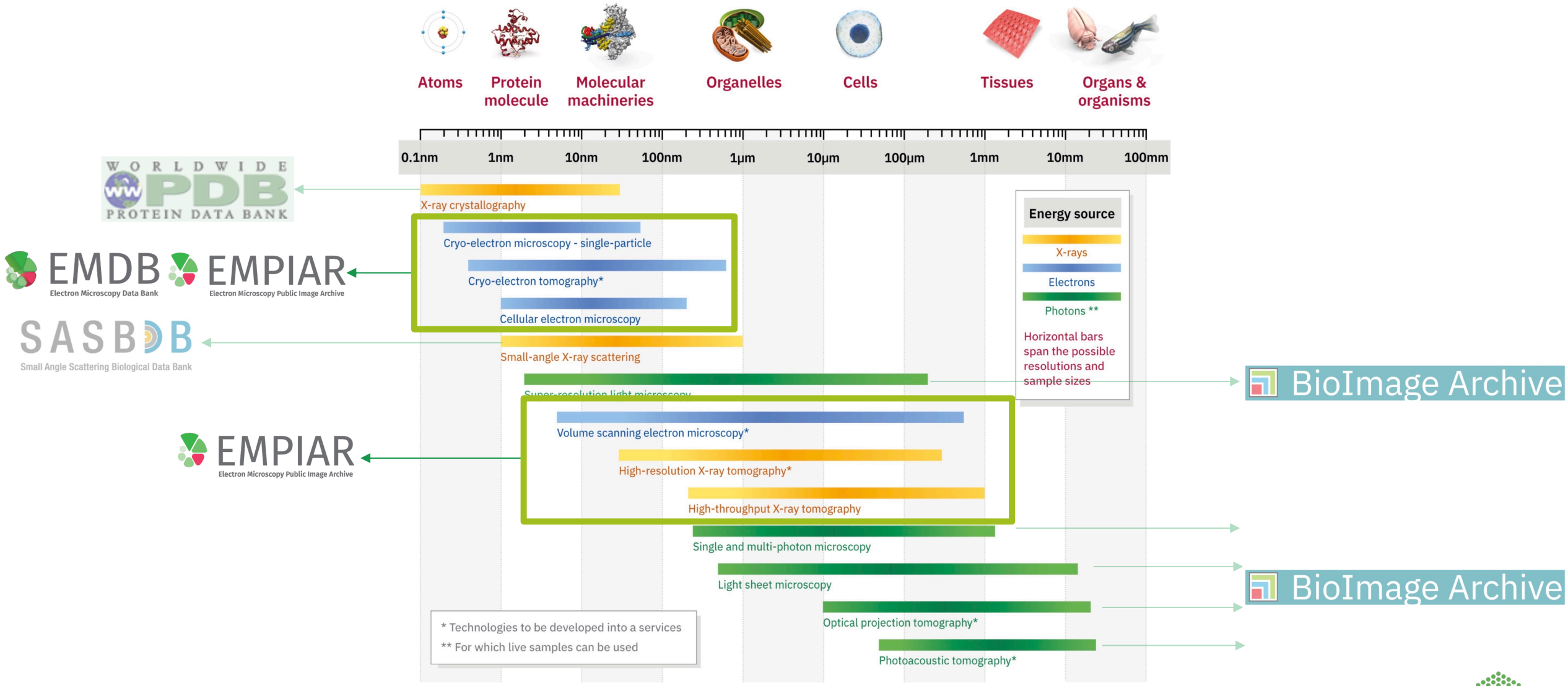


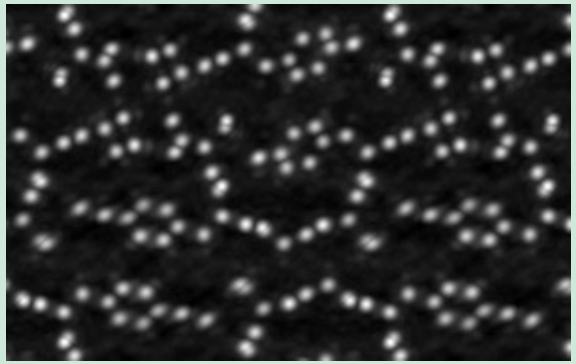
Image credit: EMBL Indicative Scheme pg150 (https://www.embl.org/documents/wp-content/uploads/2022/04/embl-programme-2022-2026.pdf)

EMBL-EBI

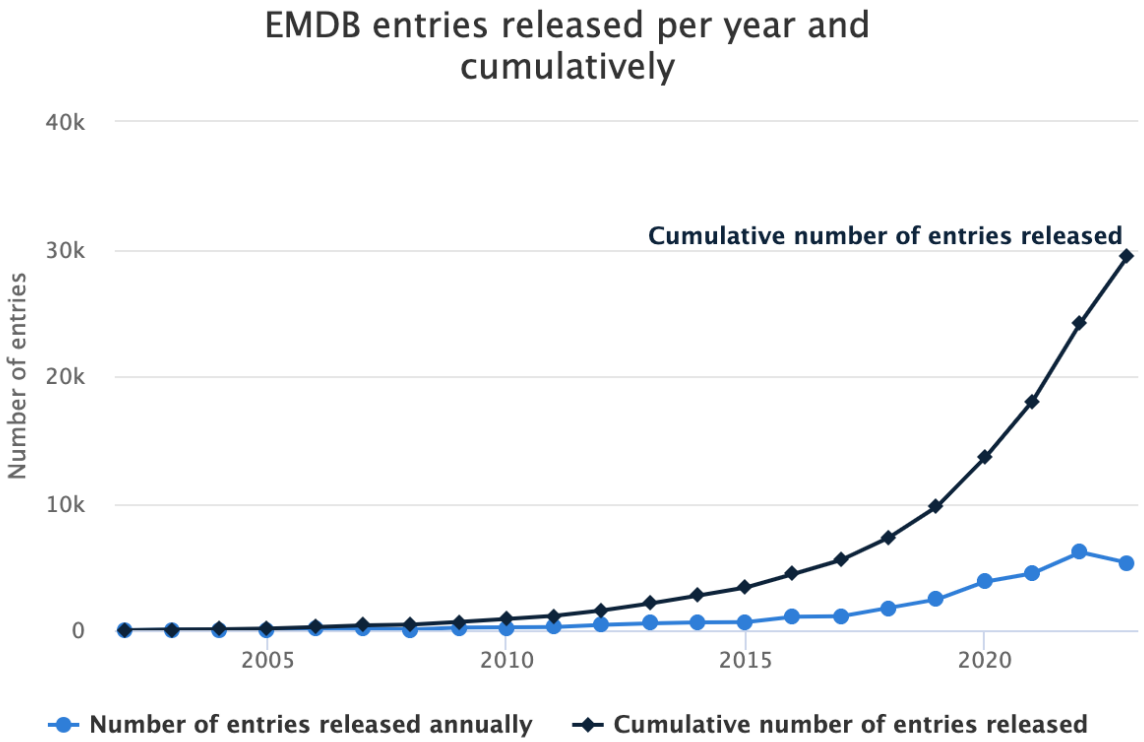
The Electron Microscopy Data Bank - EMDB

- Established in 2002 at EMBL-EBI
- wwPDB Partner
- Electron microscopy volume maps and Tomograms
- All data is freely and openly available under a CC0 license

EMD-15182



EMD-0698



EMDB: emdb-empiar.org
EMPIAR: empiar.org

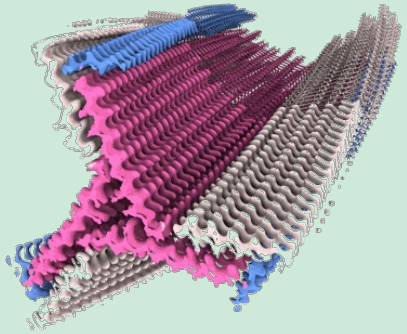
EMDB – continued growth of the archive

EMDB	2022	2021	2020
Released entries	6,131 (+37%)	4,488 (+17%)	3,831
Deposited entries	8,366 (+35%)	6,210 (+36%)	4,568
Total entries	24,187 (+34%)	18,058 (+33%)	13,583
Total size (TB)	8.9	4.5	2.9

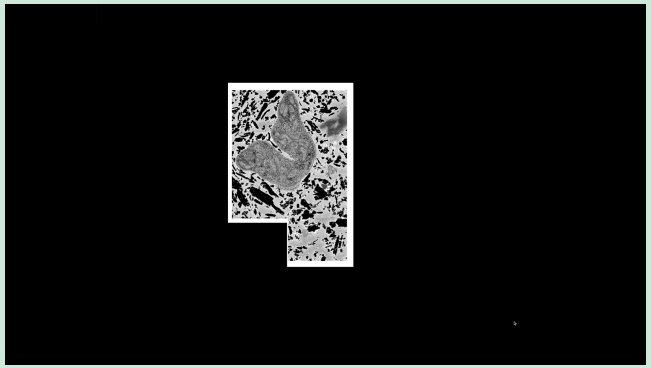
- Milestone of **30,000** release entries reached in February 2023
  - Exponential growth, doubling time **~2 – 2.5 years**
    - If this continues **50k** entries in **2025** and **100k** entries in **2028**

