2022 wwPDB AC Meeting

Friday, October 14, 2022

08:30 - 11:30 US EDT (UTC+4)

Agenda

8:30 (US EDT)	Welcome and Introductions (10')	Denise Hien and Stephen K. Burley
8:40	Executive Session No. 1 (20')	AC Only-Breakout Room *via* Host
09:00	Discussion/Questions for AC (60')	All
10:00	Executive Session No. 2 (45')	AC Only-Breakout Room *via* Host
10:45	Feedback to wwPDB Leadership (30')	wwPDB PIs
11:15	Acknowledgements (10')	Stephen K. Burley
11:25	Group Photo/Meeting Close (5')	All, Photo *via* Host

Rutgers University Welcome

Denise Hien, PhD, ABPP

Vice Provost for Research, Chancellor-Provost's Office, Rutgers-New Brunswick
Center Director and Helen E. Chaney Endowed Chair in Alcohol Studies
Professor, Graduate School of Applied and Professional Psychology

Introductions

Stephen K. Burley, RCSB PDB

Executive Session No. 1

Host will add Advisors into Zoom Breakout Room

Please rejoin us at the end of your discussion by selecting
Leave the breakroom (lower righthand corner)

Discussion/Questions for AC

Questions for Advisors Overview I

1. Does the Advisory Committee have any questions or concerns regarding the performance of the organization since the 2021 AC Meeting (AC report arrived shortly after slide deck was drafted)? (SKB)

2. Does the Advisory Committee have any questions/concerns regarding individual wwPDB member 2021 AC reports? (SKB, SV, GK, JCH, AP)

3. Does the the Advisory Committee have any questions/concerns regarding training of PDBc Biocurators? (GK, WX)

4. Does the Advisory Committee concur with wwPDB plans to collaborate with structure prediction/docking challenges (*e.g*., CASP, CAMEO, CELPP)? (SV)

5. Does the Advisory Committee concur with wwPDB plans for managing grouped depositions as Investigations? (SKB)

6. Does the Advisory Committee have feedback regarding the ongoing campaign for renewal of BMRB funding? (JCH)

Questions for Advisors Overview II

7.   Does the Advisory Committee have any questions or concerns regarding deprecation
        of FTP for distributing PDB archive content? (SKB)

8.   Next Advisory Committee Meeting Hosted by PDBe: format/scheduling? (SV)

Advisor Question #1

Does the Advisory Committee have any questions or concerns regarding the performance of the organization since the 2021 AC Meeting (AC report arrived shortly after slide deck was drafted)?

Advisor Question #2

Does the Advisory Committee have any questions/concerns about individual wwPDB member 2022 AC reports?

Advisor Question #3

Does the the Advisory Committee have any questions/concerns regarding training of PDBc Biocurators?

Advisor Question #4

Does the Advisory Committee concur with wwPDB plans to collaborate with structure prediction/docking challenges (e.g., CASP, CAPRI)?

Advisor Question #5

Does the Advisory Committee concur with wwPDB plans for managing grouped depositions as Investigations?

Advisor Question #6

Does the Advisory Committee have feedback regarding the ongoing campaign for renewal of BMRB funding?

Advisor Question #7

Does the Advisory Committee have any questions or concerns regarding deprecation of FTP protocol for downloading PDB archive content?

Advisor Question #8

2023 wwPDB AC will be hosted by PDBe

Would the Advisory Committee prefer a face-to-face meeting in Hinxton, UK (pandemic permitting)?

Which date would the Advisory Committee prefer?
*   Friday October 13th 2023*
*   Friday October 20th 2023*

Which format would the Advisory Committee prefer?
*   Virtual-appears to be the Advisor Preference*
*   Hybrid*

Executive Session No. 2

Host will add Advisors into Zoom Breakout Room

Please rejoin us at the end of your discussion by selecting
Leave the breakroom (lower righthand corner)

Feedback to wwPDB Leadership

Acknowledgements

Stephen K. Burley, RCSB PDB

Group Photo

Meeting Close

Thank you

Poliovirus Neutralization (Artist: David S. Goodsell)

Question Briefing Slides

Advisor Question #3

Does the the Advisory Committee have any questions/concerns regarding training of PDBc Biocurators? Reference: Appendix #3.

wwPDB is Training PDB China Biocurators:

- PDBc agreement to the terms of the extant wwPDB Charter became effective, which enabled PDBc biocurators to access OneDep via the PDBc training instance physically located at Osaka (PDBcOneDep@Osaka).

  wwPDB Plan: wwPDB Core Members provided two types of formal training; (i) remote on-line orientation and policy training provided by RCSB PDB and PDBe (done), and (ii) on-site hands-on biocuration training in Osaka provided by PDBj (on-going until November).

- PDBc biocuration performance will be assessed during an initial six month evaluation period that commences after completion of on-site training including the production processing assigned step-by-step by PDBj lead annotator.

  wwPDB Plan: wwPDB Core Members will determine whether or not PDBc biocuration performance meets the global standard, defined in section 4 of the wwPDB Charter, and PDBc should be allowed to install and operate a OneDep system locally in Shanghai. During remote processing using PDBcOneDep@Osaka, depositions for PDBc biocurators will be assigned by the lead PDBj biocurator. Formal geographic direction of depositions to the PDBc server will only begin once the PDBcOneDep@Shanghai is in production.

Advisor Question #4

Does the Advisory Committee concur with wwPDB plans to collaborate with structure prediction/docking challenges (e.g., CASP, CAPRI)?

wwPDB wishes to work more closely with CASP, CAPRI, *etc*.:

- wwPDB partners have worked with community efforts such as CASP and CAPRI to encourage depositors to contribute target structures. Support provided has included amplifying the request for targets using wwPDB channels and providing a mechanism in OneDep for depositors to identify a submission as a CASP/CAPRI target.

  wwPDB Plan: Explore the possibility of a more active role in supporting these community efforts by

  - *Working with the CASP and CAPRI teams to identify requirements for target selection;*

  - *Exploring ways to extend OneDep functionality to identify targets (rather than depend on depositors to nominate their depositions);*

  - *Developing joint funding opportunities with the CASP and CAPRI teams to implement additional support; and*

  - *Using CASP and CAPRI related activities to establish the process to support community driven efforts.*

Advisor Question #5

Does the Advisory Committee concur with wwPDB plans for managing grouped depositions as Investigations?

wwPDB wishes to Improve the OneDep Depositor Experience and Group Related Structures together as "Investigations:"

- Multiple related structures have been reported in one publication (*e.g.*, fragment screening study, PanDDA method, *etc.*).

  wwPDB Plan: Enable single deposition of grouped structures as an Investigation defined by depositors.

- Refinement software packages now support PDBx/mmCIF formatted files for PDB depositions (*e.g.*, Phenix, CCP4, Global Phasing, CCPEM, *etc.*).

  wwPDB Plan: Work with structure determination software developers to further automate file generation for investigations to support parallel deposition of multiple related structures

Advisor Question #6

Does the Advisory Committee have feedback regarding the ongoing campaign for renewal of BMRB renewal?

wwPDB is Responding to a BMRB Funding Renewal Challenges:

- BMRB re-submitted a U24 proposal on October 3, 2022

- More than 100 letters were submitted in support of the proposal, including from the wwPDB PIs, two Nobel Prize winners, editors of majors journals, scientists and executives from biotechnology companies, early-career researchers, trainees, and students.

- wwPDB is collaborating with BMRB on an NSF Research Collaboration Network funding proposal to develop community standards for statistically robust structure validation.

"NIGMS does not view NMR as significant."

NIH/NIGMS Program Officer, 4/22/22

NIH funding for BMRB is set to expire 3/31/23. If you are willing to write a support letter, visit https://bmrb.io/bmrb/support.shtml

Advisor Question #7

Does the Advisory Committee have any questions or concerns regarding deprecation of FTP protocol for downloading PDB archive content?

wwPDB is Responding to Changes in Internet Browser Technology:

- The FTP protocol has been losing its popularity and the support by main web browsers (Chrome and Firefox)
- HTTP/S is recommended for the speed, statelessness, security, and is well-supported by a broad array of clients.

wwPDB plans to deprecate FTP protocol effective 11/1/2024, following multiple warning announcements (beginning 11/1/2022).

