

# 2024 wwPDB AC Meeting

**Friday, October 18, 2023**

**21:00 - 24:00 Japan (UTC+9)**



# Agenda

Time Zones							Session	Participants
Europe UTC+3	UK UTC+1	US NJ/CT UTC-4	US CA UTC-7	Australia Sydney UTC+10	Japan UTC+9	India UTC+5: 30		
15:00	13:00	8:00	5:00	23:00	21:00	17:30	Welcome and Introductions (10 min)	PDBj
15:10	13:10	8:10	5:10	23:10	21:10	17:40	Discuss questions/answers/any other issues (90 min)	All
16:40	14:40	9:40	6:40	00:40 15 <sup>th</sup> Oct.	22:40	19:10	Executive Session (35 min)	wwPDB-AC
17:15	15:15	10:15	7:15	01:15	23:15	19:45	Summarise wwPDB-AC feedback and further discussion (35 min)	All
17:50	15:50	10:50	7:50	01:50	23:50	20:20	Acknowledgements and photo (10 min)	All
18:00	16:00	11:00	8:00	02:00	24:00	20:30	Meeting End	



# Welcome and Introductions

*Genji Kurisu, PDBj*



**Discuss questions/answers/  
any other issues**

# Executive Session

Host will add Advisors into Zoom Breakout Room

Please re-join us at the end of your discussion by selecting  
*Leave the breakroom* (lower right hand corner)

# Request for Advisor Feedback

1. Funding Update/Issues
  - All Full Member Reports plus PDBc
  - Additional Funding Logistics/Timeline
  - EMDB Funding Renewal/Potential Shortfall
2. Future of OneDep
  - Growing complexity of challenges (cryoET, Time-resolved, variability analyses)
  - Collaboration(s) with Software Developers
  - Automation of Deposition and Annotation
  - Maintenance/Scaling of Software
  - Centralization of Hardware Instances

# Request for Advisor Feedback (cont.)

## 3. PDBc

- Report on progress towards 100% (“quality”, “trust”)
- Commitment/timeline re processing of incoming depositions
- Commitment/timeline re protections for data security/privacy
- Commitment/timeline re contributions to OneDep maintenance, *etc.*

## 4. Next Advisory Committee Meeting

- Hosted by BMRB
- October 17 2025 or October 24 2025

# 1. Funding update - wwPDB Core Members

- RCSB PDB: NSF, NIH, DOE (2024-2028)
- BMRB: NIH (2023-2028)
- PDBe: EMBL-EBI, Wellcome Trust (2021- May 2026)
- PDBj: JST-NBDC, AMED-BINDS (2022-March 2027)
- EMDB: EMBL-EBI, Wellcome Trust (December 2024)
  - Ongoing discussion with Wellcome trust regarding continuation
- RCSB PDB/PDBe: Joint NSF/BBCRC Project No. 2 - Mol\* Visualization
  - RCSB PDB until end of November 2024
  - PDBe until the end of December 2025



# 1. Funding update - wwPDB Associate Member

Basic support of PDB China (basic PDBc data-in and data-out operations) can be provided by guaranteed funds for NFPS. “Data-base and Computation” for protein science, with approved support for 9 FTEs (personnel), is one of these nine technical systems in NFPS.

For the current funding period (years 2023-2025), this basic yearly grant is CNY 55,670,000 (~USD \$7,852 k/year).

On top of this basic grant, we also receive 0-15% merit-based bonus operation fund every year, and have supplementary salary support for personnel from the Chinese Academy of Sciences (CAS).

# 1. Funding issues

- Current funding is constrained to support of Operations at all wwPDB Core Member sites
- Very few funding calls to support development of new services or software tools (Joint NSF/BBSRC mechanism is a notable exception)
- Time between application to receipt of funds/starting work can be one or more years (difficult to keep up with pace of change)
- There are essentially no funding calls to address technical debt (*e.g.*, refactoring of OneDep) and necessary infrastructure updates (*e.g.*, operating system changes, new hardware needs)

# Discussion/Feedback

## 2. Future of OneDep Development

- Complexity of depositions to OneDep continues to increase
  - Number of new structures increasing ('24 proj. 19118 vs. 17064 in '23)
  - Average structure is getting bigger and more challenging to process (<# residues/structure>: 2169 in '23 vs. 1860 in '22)
  - Average structure contains more polymer and ligand components
  - Experimental methods becoming more complex (3DEM, XFEL, *etc.*)
  - Scope of Investigations becoming broader (dynamics, screening, *etc.*)
  - Structure validation becoming more challenging
- Increased collaborations with external resources required
  - CCDC collaboration continues to be highly productive
  - Additional funds required to support collaborations with key software providers (NSF/BBSRC application with CCP4, CCP-EM, Phenix, GPL)
  - Need to expand collaborations with other 3DEM software providers
  - Need for enhanced project management and change control processes to manage existing/expanded/new collaborations

## 2. Future of OneDep Development (cont.)

- Streamline data deposition to OneDep
  - Reduce data entry burden by coordinating software development that supports automated harvesting and uploads from data collection facilities and structure-determination software providers as nearly complete PDBx/mmCIF format files
  - Enable parallel deposition of related structures, experimental data, metadata as Investigations (building on GroupDep experience)
- Streamlining biocuration
  - Incorporate AI/ML-based methods into OneDep
    - Validation: Better detection of outliers, anomalies, inconsistencies, *etc.*
    - Biocuration: Increase efficiency of wwPDB annotators
      - Example: Using natural language processing of deposited metadata to automatically connect PDB structures to publications appearing after release (PDB archive would be the training set)

## 2. Future of OneDep Development (cont.)

- Software infrastructure challenges
  - Technical debt needs to be addressed, while continuing to fix bugs and make user required/requested improvements
  - Need to make our software more modular and more maintainable
  - Move to Cloud Computing: opportunities vs. challenges
  - Operating system evolution/migration is repeatedly diverting efforts away from maintenance and refactoring
- Hardware infrastructure challenges
  - Currently operating and maintaining four instances of OneDep (PDBj, PDBc@PDBj, RCSB PDB, PDBe)
  - Long-term Goal: Single scalable OneDep instance that supports all of wwPDB (divide-and-conquer software management, centralized system management, warm failover, flexible machine provisioning, long lease)

# Discussion/Feedback

# 3. PDBc Status Report

- Report on progress towards 100%
  - Number of entries from Mainland China annotated in Shanghai

	Total	XRAY	EM (m+m)	EM (model)	EM (map)	NMR (l+s)	PDBc	Percentage
2023-Jun	393	118	163	7	99	6	168	42.7
2023-Jul	325	127	155	0	37	6	142	43.7
2023-Aug	414	146	177	2	85	4	147	35.5
2023-Sep	395	149	221	1	19	5	137	34.7
2023-Oct	402	140	186	2	64	10	81	20.1
2023-Nov	327	105	153	2	60	7	56	17.1
2023-Dec	383	122	221	2	31	7(6+1)	59	15.4
2024-Jan	536	188	250	7	86	5	213	39.7
2024-Feb	407(-1)	109	228	0	69	1	106	26.0
2024-Mar	364	126	164	7	66	1	160	44.0
2024-Apr	443	175	214	11	40	3	291	65.7
2024-May	440	118	205	1	116	0	272	61.8
2024-Jun	355	139	177	0	32	7	271	76.3
2024-Jul	409	149	177	0	82	1	285	69.9



# 3. Transition to Core Membership

- Commitment/timeline re processing of incoming depositions
  - 100% coverage of depositions from Mainland China (currently ~70%)
  - Development of new function that enable depositors in Hong Kong/Taiwan to choose PDBc or PDBj
  - Installation of OneDep deposition server for PDBc@Osaka
- Start of non-annotation activities
  - Testing new releases of updated OneDep software
  - Assistance with problems during the deposition process
  - Full maintenance of PDBc servers at Osaka
  - Supporting OneDep software development

# 3. Transition to Core Membership (cont.)

- Commitment/timeline re protections for data security/privacy
  - Going forward wwPDB sites will document completion of research ethics training by each team member on an annual basis.
- Commitment/timeline re contributions to OneDep maintenance
  - Trusted weekly data transfer to the wwPDB Core Archive Keepers
  - Globus based data transfer to/from P.R.C.

# 3. Transition to Core Membership (cont.)

The wwPDB Charter states

“A wwPDB Associate Member may be invited to apply to become a wwPDB Core Member, following successful completion of a term of no less than five years as a wwPDB Associate Member, with evidence of sufficient technical expertise, adequate infrastructure, and sustainable funding”

- After becoming a Core Member, PDBc will
  - Install a OneDep production server at Shanghai
  - PDBj will transfer all data remotely processed by PDBc to Shanghai

# Discussion/Feedback

# 4. Draft wwPDB Policy re Obsoleting Entries

DRAFT 09/21/2024

## Common Policy for Obsoleting wwPDB Core Archive Entries

### Entry Removal in Unusual Circumstances

Circumstances may arise in which the integrity, correctness, ownership, or provenance of data deposited as part of a wwPDB Core Archive Entry (*i.e.*, experimental data and related metadata and/or atomic coordinates for a three-dimensional structure of a biomolecule with an issued BMRB, EMDB, or PDB ID) are called into question.

In such circumstances, the wwPDB may need to remove (hereafter obsolete) the wwPDB Core Archive Entry (hereafter Entry) to help ensure the accuracy of the scientific record.

Examples of cases in which this might occur include (but are not limited) to the following:

- The Principal Investigator responsible for the Entry requests that the Entry be obsoleted.
- Retraction of a published paper describing the collection and analyses of data contributed as part of the Entry by some or all of the publication authors or by the Editor of the Journal in which it appeared.
- A qualified third-party (*e.g.*, Research Integrity Officer at the Host Institution of the Principal Investigator responsible for the Entry) is advised in writing by a formally-appointed independent investigative body (*e.g.*, External Expert Review Panel) that the Entry should be obsoleted, and informs wwPDB leadership of the circumstances by providing supporting documentation (*e.g.*, External Expert Review Panel Report).
- A governmental or supranational agency with jurisdiction in the region (*e.g.*, US Department of Health and Human Services Office of Research Integrity) is advised in writing by an official investigative body (*e.g.*, US Administrative Law Court Judgement published in the US Federal Register) that the Entry should be obsoleted.

In such circumstances, the wwPDB will first make best efforts to advise Entry authors.

Thereafter, the wwPDB may obsolete the Entry and specify the reason for doing so using the following PDBx/mmCIF data item: `_pdbx_database_PDB_obs_spr.details` (*e.g.*, journal publication retraction).

# 4. Current wwPDB Obsoleting Policies

## PDB:

There are some rare circumstances in which an obsolete structure is not superseded by a new structure. The entry must contain a statement specifying the reason for obsoleting the structure (under `_pdbx_database_PDB_obs_spr.details`).

- The publication is retracted. The associated PDB entry will be obsoleted if requested by the journal. If a request has not been received, the wwPDB will do its best to contact the depositor and co-authors, (former) PIs, journal editors, etc. when made aware of the retraction. If the reason(s) for retraction were such that the associated PDB entry needs to be made obsolete, the wwPDB will obsolete the entry. The citation in the obsoleted entry is the published journal retraction.
- There is no associated publication and the entry author obsoletes the entry (e.g., the structure is incorrect).
- A third-party (such as the employer) requests that the entry is obsoleted (e.g., in case of malfeasance). In such cases, the wwPDB will obsolete the entry if either the primary citation for the structure is retracted or a formal report by an appropriate government or supranational agency in the region (e.g., US Office of Research Integrity) is published.

# 4. Current wwPDB Obsoleting Policies (cont.)

## **EMDB:**

Circumstances may arise in which the integrity, correctness, ownership or provenance of data is called into question. In such unusual circumstances, EMDB may receive a request to make the entry obsolete.

Examples of cases in which this might occur are:

- The publication describing an EM map is retracted by (some of) the contact authors, their home institute, or the journal in which it was published
- An official investigative body (e.g. the Office of Research Integrity in the USA) recommends retraction

In cases like these, efforts will be made to communicate with entry authors regarding the matter before a decision is made by the wwPDB PIs.

## **BMRB:**

Authors are encouraged to update their BMRB files to take new results into account. Entries may be deleted (from the actively-accessed database) at the author's request.

# Discussion/Feedback



# 5. Advisory Committee Meeting 2025

*2025 wwPDB AC Meeting (format virtual) will be hosted by BMRB*

*Which date would the Advisory Committee prefer?*

*Friday October 17<sup>th</sup> 2025*

*Friday October 24<sup>th</sup> 2025*

# Acknowledgement to Jenny

The entire wwPDB team wishes to thank Jennifer L. Martin for her untiring service to wwPDB and the international structural biology community.



# **Summarise wwPDB-AC feedback and further discussion**

# Acknowledgements

*Genji Kurisu, PDBj*

# Group Photo



# Meeting Close

# Thank you



Poliovirus Neutralization (Artist: David S. Goodsell)



# Pre-Meeting Review Slides



# Table of Contents

- Funding
- Governance
- Outreach
- OneDep
- PDB Archive
- EMDB Archive
- BMRB Archive
- Joint Projects

# Funding Update

# wwPDB Core Member Funding

- RCSB PDB: Joint NSF/NIH/DOE funding (2024-2028)
- BMRB: NIH NIGMS R24 funding (2023-2028)
  - Inadequate budget: still need to find additional support
    - NIH R35 pending, requesting \$500K for BMRB & NMRbox R & D
  - UConn funding 25% administrative assistant and 25% project manager positions
- PDBj: JST-NBDC and AMED-BINDS funding (2022-2027)
  - Additional budget from S. Korea (under consideration)
- PDBe: EMBL-EBI, Wellcome Trust (2021-2025)
- EMDB: EMBL-EBI, Wellcome Trust (2019-2024)
  - Continuation of Wellcome funding to EBI resources under discussion
- RCSB PDB/PDBe: Joint NSF/BBSRC Mol\* Visualization funding (2021-2025)
  
- Pending: RCSB PDB/PDBe/EMDB: Joint NSF/BBSRC application (2025-2028)

# wwPDB Associate Member Funding

Basic support of PDB China (basic PDBc data-in and data-out operations) can be provided by guaranteed funds for NFPS. “Data-base and Computation” for protein science, with approved support for 9 FTEs (personnel), is one of these nine technical systems in NFPS.

