Introductions and Overview of the wwPDB

Sameer Velankar

WORLDWIDE PDB
PROTEIN DATA BANK

wwpdb.org

wwPDB Vision Statement

Our vision is to:

Sustain freely accessible, interoperating Core Archives of structure data and metadata for biological macromolecules as an enduring public good to promote basic and applied research and education across the sciences.

wwPDB Mission Statement

- Manage the wwPDB Core Archives as a public good according to the **FAIR** Principles.

- Provide expert deposition, validation, biocuration and remediation services at no charge to Data Depositors worldwide.

- Ensure universal open access to public domain structural biology data with no limitations on usage.

- Develop and promote community-endorsed data standards for archiving and exchange of global structural biology data.

Developments since 2017- Meeting I

- Successful renewal for RCSB PDB 2019-2023.

- Very positive peer review of BMRB grant for 2019-2024.

- Plan for orderly leadership change at BMRB.

- Continued enhancement of the OneDep system for deposition/validation/biocuration of MX, NMR, and 3DEM structures.

- Continued growth in 3DEM structure depositions and engagement with the 3DEM community.

- Implementation of PDB archive versioning.
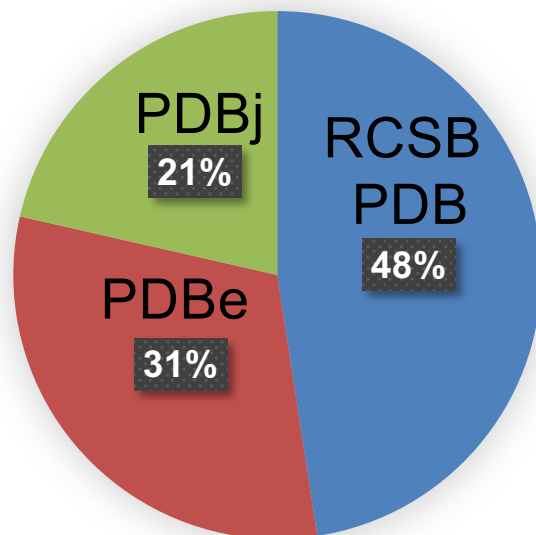
Developments since 2017- Meeting II

- BMRB designated as wwPDB Core Archive.

- EMDB to join as wwPDB Core Archive and as wwPDB Core Member.

- Ardan Patwardhan, EMDB Team leader, will represent EMDB.

- Continued depositions to PDB-Dev for archiving of other I/HM structures.
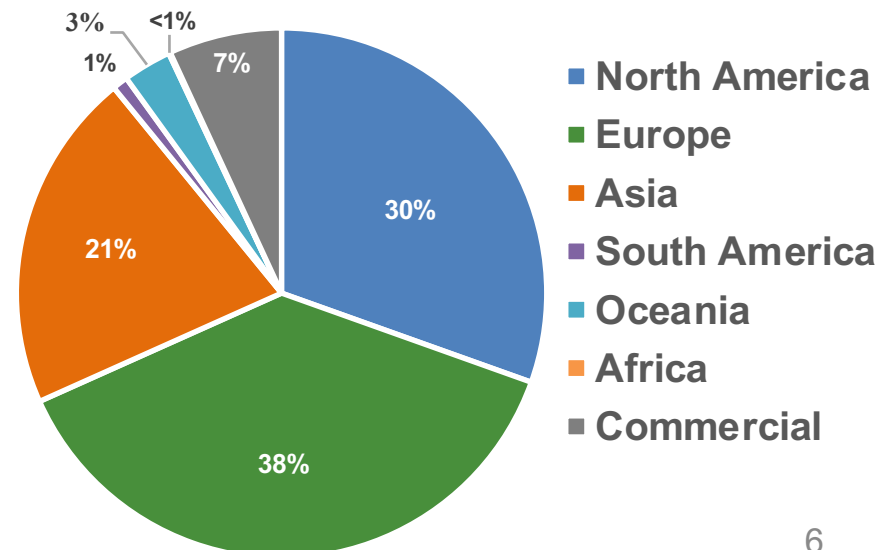
PDB Core Archive Depositions

- On track for ~12,500 depositions in 2018.
- 3DEM growth continuing in 2018.

Method	2017 Depositions	2016 Depositions
MX	11,889 (91.2%)	10583
NMR	460 (3.5%)	474
3DEM	658 (5.0%)	531
Other	44 (0.3%)	27

2017 Processing Sites

- PDBj 21%
- RCSB PDB 48%
- PDBe 31%

2017 Depositor Locations

- North America 30%
- Europe 38%
- Asia 21%
- South America 1%
- Oceania 3%
- Africa <1%
- Commercial 7%

PDB Core Archive Growth
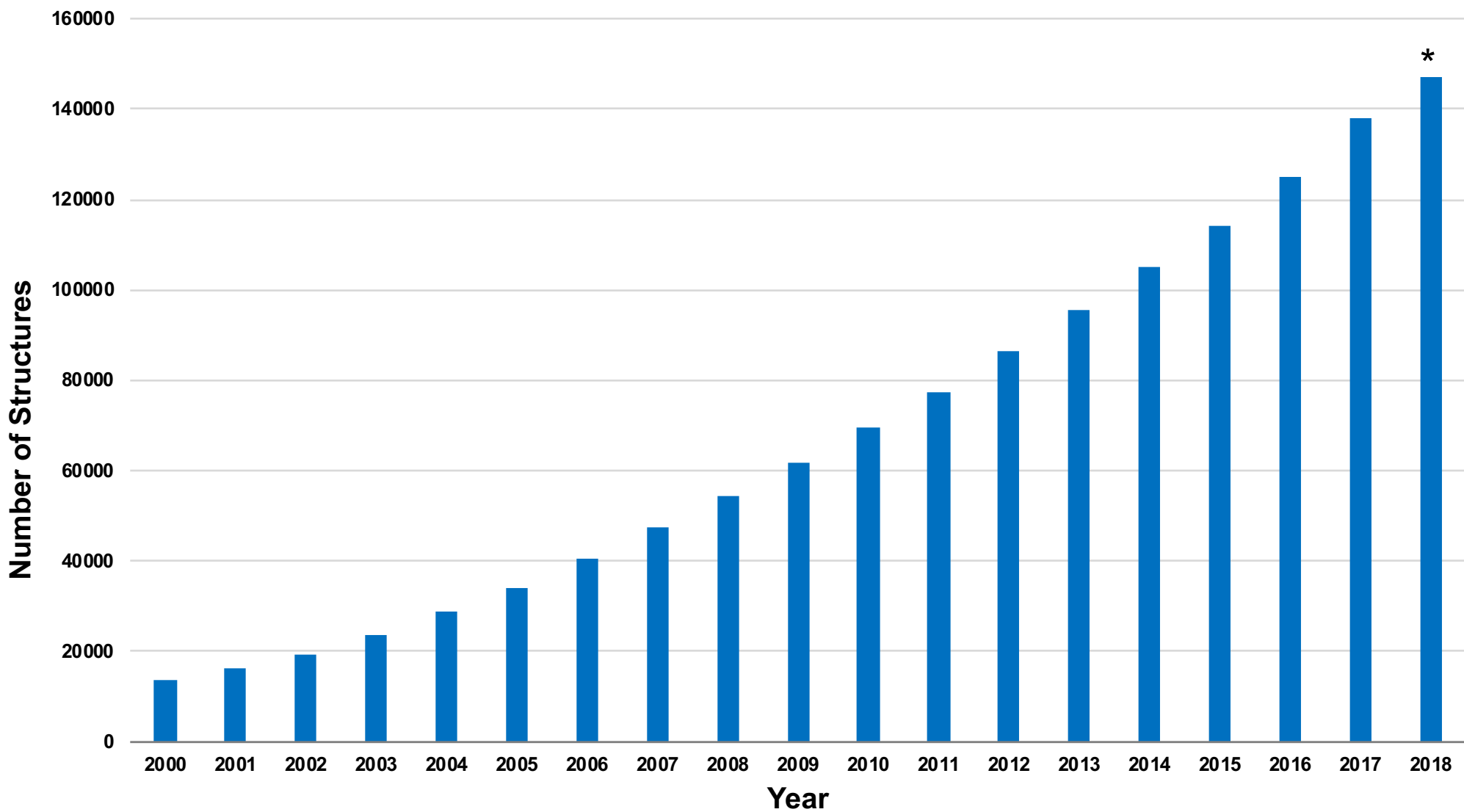
* As of 7 September 2018