Advisor Question #8

Virtual wwPDB Advisory Committee Meeting Game Plan

Wednesday: *wwPDB PIs Arrive at EBI*

Thursday: *Virtual Pre-meeting with AC Chair/Co-Chair (half-day)*

Senior Biocurators Arrive at EBI

Friday: ***Virtual AC Meeting (half-day)***

Saturday: *wwPDB PI and Senior Biocurator Team Building*

Sunday: *wwPDB PI and Senior Biocurator Face-to-Face Meetings*

Monday: *wwPDB PI and Senior Biocurator Face-to-Face Meetings*

Tuesday: *Visitors Departing EBI*

Potential Impact of Computed Structure Models (CSMs) on wwPDB

Potential issues pertaining to Computed Structure Models or CSMs to consider include
- Inappropriate use of CSM information
  - Inclusion of atomic coordinates for substantial polypeptide chain segments undetectable by experimental methods (*i.e.*, not simply short surface loops lacking experimental electron density or 3DEM density information)
- Capturing accurate information relating to the provenance of CSMs
  - Distinguishing models from the AlphaFold DB *vs*. ModelArchive *vs*. models generated by someone else (*e.g.*, Depositors) using AlphaFold2, RoseTTAFold2, *etc*.
- Validation
  - "Under-determined 3DEM structures" consisting of a 3DEM density map plus CSM(s) that have not been fitted into the map
  - Detection of fraudulent PDB depositions (*i.e.*, CSMs being represented as experimentally-determined structures)
- Other issues?

Pre-Meeting Review Slides

Table of Contents

- Funding

- Governance

- Outreach

- OneDep

- PDB Archive

- EMDB Archive

- BMRB Archive

- Joint Projects

Funding Update

Stephen K. Burley, RCSB PDB

wwPDB Full Member Core Funding

- RCSB PDB: Joint NSF/NIH/DOE funding (2019-2023)
- BMRB: NIH NIGMS funding (2019-2023)
  - Inadequate budget: still need to find additional support
  - UConn funding 25% administrative assistant and 25% project manager positions
  - NIH U24 submitted but not funded. Re-submitted 10/2022.
- PDBe: EMBL-EBI, Wellcome Trust (2021-2025)
- PDBj: NBDC-JST and AMED funding (2022-2027)
  - Additional budget from S. Korea (decision pending due to COVID-19)
- EMDB: EMBL-EBI, Wellcome Trust (2019-2024)
- RCSB PDB/PDBe: Joint NSF/BBCRC NextGen Archive funding (2020-2022)
- RCSB PDB/PDBe: Joint NSF/BBCRC Mol* Visualization funding (2021-2024)

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 2020-2022), this basic yearly grant is CNY 57,130,000 (~USD $9,021,000/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

Stephen K. Burley, RCSB PDB

wwPDB Charter Updated

- PDBc admitted as the first Associate Member!

- Preparing to admit PDBi as an Associate Member

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 China Update (PI/co-PI: Profs. Wenqing Xu, James Z. Liu, Quan Wang)

- NFPS has recruited/formed the Data-in team (5 people) and the joint Data-out team (5 people, in collaboration with the ShanghaiTech University).

- The Advisory Committee for PDB China has been formed, with balanced scientific expertise, gender and regional distribution. The PDB-China AC member list has been approved by wwPDB AC.

- The National Facility for Protein Science in Shanghai (NFPS), the primary host of PDB China, has upgraded the cooling and security systems for the PDBc computation-facility; and with support of the Shanghai municipal government, has purchased computational hardware dedicated to PDB China (with >$1.2m USD) and performed related infrastructure set up.

- PDB Japan obtained funding to purchase the computers for PDB China's first OneDep data-in system (to be sited in Osaka); equipment are used for on-site training.

- The first version of the PDB China data-out system, with primary search functions, has been developed, and is under test and further development.

- Remote data-in training for PDBc biocurators started in March. Biocurators from RCSB, PDBe and EMDB have given talks on data format, wwpdb policies and EM annotation. On-site data-in training for two PDBc biocurators started in late August at PDBj.

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

- First phase funding (Oct 2019-2022) received from NSM (India) was used for hiring PDBi staff, currently working at IISc, Bangalore & NII, New Delhi. NSM (India) has extended funding for salaries of PDBi staff till March 2024. Efforts are on to secure funding from DBT after 2024 for 5 years.

- PDBi PIs have made arrangements to utilize servers/storage of the Indian Biological Data Center (IBDC), a facility funded by DBT, India. IBDC has recently (Sept 2022) provided access to PDBi PIs for utilizing IBDC servers/storage for PDBi activity.

- Six PDBi staff are currently working on development of structural bioinformatics software for comparative analysis of Protein-DNA complexes in PDB, structure-based analysis of PPI network of M.Tb./Plasmodium. A prototype data-out server with very basic search interface has been set up by PDBi staff.

- PDBe (Dr. Velankar) has provided online remote training of four PDBi staff on annotation/curation of structural data. Three of them will visit EBI in person soon for further training on data curation (Funded by PDBe).

- National Advisory Committee for PDB-India will be constituted in consultation with NSM & DBT.

Outreach Update

Genji Kurisu, PDBj

2022 wwPDB Outreach Calendar

- BMRB Poster and Booth/Biophysical Society: February 19-23
- BMRB Workshop, Poster, and Booth/ICMRBS: August 21-25
- RCSB PDB-PDBe Poster/ISMB: July 10-14
- RCSB PDB Talk/ISMB: July 10-14
- RCSB PDB Poster, Talks, and Booth/ACA*: July 29-August 2
- RCSB PDB Workshop and Talks/ASCB: December 6
- EMDB Talk and Posters/CCPEM Symposium: May 3-5
- EMDB Posters/EMBL Imaging Centre Symposium: May 31
- EMDB Talk/South-West Electron Microscopy Plymouth: July 18-19
- EMDB Talk/CCPEM Icknield workshop: September 9
- EMDB Poster/Tomography Congress 2022: September 11-14
- PDBe Posters and Booth/ECM: August 23-27
- PDBj Talk/AsCA: October 30-November 2

* In-Person/Virtual Hybrid Event

OneDep Update

Stephen K. Burley, RCSB PDB

wwPDB Collaboration Resource November 2021-October 2022

wwPDB Partner	Software Development	Production Maintenance/ Project Management	Requirements Setting/ Testing	Core Archive Keeping*	Outreach	Biocuration/ Remediation	Total FTE Commitments
RCSB PDB	2.8**	1.3	0.35/0.35	2.0	0.3	6.3	13.4
PDBe	1.4**+0.8#	0.6	0.2/0.2	-	0.2	4	7.4
PDBj	0.9***	0.4	0.2/0.2	-	0.1	4.5	6.3
BMRB	0.95	-	0.1/0.1	0.5	-	0.2	1.85
EMDB	1.15	0.5	0.1/0.1	0.35	-	1.2	3.4
Total wwPDB	8.0	2.8	1.9	2.85	0.6	16.2	32.35

*RCSB PDB; EMDB; BMRB

**including additional resource from BBSRC/NSF joint grant, 1.0 FTE at PDBe and 1.3 FTE at RCSB PDB

***PDBj added 0.5 FTE for OneDep development from April 2022

#A new hire at PDBe working on PTM remediation project started in April 2022

OneDep 2021/2022 Progress *versus* Goals I

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

Delivered,
To be delivered,
Delayed

	Projects	Timeline			
		2021	2022		
		Q4	Q1	Q2	Q3
1. Validation	1.1 EDS software upgrade				
	1.2 Improve chemical shift validation code				
2. Backend Stabilization	2.1 Upgrade Django software				
	2.2 Improve OneDep software packaging and release				
	2.3 Centos upgrade				
	2.4 Enable PDB China biocuration training				
	2.5 Provide API for file access/exchange				
	2.6 Migrate legacy EMDB entries into OneDep				
3. Public facing (OneDep or wwPDB.ORG)	3.1 Enable author-annotated assembly				
	3.2 Improve collection of EM auxiliary data				
	3.3 Improve EM mandatory items and value boundary				
	3.4 Improve EM data processing via single dictionary				
	3.5 Present graphical analysis from NMR validation report				
4. Biocuration	4.1 Provide PDB China biocuration training				
	4.2 Improve error message from EM XML file generation				
	4.3 Support depositor-annotated assembly				
	4.4 Ongoing biocuration improvements				
5. Archive Improvements	5.1 Regenerate X-ray validation reports with new stats				
	5.2 Distribute assembly files in mmCIF format (PDB)				
	5.3 NMR restraint remediation (PDB & BMRB archives)				
	5.4 PTM remediation (PDB archive)				
	5.5 Distribute EMDB status files at EMDB archive				

45

OneDep 2021/2022 Progress *versus* Goals II

Ad hoc projects:
- Supported EBI server upgrade
- Improved support for extended PDBx/mmCIF structure factor files
- Made the types of validation reports (For Manuscript Review) more prominent
- Added map-model fitness using Q-score in the 3DEM validation reports
- PDB India training