For the current funding period (years 2023-2025), this basic yearly grant is CNY 55,670,000 (~USD \$7,852 k/year).

On top of this basic grant, we also receive 0-15% merit-based bonus operation fund every year, and have supplementary salary support for personnel from the Chinese Academy of Sciences (CAS).

# Governance Update

# wwPDB Charter Updated

- Ready to admit PDBi as an Associate Member when they make their application

## CHARTER OF THE WORLDWIDE PROTEIN DATA BANK

Agreement between

RESEARCH COLLABORATORY FOR STRUCTURAL BIOINFORMATICS PROTEIN DATA BANK (RCSB PDB Organization), Rutgers, The State University of New Jersey, Piscataway, New Jersey, United States

and

The EUROPEAN MOLECULAR BIOLOGY LABORATORY ("EMBL"), an intergovernmental institution established by treaty, headquartered at Meyerhofstrasse 1, 69117 Heidelberg, Germany acting through its UK Outstation the European Bioinformatics Institute ("EMBL-EBI"), located on the Wellcome Genome Campus in Hinxton, Cambridgeshire, UK, operating both the PROTEIN DATA BANK IN EUROPE (PDBe Organization) and the ELECTRON MICROSCOPY DATA BANK (EMDB Organization)

and

PROTEIN DATA BANK JAPAN (PDBj Organization), Osaka University, Osaka, Japan

and

BIOLOGICAL MAGNETIC RESONANCE DATA BANK (BMRB Organization), University of Connecticut, Farmington, Connecticut, United States

Effective from January 1, 2021

## CHARTER OF THE WORLDWIDE PROTEIN DATA BANK (wwPDB)

### 1. Rationale

The Worldwide Protein Data Bank (wwPDB) was created to provide an enduring organizational framework for global management and dissemination of public-domain structural biology data.

It is essential for the progress of international science that structural biology data be maintained within a limited number of unfragmented archives, identified herein as *wwPDB Core Archives*.

At present, the wwPDB manages three *wwPDB Core Archives*, including the *Protein Data Bank (PDB)*, the *Biological Magnetic Resonance Data Bank (BMRB)*, and the *Electron Microscopy Data Bank (EMDB)*.

It is paramount that facilities for deposition, validation, biocuration, remediation, and storage of public-domain structural biology data in the *wwPDB Core Archives* be managed jointly by the **wwPDB Core Members** with all services provided at no charge to **wwPDB Data Depositors**.

It is equally important that structural biology data stored in the *wwPDB Core Archives* be freely and publicly disseminated by the **wwPDB Core Members** and the **wwPDB Associate Members** without charge or limitations on usage.

**wwPDB Core Members** and **wwPDB Associate Members** are fully committed to the FAIR Principles of Findability-Accessibility-Interoperability-Reusability, emblematic of responsible stewardship of public domain information.

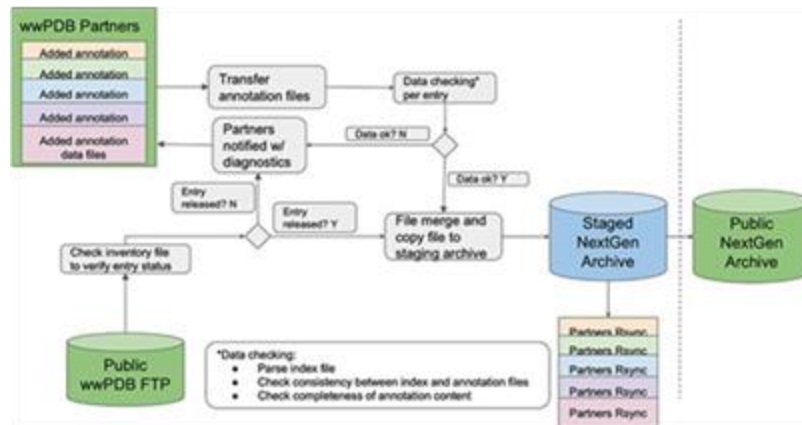
# PDB India Update (PI/Co-PI: Prof. M. Bansal, Prof. K. Sekar, Dr. D. Mohanty)

- The first phase of funding for PDBi from NSM (India) ended in March 2024. PDBi PIs are making efforts to secure funding from Department of Biotechnology (DBT), India for a period of 5 years as a part of Indian Biological Data Center (IBDC).
- As per the requirement of DBT (India), wwPDB member PIs provided a detailed letter clarifying criteria for PDBi to join wwPDB as associate member. The said letter has been submitted to DBT and IBDC in April 2024 and they have indicated that they would consider support for PDBi in the next funding cycle starting from April 2025.
- In absence of funding for PDBi from March 2024, the staff have been shifted to other projects available with the PDBi PIs and hence, Dr. Vetri has not been able to contribute to annotation of structures for PDBe.
- PDBi PIs have requested NSM (India) to extend their support till DBT/IBDC provide full support for PDBi in 2025. If funds are received from NSM (India), Dr. Vetri will be able to continue with the annotations for PDBe and a trained/skilled staff like her can be retained till the funding for next phase is received.

# Outreach Update

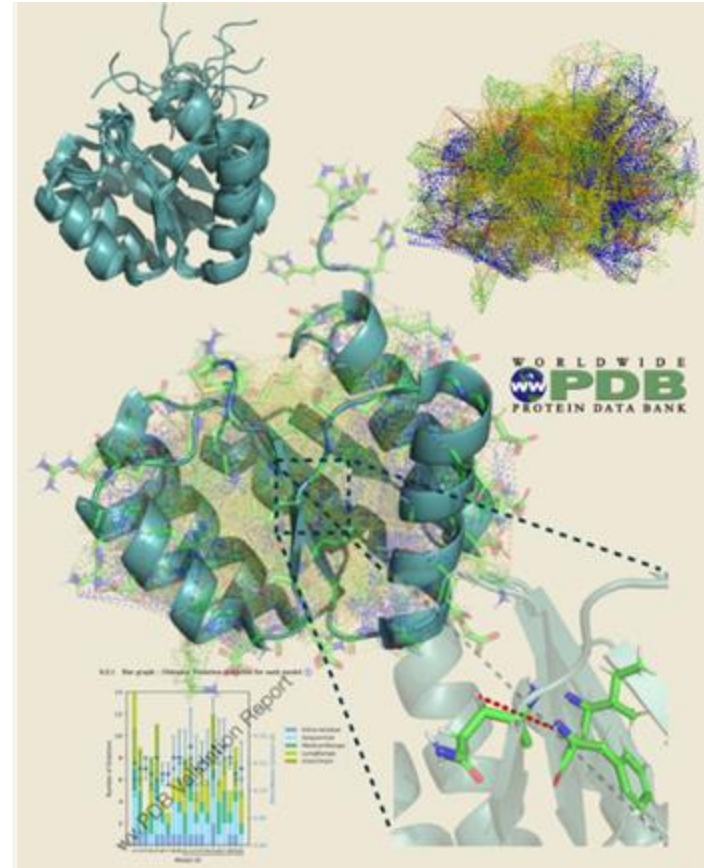


# Recent Publications



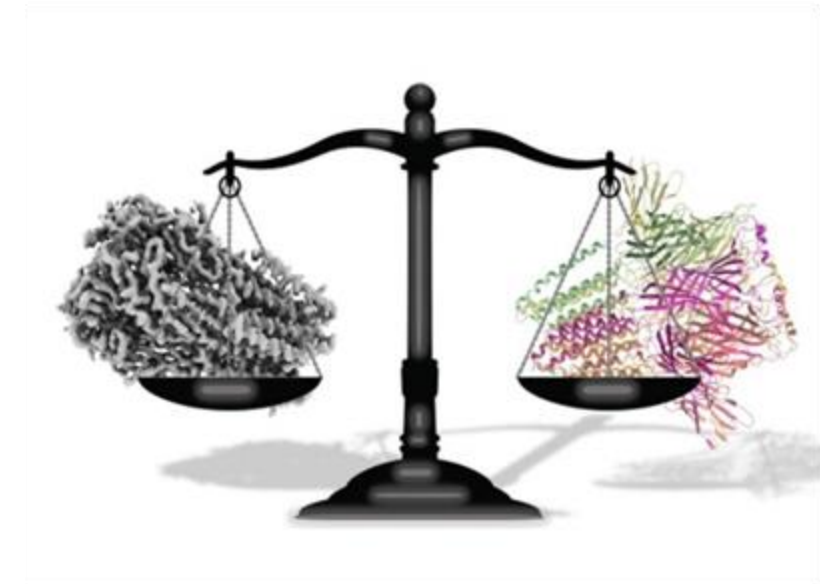
## NextGen Archive: Centralising Access to Integrated Annotations and Enriched Structural Information by the Worldwide Protein Data Bank

Preeti Choudhary, Zukang Feng, John Berrisford, Henry Chao, Yasuyo Ikegawa, Ezra Peisach, Dennis W. Piehl, James Smith, Ahsan Tanweer, Mihaly Varadi, John D. Westbrook, Jasmine Y. Young, Ardan Patwardhan, Kyle L. Morris, Jeffrey C. Hoch, Genji Kurisu, Sameer Velankar, Stephen K. Burley (2024) Database 2024: baae041  
doi: 10.1093/database/baae041



## Restraint Validation of Biomolecular Structures Determined by NMR in the Protein Data Bank

Kumaran Baskaran, Eliza Ploskon, Roberto Tejero, Masashi Yokochi, Deborah Harrus, Yuhe Liang, Ezra Peisach, Irina Persikova, Theresa A Ramelot, Monica Sekharan, James Tolchard, John D Westbrook, Benjamin Bardiaux, Charles Schwieters, Ardan Patwardhan, Sameer Velankar, Stephen K Burley, Genji Kurisu, Jeffrey C Hoch, Gaetano T Montelione, Geerten W Vuister, Jasmine Y Young (2024) Structure 32, 1–14:  
doi: 10.1016/j.str.2024.02.011



## Community Recommendations on CryoEM Data Archiving and Validation

Gerard J. Kleywegt, Paul D. Adams, Sarah J. Butcher, Cathy Lawson, Alexis Rohou, Peter B. Rosenthal, Sriram Subramaniam, Maya Topf, Sanja Abbott, Philip R. Baldwin, John M. Berrisford, Gérard Bricogne, Preeti Choudhary, Tristan I. Croll, Radostin Danev, Sai J. Ganesan, Timothy Grant, Aleksandras Gutmanas, Richard Henderson, J. Bernard Heymann, Juha T. Huiskonen, Andrei Istrate, Takayuki Kato, Gabriel C. Lander, Shee-Mei Lok, Steven J. Ludtke, Garib N. Murshudov, Ryan Pye, Grigore D. Pintilie, Jane S. Richardson, Carsten Sachse, Osman Salih, Sjors H.W. Scheres, Gunnar F. Schroeder, Carlos Oscar S. Sorzano, Scott M. Stagg, Zhe Wang, Rangana Warshamanage, John D. Westbrook, Martyn D. Winn, Jasmine Y. Young, Stephen K. Burley, Jeffrey C. Hoch, Genji Kurisu, Kyle Morris, Ardan Patwardhan, Sameer Velankar (2024) IUCrJ 11: 140–151  
doi: 10.1107/S2052252524001246

# Biocurator Milestones: >10K Processed



*Chairman of the Protein Research Foundation,  
Prof. Toshiharu Hase, and Dr. Minyu Chen (PDBj)*



*Yuhe Liang (RCSB PDB)*



*Irina Persikova (RCSB PDB)*

# PDBe Outreach

- PDBe workshops with a specific focus on data deposition and access:
  - Hot Topics in Contemporary Crystallography (HTCC5) - Dubrovnik, Croatia (April 16-19 2023)
  - Crystallography Course - Oulu, Finland (virtual, May 29-June 2, 2023)
  - European Crystallographic Meeting 8 - Berlin, Germany (June 18-24, 2023)
  - CCP4 Summer School - York, UK (Aug 23-25 2023)

# PDBj Outreach

- PDBj Talk/PSSJ2024: June 13
- PDBj booth: June 29  
OsakaU@Expocity2024
- PDBj booth/IUPUB2024: June 24-28
- PDBj Talk/ICMRBS2024@Korea:  
August 22
- PDBj Talk/APBJC2024: October 24
- PDBj booth/AsCA2024@Malaysia:  
December 1-6



# RCSB PDB Outreach

- International Biocuration Conference  
*PDB: Improved and enriched biodata repository serving many millions of users worldwide* (talk)
- [American Crystallographic Association](#) (ACA)  
*Sixth Decade of Protein Data Bank Operations: Transition to Extended PDB IDs and PDBx/mmCIF Format* (talk)
- Microscopy & Microanalysis (M&M)  
*Frozen in Motion: FAIR and Sustainable Data Management in Cryo-EM at the Worldwide Protein Data Bank* (talk)
- Virtual Office Hours:
  - [Ask a Biocurator](#)
  - [Streamlining PDB Deposition](#)
- Virtual Crash Course  
[Understanding PDB Validation: Which experimental structures should I rely on?](#)



Register for  
**March 28, 2024**  
1pm ET | 10am PT

Virtual Office Hour:  
**Ask A Biocurator**

   
Brian P. Hudson  
and Ezra Peisach

The graphic features a stylized background with green and blue shapes, including a large 'PDB' logo in the top right corner.