PDB Core Archive Downloads

| Year	Total	Total FTP Archive	Total Website	RCSB PDB FTP Archive	RCSB PDB Website	PDBe FTP Archive	PDBe Website	PDBj FTP Archive	PDBj Website
2017	679,421,200	454,723,083	224,698,117	369,868,918	143,839,718	40,168,072	65,420,002	44,686,093	15,438,397
2016	591,876,087	366,677,897	225,198,190	293,648,366	161,208,456	30,274,284	44,432,830	42,755,247	19,556,904
2015	534,339,871	368,244,766	166,095,105	255,346,630	111,802,897	48,544,330	41,127,219	64,353,806	13,164,989
2014	512,227,251	339,193,721	173,033,530	237,168,615	110,115,316	52,362,370	48,031,414	49,662,736	14,886,800
2013	441,262,210	296,176,290	145,085,920	215,331,908	97,549,580	43,684,850	37,762,496	37,159,532	9,773,844
2012	376,944,070	255,837,735	121,106,335	213,510,347	90,438,501	21,601,103	23,982,801	20,726,285	6,685,033
2011	383,131,048	276,952,286	106,178,762	204,939,406	81,560,098	40,960,368	18,515,245	31,052,512	6,103,419
2010	294,326,976	213,180,966	81,146,010	159,248,214	64,569,658	34,383,219	14,017,349	19,549,533	2,559,003

More than 1.8 million/day!

N.B.: Some 2018 data lost due to GDPR.
Hope to be back on track for 2019→.

Geographic Origins of FTP downloads; 2012-2015

BMRB Core Archive Growth

- BMRB is on track for ~870 new entries released for 2018.

- Total released entries estimated to reach ~12,700 by the end of 2018.

BMRB TOTAL RELEASED ENTRIES
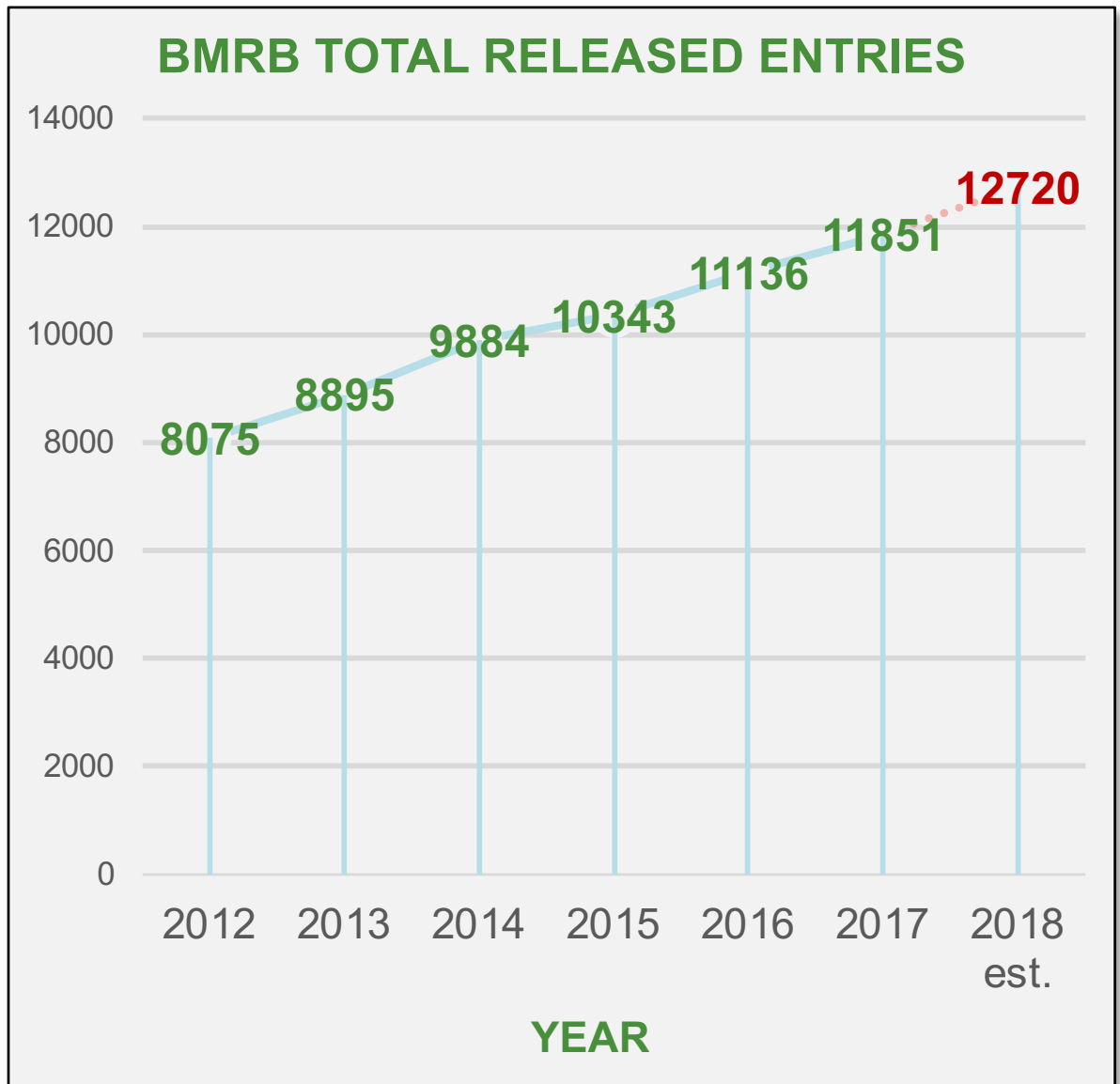
BMRB Core Archive Growth
Total Released Entries

Year	Total released	Yearly increase	Structures	Yearly increase	Non-structures	Yearly increase
2012	8075	814	4852	770	3223	44
2013	8895	820	5647	795	3248	25
2014	9884	989	6614	967	3270	22
2015	10343	459	7025	411	3318	48
2016	11136	793	7649	624	3487	169
2017	11851	715	8125	476	3726	239
2018	12430	579	8411	286	4019	293
2018 est.	12720	869	8554	429	4166	440

BMRB Core Archive Growth
Internet Servers Traffic (Website) – All Mirrors

Year	Server requests	Page requests	File requests	Distinct hosts served	Total data transferred
2012	12,703,408	3,290,436	1,730,857	310,074	8.56 TB
2013	10,521,902	2,176,005	1,508,510	342,291	4.85 TB
2014	11,314,520	2,193,844	1,561,558	383,052	6.99 TB
2015	12,869,845	2,348,759	1,991,915	444,873	12.6 TB
2016	15,986,881	2,766,144	2,496,663	445,679	12.0 TB
2017	42,071,167	10,541,522	6,633,980	460,550	13.1 TB
2018 (by 9/12)	40,014,144	10,303,509	3,949,570	344,560	7.61 TB

~150K/day Server and page requests
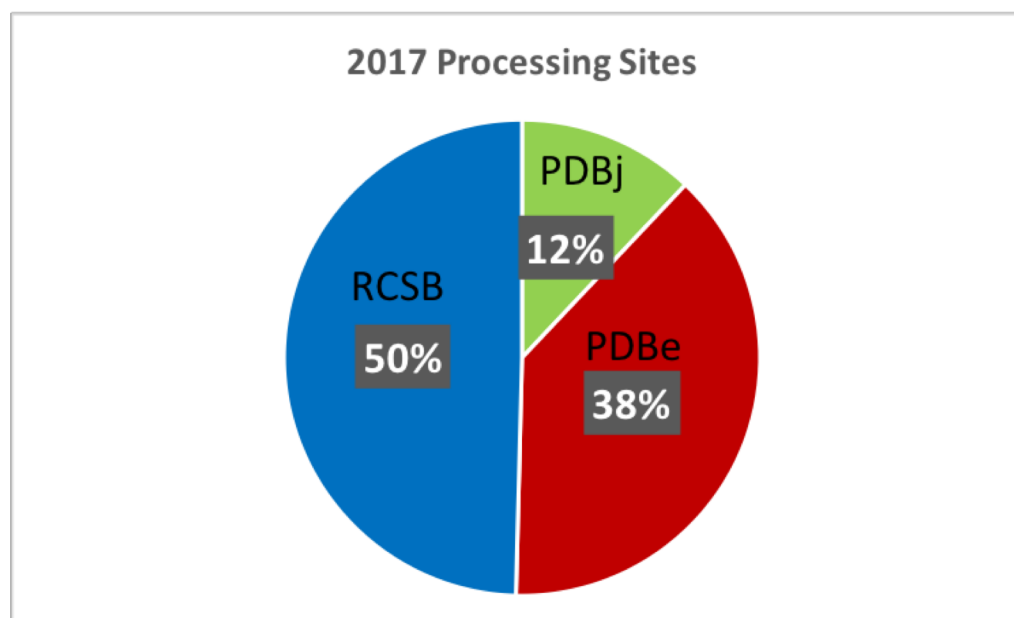
BMRB Core Archive Growth
Internet Servers Traffic (FTP servers) – All Mirrors