Bold: re-forecasted to 2022-2023
** Not enough incoming cases

Components	Major Projects to be Completed	Primary resource
Validation	**Upgrade 3rd party EDS software**	**PDBe**
	Refactor NMR chemical shifts validation software	**BMRB**
Deposition	Extend/improve collection of auxiliary data	EMDB
	Streamline EM data model	EMDB
	Improve NMR and EM depositions: mandatory half-map deposition for 3DEM	EMDB
	Present graphical outliers of NMR validation at deposition**	BMRBj/BMRB
	Enable author-annotated assemblies	**PDBe**
Infrastructure	Support PDB China: OneDep instance installation and global Biocuration training	PDBj/RCSB PDB/PDBe
	Migrate legacy EMDB entries to OneDep system	**EMDB**
	Support new data representation for unmerged data for X-ray entries	RCSB PDB
Core Archives	Distribute mmCIF formatted assembly files to PDB archive	RCSB PDB
	Distribute EMDB status files to the EMDB archive	EMDB
	Provide mmCIF formatted validation reports to PDB archive	RCSB PDB
	Remediate NMR restraint files for BMRB and PDB archives	BMRBj/BMRB

wwPDB Biocurator Productivity

- 14573 depositions in 2021

- Better automation for biocurating incoming depositions

  - Based on existing structure annotations (where possible)

  - with > 100 ligand instances

  - Median processing time/entry reduced for large and/or complex entries (3-4 days to <2 days)

- Better automation for citation update based on DOI or PubMed ID

New Structures/wwPDB Biocurator
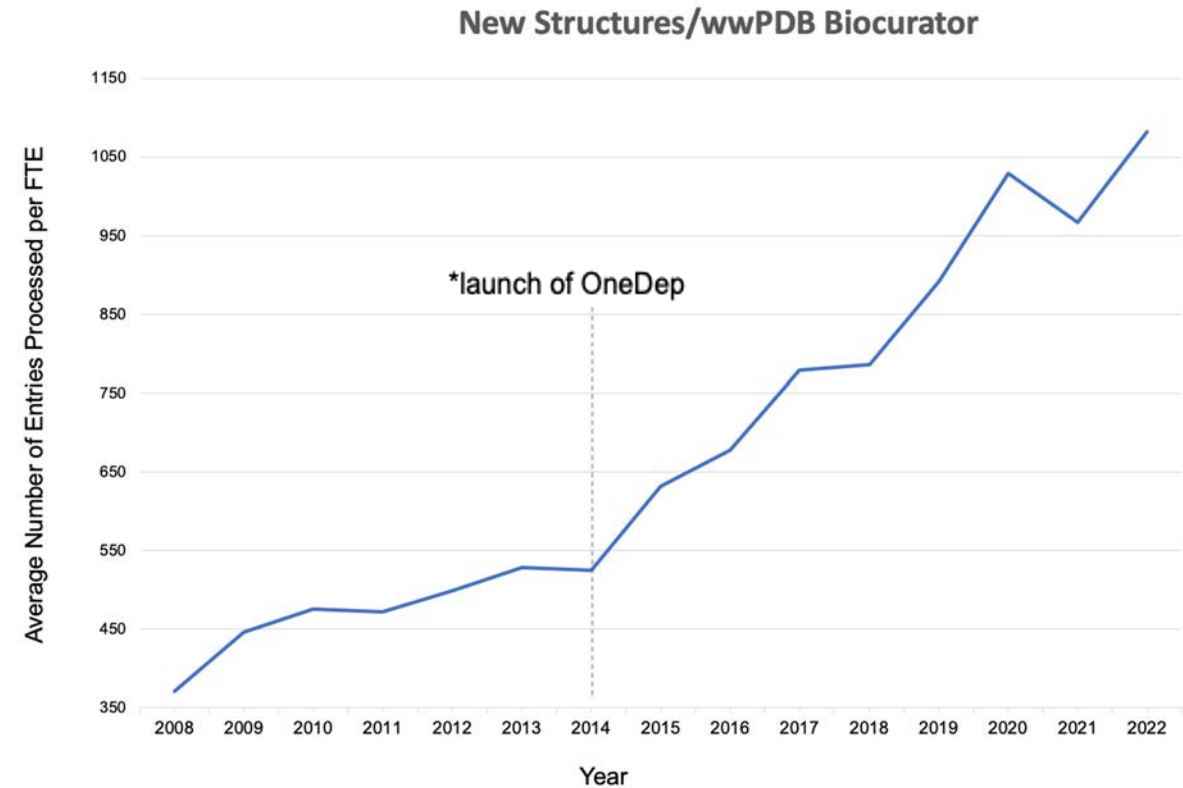
wwPDB Collaboration Resources November 2022-October 2023

wwPDB Partner	Software Development	Production Maintenance/ Project Management	Requirements Setting/ Testing	Core Archive Keeping*	Outreach	Biocuration/ Remediation	Total FTE Commitments
RCSB PDB	2.3**	1.3	0.35/0.35	2.0	0.3	6.2	12.8
PDBe	2.7**	0.6	0.2/0.2	-	0.2	5.5#	9.4
PDBj	1.1	0.5	0.2/0.2	-	0.1	4.6	6.7
BMRB	0.95	-	0.1/0/1	0.5	-	0.2	1.85
EMDB	2.1	0.65	0.25	0.15	-	0.9	4.05
Total wwPDB	9.15	3.05	1.95	2.65	0.6	17.4	34.8

*RCSB PDB; EMDB; BMRB

**including additional resource from BBSRC/NSF joint grant from PDBe and RCSB PDB and 0.8 FTE on PTM project from PDBe

#including 2.0 FTE PDB-India annotators in training

OneDep 2022/2023 Goal Setting I

Bold: re-forecasted from 2021-2022

Components	Major Projects to be Completed	Primary resource
Validation	**Upgrade 3rd party EDS software**	**PDBe**
	Refactor NMR chemical shifts validation software	**BMRB**
	Improve performance on validation calculations within OneDep	RCSB PDB
	Upgrade 3rd party MolProbity software	PDBj
	Enhance 3DEM map validation: False color images and symmetry analysis	EMDB
	Provide outlier detection and quality slider for 3DEM map-model fitness	EMDB w/Greg&Chenghua
Deposition	Improve composite map support for 3DEM	EMDB
	Support metadata upload in mmCIF for map-only 3DEM depositions	EMDB
	Improve EM metadata checking/validation	EMDB
	Enable author-annotated assembly	**PDBe**
	Enable geographic redirection to PDBc	PDBj
	Provide a Review page on the summary of author provided metadata at DepUI	PDBe
	Improve collection of starting models	PDBe/EMDB
	Improve file re-upload process using file access API	PDBe

OneDep 2022/2023 Goal Setting II

Bold: re-forecasted from 2020-2021

Components	Major Projects to be Completed	Primary resource
Infrastructure	Centos upgrade	RCSB PDB
	Support CCD ID extension	ALL sites
	Migrate EMDB legacy entries into OneDep	**EMDB**
	Session login via ORCiD	PDBe
	Upgrade OneDep Third party packages/libraries	EMDB/RCSB PDB/PDBe
	Increase efficiency on validation report generation at release time	RCSB PDB
Core Archives	Remediate NMR restraint files for BMRB and PDB archives	BMRBj
	Archival re-calculation of validation reports for PDB and EMDB archives	PDBe
	Remediate entries with non-crystal frame (PDB)	RCSB PDB
	Create Next Generation PDB Archive (NextGen)	RCSB PDB/PDBe
	Distribute mmCIF formatted header files to the EMDB archive	EMDB
	PTM remediation (PDB)	PDBe