Electron Microscopy Public Image Archive - EMPIAR

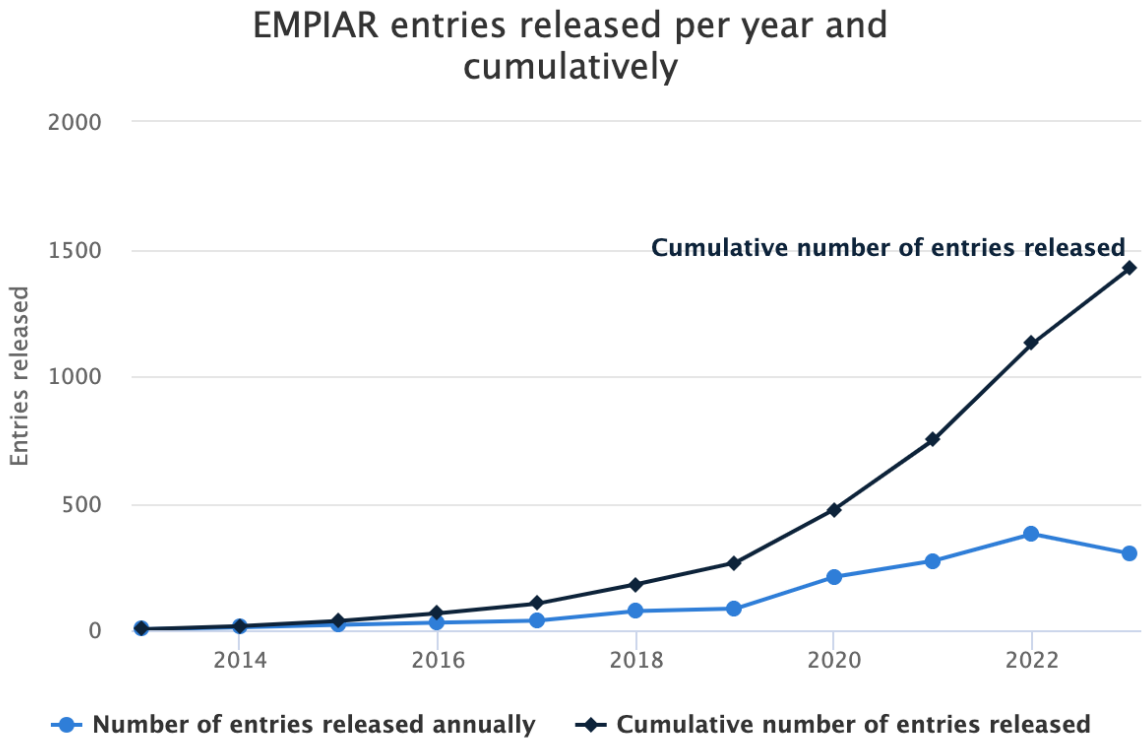
- Raw imaging data used to produce maps found in the EMDB
- Volume EM techniques and X-ray tomography and microscopy
- All data is freely and openly available under a CC0 license



EMPIAR-11278/EMD-14167



EMPIAR-10310



EMDB Electron Microscopy Data Bank

EMPIAR Electron Microscopy Public Image Archive

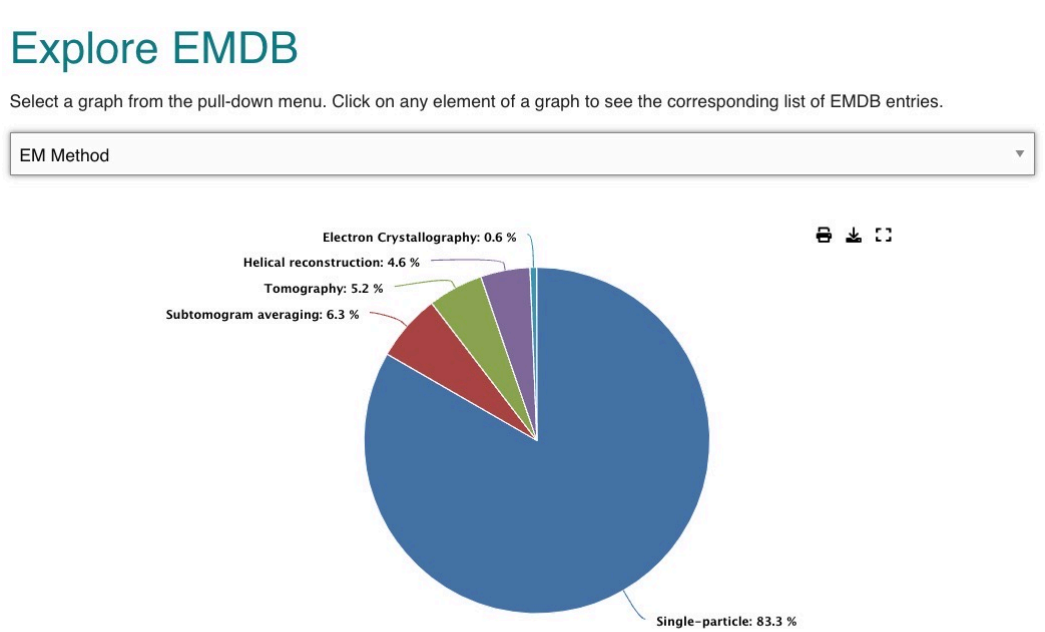
WORLDWIDE PDB PROTEIN DATA BANK

EMDB: emdb-empiar.org

EMPIAR: empiar.org

EMBL-EBI

The Electron Microscopy Data Bank - Website



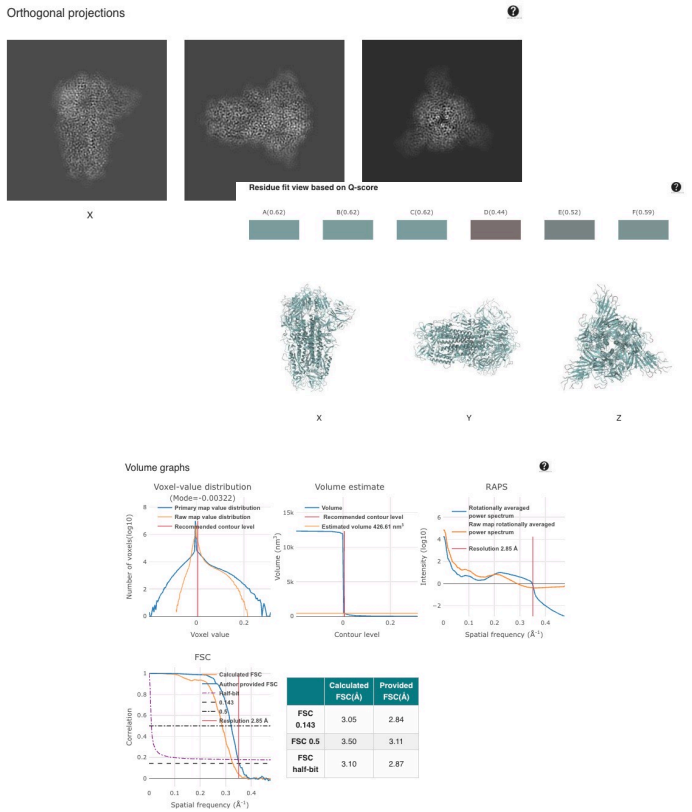
https://www.ebi.ac.uk/emdb/

EMBL-EBI

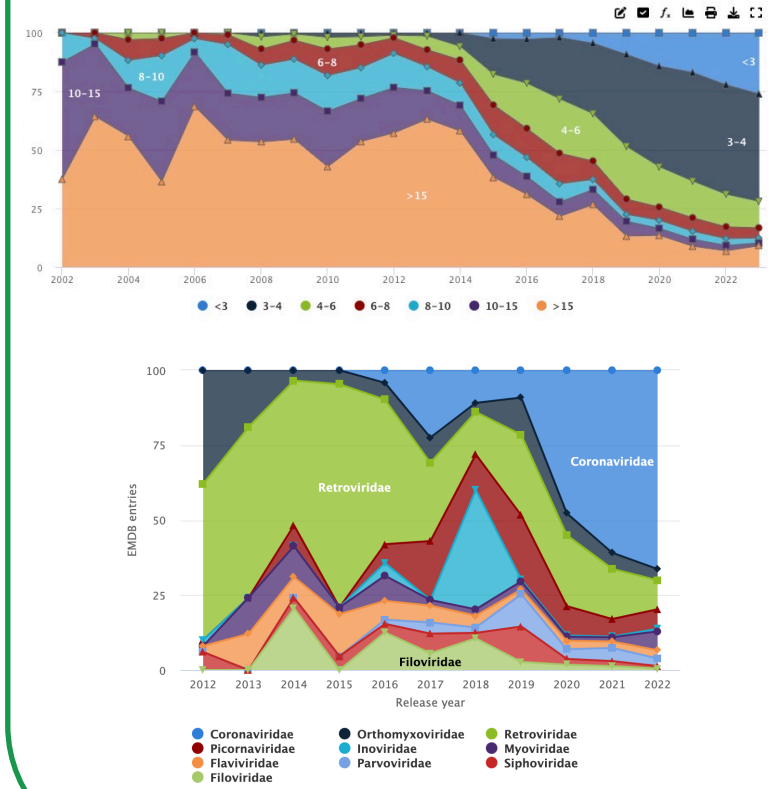
The Electron Microscopy Data Bank - Website