| Year	Server requests	Distinct files requested	Distinct hosts served	Total data transferred
2012	2,304,687	1,820,240	4,788	1.15 TB
2013	1,992,706	1,481,995	5,253	1.40 TB
2014	1,994,184	1,488,431	4,645	1.59 TB
2015	2,197,159	1,666,353	3,742	0.905 TB
2016	5,459,161	1,655,031	5,774	1.66 TB
2017	5,502,929	2,461,117	4,126	5.02 TB
2018 (by 9/12)	8,168,965	2,950,997	3,778	4.72 TB

EMDB Core Archive Depositions

- On track for ~1400 3DEM depositions in 2018

Processing Site	2018 Depositions	2017 Depositions
PDBj	90	162
PDBe	336	516
RCSB	828	667

2017 Processing Sites

PDBj 12%
PDBe 38%
RCSB 50%

EMDB Core Archive Growth
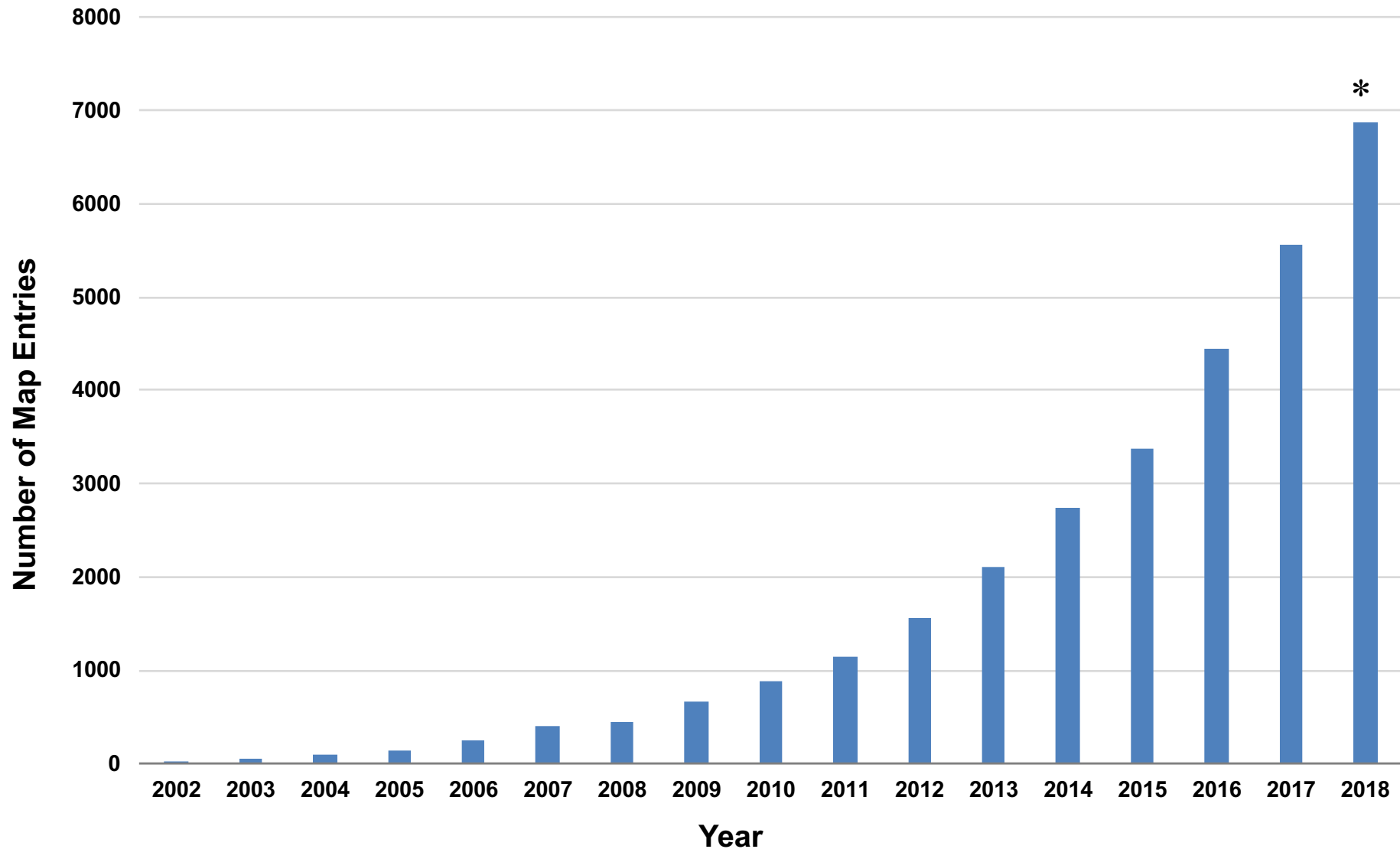
* As of 2 October 2018

Core Member Funding Status

- RCSB PDB: NSF/NIH/DOE funding renewed →12/31/2023.

- BMRB: NIH NIGMS funding→03/31/2019; renewal pending.

- PDBe: EMBL-EBI, Wellcome Trust→01/01/2020.

- PDBj: NBDC-JST and AMED funding→03/31/2022.

- EMDB: EMBL-EBI, Wellcome Trust→01/01/2024.

wwPDB Outreach

2018 OneDep Developer Summit

IUCr

ACA

ECM

NMR

3DEM

**AsCA –
2-5 December 2018**

wwPDB 2017-2018 Publications

Worldwide Protein Data Bank biocuration supporting open access to high-quality 3D structural biology data

Original article

Jasmine Y. Young[1,*], John D. Westbrook[1], Zukang Feng[1], Ezra Peisach[1], Irina Persikova[1], Raul Sala[1], Sanchayita Sen[2], John M. Berrisford[2], G. Jawahar Swaminathan[2], Thomas J. Oldfield[2], Aleksandras Gutmanas[2], Reiko Igarashi[3], David R. Armstrong[2], Kumaran Baskaran[4], Li Chen[1], Minyu Chen[3], Alice R. Clark[2], Luigi Di Costanzo[1], Dimitris Dimitropoulos[1], Guanghua Gao[1], Sutapa Ghosh[1], Swanand Gore[2], Vladimir Guranovic[1], Pieter M. S. Hendrickx[2], Brian P. Hudson[1], Yasuyo Ikegawa[3], Yumiko Kengaku[3], Catherine L. Lawson[1], Yuhe Liang[1], Lora Mak[2], Abhik Mukhopadhyay[2], Buvaneswari Narayanan[1], Kayoko Nishiyama[3], Ardan Patwardhan[2], Gaurav Sahni[2], Eduardo Sanz-García[2], Junko Sato[3], Monica R. Sekharan[1], Chenghua Shao[1], Oliver S. Smart[2], Lihua Tan[1], Glen van Ginkel[2], Huanwang Yang[1], Marina A. Zhuravleva[1], John L. Markley[4], Haruki Nakamura[3], Genji Kurisu[3], Gerard J. Kleywegt[2], Sameer Velankar[2], Helen M. Berman[1] and Stephen K. Burley[1,5,6,7]