Register for  
**September 12, 2024**  
2pm ET | 11am PT

Virtual Office Hour:  
**Streamlining PDB Deposition**

    
Yuhe Liang,  
Irina Persikova, and  
Monica Sekharan

The graphic features a stylized background with green and blue shapes, including a large 'PDB' logo in the top right corner.

# EMDB Outreach

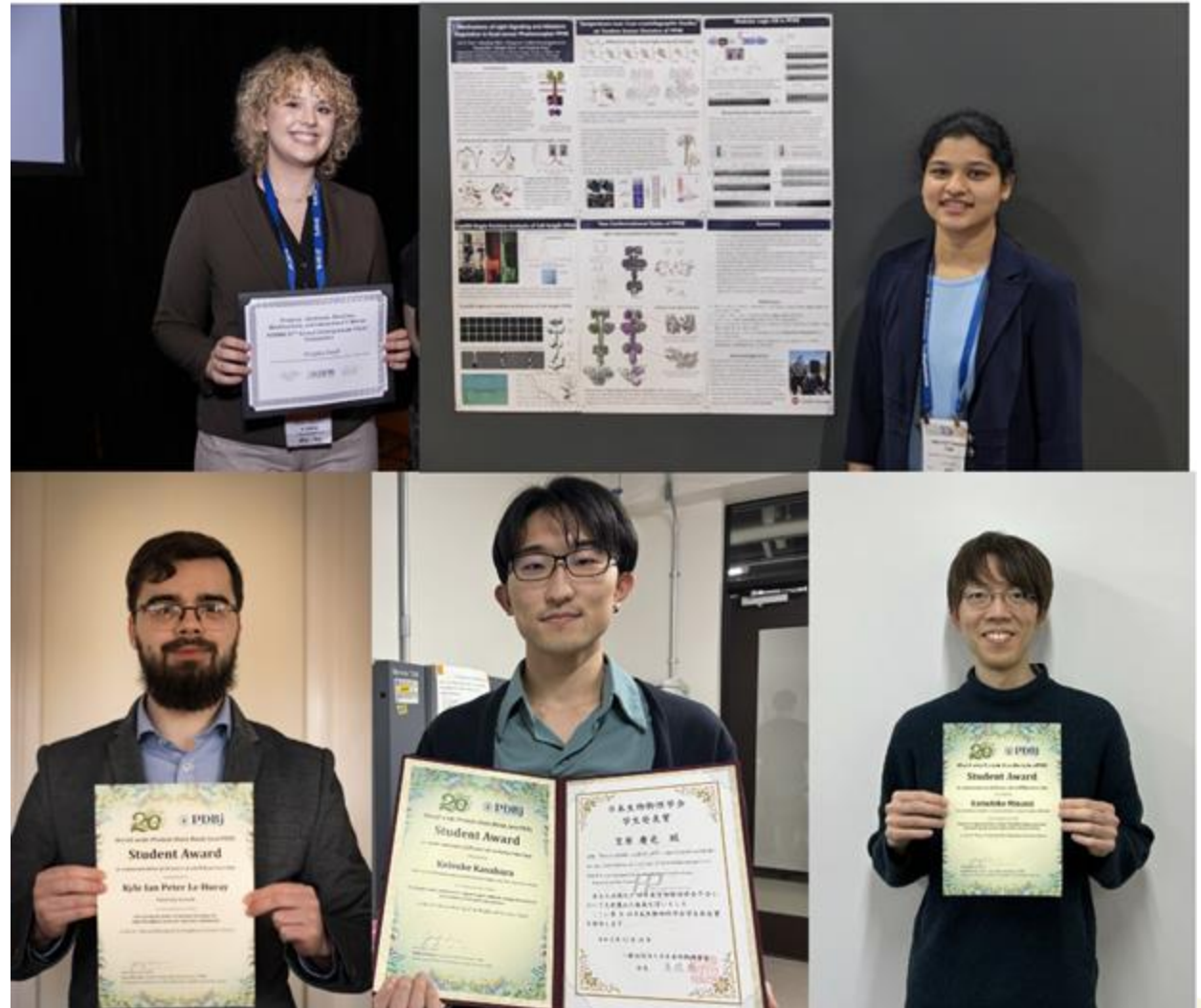
- Workshops and Conferences:
  - CSHL CryoEM course (15 Jan 24)
  - OSCEM standards workshop (22 Feb 24)
  - CryoET community standards workshop (17 Apr 24)
  - CCP-EM symposium (30 Apr 24)
  - Cryo-NET (20 May 24)
  - EMBO cryoEM practical course (18 Aug 24)

# BMRB Outreach

- BMRB hosted a workshop of the North East Consortium on Dynamics and Disorder in Solution. During the workshop, participants helped develop recommendations for relaxation experiments to be curated by BMRB, a controlled vocabulary for describing relaxation experiments, and associated metadata.
- BMRB had a presence in a shared booth at the 2024 Biophysical Society meeting (February 10-14, Philadelphia). BMRB also had presence in a shared booth at the Experimental NMR Conference (April 6-11, Pacific Grove, CA). Both meetings provided an opportunity to engage BMRB users, depositors, and potential users.
- Dr. Baskaran made presentations on BMRB at the 2023 Metabolomics Society meeting (June 18-22 , Niagara Falls, ON) and the NMRbox summer workshop (June 10-11, Farmington CT).

# wwPDB Foundation

- [wwPDB Foundation-sponsored poster prizes](#)
  - Biophysical Society Japan
  - DiscoverBMB
  - Biophysical Society
  - Protein Society
  - Planned: Latin American Crystallographic Association Meeting (September 2024)
- News and social media posts
- [PDBx/mmCIF User Guide](#)





# OneDep Update

# wwPDB Collaboration Resources November 2023-October 2024

wwPDB Partner	Software Development	Production Maintenance/ Project Management	Requirements Setting/ Testing	Core Archive Keeping*	Outreach	Biocuration/ Remediation	Total FTE Commitments
RCSB PDB	2.3	1.4	0.2/0.2	2.0	0.3	6.0	12.4
PDBe	0.8**+0.8#	0.6	0.1/0.1**	-	0.1	4.2##	6.7
PDBj	1.3	0.6	0.2/0.2	-	0.1	3.2	5.6
BMRB	0.85	0.1	0.1/0.1	0.15	-	0.6	1.9
EMDB	1.5	0.5	0.2	1.0	0.1	0.7	4.0
Total wwPDB	7.875	3.1	1.4	2.55	0.65	14.625	30.20

\*RCSB PDB; EMD; BMRB

\*\* If PDBe receives new BBSRC/NSF grant, will have additional 0.8 and 0.1/0.1 FTE

# BBSRC/BBR grant for the PTM remediation project started in April 2022

## Including 0.2 FTE PDB-India annotator in training

# OneDep 2023/2024 Progress *versus* Goals I

- Our familiar Table is here, and details are in Reference.

[Ref. Quarterly Report Appendix](#)

Delivered,  
To be delivered,  
Delayed,  
Project change  
(on hold)

## 2023-2024 Project Roadmap

Delivered, To be delivered, Delayed, Project change (On hold). Arrows represent projects continued either before or after this report cycle.

	Projects	Timeline			
		2023	2024		
		Q4	Q1	Q2	Q3
1. Validation	1.1 EDS software upgrade	Delivered	Delivered	Delivered	
	1.2 MolProbity software upgrade	Delivered	Delivered	Delivered	
	<del>1.3 Improve EM map model fitting in the validation with overall Qscore slider</del>	Delivered	Delivered	Delivered	
	1.4 Modularize validation software for easy maintenance and parallel execution	Delivered	Delivered	Delivered	Delivered
2. Public facing (OneDep or wwPDB.ORG)	<del>2.1 Improve composite map deposition</del>		Delivered		
	2.2 Improve EM cross-file checks at file upload		Delivered		
	2.3 Improve file re-upload process w/ better tracking	Delivered	Delayed		
	<del>2.4 Capture metadata for MicroED</del>			Delivered	Delivered
	<del>2.5 Show EM maps with 3D Mol<sup>®</sup> at the depUI</del>			Delivered	Delivered
	2.6 Community outreach activities to promote mmCIF and transition to extended PDB ID	Delivered	Delivered	Delivered	Delivered
3. Biocuration	3.1 Annotation improvement- automation	Delivered	Delivered		
	<del>3.2 Provide revision history for EMDB files</del>		Delivered		
	<del>3.3 Metadata annotation to support MicroED</del>			Delivered	Delivered
	3.4 Annotation improvement- workflow for large structures			Delivered	
	3.5 Annotation improvement - sequence search	Delivered	Delivered	Delivered	Delivered
4. Backend Stabilization	4.1 wwPDB data transfer changes due to PDBe and EMDB moving to Globus service	Delivered	Delivered		
	4.2 Remove dependency on email handler at RCSB PDB			Delivered	
	4.3 Third party S/W updates - OpenBabel			Delivered	
	<del>4.4 Migrate legacy EMDB entries into OneDep</del>		Delivered		
	4.5 Third party S/W updates - Phenix				Delayed
5. Archive Improvements	<del>5.1 mmCIF extension to support MicroED</del>	Delivered	Delivered		
	5.2 Recalculation of validation reports at PDB Archive				Delayed
	5.3 Provide holdings files at EMDB Archive		Delivered		
	<del>5.4 Metadata remediation (EMDB archive)</del>	Delivered	Delivered		
	5.5 Unifying PDB-dev into PDB Archive	Delivered	Delivered	Delivered	Delivered
	5.6 PTM remediation (PDB archive)	Delivered	Delivered	Delivered	Delivered
	5.7 Metalloprotein remediation (PDB Archive)			Delivered	Delivered
	5.8 PDB archive-wide remediation			Delivered	Delivered
	<del>5.9 NMR remediation with peak lists</del>			Delivered	Delivered

# OneDep 2023/2024 Progress *versus* Goals I

Components	Major Projects to be Completed	Primary resource
Validation	Ongoing system improvements & bug fixing	All sites
	Upgrade 3rd party EDS software	PDBe
	Modularize validation software for easy maintenance and parallel execution	RCSB PDB
	<b>Upgrade 3rd party MolProbity software</b>	<b>PDBj</b>
	<b>Improve EM map-model fit-validation with overall Qscore slider</b>	<b>EMDB</b>
Deposition	Ongoing system improvements & bug fixing	PDBe/PDBj/EMDB
	<b>Improve 3DEM composite map deposition</b>	<b>EMDB</b>
	Improve EM cross-file checks at file upload	EMDB
	Show EM maps with 3D Mol* at the depUI	<b>EMDB</b>
	<b>Capture metadata for MicroED</b>	<b>PDBe</b>
	<b>Improve file re-upload process with better tracking</b>	<b>EMDB/BMRB</b>

**Bold: re-forecasted to 2024-2025**

# OneDep 2023/2024 Progress *versus* Goals II

Components	Major Projects to be Completed	Primary resource
Biocuration	Ongoing system improvements & bug fixing	RCSB PDB
	Annotation improvement- automation	RCSB PDB
	Annotation improvement- workflow for large structures	RCSB PDB
	Annotation improvement - sequence search	RCSB PDB
	<b>Provide revision history for EMDB files</b>	<b>RCSB PDB</b>
Infrastructure	Third party S/W updates - OpenBabel	RCSB PDB
	<b>Third party S/W updates - Phenix</b>	<b>RCSB PDB</b>
	<b>Migrate EMDB legacy entries into OneDep</b>	<b>EMDB</b>
	wwPDB data transfer changes due to PDBe and EMDB moving to Globus service	RCSB PDB/PDBj
	Remove dependency on email handler at RCSB PDB	RCSB PDB

**Bold: re-forecasted to 2024-2025**

# OneDep 2023/2024 Progress *versus* Goals III

Components	Major Projects to be Completed	Primary resource
Core Archives	<b>mmCIF extension to support MicroED</b>	<b>RCSB PDB</b>
	<b>Archival re-calculation of validation reports for PDB and EMDB archives</b>	<b>PDBe</b>
	<b>Metadata remediation at EMDB archive</b>	<b>EMDB</b>
	<b>NMR data remediation with peak lists at PDB and BMRB archives</b>	<b>PDBj</b>
	Provide holdings files at EMDB Archive	EMDB
	PDB archive-wide remediation	RCSB PDB/PDBe/PDBj
	Unifying PDB-dev into PDB Archive	RCSB PDB
	PTM remediation (PDB)	PDBe
	Metalloprotein remediation planning	RCSB PDB
	Ongoing mmCIF dictionary development and maintenance	RCSB PDB
	Community outreach activities to promote mmCIF and transition to extended PDB ID	All sites