Powerful Search

Validation



Statistics



https://www.ebi.ac.uk/emdb/

EMBL-EBI

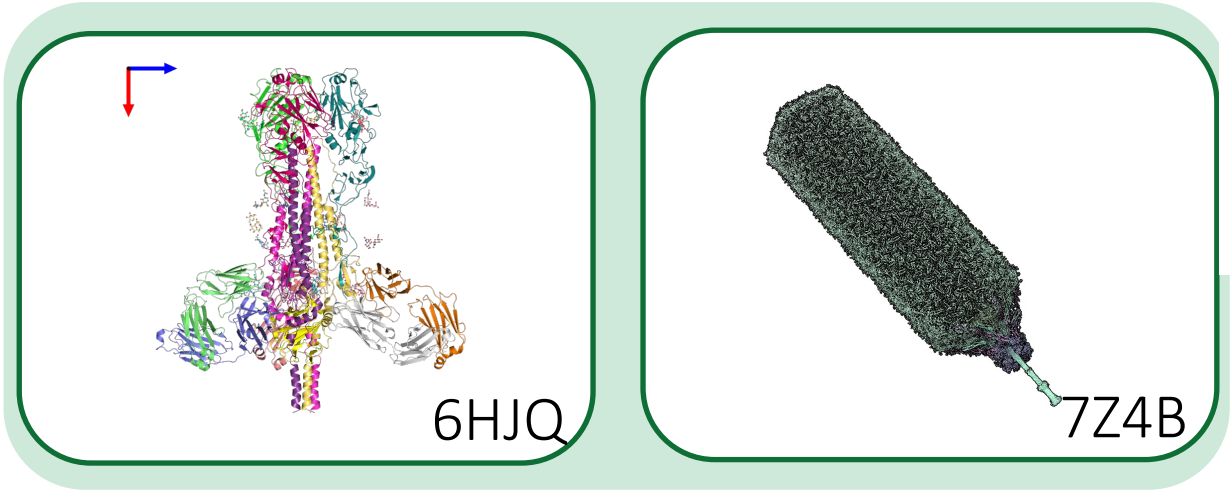
Electron Microscopy Public Image Archive - EMPIAR

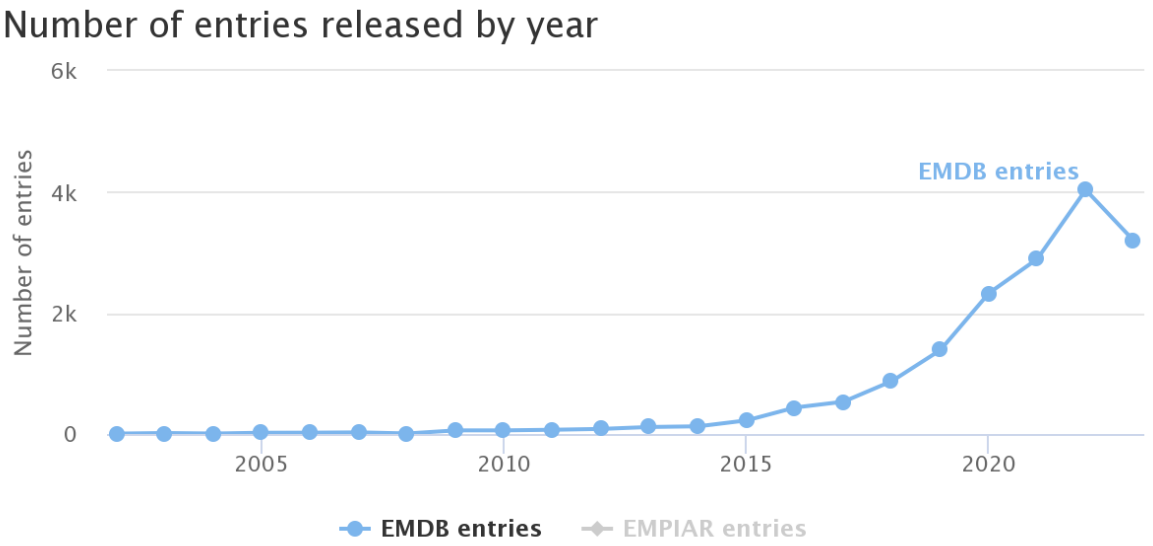
EMPIAR	2022	2021	2020
Released entries	376 (+39%)	273 (+30%)	210
Deposited entries	436 (+51%)	288 (+3%)	280
Released size (TB	903 (+15%)	788 (+75%)	450
Total entries	1129 (+51%)	750 (+57%)	477
Total size (TB)	2,399	1,496	707

- 1 PB in June 2021
- 2PB and 1,000th entry in 2022
- **3 PB 2023**

Protein Data Bank - PDB

- Archiving Coordinate models from EM, X-ray and NMR
- All data is freely and openly available under a CC0 license





~16,500 entries

www.ebi.ac.uk/pdbe/ - www.rcsb.org - www.pdbj.org

EMDB, EMPIAR & PDB Summary

- Archive the full electron microscopy (EM) processing workflow
- EMPIAR – Supporting development of software and scientists
- EMDB – Refined EM structures and tomograms
- PDB - Coordinate models from X-ray, EM and NMR