[1]RCSB Protein Data Bank, Center for Integrative Proteomics Research, Rutgers, The State University of New Jersey, 174 Frelinghuysen Road, Piscataway, NJ 08854, USA, [2]Protein Data Bank in Europe (PDBe), European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK, [3]PDBj, Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita-shi, Osaka 565-0871, Japan, [4]BMRB, BioMagResBank, University of Wisconsin-Madison, 433 Babcock Drive, Madison, WI 53706, USA, [5]RCSB Protein Data Bank, San Diego Supercomputer Center and Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, 9500 Gilman Dr., La Jolla, CA 92093, USA, [6]Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, 174 Frelinghuysen Road, Piscataway, NJ 08854, USA and [7]Rutgers Cancer Institute of New Jersey, Rutgers, The State University of New Jersey, Little Albany St, New Brunswick, NJ 08901, USA

*Corresponding author: Tel: (848) 445-0103 ext. 4920; Fax: (732) 445-4320; Email: jasmine.young@rcsb.org

Citation details: Young, J.Y., Westbrook, J.D., Feng, Z. et al. Worldwide Protein Data Bank biocuration supporting open access to high-quality 3D structural biology data. Database (2018) Vol. 2018: article ID bay002; doi: 10.1093/database/bay002

Received 2 November 2017; Revised 13 December 2017; Accepted 2 January 2018

Abstract

The Protein Data Bank (PDB) is the single global repository for experimentally determined 3D structures of biological macromolecules and their complexes with ligands. The

Protein Data Bank: the single global archive for 3D macromolecular structure data

wwPDB consortium

ABSTRACT

The Protein Data Bank (PDB) is the single global archive of experimentally determined three-dimensional (3D) structure data of biological macromolecules. Since 2003, the PDB has been managed by the Worldwide Protein Data Bank (wwPDB; wwpdb.org), an international consortium that collaboratively oversees deposition, validation, biocuration, and open access dissemination of 3D macromolecular structure data. The PDB Core Archive houses 3D atomic coordinates of more than 144 000 structural models of proteins, DNA/RNA, and their complexes with metals and small molecules and related experimental data and metadata. Structure and experimental data/metadata are also stored in the PDB Core Archive using the readily extensible wwPDB PDBx/mmCIF master data format, which will continue to evolve as data/metadata from new experimental techniques and structure determination methods are incorporated by the wwPDB. Impacts of the recently developed universal wwPDB OneDep deposition/validation/biocuration system and various methods-specific wwPDB Validation Task Forces on improving the quality of structures and data housed in the PDB Core Archive are described together with current challenges and future plans.

INTRODUCTION

The Protein Data Bank (PDB, pdb.org) was established in 1971 as the first open-access, molecular data resource in biology (1). More than 47 years later, the PDB continues to serve as the single global repository for atomic-level, 3D structure data, making >144 000 experimentally-determined structures of proteins, DNA, and RNA, and their complexes with metal ions, drugs, and other small molecules freely available without restrictions on use. Since 2003, the PDB has been managed jointly by the Worldwide Protein Data Bank (wwPDB) consortium (2), in-

(6). The wwPDB partners are committed to ensuring adherence to the **FAIR** Principles of **F**indability-**A**ccessibility-**I**nteroperability-**R**eusability (7).

Today, the PDB is universally regarded as a core data resource essential for understanding the functional roles that macromolecules play in biology and medicine. Publication of new macromolecular structures in most scientific journals is contingent on mandatory deposition to the PDB of the 3D atomic coordinates comprising the structural model plus experimental data used to derive the structures and associated metadata. Many governmental and non-governmental research funders also require PDB deposition of unpublished macromolecular structure data. All of these 3D structural data are stored in one of two wwPDB Core Archives. The PDB Core Archive houses 3D atomic coordinates of >144 000 structural models of proteins, DNA/RNA, and their complexes with metals and small molecules. The PDB Core Archive also houses related experimental data/metadata from Macromolecular Crystallography (MX). The BioMagResBank (BMRB; www.bmrb.wisc.edu) Core Archive houses related experimental data/metadata from Nuclear Magnetic Resonance spectroscopy (NMR). The wwPDB partners work closely with the Electron Microscopy Data Bank (EMDB; emdb-empiar.org), which houses related experimental data/metadata from 3D Electron Microscopy (3DEM) and Electron Tomography (ET).
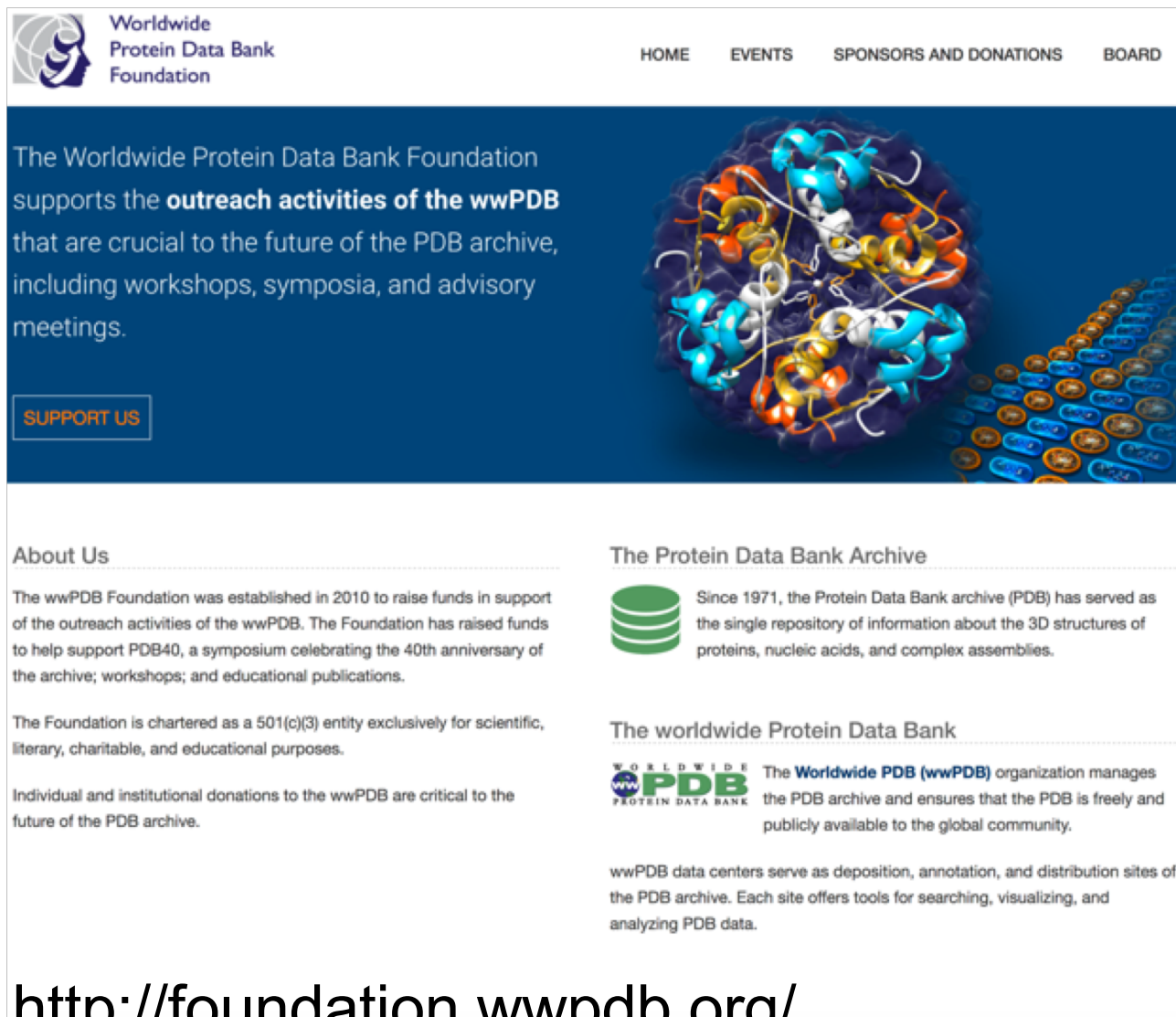
The PDB Core Archive has seen steady growth since its inception, with over 11,000 new structures plus experimental data/metadata released in 2017 (Figure 1A). In aggregate, most of the 3D structures (89.5%) in the PDB Core Archive were determined using macromolecular crystallography (MX), with the remainder determined by NMR (8.5%), 3DEM (1.6%), and other techniques (0.4%). These overall metrics mask recent trends, which show that in 2016 3DEM overtook NMR as the second most popular technique for determining atomic level structures (Figure 1B).

While the PDB Core Archive has grown enormously in scale and scope over the past 47 years and its management has evolved concurrently, adherence to the principle of open access and commitment to community engagement (1) continue to this day.