OneDep 2022/2023 Roadmap

- will be further adjusted based on requirement setting
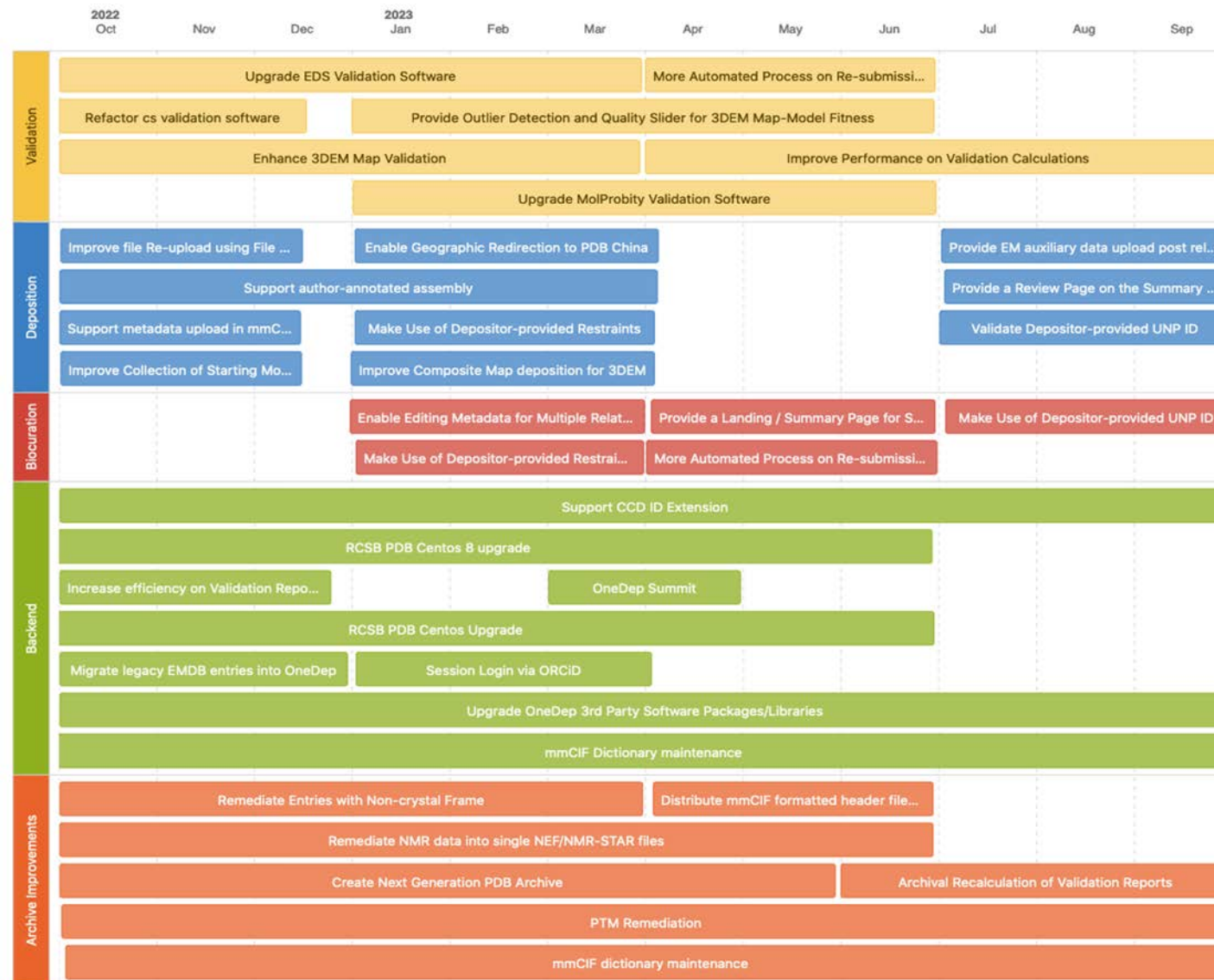
PDB Archive Update

Stephen K. Burley, RCSB PDB

Current PDB Archive Status

- Total Entries=194,820 (as of 9/02/2022)

- Core Archive Storage
  - OneDep Sessions: ~38.8 TB
  - ftp(legacy + versioned): ~1.3TB
  - EMDB ftp: ~6.3 TB

- CoreTrustSeal certification renewed through April 2024 (CoreTrustSeal.org)

- PDB archival data now being housed and delivered by Amazon Web Services (AWS) with no storage or egress fees
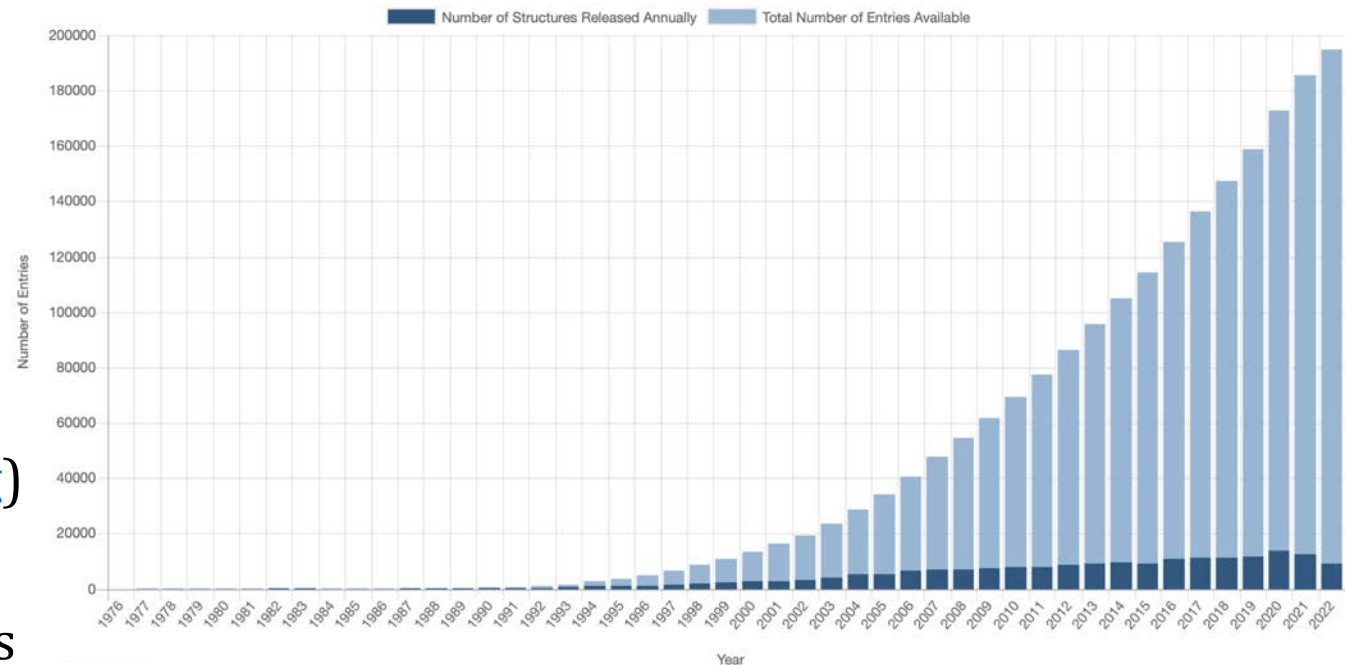
Anticipating 200,000 structures in early 2023!

PDB Archive Growth in 2021

- Year-end holdings 185,472
- 12,592 new entries released
  - ~10% increase *versus* 2019
  - ~7% growth in the archive
- Archival entries growing in both size and complexity
- Record 2,952 new 3DEM entries released
  - ~103% increase *versus* 2019!
  - ~23% of new 2021 entries!

Structures by Method

— MX — EM — NMR

Average Number of Residues per Structure
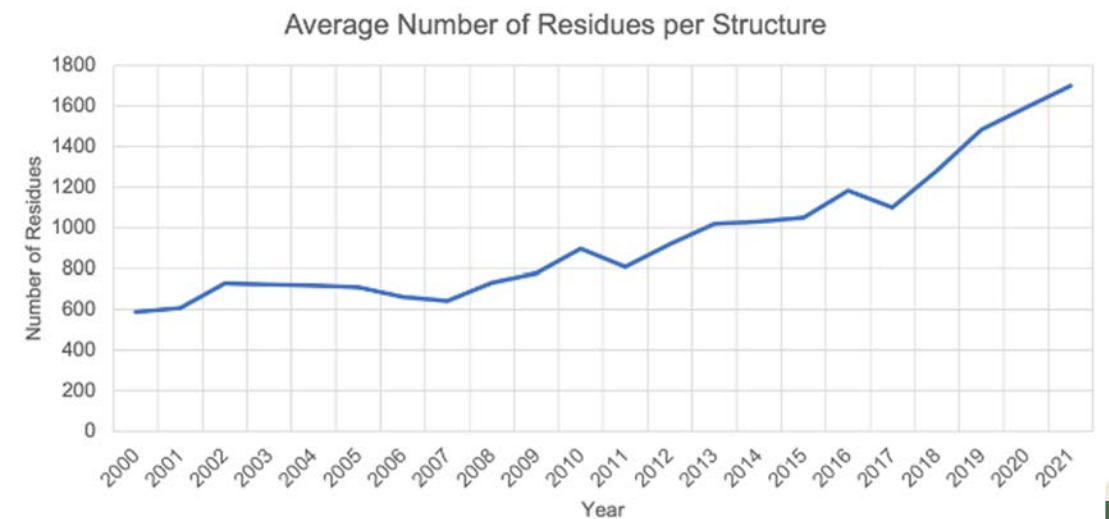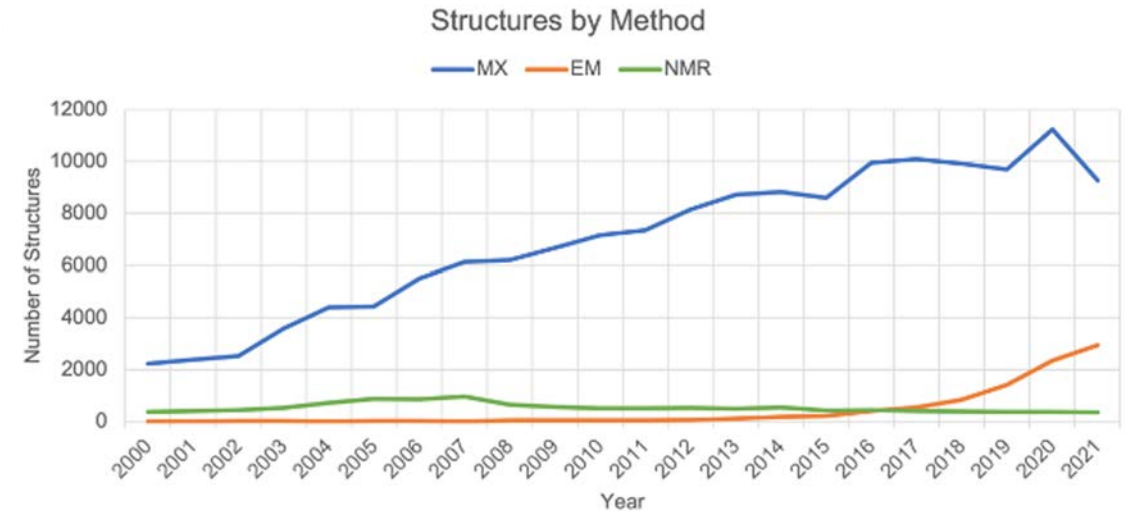
Global Biodata Coalition (GBC)