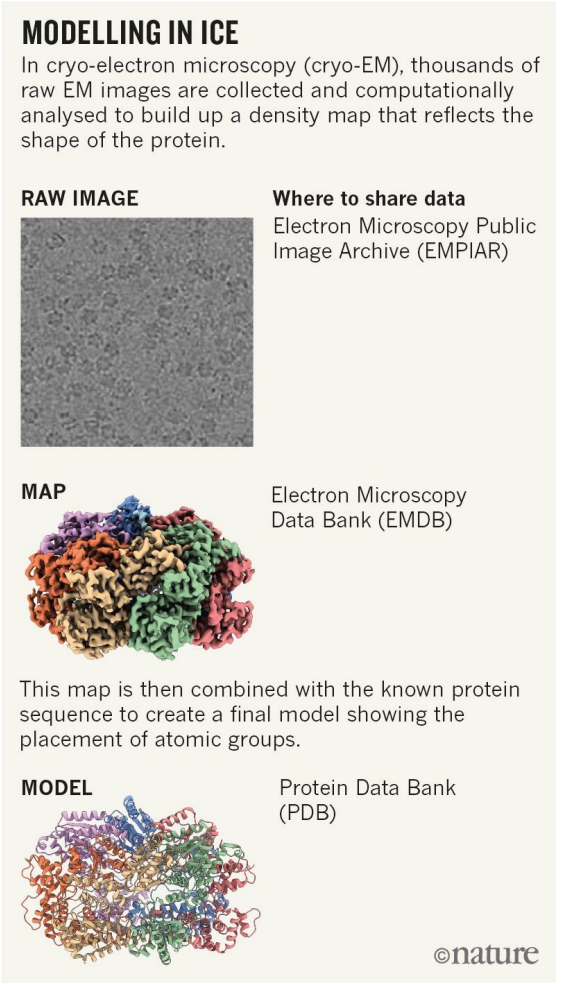


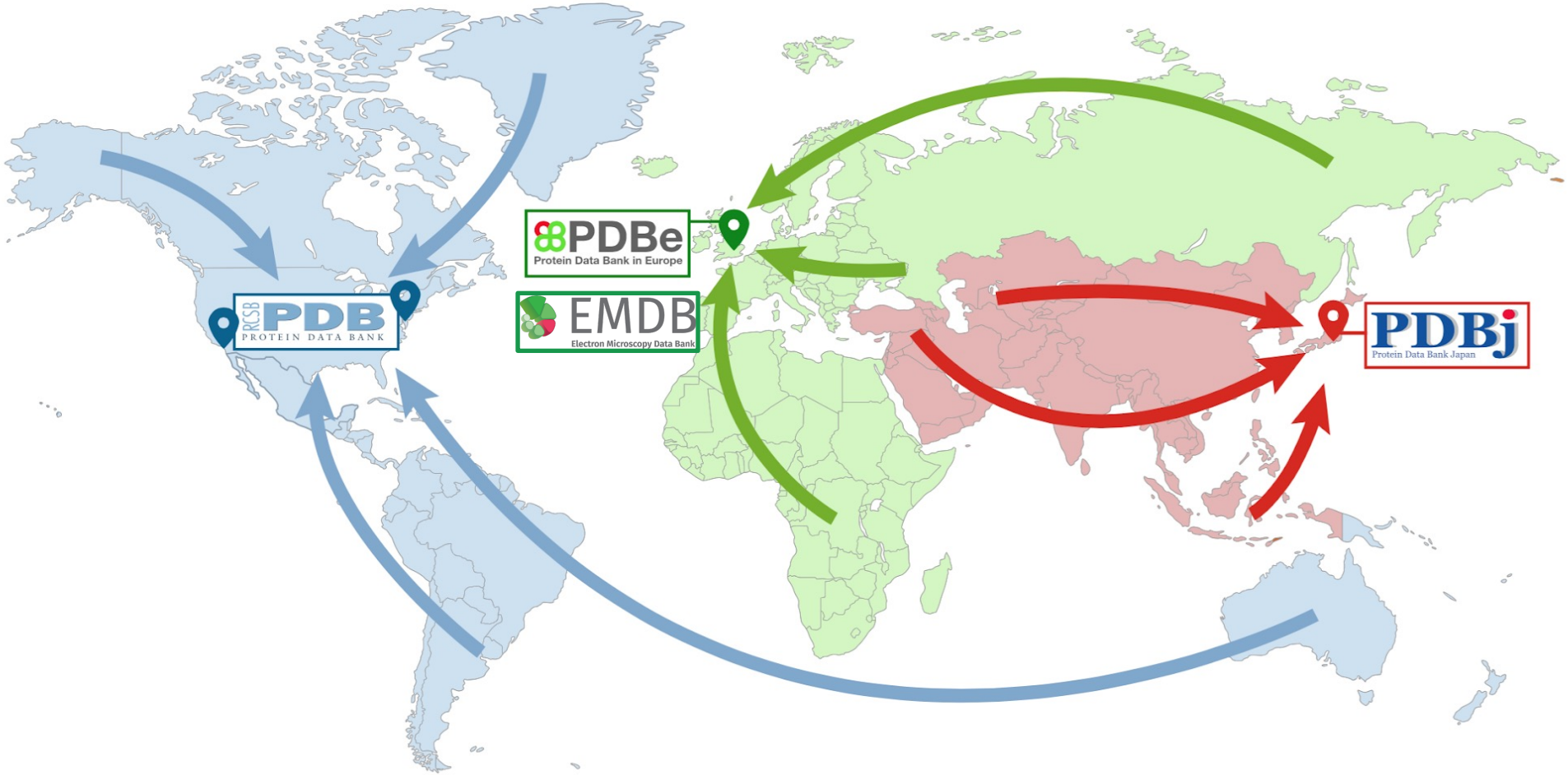
Image credit: *Nature* **561**, 565-567 (2018)

EMBL-EBI

Deposition

EMBL-EBI

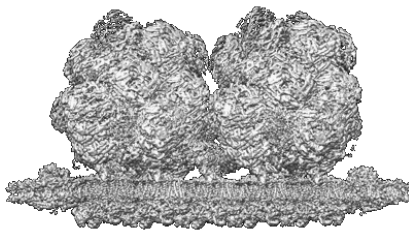
Onedep – A unified deposition system for EM, X-ray and NMR



EMDB - Methodologies

Single Particle

Double-PBS-PSII-PSI-LHCs megacomplex



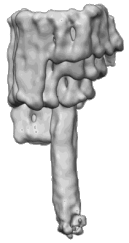
Author List: You X, Zhang X, Cheng J, Xiao YN, Sun S, Sui SF

EMD-33658

[1]

Subtomogram Averaging

DNA Origami Signpost



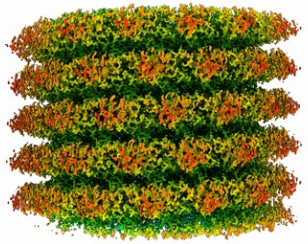
Author List: Silvester E, Vollmer B, Prazak V, Vasishtan D, Machala EA, Whittle C, Black S, Bath J, Turberfield AJ, Gruenewald K, Baker LA

EMD-12188

[2]

Helical Reconstruction

Tobacco Mosaic Virus

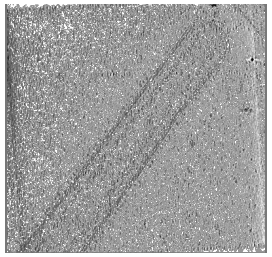


Author List: Bhella D, Love AJ, Streetley J, Taliansky M, McGeachy K, Bukharova T

EMD-16572

Tomography

Syncytial Virus Filamentous Virion



Author List: Conley MJ , Vijayakrishnan S , Bhella D

EMD-13856

[3]

Electron Crystallography

Metarhodopsin I

Author List: Ruprecht JJ, Mielke T, Vogel R, Villa C, Schertler GFX

EMD-1079

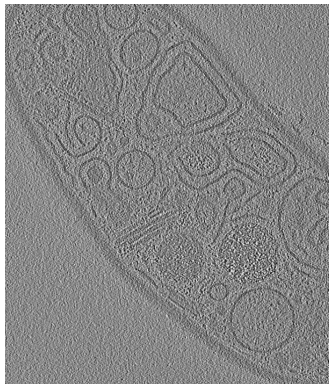
[4]

[1] You, X., Zhang, X., Cheng, J. et al. In situ structure of the red algal phycobilisome–PSII–PSI–LHC megacomplex. Nature 616, 199–206 (2023). https://doi.org/10.1038/s41586-023-05831-0
[2] Silvester, E., Vollmer, B., Pražák, V., Vasishtan, D., Machala, E. A., Whittle, C., Black, S., Bath, J., Turberfield, A. J., Grünewald, K., & Baker, L. A. (2021). DNA origami signposts for identifying proteins on cell membranes by electron cryotomography. Cell, 184(4), 1110-1121.e16. https://doi.org/https://doi.org/10.1016/j.cell.2021.01.033
[3] The EMBO Journal (2022) 41: e109728 https://doi.org/10.15252/embj.2021109728
[4] Ruprecht, J.J., Mielke, T., Vogel, R., Villa, C. and Schertler, G.F. (2004), Electron crystallography reveals the structure of metarhodopsin I. The EMBO Journal, 23: 3609-3620. https://doi.org/10.1038/sj.emboj.7600374