PDB NAR Database Issue Paper published on 24th October 2018
https://doi.org/10.1093/nar/gky949

wwPDB Foundation Progress

- Website improved.
- Fundraising ongoing.
- 2017 events:
  - WT meeting on Sustainability.
- 2018 events:
  - Tokyo meeting on Sustainability.

http://foundation.wwpdb.org/

wwPDB Collaboration Resource Commitments
October 2017- September 2018

wwPDB Partner	Software Development	Production Maintenance/ Production Management	Requirements Setting/ Testing	Archive Keeping/ Outreach	Biocuration/ Remediation	Total FTE Commitments
RCSB PDB	3.2	1.5	1.0	2.1	~~7.0~~ 6.0	~~14.8~~ 13.8
PDBe	~~2.9~~ 2.1	1.0	0.5	0.1	4.0	~~8.5~~ 7.7
PDBj	0.1	0.9	0.5	0.5	4.3	6.3
BMRB	1.25	--	--	--	--	1.25
wwPDB	~~7.45~~ 6.65*	3.4	2.0	2.7	~~15.3~~ 14.3*	~~30.85~~ **29.05***

* loss of committed resource

OneDep 2017/2018 Progress *vs.* Goals

|	Projects	Timeline				
		2017	2018			
		Q4	Q1	Q2	Q3	Q4
1. Validation	1.1: Annual recalculation of validation reports		Delivered			
	1.2: NMR restraint validation	Delivered	Delivered	Re-forecast		
	1.3: Provide ligand 2D geometrical quality depiction	Re-forecast				
	1.4: Provide ligand ED maps	Re-forecast	Re-forecast			
2. Backend Stabilization	2.1: Towards automated testing framework	Delivered	Delivered			
	2.2: Improve session management		Delivered	Delivered	Delivered	Delivered
	2.3: User account management			Re-forecast	Re-forecast	Re-forecast
3. OneDep public facing	3.1: Retain Previous Annotation During Coordinate Replacement Post Submission	Delivered	Delivered			
	3.2: Enable NEF upload for NMR restraint validation	Re-forecast	Re-forecast	Re-forecast	Re-forecast	
	3.3 Extend metadata collection to support SFX/XFEL			Work in progress		
	3.4 Mandatory ORCID for contact authors				Delivered	
	3.5: Enable depositor-initiated coordinate replacement post release			Re-forecast	Re-forecast	Re-forecast
	3.6 Enhance cross validation and simplify DepUI			Work in progress	Work in progress	Work in progress
4. Biocuration	4.1: Retain previous annotation for entry reprocessing	Delivered	Delivered			
	4.2: Improve overall efficiency	Delivered	Delivered	Delivered	Work in progress	Work in progress
	4.3: Enable WF to use external computing resources			Work in progress	Work in progress	
	4.4: Apply same annotation to identified multi-related entries				Work in progress	Work in progress
5. Archive Improvements	**5.1: XFEL remediation**				Re-forecast	
	5.2: Carbohydrate remediation	Work in progress	Work in progress	Work in progress	Work in progress	Work in progress
	5.3: Protein Modification remediation	Work in progress	Work in progress	Work in progress	Work in progress	Work in progress
	5.4: Expand content at versioned FTP				Work in progress	Work in progress
	5.5: Improve release process					Work in progress
6. EMDB related projects	6.1 Improve data field validation with new EMDB schema			Delivered		
	6.2 Extend EMDB accession code				Delivered	
7. Software Upgrade	7.1 Deposition Software					Work in progress

Delivered, Work in progress, **Bold**: Re-forecast to 2019/2020, Added to original plan

20

wwPDB Progress

- SFX/XFEL
  - The wwPDB has recently produced an extension to the PDBx/mmCIF dictionary to incorporate multiple crystal data collection techniques used in serial femtosecond crystallography (SFX) and X-ray free electron laser (XFEL) experiments.
- PDBx/mmCIF dictionary management
  - An agreed process for dictionary updates.
  - All wwPDB sites involved in dictionary updates.
- Invitation to EMDB.
- Restructuring of wwPDB agreement.

Biocuration Performance

New Structures/wwPDB Biocurator
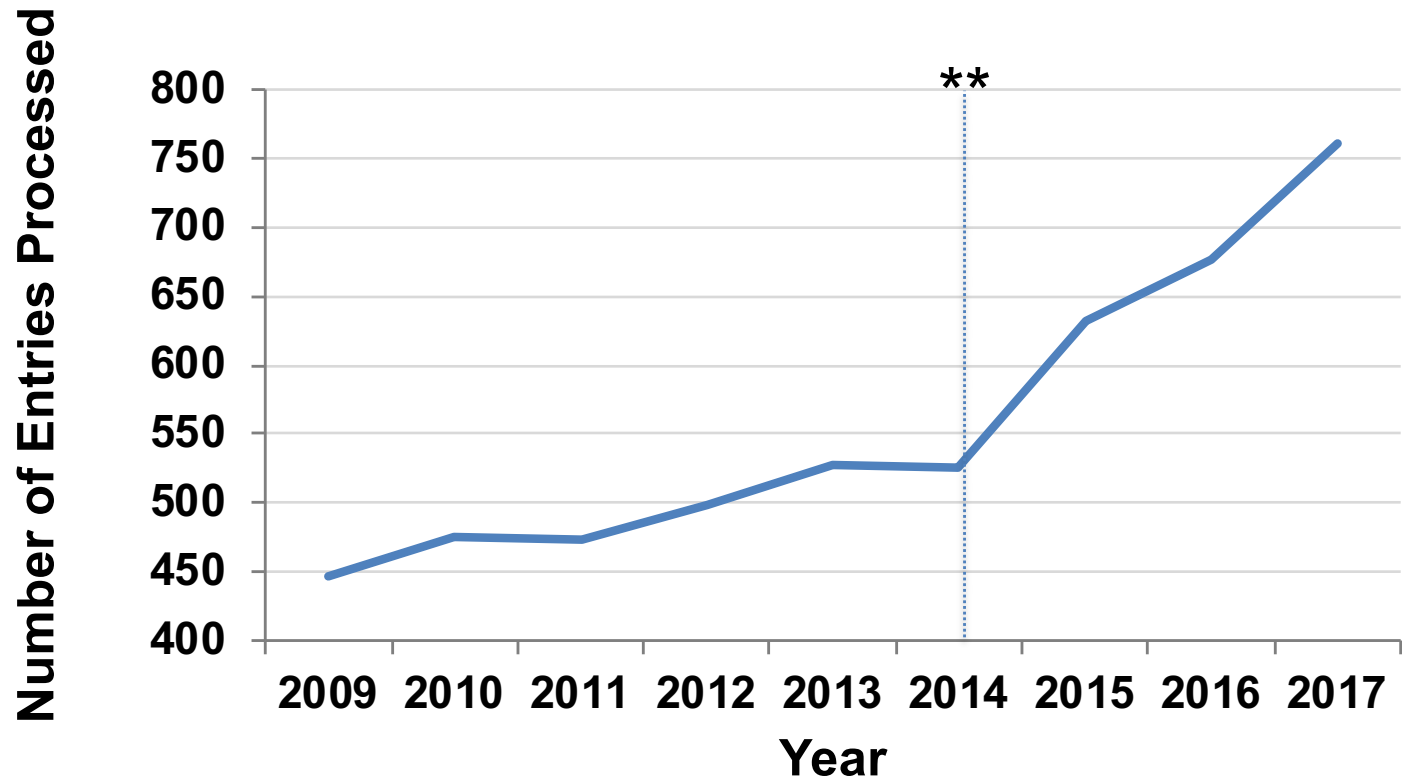
** launch of OneDep

2017/2018 Progress *vs.* Goals I

OneDep projects re-forecasted to 2018/2019

- Implement Ligand Validation Workshop recommendations (2D depictions and ED maps).
- Implement NMR restraint validation.
- Implement EM map validation.
- Implement Author-initiated Coordinate Replacement.
- Use ORCiD for User Authentication.

Mitigations

- Upgrade third party software in 2018 to enable new feature development in 2019.
- Cross-site resource sharing for validation projects.

2017/2018 Progress *vs.* Goals II

- Remediation work continuing
  - Carbohydrates (Lead: RCSB PDB).
  - Post-translational modifications (Lead: PDBe).