- wwPDB seeking formal recognition of PDB as a Global Core Biodata Resource
  - Expression of Interest submitted 04/13/2022
  - Full Application submitted 08/04/2022
  - Positive feedback received
  - Awaiting decision

- wwPDB plans to seek formal recognition for EMDB and BMRB in response to subsequent GBC calls

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)

- AWS can deliver PDB data faster than
  RCSB PDB, PDBe, or PDBj!

EMDB Archive Update

Ardan Patwardhan, EMDB

2021-2022 Achievements

- Improvements to 3DEM validation reports

- Improvements to 3DEM deposition

- Improvements to 3DEM annotation

- 20,000th EMDB entry released

- Youtube Channel Launched

Improvements to 3D EM Validation Reports

- Reports now display the full model assembly in the 3DEM map

- Atom inclusion and Q-score mapped to the coordinate model is now shown in the validation reports

- A summary table containing information on chain and full structure atom inclusion and Q-score is presented in the reports

Improvements to 3DEM Validation Reports

Q-score

- Provided for entries with maps that are below 1.25 Å in resolution

- The Q-score per residue is displayed on the atomic coordinate model

- The Q-score for the whole structure and per chain is presented in a summary table.
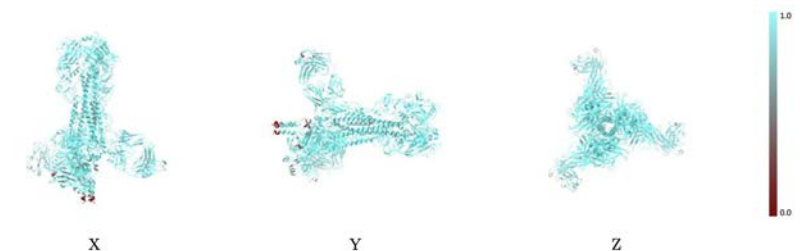
Atom Inclusion

- For entries with a map and model residue atom inclusion is displayed on the coordinate model is displayed

- The atom inclusion for the whole structure and per chain is presented in a summary table

9.2   Q-score mapped to coordinate model ⓘ

X   Y   Z
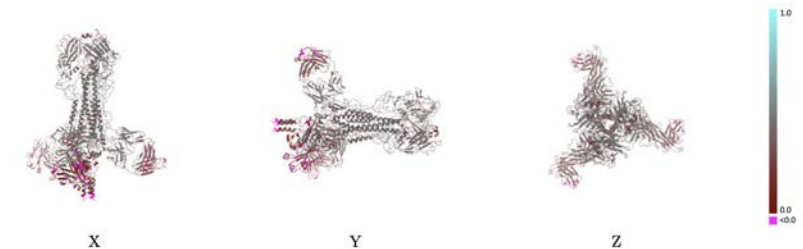
The images above show the model with each residue coloured according its Q-score. This shows their resolvability in the map with higher Q-score values reflecting better resolvability. Please note: Q-score is calculating the resolvability of atoms, and thus high values are only expected at resolutions at which atoms can be resolved. Low Q-score values may therefore be expected for many entries.

9.3   Atom inclusion mapped to coordinate model ⓘ

X   Y   Z

The image above show the model with each residue coloured according its atom inclusion. This shows to what extent they are inside the map at the recommended contour level (1.0).

Improvements to 3DEM Deposition

- Mandatory deposition of half-maps for relevant modalities

- Improved support for large files during file upload

- Improved metadata cross-checks during deposition

- Collection of ORCiDs for 3DEM entries

- Improved composite map support

- Better 3DEM-data-related integrity checks

- Tomography deposition is undergoing rounds of review and improvement

Improvements to 3DEM Annotation

- New metadata checks during annotation

- Molstar has been implemented to allow annotator review of 3DEM maps and models from within the OneDep annotation workflow

Awaiting Community Feedback

- Composite Map Deposition
  - First stage of implementation fully planned
  - Awaiting community feedback on the map naming conventions

- New Content Types
  - Planned and some implementation complete
  - Awaiting community feedback on names used for content types

Social Media

- Twitter
  - Group of EMDB and EMPIAR employees responsible for weekly tweets
    - 5600+ followers
    - 3+ tweets a week

- YouTube
  - Youtube channel hosting tutorials, talks and structure shorts launched June 2022
    - 1300+ views
    - 29 watch hours
    - 59 Subscribers

Over 20,000 EMDB Entries

- Reached in May 2022 after 20 years
- Twitter posts organised to celebrate the occasion
- Over 22,000 entries by early September 2022
- 4,488 new entries released in 2021
- 4,033 released in 2021 (until early September)

BRMB Archive Update

Jeffrey C. Hoch, BMRB

Developments since 2021 Meeting

- NMR Restraints validation software was successfully integrated into the OneDep system.

- OneDep system now accepts all experimental data (Chemical shifts, restraints and peak lists) in NMR-STAR v3.2 or NEF v1.2.

- The chemical shift validation code has been refactored and is now under testing.

- Working with PDBj on restraints remediation project.

- Deployed 1000-core CS-Rosetta server.

Developments since 2021 Meeting (cont.)

- Modernized the BMRB website (bmrb.io).

- BMRBdep upgraded to newest version of Angular.

- ReBoxitory Data Lake (on NMRbox Platform-as-a-Service, nmrbox.org) now contains monthly snapshots of BMRB, PDB, NCBI, and the full AlphaFoldDB archive.

- Contribution to NAR Database Issue submitted September 22 2022.

- Invitation from *Science Advances* to contribute a perspective on biomolecular NMR (and BMRB) in the 21st Century.