**Bold: re-forecasted to 2024-2025**

# wwPDB Collaboration Resources November 2024-October 2025

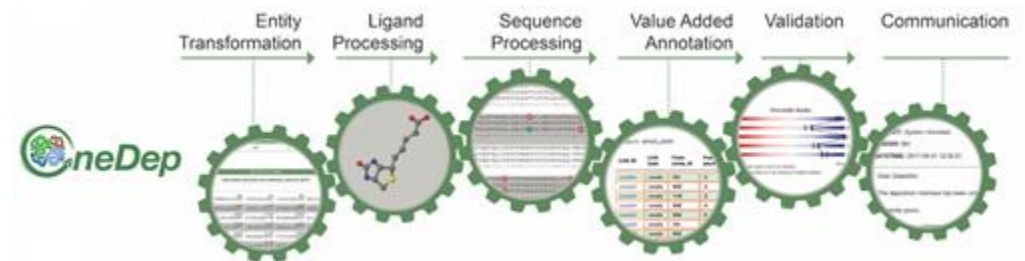
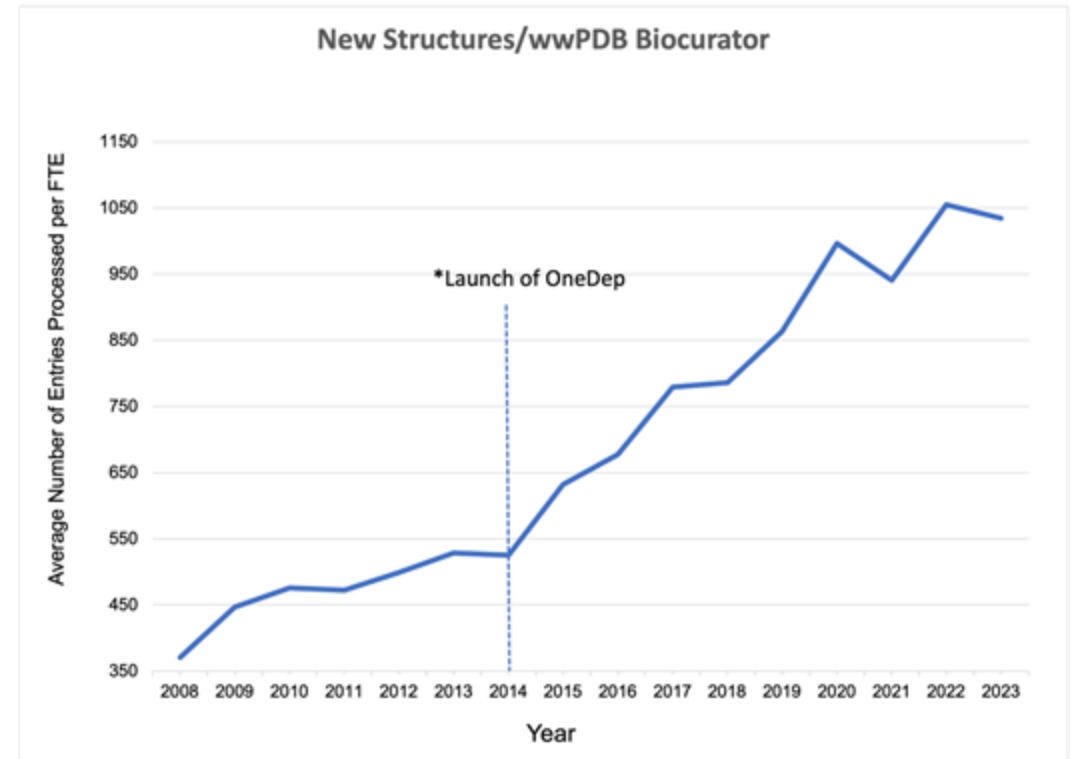
wwPDB Partner	Software Development	Production Maintenance/ Project Management	Requirements Setting/ Testing	Core Archive Keeping*	Outreach	Biocuration/ Remediation	Total FTE Commitments
RCSB PDB	2.3	1.4	0.2/0.2	2.0	0.3	6.0	12.4
PDBe	1.15	0.6	0.15/0.15	-	0.1	3.9	6.05
PDBj	1.3	0.6	0.2/0.2	-	0.1	7.2 +	9.6
BMRB	0.85	0.1	0.2/0.2	0.15	0.1	0.6	2.1
EMDB	1.5	0.5	0.1/0.1	1.0	0.1	0.7	4.0
Total wwPDB	7.1	3.2	1.1/0.6	3.15	0.7	18.4	34.15

\*RCSB PDB; EMDB; BMRB

+PDBj Biocuration/Remediation includes 4.0 FTEs from PDBc

# wwPDB Biocurator Productivity

- 17,064 depositions in 2023
- Improved efficiency for biocurating incoming depositions
  - Auto-approve entries w/o corrections
  - Automated validation runs without UI access
  - Provide site specific setting on automated assembly annotation without UI access
  - Improved automation of ligand processing for unambiguous ligand cases
  - Improve efficiency by skipping PISA calculations for NMR, EM and large X-ray entries
- New sequence builder tool supports chimeric proteins
- Improved use of author-provided ligand restraints for graph search





# OneDep 2024/2025 Goal Setting I

Components	Major Projects to be Completed	Primary resource
Validation	Ongoing system improvements & bug fixing	All sites
	Upgrade 3rd party Phenix software	RCSB PDB
	Modularize validation software for easy maintenance and parallel execution	RCSB PDB
	<b>Upgrade 3rd party MolProbity software</b>	<b>PDBj</b>
	<b>Provide quality slider of map-model fitness for 3DEM based on Q-scores</b>	<b>EMDB</b>
Deposition	Ongoing system improvements & bug fixing	PDBe/PDBj/EMDB
	<b>Improve (refactor) file upload for 3DEM depositions</b>	<b>EMDB</b>
	Display 3DEM validation data and warnings to depositors and biocurators	EMDB
	Support EM deposition of variability analysis data	EMDB
	<b>Improve 3DEM composite map deposition</b>	<b>EMDB</b>
	<b>Provide synthetic sources for EM entries</b>	EMDB
	<b>Improve file upload for NMR depositions</b>	<b>PDBj</b>
	<b>Improve MicroED deposition</b>	<b>PDBe</b>
	Display extended PDB IDs throughout OneDep pipeline	PDBe
	Improve capturing author's assembly	PDBe

**Bold:** re-forecasted from 2023-2024

# OneDep 2024/2025 Goal Setting II

Components	Major Projects to be Completed	Primary resource
Biocuration	Ongoing system improvements & bug fixing	RCSB PDB
	<b>Improve microED deposition</b>	<b>RCSB PDB</b>
	Display extended PDB IDs throughout OneDep pipeline	RCSB PDB
	Improve CCD file checking	RCSB PDB
	Enhance biocuration with a new feature	RCSB PDB
	<b>Provide revision history for EMDB files</b>	<b>RCSB PDB</b>
Infrastructure	EBI infrastructure migration	PDBe/RCSB PDB
	Move Tools source code from CVS to github repository	RCSB PDB
	<b>Migrate EMDB legacy entries into OneDep</b>	<b>EMDB/RCSB PDB</b>
	Infrastructure planning for supporting simultaneous depositions with shared metadata	All sites
	Requirements setting for improving EM composite map depositions	EMDB
	Standalone NMR Data Conversion Service	BMRBj/BMRB

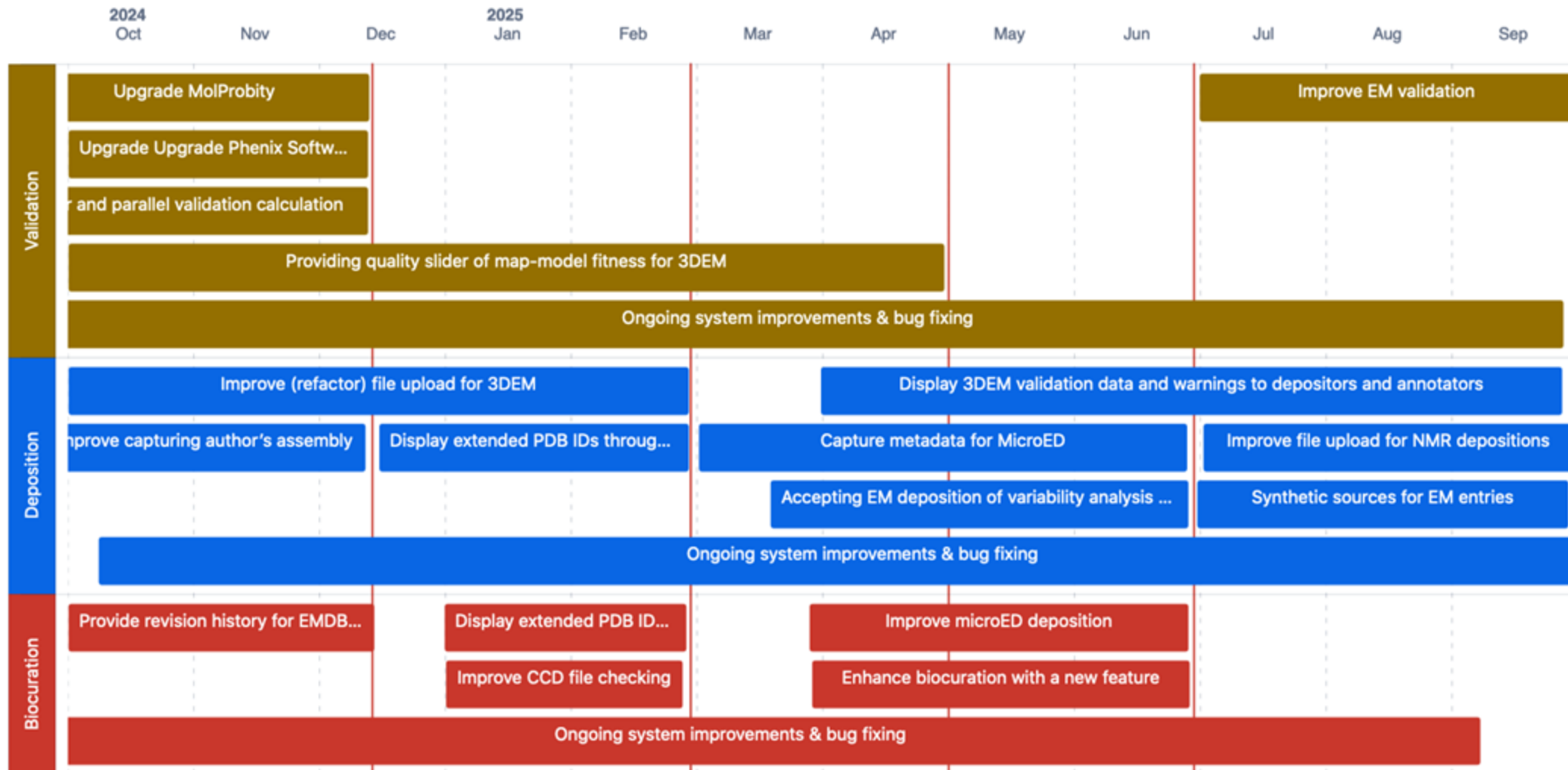
**Bold: re-forecasted from 2023-2024**

# OneDep 2024/2025 Goal Setting III

Components	Major Projects to be Completed	Primary resource
Core Archives	<b>mmCIF extension to support MicroED</b>	<b>RCSB PDB</b>
	<b>Archival re-calculation of validation reports for PDB and EMDB archives</b>	<b>PDBe</b>
	<b>Archival EM remediation including starting model remediation</b>	<b>EMDB</b>
	NMR peak lists remediation at PDB and BMRB archives	PDBj
	NMR refinement method remediation at PDB archive	PDBe
	Inclusion of IHM data in SIFTS	PDBe
	Remediate SF file for missing or inconsistent cell information	RCSB PDB
	Metalloprotein remediation	RCSB PDB
	Planning for beta PDB archive	All sites
	Ongoing mmCIF dictionary development and maintenance	RCSB PDB
	Community outreach on PDB-IHM deposition, extended PDB ID, and mmCIF adoption	All sites
	Create longer-term plans to support the growth of cryo-ET	All sites