- wwPDB Partnership (re-forecasted to 2018/2019)
  - Weekly release process auditing and automation.
  - Resolution of DOI to a single web page on wwPDB site with links to all wwPDB partner sites.
  - Exploring possible new Regional wwPDB Partners.

Updating the wwPDB
Vision, Mission, and Scope

Stephen K. Burley

WORLDWIDE
PDB
PROTEIN DATA BANK

wwpdb.org

Drivers for Updating the wwPDB Vision, Mission, and Scope

- Structural biology is much more than just MX/NMR.

- Structural biology techniques are no longer being used one at a time.

- Contributions from Integrative/Hybrid Methods to PDB-Dev are increasing.

- Data archiving to multiple specialized structural biology data resources needs to be coordinated.

- Role of BMRB archive in wwPDB was never explicitly defined when wwPDB was expanded in 2005.

- wwPDB and EMDB are exchanging data without a formal agreement that protects their Depositors.

- wwPDB and SASBDB are exchanging data without a formal agreement that protects their Depositors.

wwPDB Guiding Principles

- Structural biology is an international science.

- Secure storage and preservation of global structural biology data is essential to the long-term success of the discipline.

- Open access to global structural biology data without limitations on usage is essential to basic and applied research and education across the sciences.

- Global structural biology data represent public goods that require joint international management.

- wwPDB partnership provides an effective mechanism for joint international management of global structural biology data.

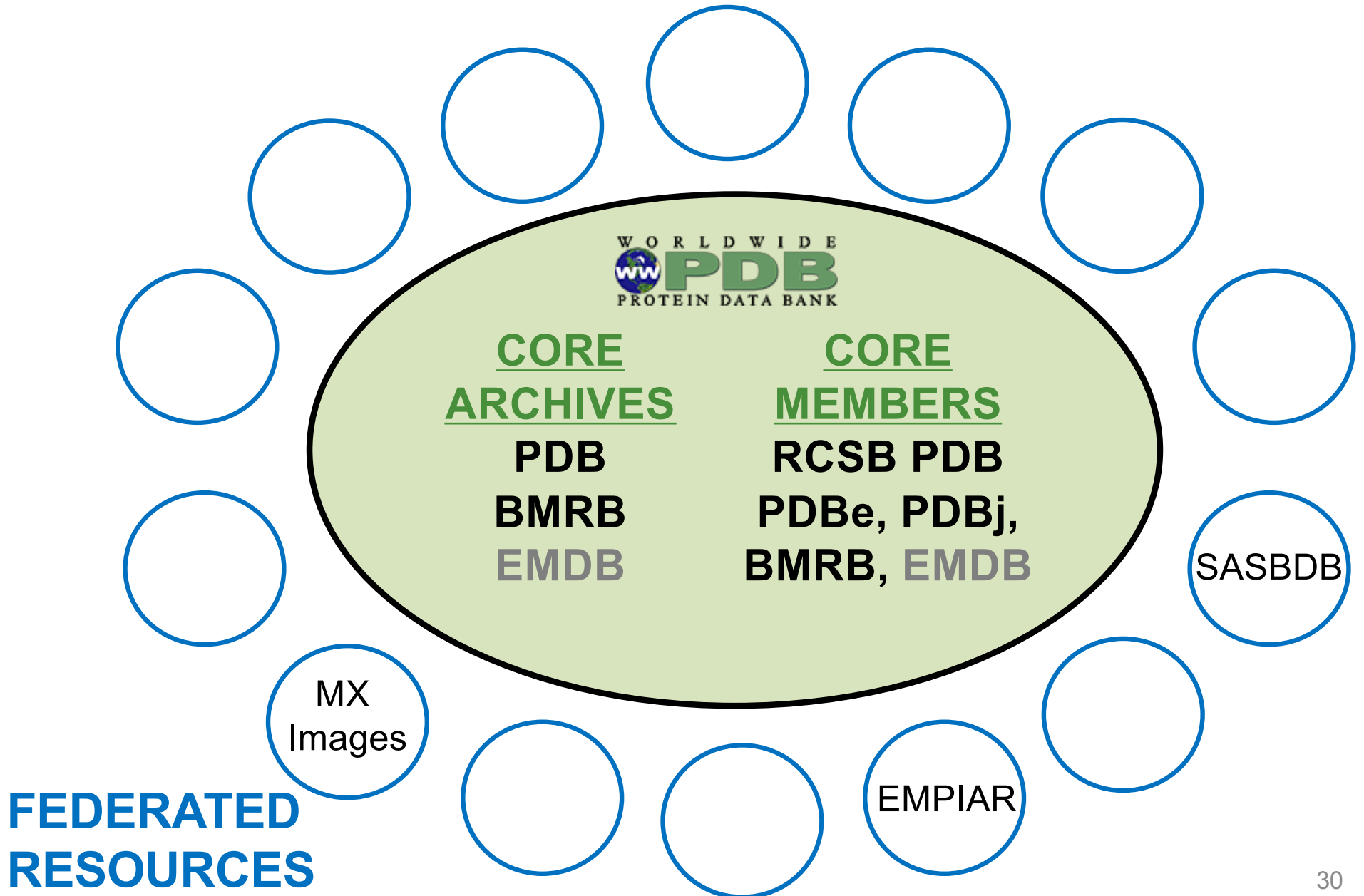- The multilateral wwPDB Agreement is periodically reviewed, revised, and reaffirmed.

The New Vision of the wwPDB is to

Sustain freely accessible, interoperating Core Archives of structure data and metadata for biological macromolecules as an enduring public good to promote basic and applied research and education across the sciences.

The New Mission of the wwPDB is to

- Manage the wwPDB Core Archives as a public good according to the **FAIR** Principles.

- Provide expert deposition, validation, biocuration, and remediation services at no charge to Data Depositors worldwide.

- Ensure universal open access to public domain structural biology data with no limitations on usage.

- Develop and promote community-endorsed data standards for archiving and exchange of global structural biology data.

wwPDB Aspirational Architecture

CORE ARCHIVES
PDB
BMRB
EMDB

CORE MEMBERS
RCSB PDB
PDBe, PDBj,
BMRB, EMDB

WORLDWIDE PDB PROTEIN DATA BANK

SASBDB

MX Images

FEDERATED RESOURCES

EMPIAR

wwPDB Core Archives

Definition: A wwPDB "Core Archive" is a global structural biology data resource jointly managed by wwPDB Core Members, with one member designated as the "Archive Keeper".

- Current wwPDB Core Archives:

  - PDB Core Archive: **3D Structure Data Resource** housing multiscale/atomic structural models plus molecular data and metadata, MX experimental data and metadata, and other experimental data.

  - BMRB Core Archive: **Biomolecular NMR Data Resource** housing molecular data and metadata, NMR experimental data and metadata, and other experimental data.

- Next Core Archive expected to join wwPDB:

  - EMDB Core Archive: **Molecular and Cellular EM Data Resource** housing molecular/biological data and metadata, experimental electric potential map data, and other experimental data.

wwPDB Archive Keepers

Role: A wwPDB "Archive Keeper" is a wwPDB Core Member appointed, according to the terms of the wwPDB Agreement, as having primary responsibility for data storage, preservation, composition, representation, and maintenance of a particular Core Archive, carried out in concert with other wwPDB Core Members.

- <u>Current wwPDB Archive Keepers</u>
  - PDB Archive Keeper: RCSB PDB
  - BMRB Archive Keeper: BMRB

- <u>Expected addition to wwPDB</u>
  - EMDB Archive Keeper: EMDB

wwPDB Core Members

- Role: A wwPDB "Core Member" contributes to all wwPDB partnership activities according to the terms of the wwPDB Agreement.

- Current wwPDB Core Members:
  - RCSB PDB
  - PDBe
  - PDBj
  - BMRB

- Expected addition to wwPDB:
  - EMDB

wwPDB Associate Members

- Role: A wwPDB "Associate Member" contributes to some of the wwPDB partnership activities according to the terms of the wwPDB Agreement, with the goal of becoming a wwPDB Core Member.