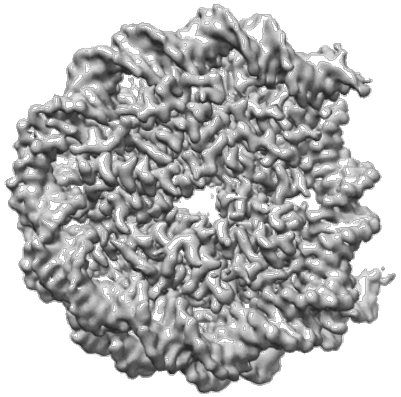
EMBL-EBI

EMDB Deposition - Types

Map Only



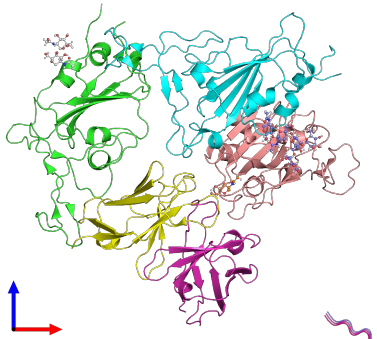
EMD-16202



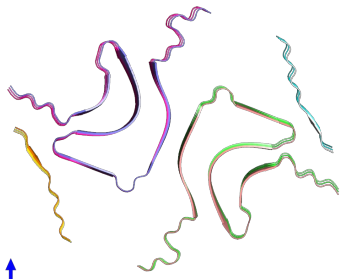
EMD-29854

All Methodologies

Model Only



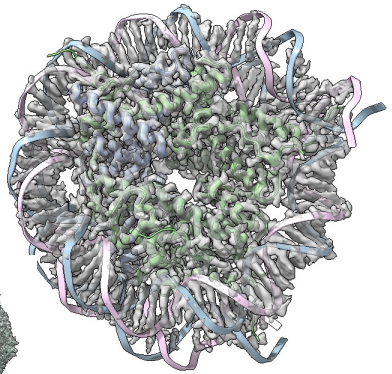
7X93



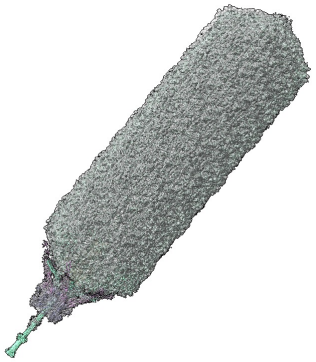
7YK8

All methodologies
except Tomography

Map + Model



EMD-29837



EMD-14492

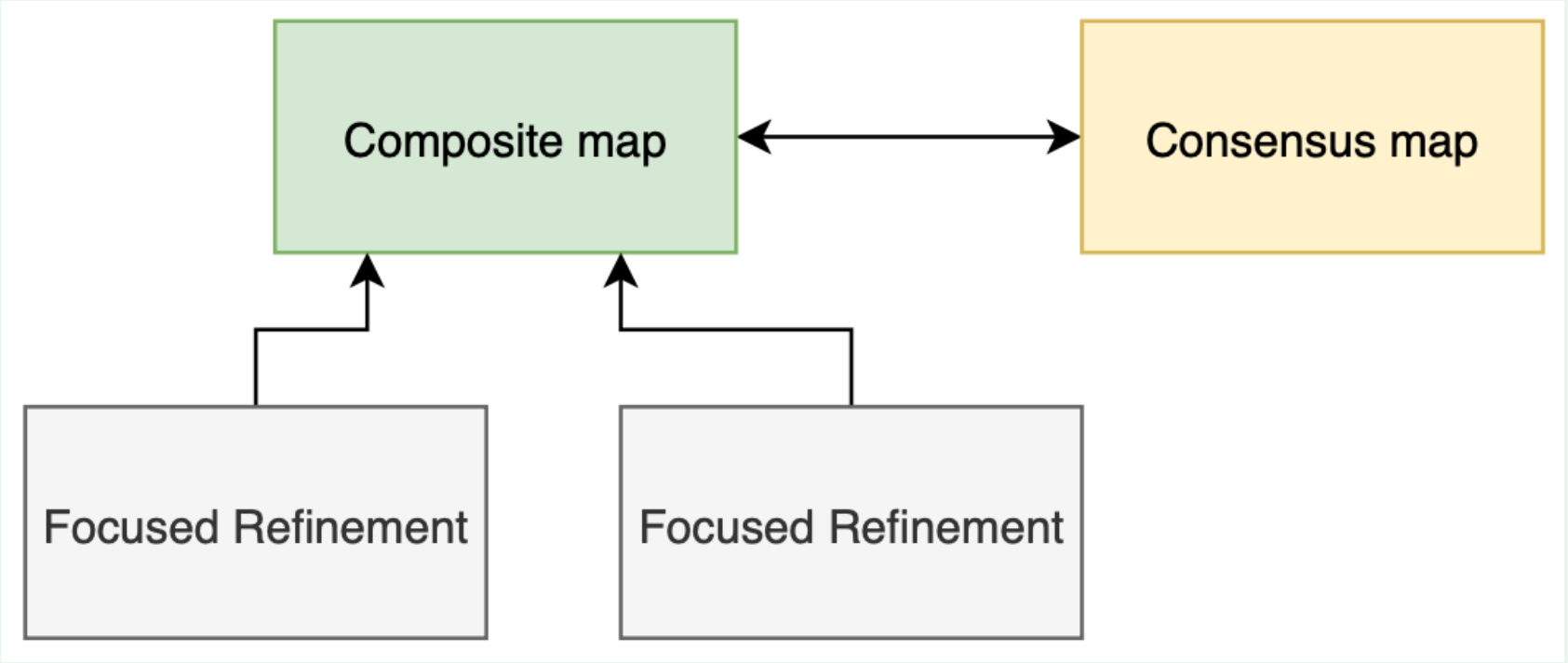
All methodologies
except Tomography

Must have associated map!

EMBL-EBI

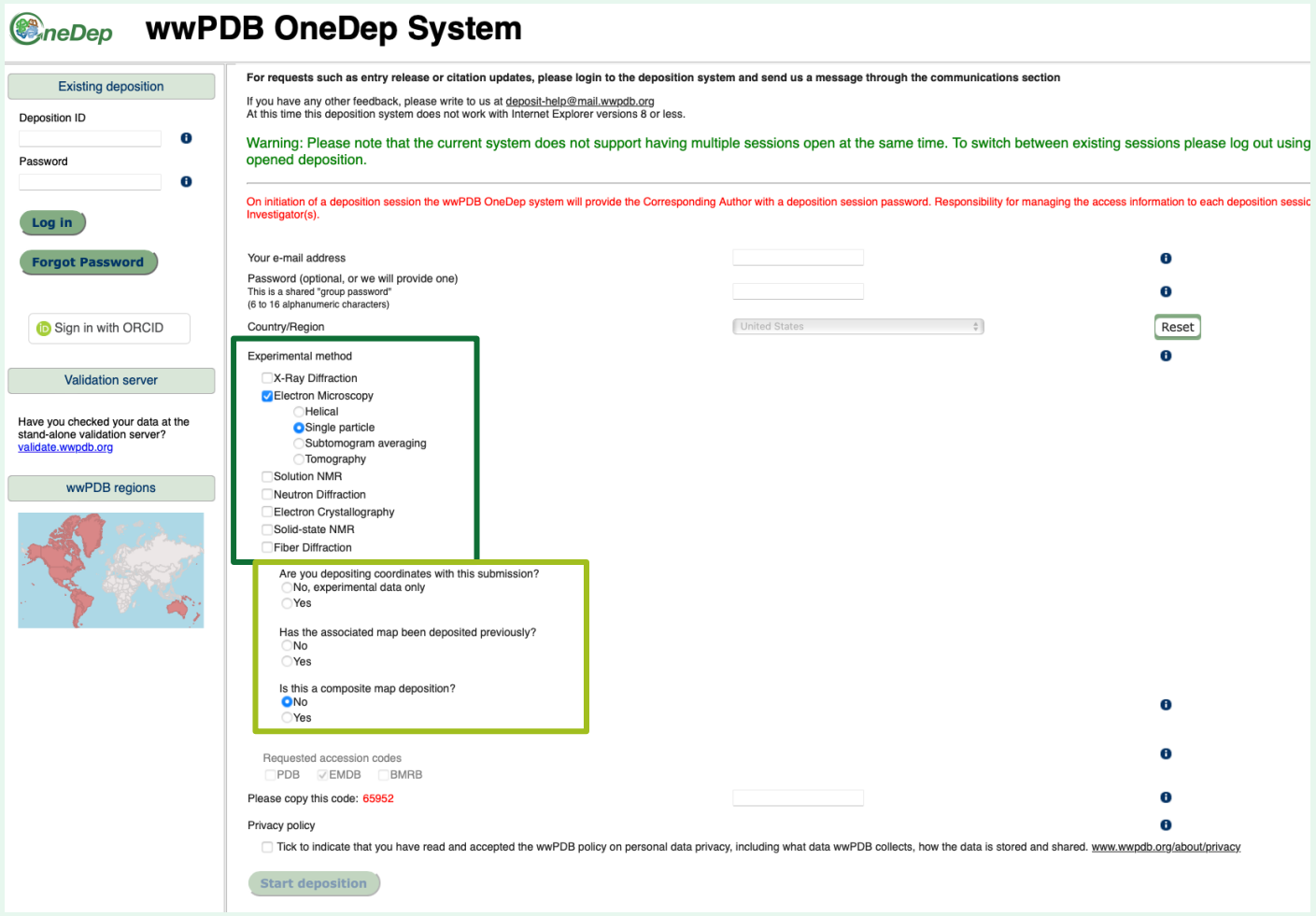
EMDB Deposition – Composite Map

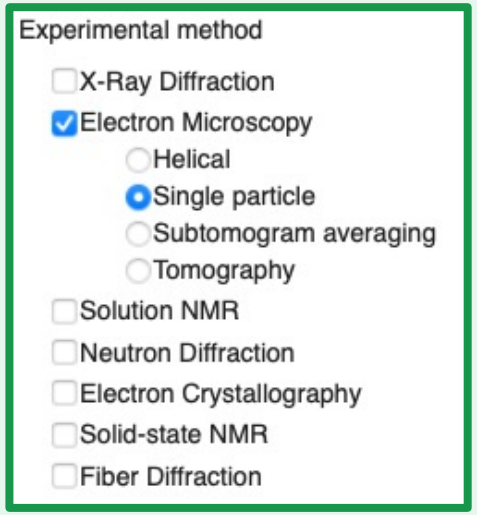
- Composite Map - Generated by combining multiple focused refinements
- Focused Refinement – A refinement focused on an area of map
- Consensus map – A refinement of the whole map without focusing on any region