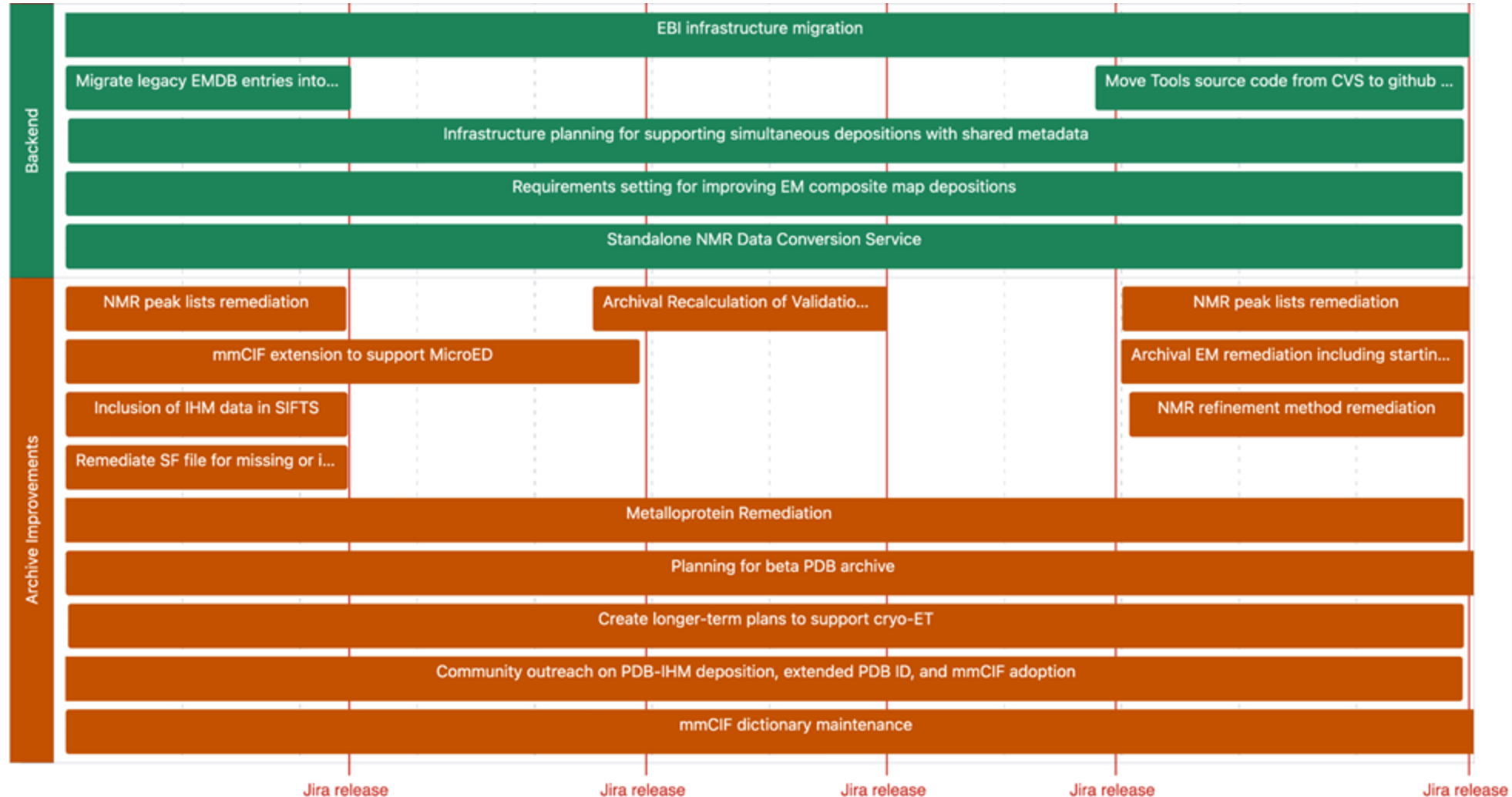
**Bold:** re-forecasted from 2023-2024

# OneDep 2024/2025 Roadmap I



- will be further adjusted based on requirement setting

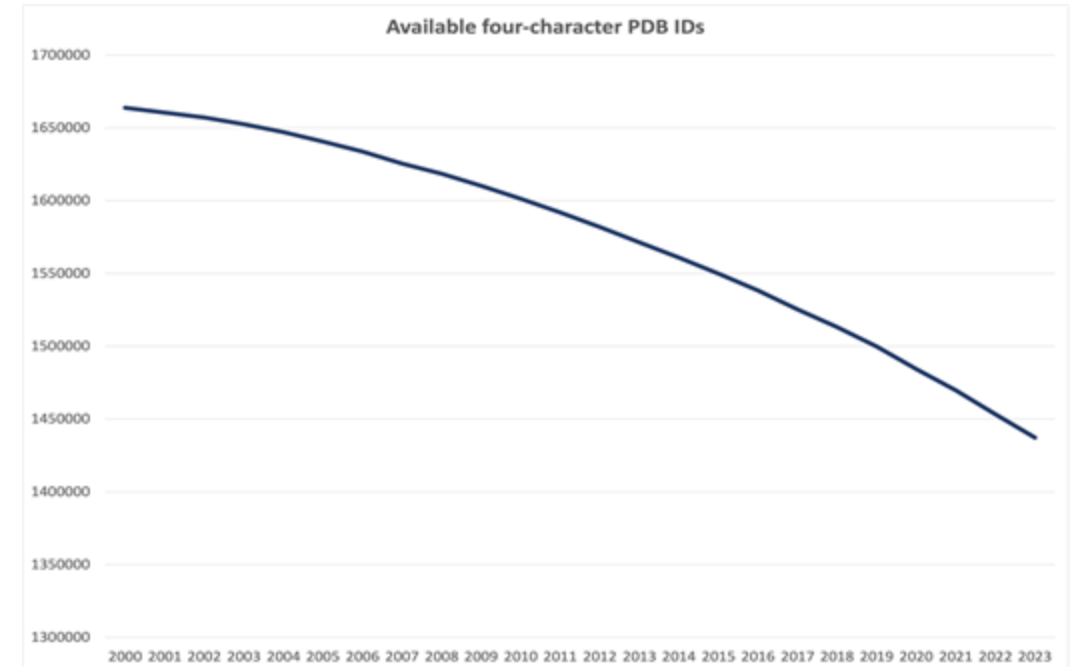
# OneDep 2024/2025 Roadmap II



- will be further adjusted based on requirement setting

# Five-year Plan for Transitioning to New PDB ID Format

- Anticipating change required in the 4 years
- [Community outreach on transitioning to new PDB ID and PDBx/mmCIF formats](#)
  - **2023: provide an [FAQ page](#) on new PDB IDs**
  - **2024: create a [PDBx/mmCIF documentation](#) for user training**
  - **2024-2025: update archival files to include new PDB IDs**
  - 2026: create a PDB “beta” archive with new PDB IDs in file naming
  - 2027: modify OneDep software to issue extended PDB IDs with new file naming
  - 2028: switch PDB beta archive to PDB archive when four-character PDB IDs are consumed
  - Ongoing: provide PDBx/mmCIF training courses at regional data centers
  - Ongoing: advertise extended PDB ID format and transition plan at professional society meetings and social media



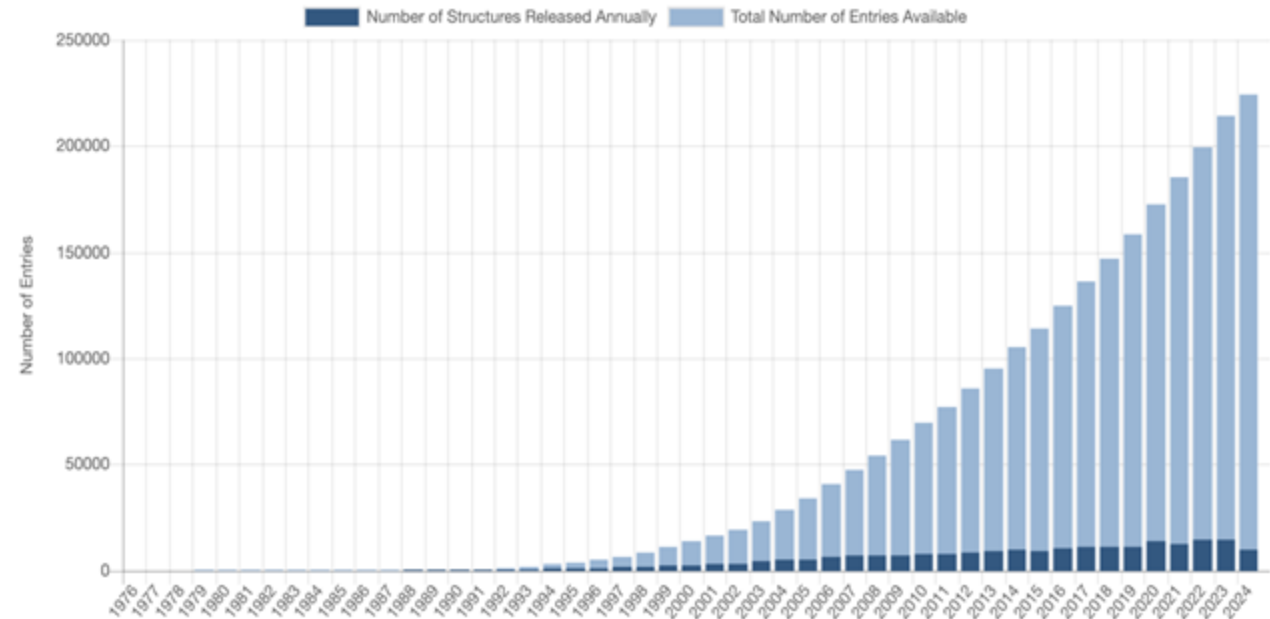
```
loop_  
_database_2.database_id  
_database_2.database_code  
_database_2.pdbx_database_accession  
_database_2.pdbx_DOI  
PDB pdb_00001abc pdb_00001abc  
10.2210/pdb_00001abc/pdb
```

# PDB Archive Update

# Current PDB Archive Status



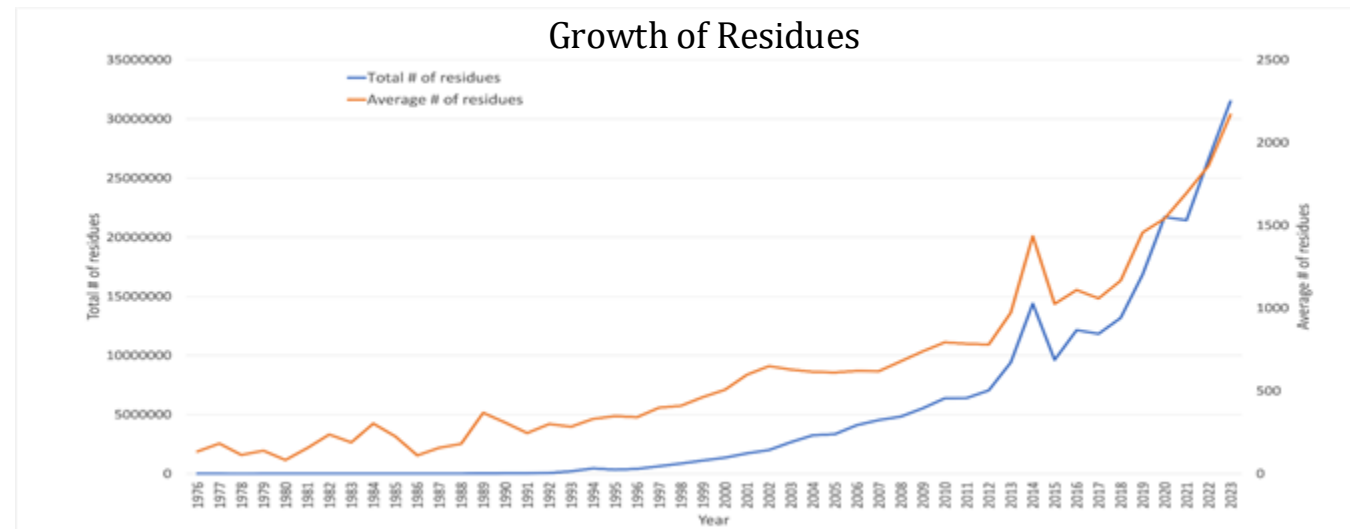
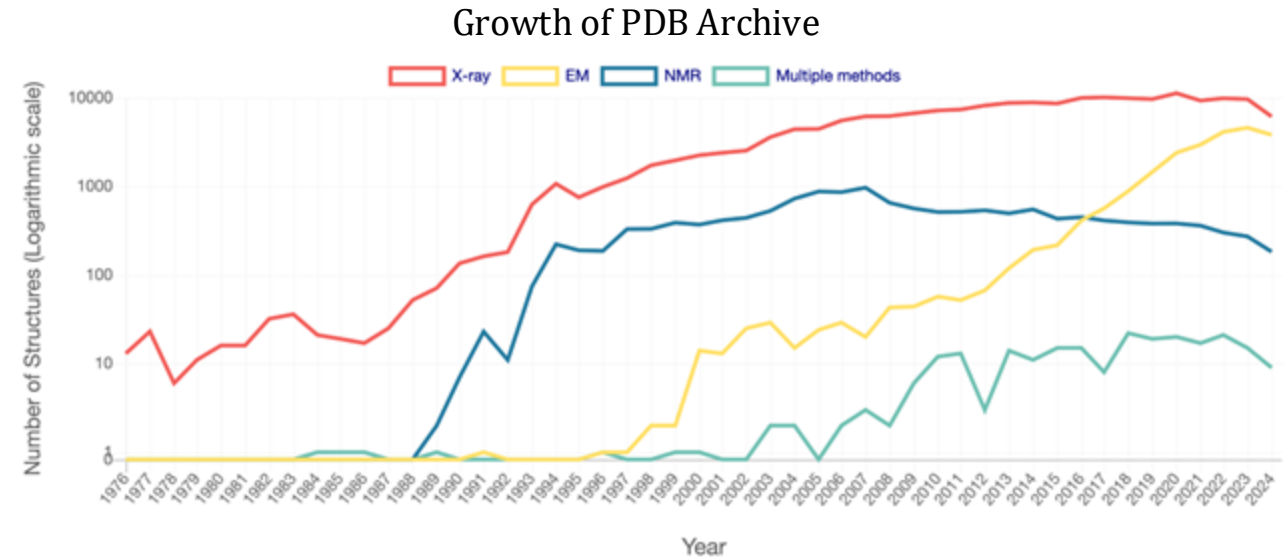
- Total Entries=224,201  
(as of 8/29/2024)
- Core Archive Storage
  - OneDep Sessions: ~96 TB
  - ftp(legacy + versioned): ~1.7 TB
  - ftp snapshots: ~13.6TB
  - EMDB ftp: ~18.5 TB
- **PDB-IHM data now being served alongside PDB structures**
- Both PDB and PDB-IHM data also being housed and delivered by Amazon Web Services (AWS) with no storage or egress fees
- NextGen archive now serving enriched annotation in the atomic coordinate files  
(<https://files-nextgen.wwpdb.org>)





# PDB Archive Growth in 2023

- Year-end holdings 224,201
- 14,468 new entries released
- Archival entries growing in both size and complexity
- Record 4,579 new 3DEM entries released
- ~12% increase *versus* 2022



# wwPDB Accreditation

- PDB has been formally recognized as a Global Core Biodata Resource
  - Went through a rigorous two-stage process
  - Assessed by a panel of more than 50 independent expert reviewers
- wwPDB plans to seek formal recognition for EMDB and BMRB in response to subsequent GBC calls
- wwPDB was highlighted in the WDS newsletter



[HOME](#) [SCIENTIFIC ACTIVITIES](#) [GOVERNANCE](#) [RESOURCES](#) [CONTACT](#)

## GLOBAL BIODATA COALITION

The Global Biodata Coalition (GBC) is a forum for research funders to better coordinate and share approaches for the efficient management and growth of biodata resources worldwide. The GBC aims to stabilize and ensure sustainable financial support for the global biodata infrastructure and in particular to identify for prioritized long-term support a set of Global Core Biodata Resources that are crucial for sustaining the broader biodata infrastructure.