- Current wwPDB Associate Members:
  - None

- Expected additions to wwPDB:
  - PDB-China
  - PDB-India

wwPDB Advisory Committee

- Role
  - Existing Terms of Reference under revision (Appendix 4)
- Current wwPDB AC Roster
  - Core Member Representatives of RCSB PDB, PDBe, PDBj, and BMRB (2 each)
  - Organizational Representatives of IUCr, ICMRBS, and EM-TBN (1 each)
  - National Representatives: China and India (1 each)
  - Institutional Representatives (1 per organization)
- Expected changes to wwPDB AC
  - Core Member Representatives of EMDB (2)
  - Observers from PDB-China, PDB-India
    (1 or 2 each, replacing National Representatives)

wwPDB Federated Resources

- Definition: A wwPDB "Federated Resource" is a structural biology data resource that participates in data exchange with one or more wwPDB Core Archives, under a formal agreement governing data exchange and confidentiality, etc.

- Current Federated Resources:
    - None

- Federated Resources expected to align with wwPDB:
    - SASBDB
    - EMPIAR

wwPDB Federated Members

- Role: A wwPDB Federated Member manages one or more wwPDB Federated Resources and collaborates with the wwPDB Core Members in developing and maintaining data exchange infrastructure

- Current Federated Members:
  - None

- Federated Members expected to align with wwPDB:
  - SASBDB
  - EMPIAR

wwPDB Community Engagement

Sameer Velankar

WORLDWIDE
PDB
PROTEIN DATA BANK

wwpdb.org

wwPDB Community Engagement: Historical Drivers

- Establishing data standards and improving data quality of the archived data.

- Consulting with community experts

  - PDBx/mmCIF Working Group – Developing data standards for representation of structure data.

  - Validation and other Task Forces – Recommendations from community experts on improving quality of experimental data, metadata, and structures.

  - Workshops – Recommendations from community experts on improving co-crystal structures.

Impact of Changes to wwPDB

- Multiple Core Archives – PDB, BMRB and EMDB

- Each Core Archive will require community input on data standards and validation.

- Number of structures using multiple data types is increasing.

- Data standards and data deposition, validation, and biocuration are all interconnected.

- Need an integrated approach to data standards and data deposition, validation, and biocuration across all wwPDB Core Archives.

wwPDB Data Archiving Steering Group

- Plan to establish a wwPDB Steering Group responsible for oversight of data standards and data deposition, validation, and biocuration requirements across all Core Archives.

- wwPDB Core Member PIs will join the steering group

- Each wwPDB Core Member will appoint one qualified external individual and one staff member to the Steering Group.

- wwPDB Core Members will jointly appoint a Chair for a 3-year term (renewable).

- One wwPDB Core Member PI will serve as rotating Co-chair (1 year term)

wwPDB Data Archiving Steering Group

- Steering Group and wwPDB Core Members will agree on priorities and convene teams of qualified experts (Expert Advisory Groups) from the community to address issues and provide recommendations as needed:

  - The Expert Advisory Groups (EAGs) will publish recommendations in a peer reviewed journal.

- Steering Group will meet quarterly and publish meeting minutes on the wwPDB website.

- Steering Group will report annually to the wwPDB Advisory Committee on progress from all EAGs.

PDB Core Archive plans

Genji Kurisu

wwpdb.org

wwPDB Collaboration Resource Commitments
November 2018- October 2019

wwPDB Partner	Software Development	Production Maintenance/ Management	Requirements Setting/ Testing	*Core Archive Keeping	Outreach	Biocuration/ Remediation	Total FTE Commitments
RCSB PDB	3.0	1.8	0.35/0.35	2.0	0.2	6.0	13.7
PDBe	1.6	1.0	0.35/0.35	-	0.2	4.2	7.7
PDBj	0.4	0.4	0.2/0.2	-	0.2	4.2	5.6
BMRB	1.5	-	-/-	1.5	0.2	0.2	3.9
EMDB	1	0.1	-/-	0.9	-	0.5	2.5
Total wwPDB	7.5	3.3	0.9/0.9	4.4	0.8	15.1	33.4

* Resource from Archive Keeper

Additional resources for software development made available by PDBj, EMDB and BMRB.
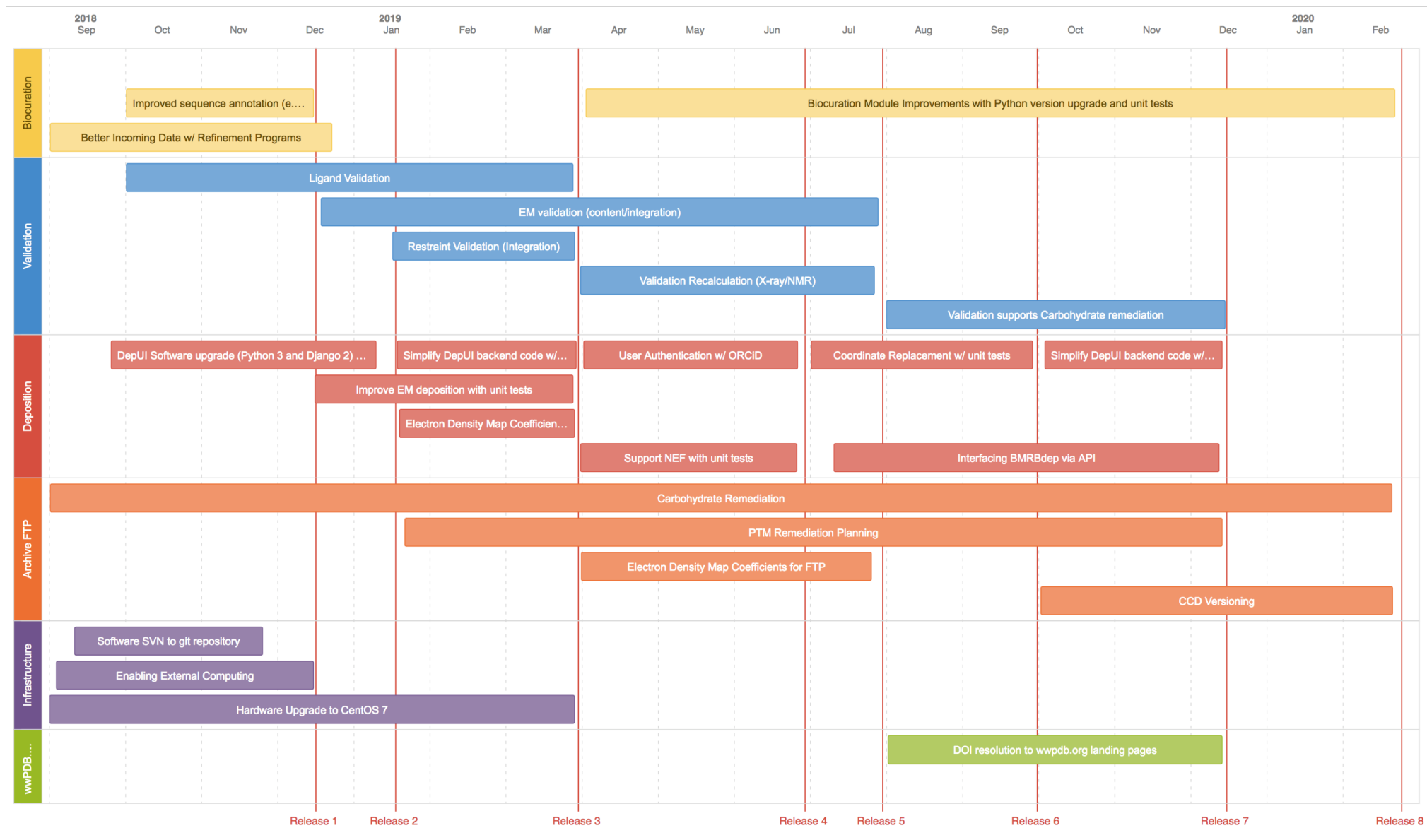
OneDep 2018/2019 Goal Setting I

Major Projects	Primary resource*
Supporting NMR data in NEF format within OneDep	BMRB
NMR restraint validation (Includes modularizing validation software with unit tests)	BMRB
Content improvement in the EM validation report	EMDB
PTM remediation (Setting remediation plan)	PDBe
Coord replacement post release (Includes User Authentication with ORCiD)	PDBe: DepUI RCSB PDB: API
Annual validation recalculation	PDBe
ED map coefficients	RCSB PDB
Simplify DepUI backend software code and add unit tests	PDBe
Carbohydrate remediation (Software development)	RCSB PDB
Biocuration software upgrade (python 3, unit tests, etc.)	RCSB PDB
CCD versioning	RCSB PDB
Enhance ligand validation (Includes modularizing validation software with unit tests)	RCSB PDB
Biocuration and DepUI improvements (Includes Biocuration software upgrade and unit tests)	RCSB PDB: Biocuration PDBe: DepUI