BMRB Core Archive Growth
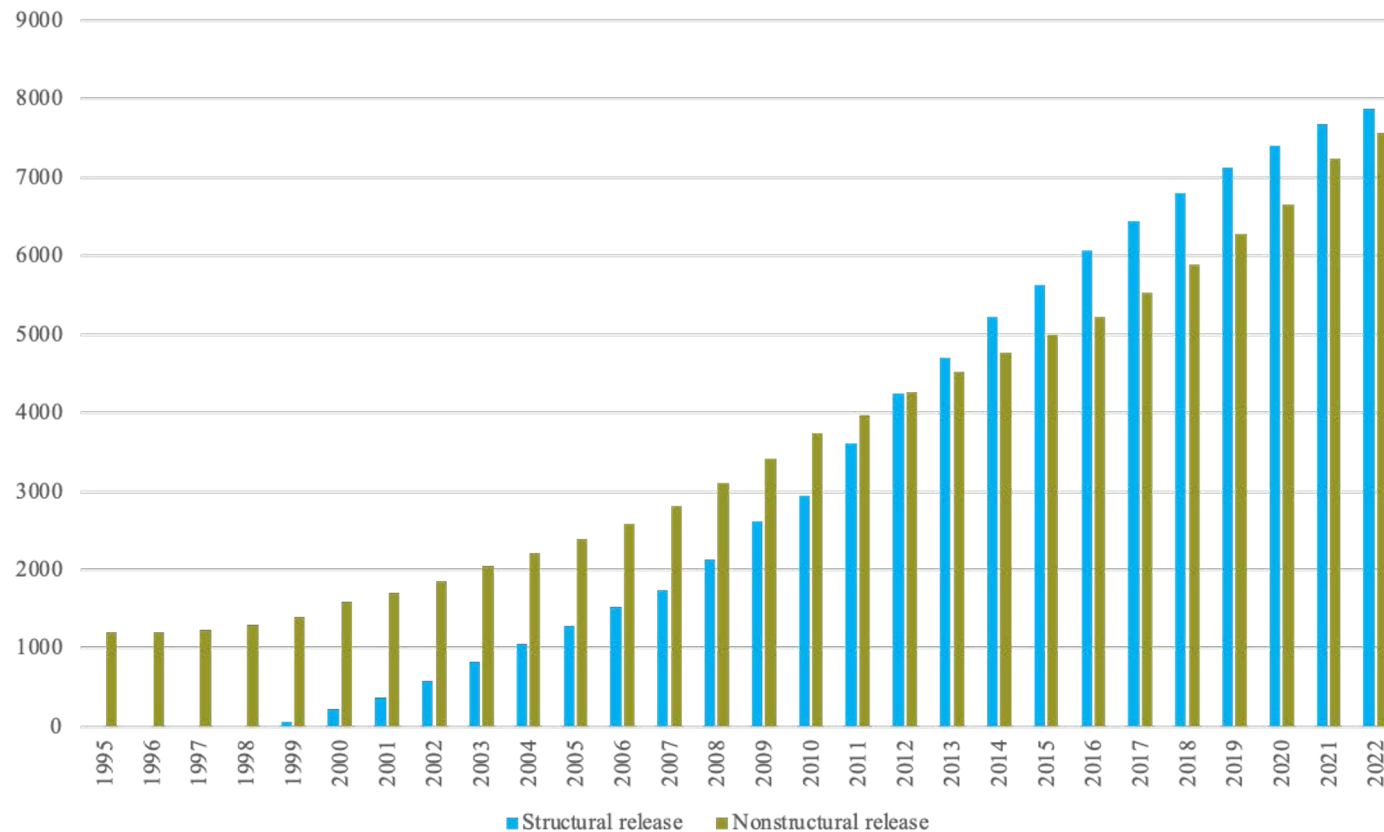
As of Sep 18, 2022
Total structural entries: 7871
Total Non-Structural entries: 7569

BMRB Core Archive Growth

Total Released Entries

Year	Yearly Totals					Accumulative Totals				
	Structural release	Nonstructural release	Original release	Total release	Withdrawn during year	First release	Total release	Structural release	Nonstructural release	Eventually withdrawn from this year
2012	632	293	925	1168	5	8510	10144	4240	4270	113
2013	468	248	716	1007	1	9226	11151	4708	4518	39
2014	522	242	764	1068	1	9990	12219	5230	4760	20
2015	402	234	636	696	0	10626	12915	5632	4994	75
2016	441	234	675	918	1	11301	13833	6073	5228	22
2017	379	307	686	839	3	11987	14672	6452	5535	70
2018	348	361	709	855	3	12696	15527	6800	5896	33
2019	320	378	698	968	4	13394	16495	7120	6274	37
2020	291	390	681	836	5	14075	17331	7411	6664	69
2021	276	577	853	1137	2	14928	18468	7687	7241	62
2022	184	328	512	985	0	15440	19453	7871	7569	23

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
2012	42,555,193	6,395,316	959,834	854,837	279,922	17.62 TB
2013	43,503,533	6,489,163	873,911	1,006,097	374,274	19.32 TB
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	43,806,131	12,121,460	816,086	1,204,987	452,012	11.92 TB

- BMRB has mirror sites in Italy and Japan, and PDBj-BMRB branch for deposition
- Updates to accounting methods resulted in slight changes to historical data from previous reports

Network for Advanced NMR: update

- First 1.1 GHz instrument in US is scheduled for delivery fall 2023.

- NSF awarded a planning grant to U Delaware (Hoch, co-PI) to develop a new $100M Midscale-RI2 proposal.

- Tentative target is five 1.2 GHz NMR spectrometers, distributed geographically across the US, and connected to the NAN cyberinfrastructure for automatically harvesting experimental data.

Joint Projects Update

Sameer Velankar, PDBe

Joint NSF/BBSRC projects

- Three year projects
  - The US funding is provided by NSF and the UK funding is provided by BBSRC

- Project 1 started in Oct 2020 at RCSB and Jan 2021 at PDBe
  - Provides additional resources for OneDep development
  - Coordination with OneDep team

- Project 2 was started in December 2021
  - PDBe will need to complete the administrative processes at BBSRC
  - Will provide resources for development of Mol*, web-based components for displaying annotations and efficient data delivery mechanism

Project 1: OneDep and NextGen PDB archive development

- SA1. To improve the fidelity and completeness of 3D structure data deposited into the PDB by
  - Harvesting data automatically from structure determination software packages; and
  - Streamlining the wwPDB data deposition, validation, and biocuration system.

- SA2. To improve "FAIRness" of PDB data for researchers, educators, and students by
  - Extending chemical metadata for small-molecule ligands (*e.g.*, bound cofactors and inhibitors);
  - Incorporating enhanced descriptions of macromolecular assemblies;
  - Grouping related PDB structures into investigations for more efficient, parallel data delivery; and
  - Creating a "Next Generation" PDB data repository with up-to-date metadata.

- SA3. To modernize the IT infrastructure to futureproof PDB data management and weekly PDB archive release to the public domain by
  - Developing new application programming interfaces (APIs) and microservices infrastructure; and
  - Updating existing mechanisms for synchronization of data and software across wwPDB data centers in the US, Europe, and Asia.

Project 1: Progress in 2021-22 (SA1)

- SA1 - Improve fidelity and completeness of 3D structure data deposited into the PDB
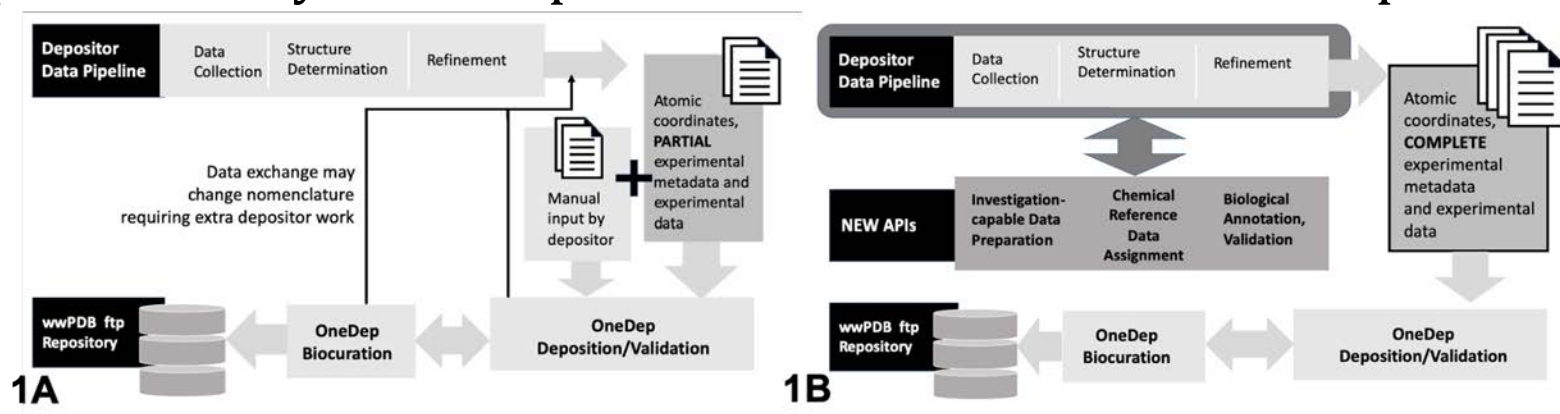
- Ongoing work towards enabling automated creation of and upload to deposition sessions

Authentication → API based creation of deposition session → API based upload files to deposition session