[Discover the benefits](#)

## WDS Member Highlights



The World Data System is proud to have these organizations as members in our continued mission to enhance the capabilities, impact, and sustainability of our member data repositories and data services by:

**creating** trusted communities of scientific data repositories,  
**strengthening** the scientific enterprise throughout the entire lifecycle of data and all related components creating first-class data that feeds first-class research output, and  
**advocating** for accessible data and transparent and reproducible science.

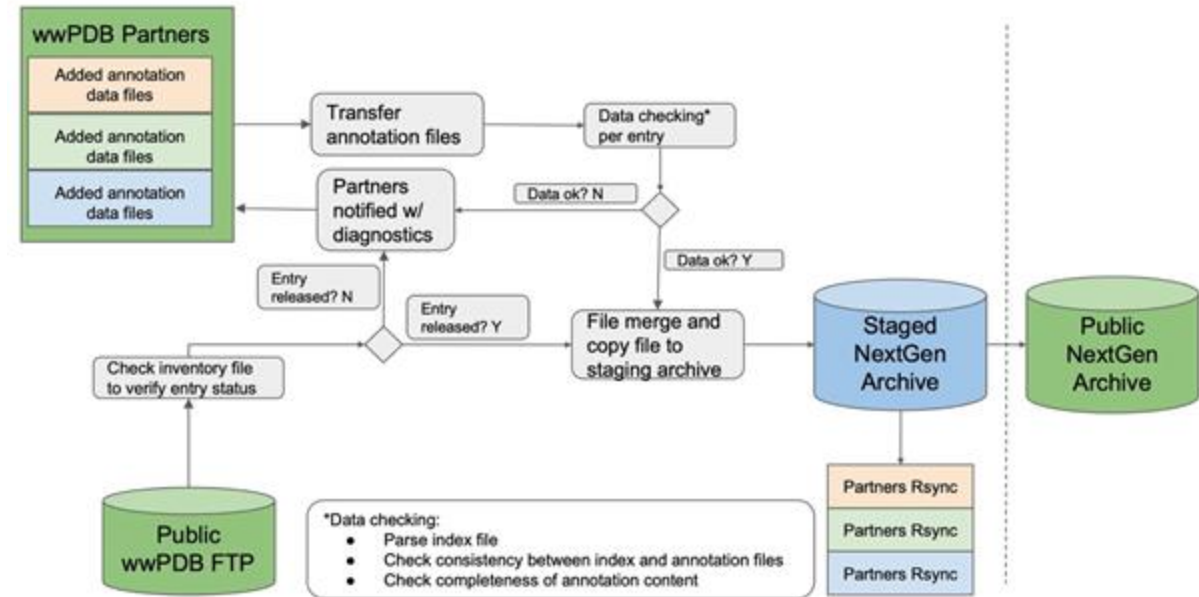
# PDB Data Delivery by AWS

- Amazon Web Services (AWS) Open Data Sponsorship Program now housing and delivering PDB data
  - No storage fees charged to RCSB PDB
  - No egress fees charged to PDB users
- Current AWS holdings include:
  - Annual PDB Archive Snapshots
  - Current PDB FTP Archive (updated weekly)
  - **PDB-IHM data now available**
- AWS can deliver PDB data faster than RCSB PDB, PDBe, or PDBj!
  - <https://s3.rcsb.org>

The screenshot shows the AWS Registry of Open Data page for the Protein Data Bank 3D Structural Biology Data. The page is titled "Registry of Open Data on AWS" and features the AWS logo in the top right corner. Below the title, there is a blue banner with the text "Open Data is now available on AWS Data Exchange" and a button labeled "Explore the catalog". The main heading is "Protein Data Bank 3D Structural Biology Data". Below this heading, there are several tags representing different data categories: amino acid, archives, bioinformatics, biomolecular modeling, cell biology, chemical biology, COVID-19, electron microscopy, electron tomography, enzyme, life sciences, molecule, nuclear magnetic resonance, pharmaceutical, protein, protein template, SARS-CoV-2, structural biology, and x-ray crystallography. The page is divided into two main sections: "Description" and "Resources on AWS". The "Description" section provides a detailed overview of the PDB archive, mentioning its establishment in 1971 and its role as a jointly-managed core archive of the Worldwide Protein Data Bank partnership. The "Resources on AWS" section is divided into two parts: "Description" and "Resources on AWS". The "Description" part describes the globally cached distribution of the dataset and its availability for browsing. The "Resources on AWS" part lists the resource type as "CloudFront Distribution", the AWS Region as "us-west-2", and provides a link to "Browse Dataset". The second part of the "Resources on AWS" section describes the "Description" of the historical snapshots of archival datasets from 2005 onwards, lists the resource type as "S3 Bucket", provides the Amazon Resource Name (ARN) as "arn:aws:s3:::pdbsnapshots", lists the AWS Region as "us-west-2", and provides the AWS CLI Access (No AWS account required) as "aws s3 ls --no-sign-request s3://pdbsnapshots/". It also includes a link to "Browse Bucket".

# NextGen PDB Archive

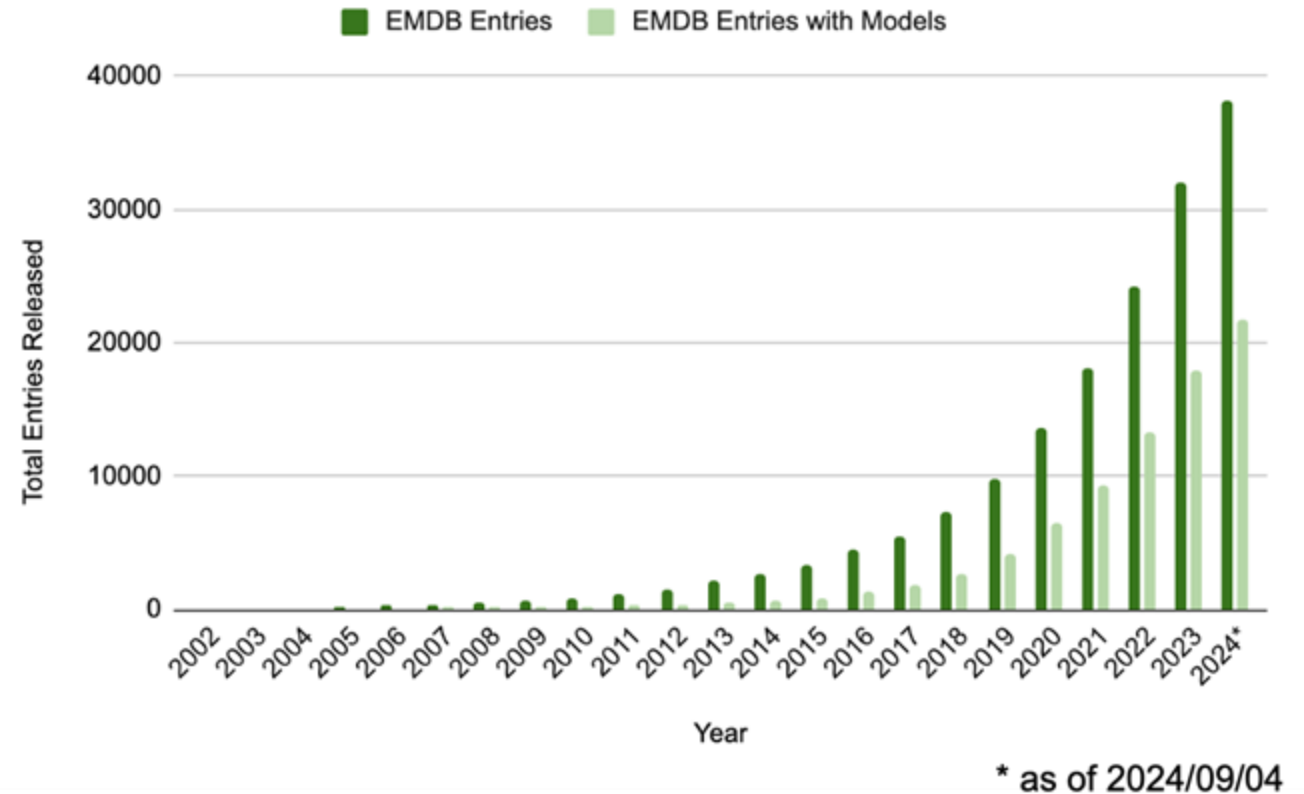
- NextGen Archive V1.0 launched in February 2023
  - <https://files-nextgen.wwpdb.org>,  
<rsync://rsync-nextgen.wwpdb.org>
- Provides enriched annotation from external database resources in the PDBx/mmCIF files
  - Sequence annotation such as UniProt, CATH, SCOP2 and Pfam from SIFTS
  - Intramolecular connectivity for each residue (atom pairs, bond order, aromatic flag, and stereochemistry)
- Automated monthly update process
  - Updated every month on the 1st Wednesday at 00:00 UTC
- Work described in a cited [Database](#) publication
- More than 14,700,000 downloads since Feb. 2023



# EMDB Archive Update

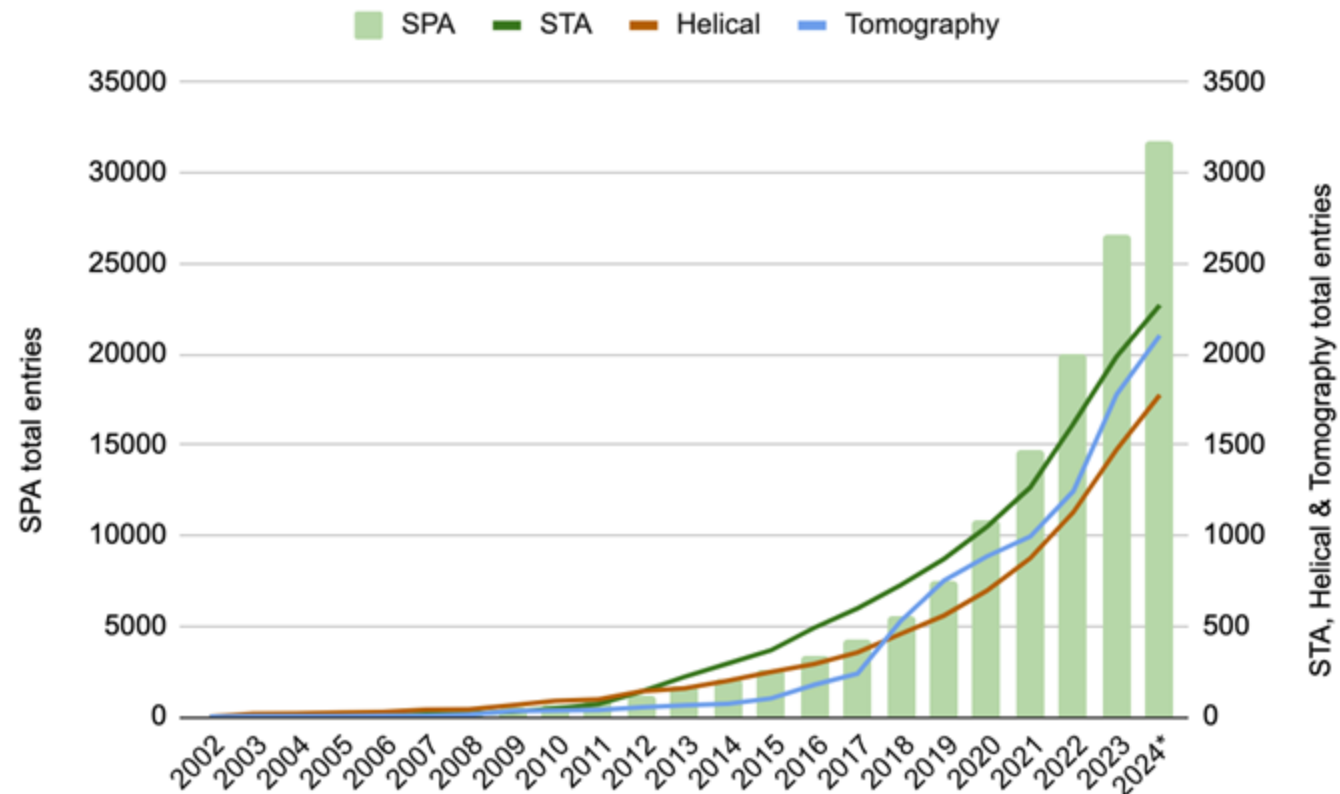
# Current EMDB Archive Status

- Total Entries = 38,141
  - (as of 2024/09/04)
- EMDB Archive Storage: 18.5 TB
  - (as of 2024/09/04)



