



Assessment Information

[CoreTrustSeal Requirements 2020–2022](#)

Repository: Worldwide Protein Data Bank
Website: <https://wwpdb.org>
Certification Date: 07 May 2021

This repository is owned by: **Rutgers, The State University of New Jersey**



Worldwide Protein Data Bank

Notes Before Completing the Application

We have read and understood the notes concerning our application submission.

True

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

CORE TRUSTWORTHY DATA REPOSITORIES REQUIREMENTS

Background & General Guidance

Glossary of Terms

BACKGROUND INFORMATION

Context

R0. Please provide context for your repository.

Repository Type. Select all relevant types from:

Domain or subject-based repository, Archive, Research project repository

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept.

Brief Description of Repository

Context and Repository Type

The Protein Data Bank (PDB) archive is the single worldwide repository of information about the 3D structures of large biological molecules, including proteins and nucleic acids. These are the molecules of life that are found in all organisms including bacteria, yeast, plants, flies, other animals, and humans. Understanding the shape of a molecule deduces a structure's role in human health, disease, and drug development. The structures in the archive range from tiny proteins and bits of DNA to complex molecular machines like the ribosome.

The PDB was established in 1971 at Brookhaven National Laboratory and originally contained 7 structures. The Research Collaboratory for Structural Bioinformatics (RCSB) became responsible for the management of the PDB in 1998. In 2003, the Worldwide Protein Data Bank (wwPDB) was formed to maintain a single PDB archive of macromolecular structural data that is freely and publicly available to the global community. It consists of organizations that act as deposition, data processing, and distribution centers for PDB data. The PDB archive is available at no cost to users. Deposition of atomic coordinates, experimental data, and metadata are required by all major scientific journals when publishing a new structure determination study. Current wwPDB partners are RCSB Protein Data Bank (RCSB PDB), Protein Data Bank in Europe (PDBe), Protein Data Bank Japan (PDBj), and BioMagResBank (BMRB).

Recent articles (1-3) provide an additional general context for the important role of the PDB in the global management of 3D biological structure data.

1. wwPDB consortium (2018) Protein Data Bank: the single global archive for 3D macromolecular structure data *Nucleic Acids Research*: gky949 doi: 10.1093/nar/gky949
2. S. K. Burley, H. M. Berman, G. J. Kleywegt, J. L. Markley, H. Nakamura, S. Velankar. (2017) Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive In *Methods in Molecular Biology: Protein Crystallography Methods and Protocols* (eds. A. Wlodawer, Z. Dauter, M. Jaskolski) Springer New York DOI: 10.1007/978-1-4939-7000-1
3. H. M. Berman, S. K. Burley, G. J. Kleywegt, J. L. Markley, H. Nakamura, S. Velankar. (2016) The archiving and dissemination of biological structure data. *Curr Opin Struct Biol* 40: 17-22 DOI: 10.1016/j.sbi.2016.06.018

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept.

Brief Description of the Repository's Designated Community.

Repository Community

The Worldwide Protein Data Bank (wwPDB) is the single global repository for 3D macromolecular structure data. The wwPDB has an international community of users, including biologists (in diverse fields such as structural biology, agriculture, biochemistry, genetics, pharmacology, zoology); other scientists (in fields such as bioinformatics, software developers for data analysis and visualization); student and educators (all levels); media writers, illustrators, textbook authors; and the general public.

*Reviewer Entry***Reviewer 1**

Comments:
Accept.

Reviewer 2

Comments:
Accept.

Level of Curation Performed. Select all relevant types from:

D. Data-level curation – as in C above; but with additional editing of deposited data for accuracy

*Reviewer Entry***Reviewer 1**

Comments:
Accept.

Reviewer 2

Comments:
Accept.

Comments

Curation

Biocuration Policies and Practices:

wwPDB data depositions contain comprehensive descriptions of structural models from macromolecular X-ray, NMR, and 3DEM investigations. In addition to atomic coordinates, details regarding the chemistry of biopolymers and any bound small molecules are archived, as are metadata describing biopolymer sequence, sample composition and preparation, experimental procedures, data-processing methods/software/statistics, structure determination/refinement procedures and statistics, and certain structural features (e.g., secondary and quaternary structure).