- Will make deposition of investigations possible *via* OneDep
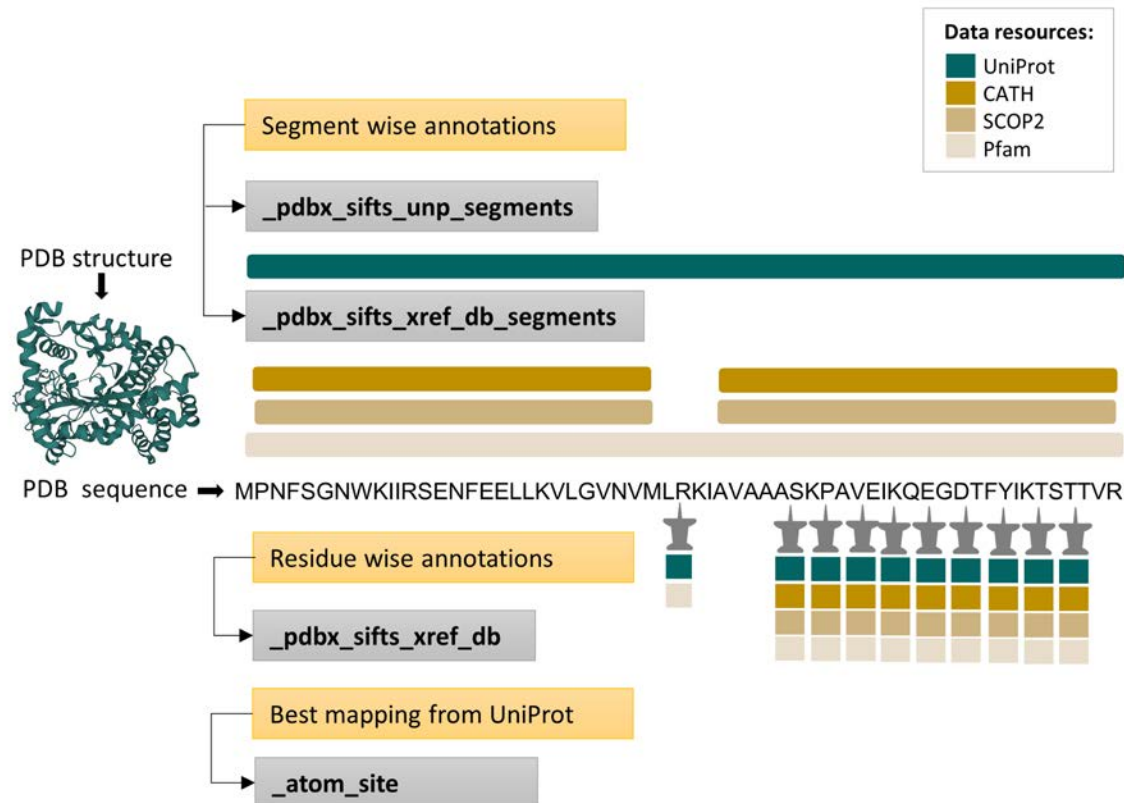
Project 1: Progress in 2021-22 (SA2)

- Background
  - Each wwPDB site adds additional data to support their data out activities
  - Aggregating of PDB entries is only possible after the PDB weekly release
  - A NextGen FTP will be introduced which will make this additional data and add aggregated data available to users

- First example of additional data
  - Cross-reference to other data resources and residue level mapping to UniProt (based on the PDBe SIFTS data)

- Added new categories to PDBx/mmCIF dictionary for SIFTS data
- Includes
  - Per residue mapping to external resource (i.e. UniProt, Pfam, CATH, SCOP)
  - UniProt numbering of residues (where appropriate)
- mmCIF files with additional SIFTS data are initially available from PDBe FTP area
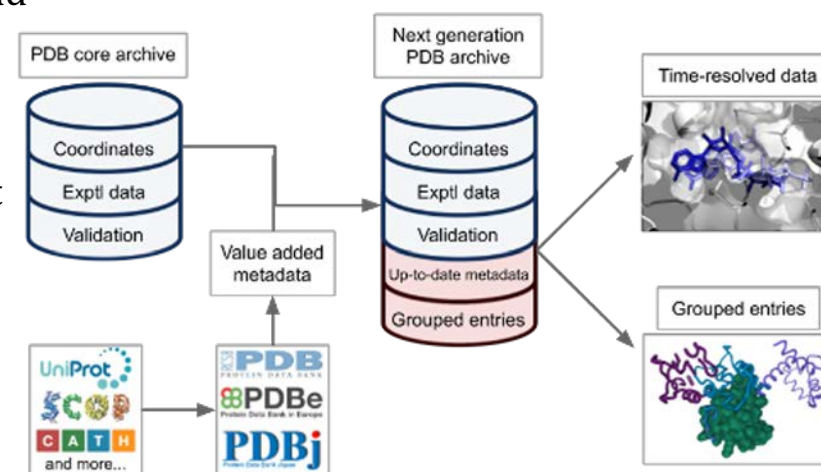
Project 1: Progress in 2021-22 (SA2)

- Three new PDBx/mmCIF categories added representing segment-wise and residue-wise SIFTS annotations mapping PDB residues to various data resources.

- _atom_site is also modified to indicate the best mapped UniProt sequence.
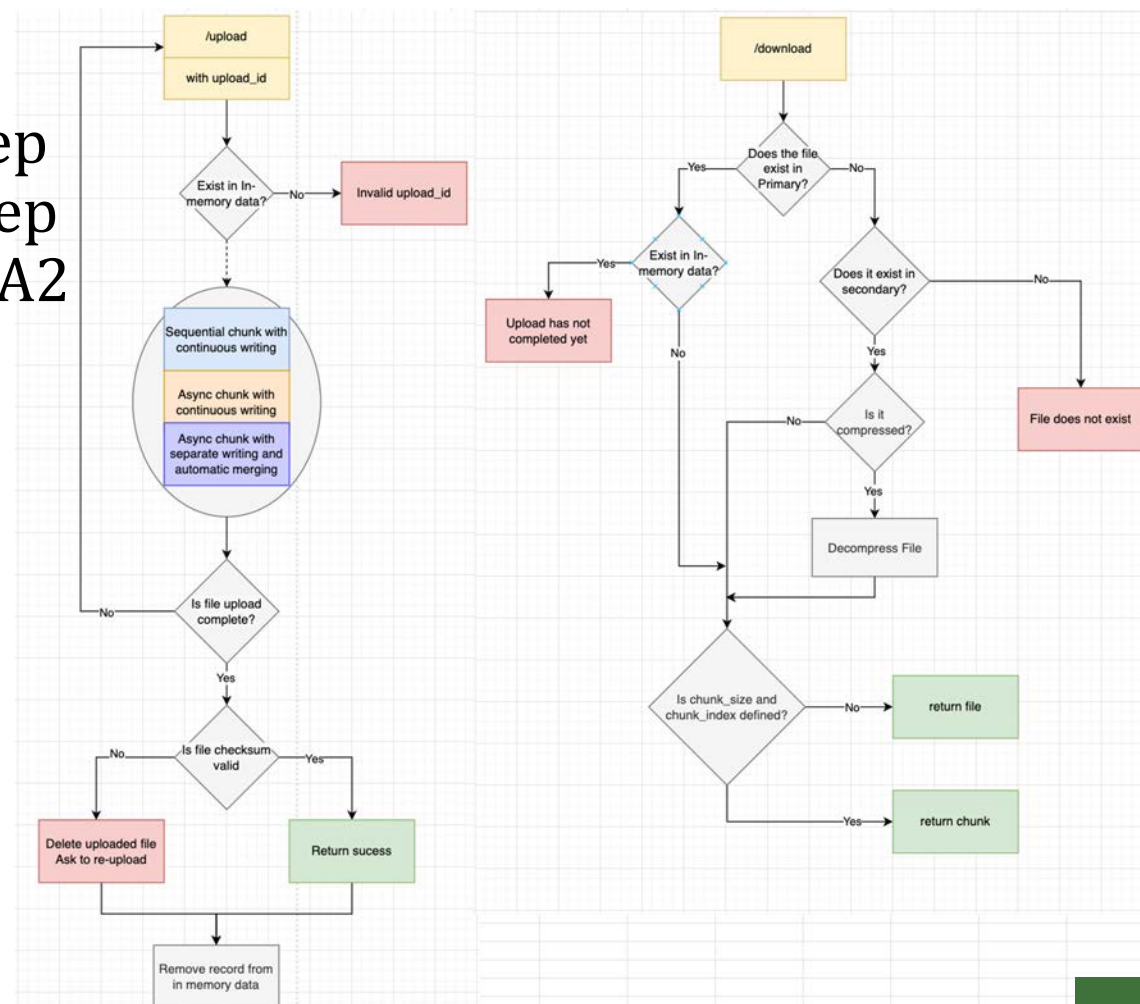
Project 1: Progress in 2021-22 (SA3)

- Installation procedure of wwPDB OneDep documented and simplified - from 133 manual steps to single installation script

- OneDep is comprised of the following components:
  - System configuration
  - Core utilities
  - Web application and software

- OneDep system configuration
  - On-going work on simplification
  - All OneDep apps changed to support a simpler configuration
  - Required to support additional wwPDB sites

- Core utilities published to PyPi Python repository

- Web applications and software
  - Docker containers created for all public facing modules
  - Docker containers automatically built using CI/CD
  - To be deployed on Kubernetes

Project 1: Progress in 2021-22 (SA3)

- File handling API to provide
    - Single access point for data in OneDep
    - Data exchange mechanism for OneDep
    - Groundwork for data exchange for SA2
    - Resumable upload functionality
    - Supporting asynchronous transfers
    - Handle file compression

- Would be rolled out initially on the deposition side.