# Current EMDB Archive Status

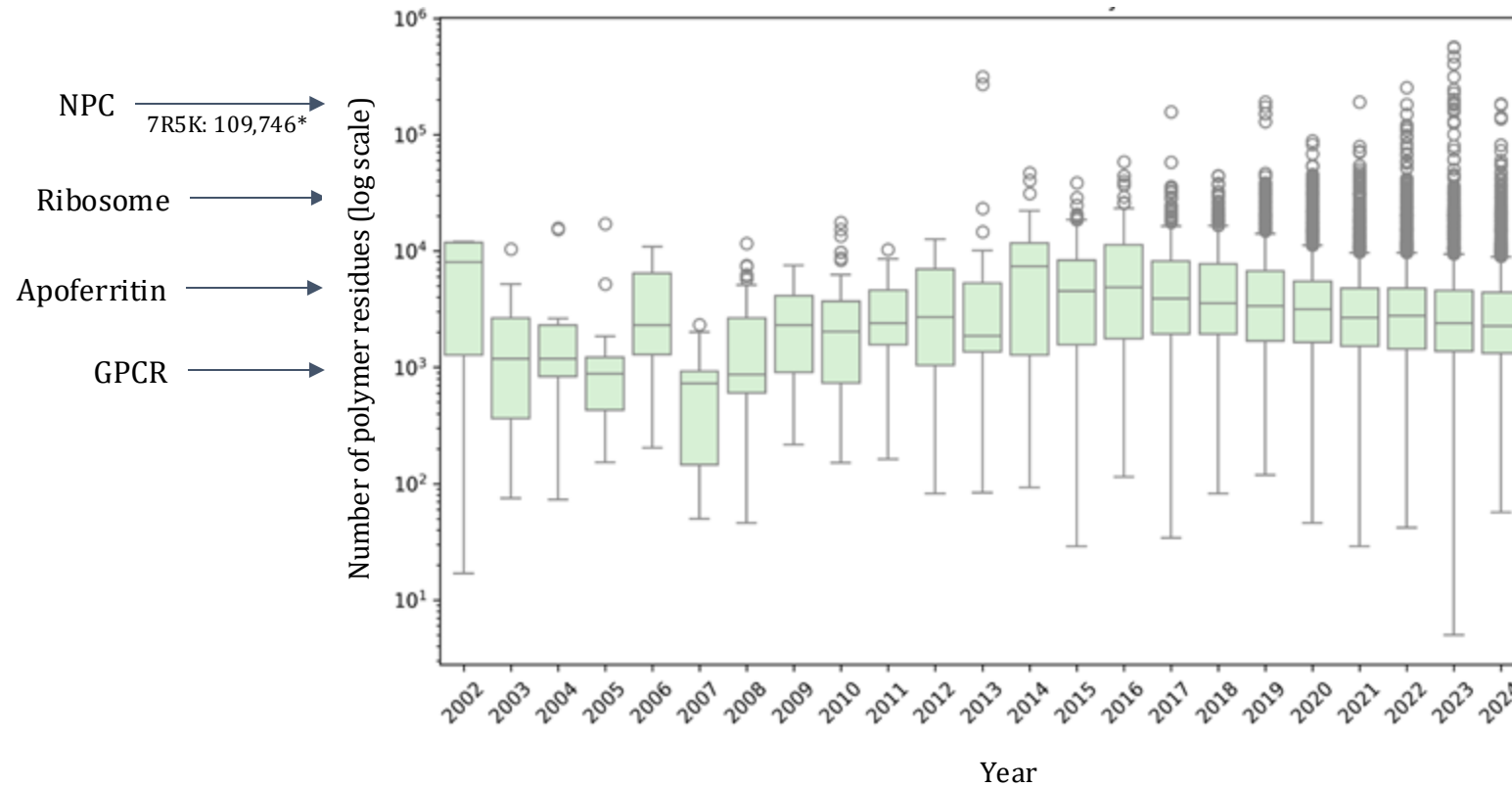
## Total entries per EM-submethod



\* as of 2024/09/04

# Increasing model complexity and size

3DEM maps used to determine increasingly larger structures



Data downloaded from RCSB PDB  
(advanced search query builder, experimental method: Electron Microscopy)

\*Note many models are asymmetric units

2024 values as of 2024/09/04



# 2023 - 2024 Achievements

- Distribution of EM metadata in PDBx/mmCIF file format
- Automated cross-checking of uploaded EM files to ensure data validity
- Review of EM entry images now supported during deposition
- Improved validation performance

# Distribution of EM metadata in PDBx/mmCIF file format

- EM Metadata is now served in the EMDE PDBx/mmCIF format
- Brings metadata format parity between
- EMDB will continue to serve metadata in

```
1 data_IHD-61888
2 #
3 _entry_id IHD-61888
4 #
5 _audit_conform.dict_name mmcif_pdbx.dic
6 _audit_conform.dict_version 3.395
7 _audit_conform.dict_location http://mmcif.pdb.org/dictionaries/ascii/mmcif_pdbx.dic
8 #
9 loop_
10 _database_2.database_id
11 _database_2.database_code
12 _database_2.pdbx_database_accession
13 _database_2.pdbx_DOI
14 WIPAC 2_138888204 ? ?
15 IHD6 IHD-61888 ? ?
16 #
17 _pdbx_database_related.db_name EMDB
18 _pdbx_database_related.details -
19 _pdbx_database_related.db_id IHD-61888
20 _pdbx_database_related.content_type 'associated EM volume'
21 #
22 _citation.abstract ?
23 _citation.abstract_id_CAS ?
24 _citation.book_id_ISSN ?
25 _citation.book_publisher ?
26 _citation.book_publisher_city ?
27 _citation.book_title ?
28 _citation.coordinate_linkage ?
29 _citation.country US
30 _citation.database_id_NeXline ?
31 _citation.details ?
32 _citation.id primary
33 _citation.journal_abbrev Biophys
34 _citation.journal_id_ASTM ?
35 _citation.journal_id_CSD ?
36 _citation.journal_id_IJSM 1692-4265
37 _citation.journal_full ?
38 _citation.journal_issue ?
39 _citation.journal_volume ?
40 _citation.language ?
41 _citation.name_First ?
```

# Automated cross-checking of uploaded EM files to ensure data validity

- Files are now checked against one another during the file upload process for EM, improving data validity and reducing annotator workload
- Examples of cross-checks include:
  - Ensuring all map files are unique
  - Ensuring the maps overlay correctly
  - Ensuring all coordinate atoms are within the map bounding box
  - Ensuring half-maps are cubic and volumes overlay

# Review of EM entry images now supported during deposition

- Entry images now displayed to the end user (or reviewer)
- Entry image requirements are highlighted to the user during submitting files
- Allows review and fixing of any issues before entry is pushed into the OneDep system



# Improved Validation Performance

- The increasing size of models built into EM maps highlighted the need to improve EM validation performance
- wwPDB recognised that Q-score generation and visualisation generation was an area which required improvement
- Working with the ChimeraX team EMDB was able to significantly increase the performance of validation visualisation generation

EMD-ID	PDB-ID	Old Processing Time (s)	New Processing Time (s)	Improvement (multiple)
EMD-29722	8G4L	8891	428	20.77x
EMD-24191	7N61	5656	363	15.5x
EMD-8742	5VY4	285	171	1.66x

# Staff changes

- New biocurator started in September 2024:
  - 3DEM/MX background from LMB and Evotec, Inc.
  - 0.5 FTE commitment to OneDep biocuration
  - Replacing biocurator (0.5 FTE) who left in July 2024

# Other achievements

## EMDB projects that will feed into future wwPDB activities

- Evaluation Activities:
  - Masking and FSC validation assessments to help biocurators
  - Performance of model-map FSC, local resolution tools
- Planning Activities:
  - Internal EMDB archive pipeline redesign to cope with future release size
  - 2024: 204 p/w
  - 2030: 985 p/w - 4.8x (6 year projection)

# **BMRB Archive Update**

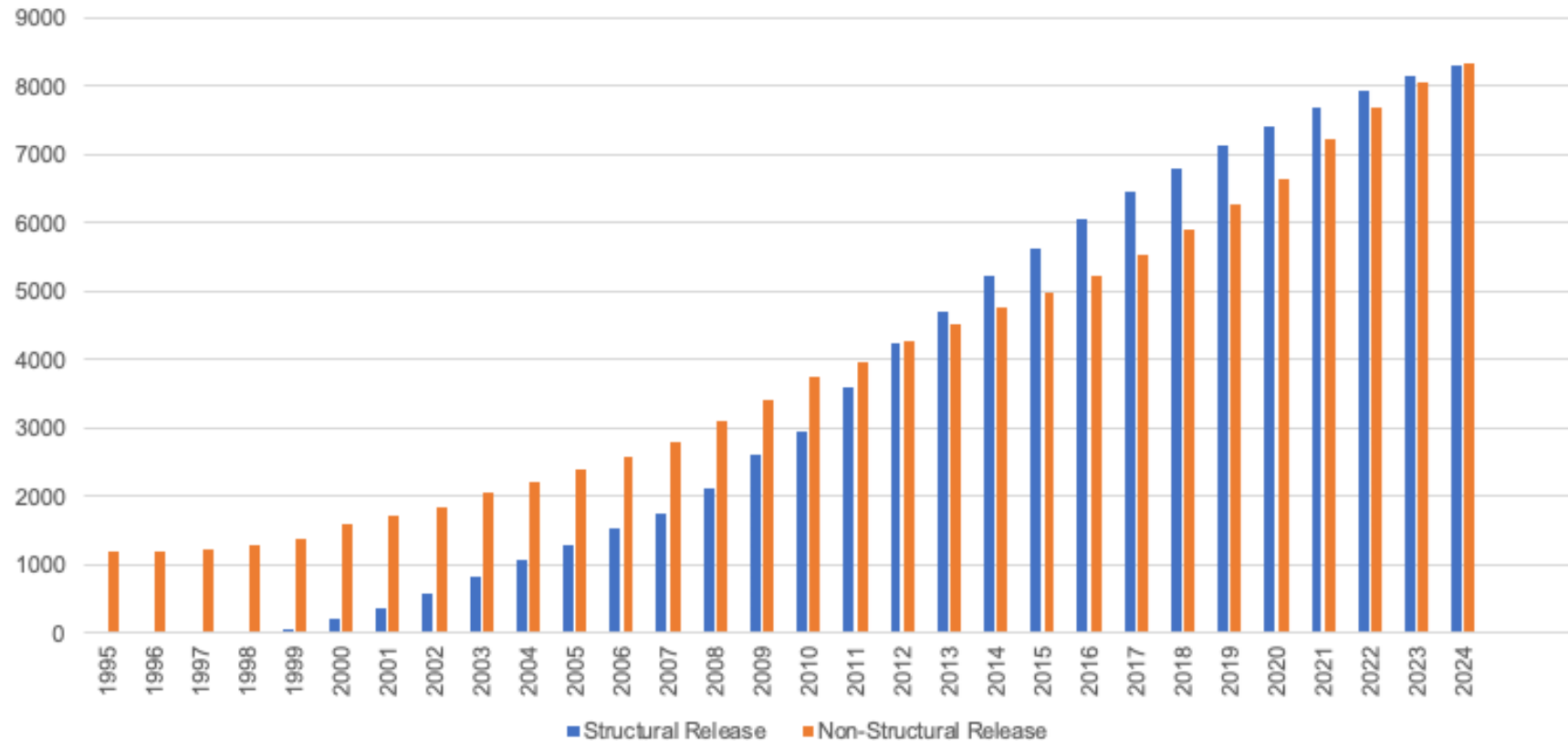


# Developments since 2023 Meeting

- Phase I of restraints remediation project has been completed.
  - Legacy restraints have been translated into NMR-STAR and NEF format
  - Remediated restraints are now available in the FTP archive.
- Fragment screening data sets are now available in BMRbig and the corresponding small molecules are linked to BMRB small molecule database
- A Simple user interface is now available to upload time domain data (FIDs) for the existing entries.
- NMR Restraints validation manuscript has been published.

# BMRB Core Archive Growth

**As of 09/06/2024**  
**Structural Release: 8301**  
**Non-Structural Release: 8351**



# BMRB Core Archive Growth

Year	In Year Totals				By Year Totals				Withdrawn	
	Structural release	Nonstructural release	Original release	Total release	First release	Total release	Structural release	Nonstructural release	Withdrawn during year	Eventually withdrawn from this year
2014	522	242	764	1068	9990	12219	5230	4760	1	20
2015	402	234	636	696	10626	12915	5632	4994	0	75
2016	441	234	675	918	11301	13833	6073	5228	1	22
2017	379	305	684	839	11985	14672	6452	5533	5	70
2018	348	361	709	855	12694	15527	6800	5894	3	33
2019	323	375	698	968	13392	16495	7123	6269	4	37
2020	294	387	681	836	14073	17331	7417	6656	5	69
2021	281	572	853	1137	14926	18468	7698	7228	2	62
2022	247	456	703	1192	15629	19660	7945	7684	0	37
2023	205	388	593	834	16222	20494	8150	8072	0	66
2024*	151	279	430	554	16652	21048	8301	8351	0	30

\*As of September 6, 2024

# BMRB Core Archive Growth

## Internet Server Traffic (Website) – All Mirrors\*

Year	Server requests	Page requests	File requests	Distinct hosts served	US hosts served	Total data transferred
2014	39,144,507	6,407,405	606,332	1,117,340	426,682	28.21 TB
2015	46,899,762	7,011,339	664,415	1,227,112	500,589	32.68 TB
2016	54,081,884	7,869,456	1,028,193	1,321,909	531,196	36.38 TB
2017	56,209,400	8,537,562	1,433,036	1,026,426	387,809	18.37 TB
2018	87,818,181	25,523,384	2,415,042	1,330,889	439,600	17.75 TB
2019	87,989,534	33,885,684	1,173,443	1,779,863	638,348	27.55 TB
2020	83,732,415	22,489,261	1,603,539	1,783,604	608,285	47.67 TB
2021	82,585,307	25,655,514	737,829	2,609,272	627,506	44.85 TB
2022	127,891,445	25,774,500	1,628,260	2,730,172	1,045,469	40.41 TB
2023	105,145,355	21,190,312	682,905	3,177,690	1,250,574	37.00 TB
2024*	183,555,553	28,567,566	2,778,606	6,659,965	4,322,040	93.26 TB

Note  
explosive  
growth

- BMRB has mirror sites in Italy and Japan, and BMRBj branch for deposition

\*As of September 6, 2024

# Upcoming BMRB Projects

- Enhanced chemical shift histogram and data visualization
  - Chemical shift histograms will be categorized based on secondary structure types by combining information from PDB entries
- Reduction of technical debt
- Remediation and enrichment of annotations: Example : relaxation data
- Work with NEF working group and the North Eastern US NMR community (NE-CODDS) to standardize the nomenclature and establish metadata recommendations for relaxation experiments (whitepaper in progress)
- Validation report improvements: both PDF and XML format report will be supported (similar to wwPDB)

# Expansion to include BMRBe

- Europe and Israel collectively contribute approximately the same number of BMRB entries annually as North America
- Europe leads the world in the deployment of ultra high-field instruments (1+ GHz)
- Labs in Europe have special expertise in relaxation experiments for characterizing biomacromolecular dynamics, and the use of NMR for screening
- With BMRB and BMRBj both facing resource constraints, BMRBe could contribute both curation of European entries and development of software for enhancing relaxation and screening data deposition
- Discussions have taken place with Harald Schwalbe (Frankfurt) and Wim Vranken (Brussels). A white paper is being drafted.