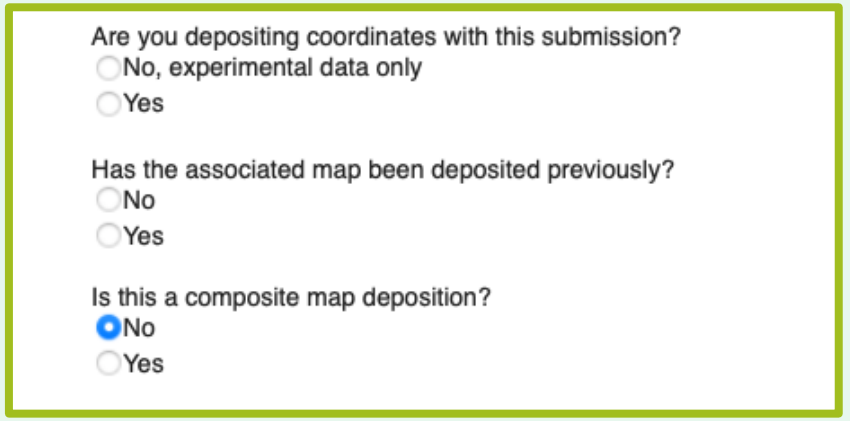


Section 1.3 of EMDB Policy

EMDB Deposition – Brain Page

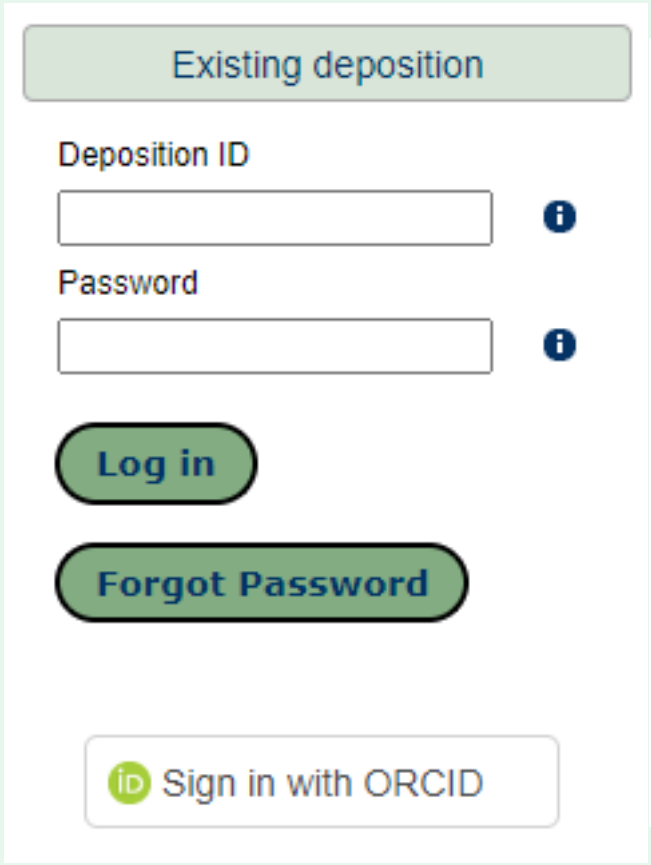






EMBL-EBI

EMDB Deposition - ORCID Login

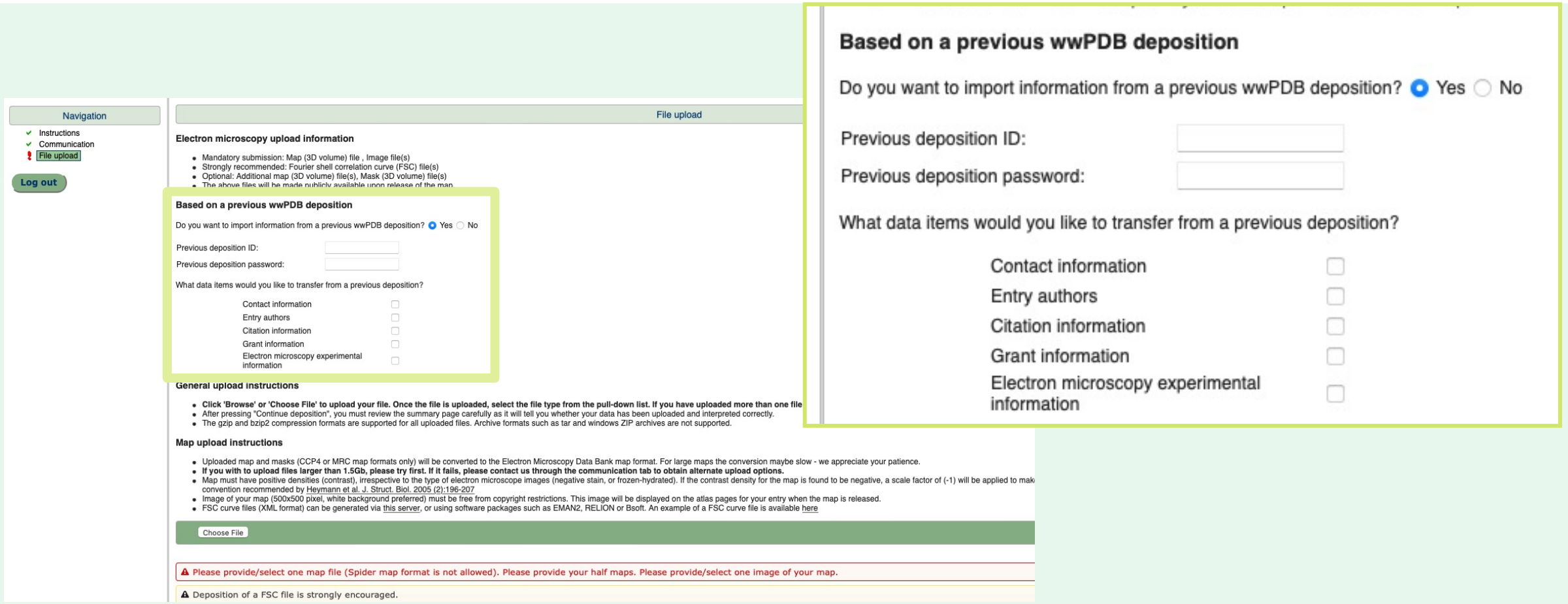


Deposition list

Depositions available to 0000-0012-3456-789X

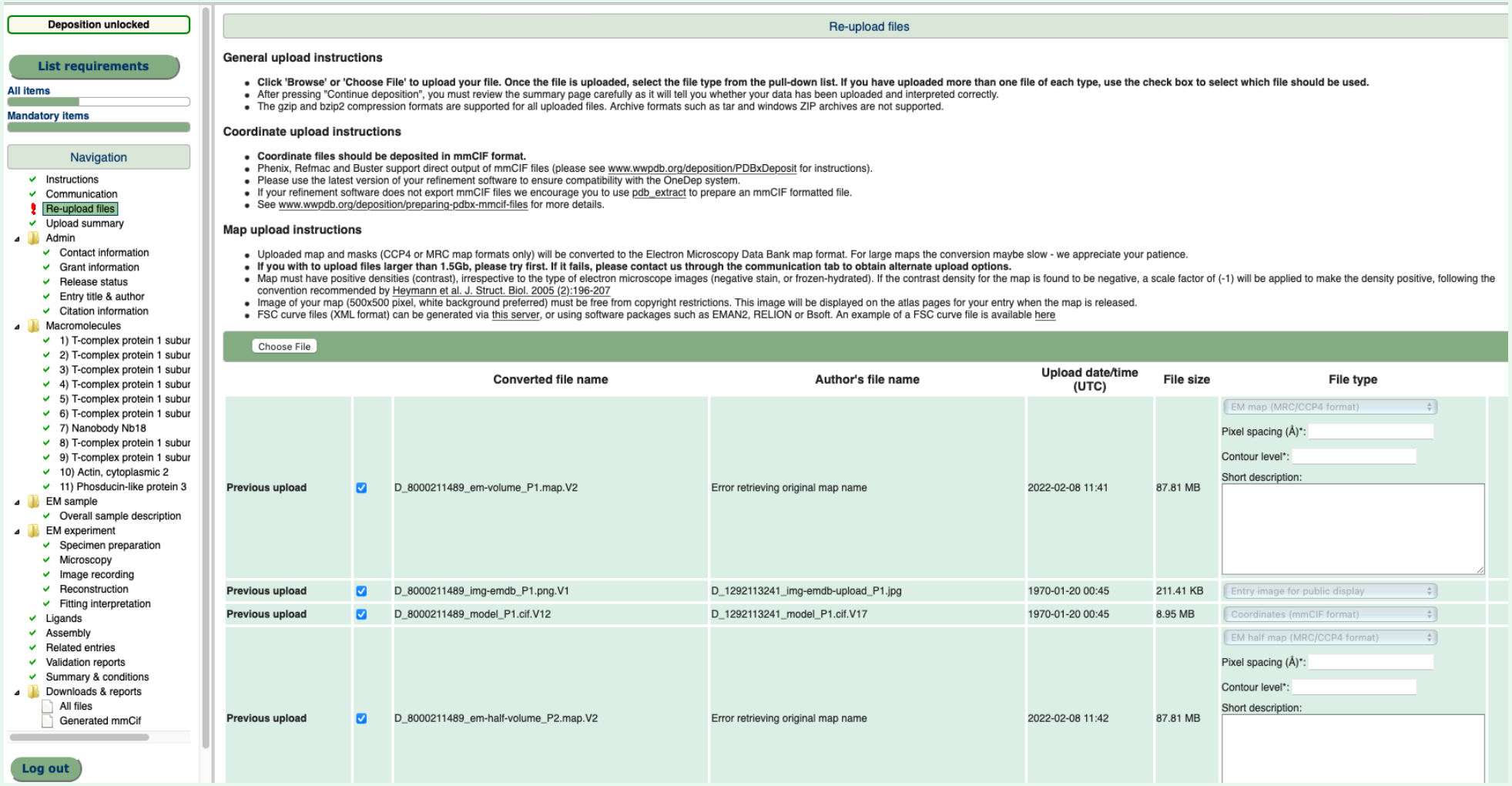
Deposition ID	Entry ID	Entry Title	Created	Site	Status	Last login
D_1292111636	7axh	Crystal structure of the hPXR-LBD in complex with alpha-zearalanol	17/03/2023	RCSB	PROC	17/03/2023
D_1292111623	7ax8	Crystal structure of the hPXR-LBD in apo form (P43212 SG)	09/01/2023	PDBe	AUTH	16/03/2023
D_1292100980	6qvt	CMP-Sialic acid bound structure the human ST6Gal1	14/07/2022	PDBj	HPUB	22/09/2022
D_1292100979	6qvs	Unliganded structure of the human wild type ST6Gal1...	04/03/2019	PDBe	REL	15/05/2020
D_1200009063	6fwu	Crystal structure of human wild type B4GalT1 in apo-closed dimeric form	07/03/2018	PDBe	REL	03/02/2019
D_1290050811	4adp	HCV-J6 NS5B POLYMERASE V405I MUTANT	02/01/2012	RCSB	REL	-
D_1290044331	2xi3	HCV-H77 NS5B Polymerase Complexed With GTP	25/06/2010	PDBj	REL	-
D_1290044262	2xhu	HCV-J4 NS5B Polymerase Orthorhombic Crystal Form	21/06/2010	RCSB	REL	-