Project 1: Progress in 2021-22 (SA3)

- OneDep workflow engine
  - Currently bespoke software which is tied to individual servers
  - Investigated open source replacements which are suitable for containerisation
  - Initial implementation developed to
    - Add resilience to regularly scheduled jobs (crons)
    - Monitor memory usage for OneDep processes
  - Will initially prototype deposition workflow and assess before making a decision about using the same approach for all of OneDep

# 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.
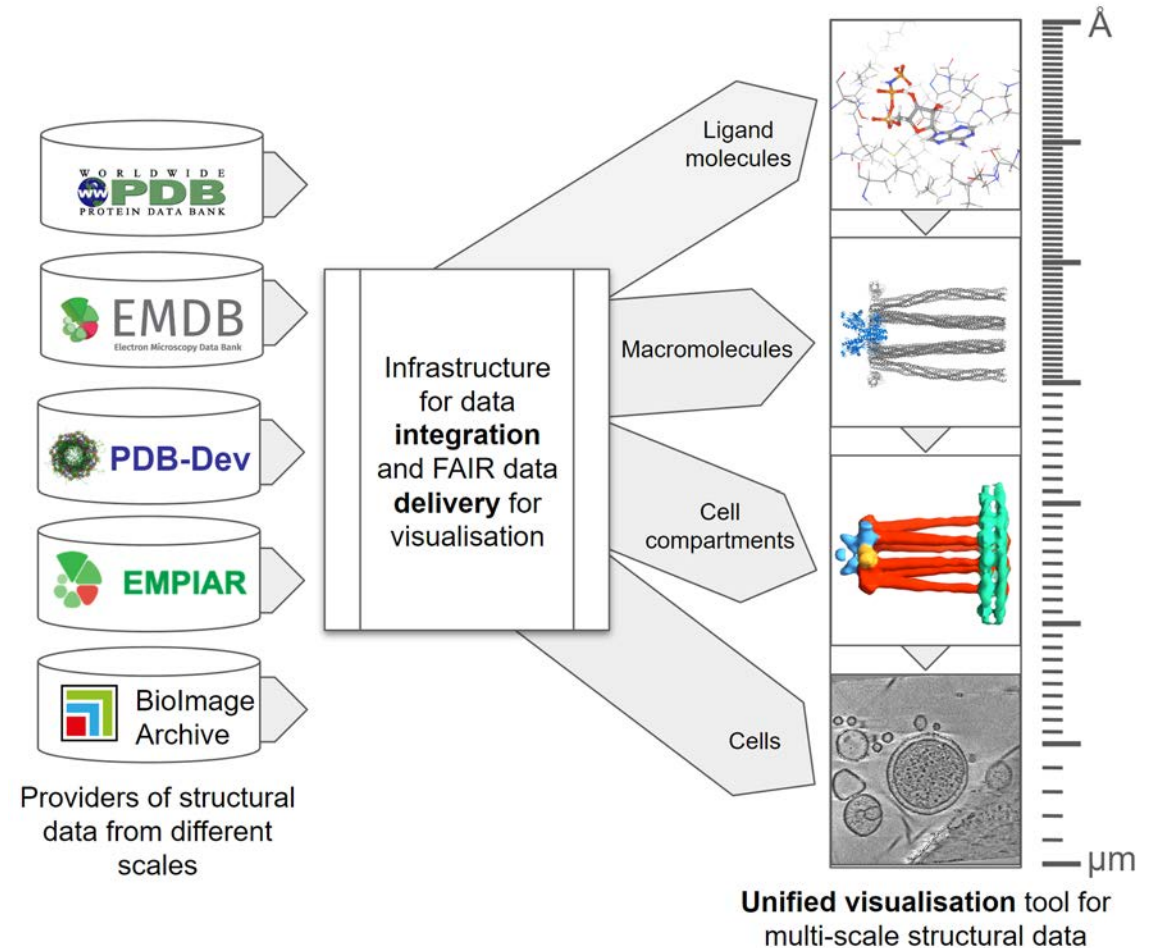
Project 2: Visualization Infrastructure

- 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.
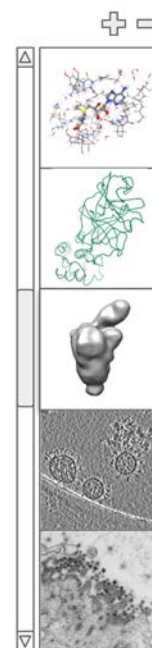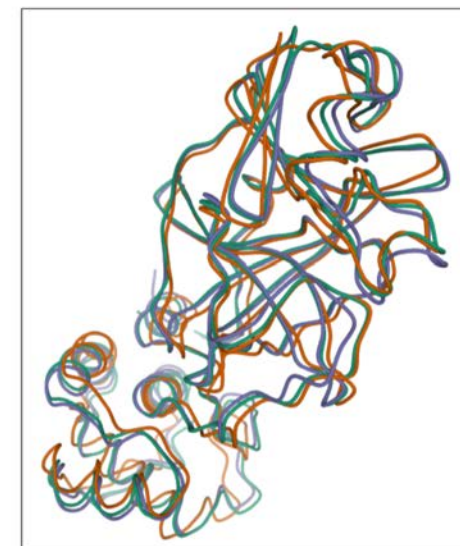

Multi-scale views

**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.
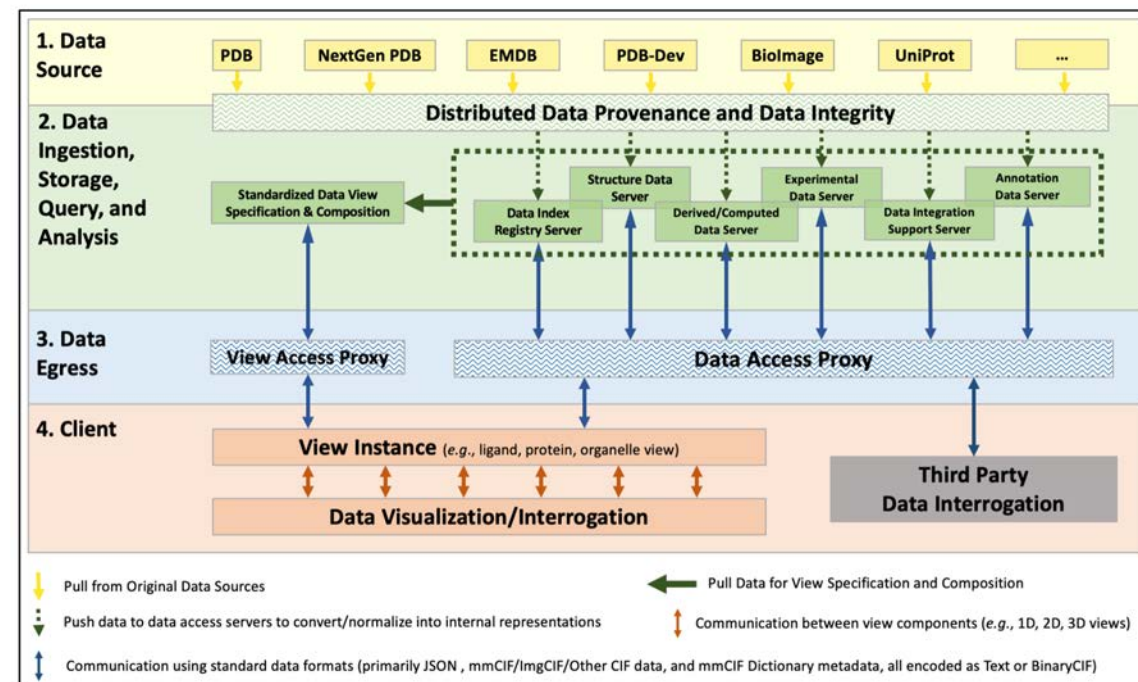
Project 2: Major Year One Activities

- SA1 - Augment Mol-star for seamless operation across length scales ranging from atoms to cells with multiscale structure capability

  a. Data standards to be developed

    - Current Mol* uses BinaryCIF to transmit structure data, volume/density data is compressed using the same technology. Coarse-grained representations of multiscale structures are supported throughout Mol*.

  b. API to support multiscale data and multi-structure alignment

- 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. 2D topology viewer for proteins and RNA is in place at PDBe. Investigating its inclusion into the project.

- 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 are being developed