# Persistence and Provenance

The NMRprime data repository (ETH Zürich) has copied BMRB entries verbatim, creating JSON metadata files that list ETH staff as “owner”. This doesn’t violate CC0, but it does create provenance and persistence issues. The data has been assigned a persistent identifier distinct from the BMRB entry, but the data appears to be identical. Given Swiss copyright laws, this claim of “ownership” could have real-world consequences.

```
JSON Raw Data Headers
Save Copy Collapse All Expand All Filter JSON
{
  "nmrprime_id": "79FRTFK",
  "name": "Structure of At3g01050, a ubiquitin-fold protei",
  "owner": {
    "name": "Henry Wetton",
    "username": "hwetton"
  },
  "authors_and_institutions": {
    "0": {
      "first_name": "Dmitriy A.",
      "last_name": "Vinarov",
      "institution": "University of Wisconsin-Madison",
      "country": "US"
    }
  }
}
```

# Joint Projects Update

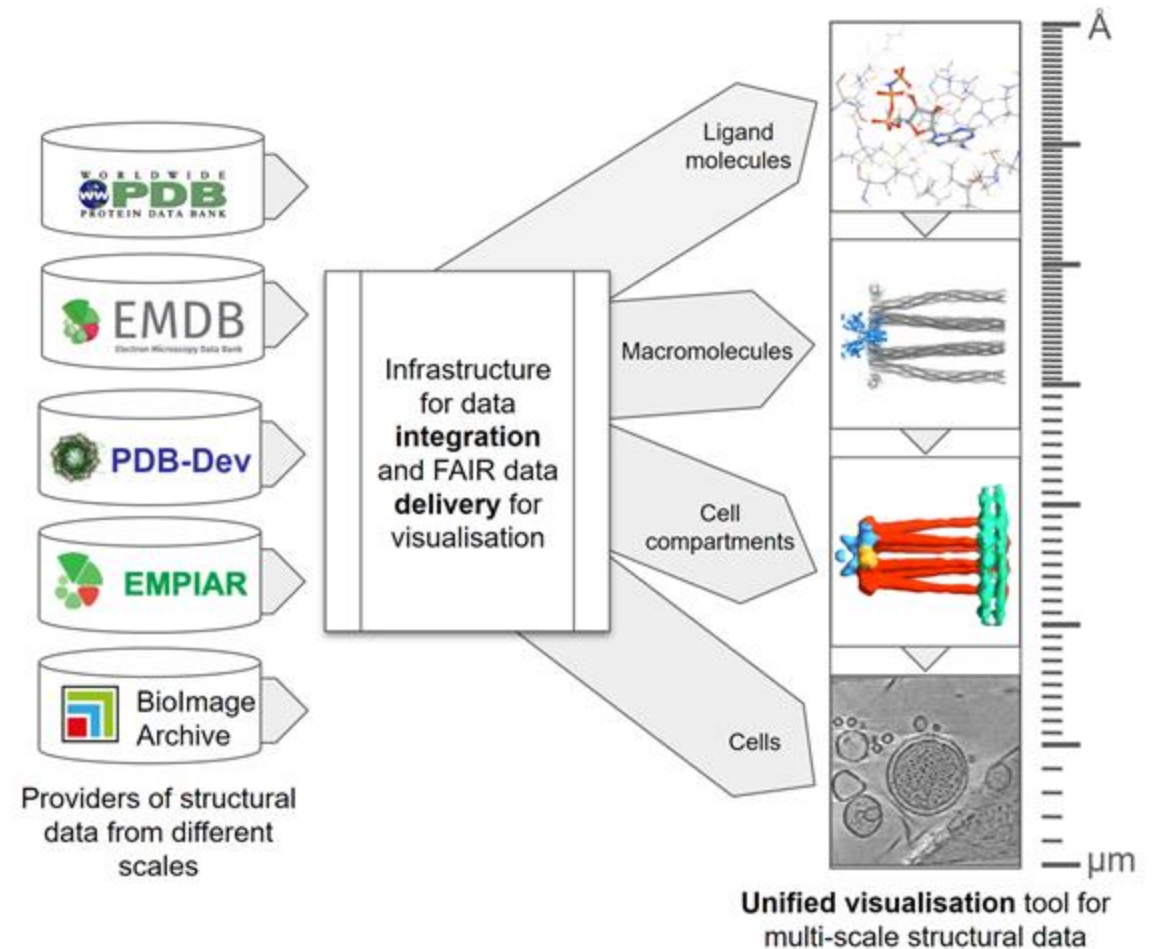


# Joint NSF/BBSRC projects

- Three year projects (two funded so far)
  - US funding provided by NSF; UK funding is provided by BBSRC
- Project 2 started in December 2021 and will end in November 2024 at RCSB; started in Jan 2023 and will end in December 2025 at PDBe
  - Provides resources for development of Mol\*, web-based components for displaying annotations and efficient data delivery mechanism
  - A standalone toolkit [MolViewSpec](#) (MVS) is published at [github](#)
  - Work from this project have been published in [Current Protocols](#)

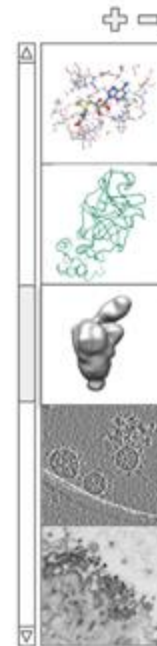
# Project 2: Visualization Infrastructure

- SA1 - To augment Mol-star for seamless operation across length scales ranging from atoms to cells with multiscale structure capability by
  - a. Extending existing data standards to combine atomic coordinates with multiscale structure information from integrative/hybrid methods and in-cell molecular details from advanced electron or visible light/super-resolution imaging techniques;
  - b. Building tools that enable superposition/simultaneous interactive display of multiscale structures; and
  - c. Enabling comparisons and analyses of multiscale structures.

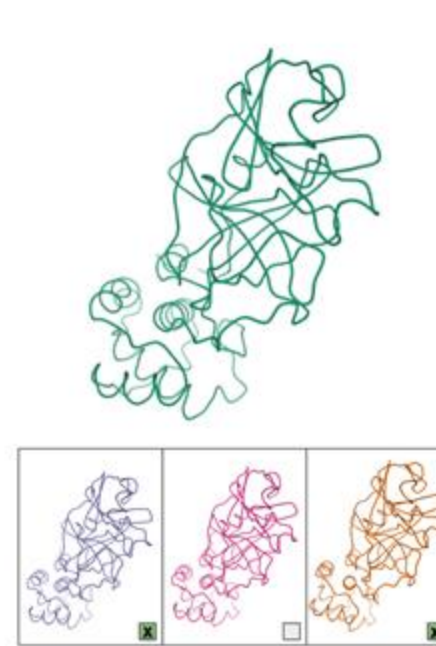


# Project 2: Visualization Infrastructure

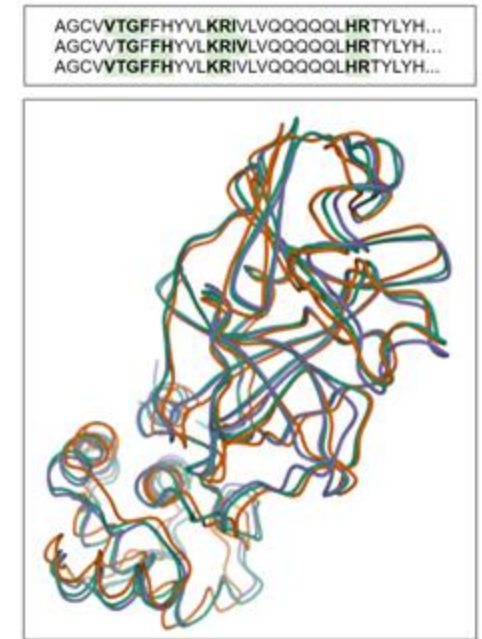
- SA2 - To extend 3D visualization beyond structure to include display of associated biological and functional annotations by
  - a. Developing a library of web-based components for displaying structural, biological and functional annotations;
  - b. Integrating Mol-star 3D display tools with complementary web-based components, such as those displaying 2D topology and 1D sequence data.



Multi-scale views



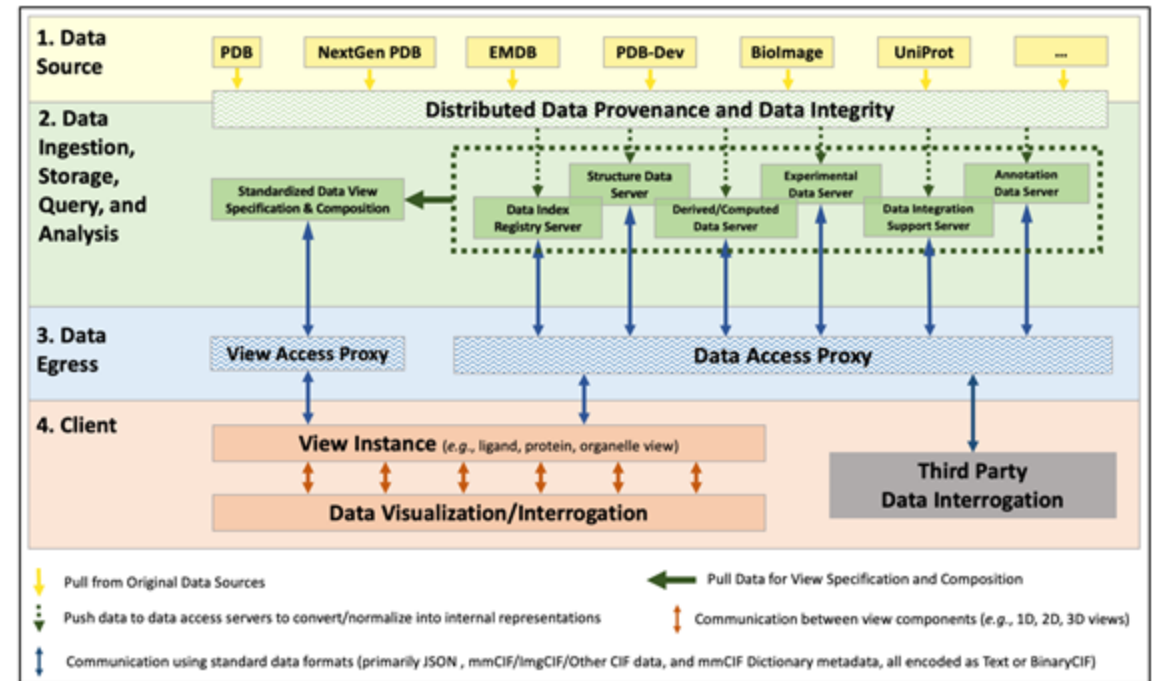
PDB structures from the same batch of experiments



Compare and contrast conformations and annotations

# Project 2: Visualization Infrastructure

- SA3 - To extend and strengthen existing IT infrastructure underpinning Mol-star by
  - a. Developing an agile data delivery system that supports high-speed interactive web-based visualization of 3D structure data; and
  - b. Enable comparison and analyses of multiscale structures across length scales ranging Å units to microns.



# Project 2: Progress 2022-2024

- SA1 - Augment Mol-star for seamless operation across length scales ranging from atoms to cells with multiscale structure capability
  - a. Data standards have been developed
    - Coarse-grained representations of multiscale structures are supported throughout Mol\*
    - Allow three formats for encoding the annotations: CIF, BCIF (BinaryCIF), and JSON
  - b. API to support multiscale data and multi-structure alignment from [MolViewSpec](#) (MVS) toolkit
- SA2 - Extend 3D visualization beyond structure to include display of associated biological and functional annotations
  - a. API to facilitate efficient mapping and data integration for chemical, polymer sequence, and 3D structure comparison.
  - b. Support annotations from external URI or within the source structure file
- SA3 - To extend and strengthen existing IT infrastructure underpinning Mol-star by developing an agile data delivery system that supports high-speed interactive web-based visualization of 3D structure data for comparison and analyses of multiscale structures across length scales ranging Å units to microns.
  - a. Views specification and infrastructure design developed
  - b. MVS library published at public [github](#)
- Results from this project have been published in part in [Current Protocols](#)