EMDB Deposition – File Upload Page



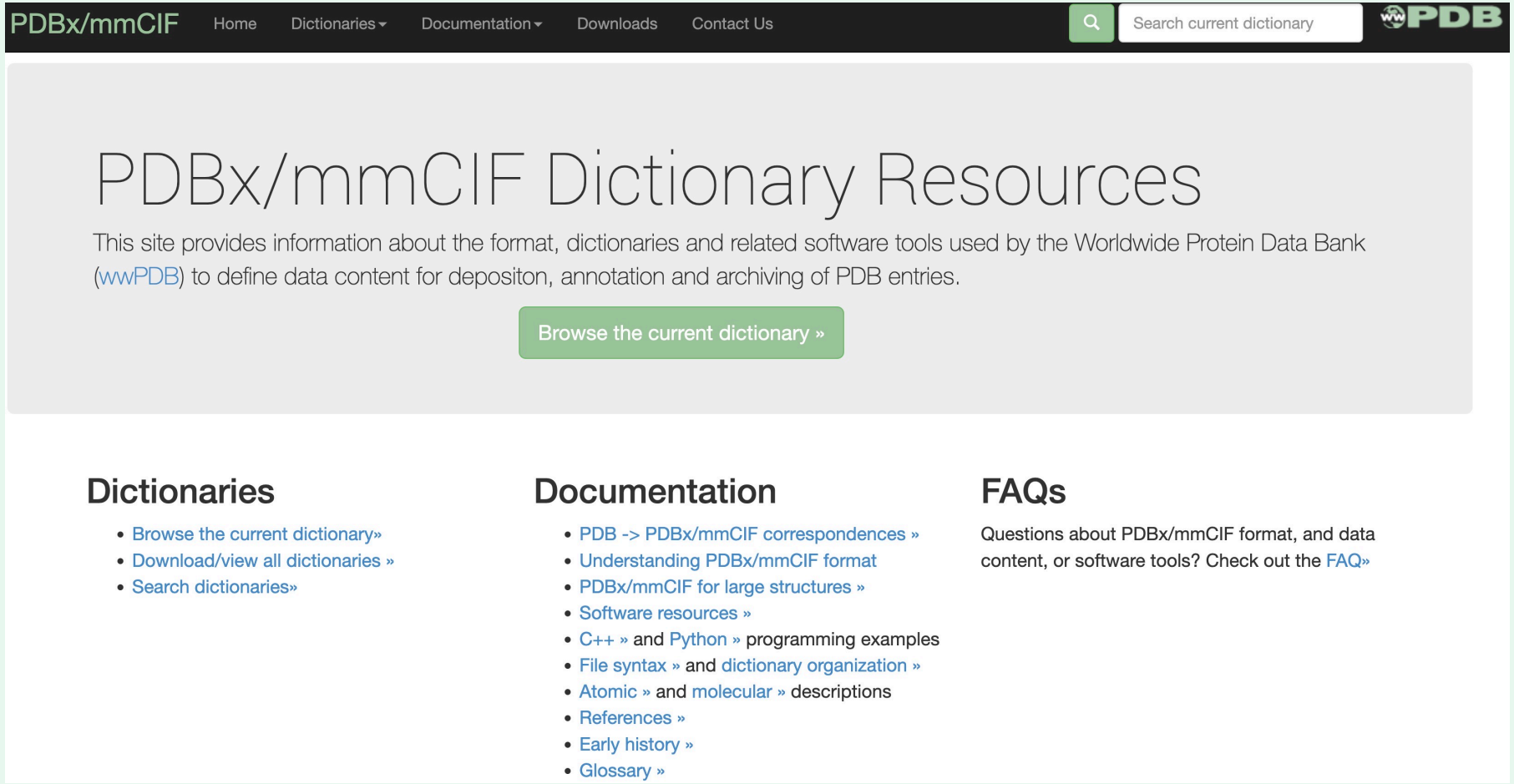
EMBL-EBI

EMDB Deposition – Deposition Interface



EMBL-EBI

EMDB Deposition - mmCIF file



https://mmcif.wwpdb.org/

EMBL-EBI

EMDB Deposition - mmCIF file

```
#
_em_imaging.entry_id                        EMD-90269
_em_imaging.id                              1
_em_imaging.astigmatism                     ?
_em_imaging.electron_beam_tilt_params       ?
_em_imaging.residual_tilt                   ?
_em_imaging.sample_support_id               1
_em_imaging.detector_id                     1
_em_imaging.scans_id                        1
_em_imaging.microscope_id                   ?
_em_imaging.microscope_model                'FEI TITAN KRIOS'
_em_imaging.specimen_holder_type            ?
_em_imaging.specimen_holder_model           'FEI TITAN KRIOS AUTOGRID HOLDER'
_em_imaging.details                         ?
_em_imaging.date                            ?
_em_imaging.accelerating_voltage            300
_em_imaging.illumination_mode               'FLOOD BEAM'
_em_imaging.mode                            'BRIGHT FIELD'
_em_imaging.nominal_cs                      2.7
_em_imaging.nominal_defocus_min             2000
_em_imaging.nominal_defocus_max             4000
_em_imaging.calibrated_defocus_min          ?
_em_imaging.calibrated_defocus_max          ?
_em_imaging.tilt_angle_min                  ?
_em_imaging.tilt_angle_max                  ?
_em_imaging.nominal_magnification           ?
_em_imaging.calibrated_magnification        ?
_em_imaging.electron_source                 'FIELD EMISSION GUN'
_em_imaging.citation_id                     ?
_em_imaging.temperature                     ?
_em_imaging.detector_distance               ?
_em_imaging.recording_temperature_minimum   ?
_em_imaging.recording_temperature_maximum   ?
_em_imaging.alignment_procedure             'COMA FREE'
_em_imaging.c2_aperture_diameter            50
_em_imaging.specimen_id                     1
_em_imaging.cryogen                         NITROGEN
#
```

```
_em_vitrification.entry_id                  EMD-90269
_em_vitrification.id                        1
_em_vitrification.sample_preparation_id     ?
_em_vitrification.specimen_id               1
_em_vitrification.cryogen_name              ETHANE
_em_vitrification.humidity                  ?
_em_vitrification.temp                      ?
_em_vitrification.chamber_temperature       ?
_em_vitrification.instrument                'FEI VITROBOT MARK IV'
_em_vitrification.method                    ?
_em_vitrification.time_resolved_state       ?
_em_vitrification.citation_id               ?
_em_vitrification.details                   ?
#
```