Bold: re-forecasted from 2017/2018

* Cross-site resource sharing identified for primary and supporting resource

OneDep 2018/2019 Goal Setting II

| | 2018 Sep | Oct | Nov | Dec | 2019 Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec | 2020 Jan | Feb |

Biocuration
- Improved sequence annotation (e….
- Better Incoming Data w/ Refinement Programs
- Biocuration Module Improvements with Python version upgrade and unit tests

Validation
- Ligand Validation
- EM validation (content/integration)
- Restraint Validation (Integration)
- Validation Recalculation (X-ray/NMR)
- Validation supports Carbohydrate remediation

Deposition
- DepUI Software upgrade (Python 3 and Django 2) …
- Simplify DepUI backend code w/…
- User Authentication w/ ORCiD
- Coordinate Replacement w/ unit tests
- Simplify DepUI backend code w/…
- Improve EM deposition with unit tests
- Electron Density Map Coefficien…
- Support NEF with unit tests
- Interfacing BMRBdep via API

Archive FTP
- Carbohydrate Remediation
- PTM Remediation Planning
- Electron Density Map Coefficients for FTP
- CCD Versioning

Infrastructure
- Software SVN to git repository
- Enabling External Computing
- Hardware Upgrade to CentOS 7

wwPDB….
- DOI resolution to wwpdb.org landing pages

Release 1 Release 2 Release 3 Release 4 Release 5 Release 6 Release 7 Release 8

* Timeline will be further refined after requirement setting.

PDB Core Archive Plans Specific to MX

- **ED map coefficients**: As part of ligand validation development, the ED map coefficients will be provided at the wwPDB FTP site, which will be implemented in 2019 for better interpretation of the ligand validation by user community.

Remediation of PDB Archive

- Remediation is a continuous process to ensure better consistency and searchability.

- Representation of **carbohydrates** led by RCSB PDB, and **post-translational modifications (PTM)** led by PDBe will be the focus of PDB Core Archive plans in 2019 and 2020.
  - <u>2018</u>: Consultation with Working Group on dictionary extension and example files for carbohydrates – Q4 2018.
  - <u>2019</u>: Requirement setting for PTM and software development for carbohydrates.
  - <u>2020</u>: Software development for PTM.

Immediate Objectives for Versioning (1)

- **Ligand improvements**: Recently, validation of ligands in the PDB Core Archive was improved by the adoption of a more robust way of flagging those molecules that do not fit electron density well. LLDF (Local Ligand Density Fit) has been replaced by a combination of RSR and RSCC since March 2018.

- **Global Phasing ligand validation code**: Their Ligand Validation Buster Report code is being evaluated by the wwPDB OneDep Software Development Team.
    - The code review has determined that the code is suitable for use in OneDep.

Immediate Objectives for Versioning (2)

- **Depositors' corrections:** Depositors will be able to make corrections to existing structures in the PDB Core Archive by updating the atomic coordinates while preserving the original PDB identifier. The recent introduction of versioning makes this long desired opportunity possible.
  - In the first year, we will restrict this to improving co-crystal structures to understand the impact.
  - Only original depositor of an entry can provide updates.
  - Number of updates for each depositor restricted:
    - 1 coordinate replacement per entry per year.
    - 3 entries update per PI per year.

Use of ORCiD in OneDep

- To enable the depositors' corrections, ORCiD ID(s) for contact author(s) became mandatory in 2018. To further improve the depositor experience and enable better management of incoming data, OneDep protocols will be changed to allow login using ORCiD in 2019.

- Receiving requests from depositors to back-populate ORCiD for released entries (updates underway).

BMRB Core Archive Plans

John Markley

BMRB Archive to Become a wwPDB Core Archive

- The BMRB will work with the other wwPDB partners to develop cross-links to data in BMRB associated with structures in the PDB archive.

- As a step towards interoperability between the BMRB small molecule archive and the wwPDB Chemical Components Dictionary (CCD), BMRB has provided the OneDep developers an analysis of the CCD that provides standard InChI strings for compounds that lack them and full InChI compliant atom designators (as generated by ALATIS[1]) as synonyms for current atom designators.

[1]Sci Data. 2017 4:170073. doi: 10.1038/sdata.2017.73

BMRBdep will be Launched

- BMRBdep will replace ADIT-NMR at the Madison and Osaka branches of BMRB and will provide an API to facilitate the acquisition of NMR data.

- The goal will be development of seamless interaction between OneDep and BMRBdep for the benefit of depositors of structures utilizing NMR data.

- The BMRBdep API will guide the collection and validation of additional NMR data that is not currently collected in OneDep.

- BMRB will lead the efforts to improve validation and scope of data currently collected in OneDep.

- Validation of NMR data against coordinates will be carried out by OneDep using standards and software developed within the OneDep team.

- BMRBdep will continue to handle NMR-based structures of peptides, natural products, and small molecules that do not fit PDB criteria (e.g., small peptides and oligosaccharides).

Goals for BMRB Archive Re-Scoping

- BMRB is working with experts in a number of growth areas with the objective of capturing data from a higher percentage of NMR publications across a wider range of biomolecular applications:
    - Dynamics (Arthur Palmer)
    - Integrative/Hybrid Methods (wwPDB I/H Methods Task Force)
    - Intrinsically Disordered Proteins (Julie Forman-Kay)
    - Solid-state NMR (Chad Rienstra)
    - Membrane Proteins (Gianluigi Veglia)
    - Metabolomics (Teresa Fan)
    - Molecular Interactions (Giuseppe Melancini)
    - Natural Products (Arthur Edison)
    - NMR Software Standards (Wim Vranken)
    - Proteins (Ichio Shimada)
    - RNA (Hashim al Hashimi)

Facilitating NMR Data Deposition

- BMRB is working with the NMRbox projects on capturing information from biomolecular NMR workflows and studies in forms that can flow seamlessly into BMRBdep.

  - BMRB is working with NMRbox to promulgate the use of NMR-STAR throughout computational workflows, not merely at the end-stage.

  - BMRB and NMRbox will collaborate with other software developers as needed, to generate NMR-STAR output.

  - Restraint data from software packages that output NEF will be archived at BMRB and translated into NMR-STAR for deposition using BMRBdep and PDB.

BMRB Leadership Transition

- The competitive grant renewal application, which was reviewed favorably, lists John Markley (UW-Madison) and Jeff Hoch (UConn Health) as Co-PIs and Co-Heads (Pedro Romero is the BMRB Director).

- Grant will be administered through UW-Madison.

- John Markley expects to become an Emeritus on 1 June 2020. At that point, Jeff Hoch will become the sole PI. The project will remain joint between UW-Madison and UConn Health with Pedro Romero continuing as BMRB Director.

- John Markley will continue to be associated with BMRB as an Emeritus Professor on a voluntary basis and will provide advice and assistance as needed.

Questions for the wwPDB AC

Sameer Velankar

WORLDWIDE PDB
PROTEIN DATA BANK

wwpdb.org

Questions for the wwPDB AC

1. Does the wwPDB AC concur with updated vision and mission statements as outlined in Appendix 1?

2. Does the wwPDB AC concur with our proposal to invite the SASBDB archive team at the EMBL-Hamburg to join the wwPDB organization as a Federated Member and for the SASBDB Archive to be recognized as a Federated Archive by establishing an agreement as outlined in Appendix 2?

3. Does the wwPDB AC concur with our proposal to establish a wwPDB Data Archiving Steering Group as outlined in Appendix 3?

Questions for the wwPDB AC (cont.)

4. Does the wwPDB AC concur with the updated Terms of Reference for the wwPDB Advisory Committee as outlined in Appendix 4?

5. What is the wwPDB AC members guidance regarding community representation for the 3DEM community (Appendix 5)?

6. What is the wwPDB AC members guidance on the comment period for the policy change of making PDBx/mmCIF files mandatory for deposition of X-ray structures as outlined in Appendix 6?

Questions for the wwPDB AC (cont.)

7. Does the wwPDB AC have any questions or concerns regarding the individual RCSB PDB, PDBe, PDBj, or BMRB Advisory Committee reports provided in Appendix 7?