https://mmcif.wwpdb.org/

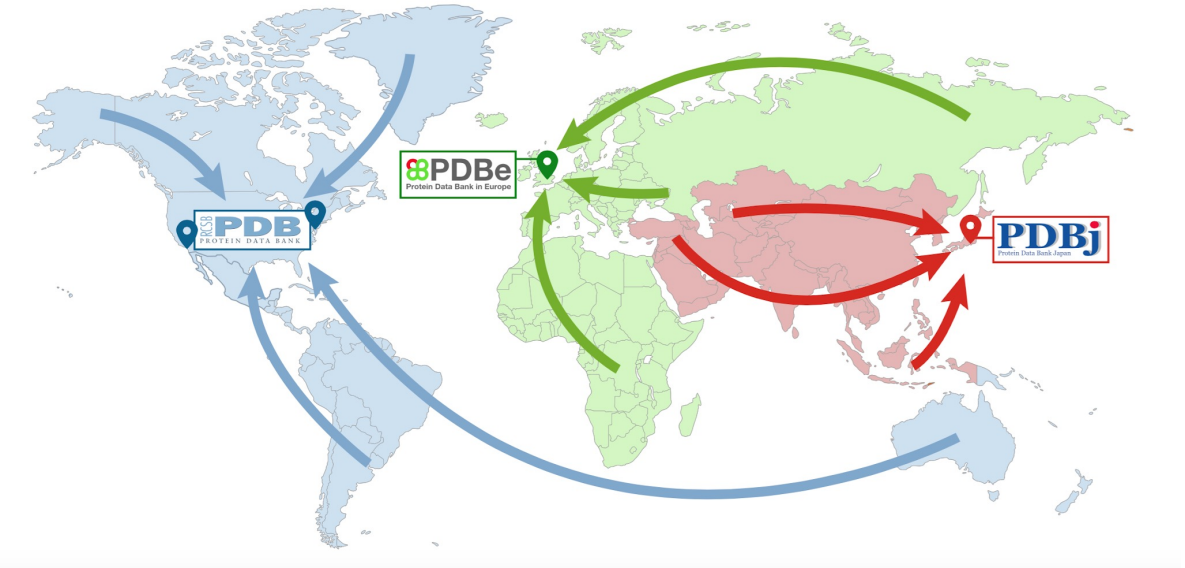
EMBL-EBI

EMDB Deposition – Common communications

- The provided sequence should be the full sequence in the sample, including any unmodeled regions.
- Half-maps must be unmasked, unfiltered and unsharpened raw maps.
- Maps and models should all be present in the same coordinate space and overlay correctly.
- Defocus values must be in nm and positive values refer to defocus.
- Please reach out if you wish to deposit variability analysis data (cryoDRGN, 3Dflex, …)

EMDB Deposition Summary

- Five methodologies
  - Single Particle
  - Subtomogram Averaging
  - Helical Reconstruction
  - Tomography
  - Electron Crystallography
- Composite map depositions must follow section 1.3 of EMDB Policy
- ORCID Login
- mmCIF files are metadata files



EMPIAR Deposition

EMBL-EBI

EMPIAR Deposition Step-By-Step

1. Register as a new user on EMPIAR's deposition system
2. Sign-in & Create a New Deposition
3. Complete the Deposition Overview Page
4. Granting rights or transferring ownership of an entry
5. Main EMPIAR Data Upload page
    a. Globus
    b. Aspera Command Line
    c. Aspera Web Interface
6. Complete the Associate image sets on the data page & Submit

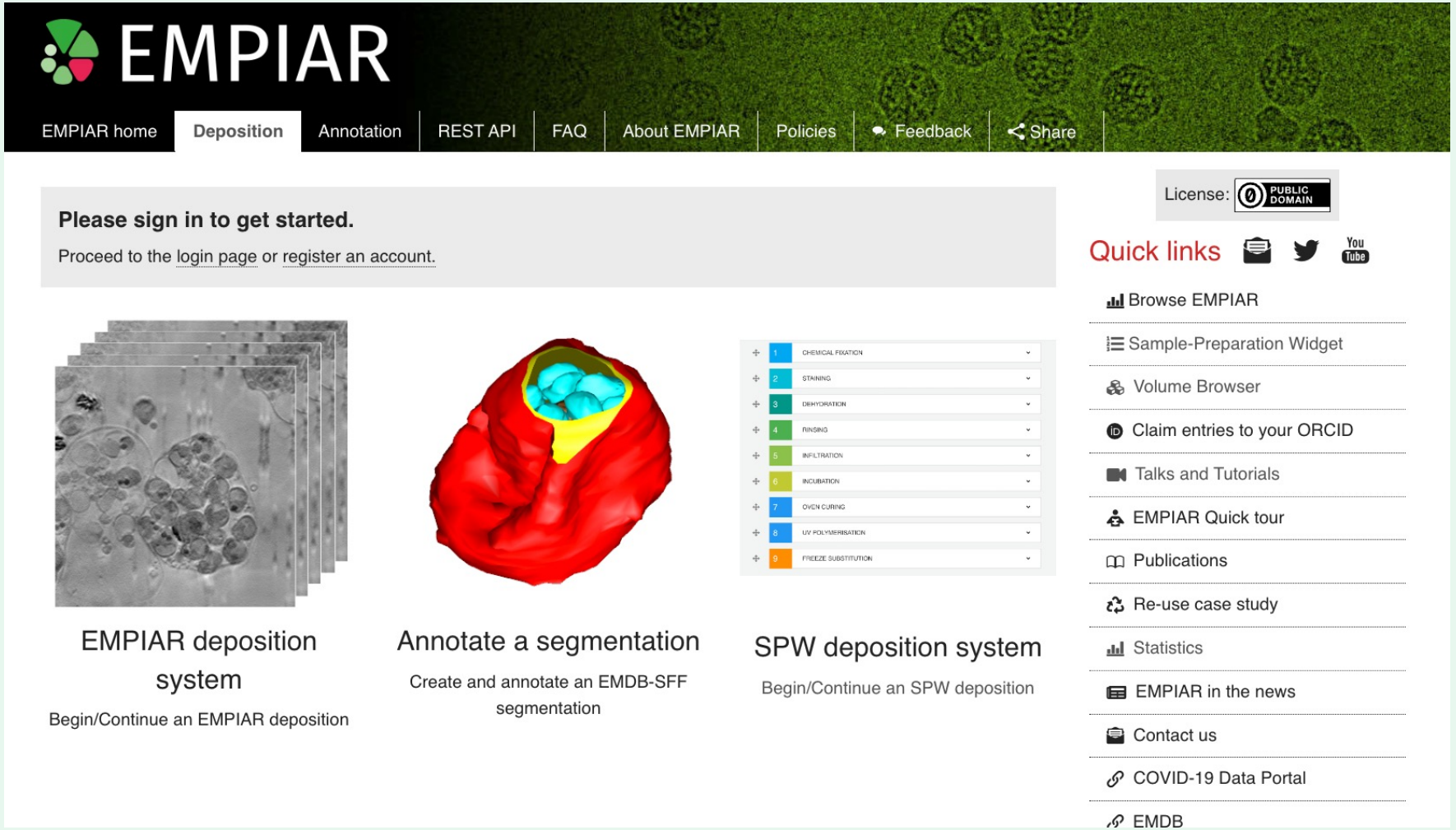
EMPIAR Deposition Policies

- EMPIAR is relatively unstructured compared to EMDB
- Polices pages should be visited to understand accepted file types
- Supported Methods (more info on the policies pages):

i. EMDB - raw image data relating to structures deposited to the EMDB
ii. SBF-SEM - image data collected using serial block-face scanning electron microscopy (like the Gatan 3View system)
iii. SXT - image data collected using soft x-ray tomography
iv. FIB SEM - image data collected using focused ion beam scanning electron microscopy
v. IHM - integrative hybrid modelling data
vi. CLEM - correlative light-electron microscopy
vii. CLXM - correlative light X-ray microscopy
viii. MicroED - microcrystal electron diffraction
ix. ATUM-SEM - Automated Tape-collecting Ultramicrotome Scanning Electron Microscopy
x. Hard X-ray/X-ray microCT - Hard X-ray/X-ray micro-computed tomography
xi. ssET - serial section electron tomography

https://www.ebi.ac.uk/empiar/policies/

EMPIAR Deposition Setup



https://www.ebi.ac.uk/empiar/deposition/choose-action/

EMBL-EBI

EMPIAR Deposition - Summary

1. Gather Information
   - EMPIAR accepts various types of data, including multi-frame micrographs, frame-averaged micrographs, particle stacks and tilt series, as well as auxiliary files (*e.g.* particle selection coordinates)
     - Make a note of dataset details (*e.g.* # of images (or tilt-series), raw or processed, single or multi-frame, image format and dimensions, pixel spacing and type)
2. Organise your Data!
   - Each data-type must be located in a separate directory (*e.g.* one for micrographs and one for particle stacks) and please sub-divide directories possessing more than 10,000 files
3. Setup Data Transfer Programs
   - Aspera (http://asperasoft.com) and Globus (https://www.globus.org) efficiently and robustly transfer large data volumes (empiar.org/faq)
4. Deposit your Data and Metadata (empiar.org/deposition)

Acknowledgements

EMDB:

Sanja Abbott

Neli Fonseca

Lucas Carrijo de Oliveira

Zhe Wang

Amudha Duraisamy

Osman Salih

EMPIAR:

Andrii Ludin

Paul Korir

Sriram Somasundharam

Simone Weyand

Osman Salih

Rosary Yao

Wider Team:

Deborah Harrus

David Armstrong

Romana Gaborova

Deepti Gupta

Genevieve Evans

Sudakshina Ganguly

Weslley Morellato Bueno

Ahsan Tanweer

& the wider PDBe team and wwPDB team

Gerard Kleywegt

Ardan Patwardhan

Sameer Velankar

UKRI Biotechnology and Biological Sciences Research Council

wellcome

WORLDWIDE PDB PROTEIN DATA BANK

EMDB Electron Microscopy Data Bank

EMPIAR Electron Microscopy Public Image Archive

EMBL-EBI