Comprehensive documentation of wwPDB Biocuration practices, archive accession procedures, and data life-cycle policies are maintained at (1).

Preservation of Primary Data:

In the context of the 3D macromolecular structure data managed by the wwPDB, the 3D coordinate data and the supporting experimental data (e.g., X-ray/Neutron structure factor amplitudes, NMR chemical shifts, NMR restraints, 3DEM mass density maps) are treated as primary data. The Biocuration process may update these data to conform with archive standard molecular and chemical nomenclature. Such changes are performed under the supervision of the wwPDB expert Biocuration staff. Changes to primary data values (e.g., atomic positions or X-ray structure factor amplitudes) are only made by the depositor of record. Once a PDB entry is released into the public archive, subsequent requests for entry change by the depositor of record which impact the primary data must be accompanied by either a re-accession of the entry or a change in the major version of the entry (2). In either of the above cases in which a released entry is changed, the prior entry remains available in the repository.

The wwPDB partners have developed the OneDep system (3-4) to provide a common and shared platform to perform PDB deposition, validation, and biocuration tasks.

1. wwPDB Biocuration Policies:

<https://www.wwpdb.org/documentation/annotation>

2. wwPDB Versioning News Item:

<http://www.wwpdb.org/news/news?year=2017#5979ef4161fd3d50915a4b00>

3. J. Young, J. D. Westbrook, Z. Feng, R. Sala, E. Peisach, T. J. Oldfield, S. Sen, A. Gutmanas, D. R. Armstrong, J. M. Berrisford, L. Chen, M. Chen, L. Di Costanzo, D. Dimitropoulos, G. Gao, S. Ghosh, S. Gore, V. Guranovic, P. M. S. Hendrickx, B. P. Hudson, R. Igarashi, Y. Ikegawa, N. Kobayashi, C. L. Lawson, Y. Liang, S. Mading, L. Mak, M. S. Mir, A. Mukhopadhyay, A. Patwardhan, I. Persikova, L. Rinaldi, E. Sanz-Garcia, M. R. Sekharan, C. Shao, G. J. Swaminathan, L. Tan, E. L. Ulrich, G. Van Ginkel, R. Yamashita, H. Yang, M. A. Zhuravleva, M. Quesada, G. J. Kleywegt, H. M. Berman, J. L. Markley, H. Nakamura, S. Velankar, S. K. Burley. (2017) OneDep: Unified wwPDB System for Deposition, Biocuration, and Validation of Macromolecular Structures in the Protein Data Bank (PDB) Archive. *Structure* 25: 536-545 DOI: 10.1016/j.str.2017.01.004

4. J.Y. Young, J. D. Westbrook, Z. Feng, E. Peisach, I. Persikova, R. Sala, S. Sen, J. M. Berrisford, G. J. Swaminathan, T. J. Oldfield, A. Gutmanas, R. Igarashi, D. R. Armstrong, K. Baskaran, L. Chen, M. Chen, A. R. Clark, L. D. Costanzo, D.

Dimitropoulos, G. Gao, S. Ghosh, S. Gore, V. Guranovic, P. M. S. Hendrickx, B. P. Hudson, Y. Ikegawa, Y. Kengaku, C. L. Lawson, Y. Liang, L. Mak, A. Mukhopadhyay, B. Narayanan, K. Nishiyama, A. Patwardhan, G. Sahni, E. Sanz-García, J. Sato, M. R. Sekharan, C. Shao, O. S. Smart, L. Tan, G. v. Ginkel, H. Yang, M. A. Zhuravleva, J. L. Markley, H. Nakamura, G. Kurisu, G. J. Kleywegt, S. Velankar, H. M. Berman, S. K. Burley (2018) Worldwide Protein Data Bank biocuration supporting open access to high-quality 3D structural biology data, Database, 2018: bay002 doi: 10.1093/database/bay002

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept.

Insource/Outsource Partners. If applicable, please list them.

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept.

Summary of Significant Changes Since Last Application (if applicable).

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept.

Other Relevant Information.

Other Relevant Information

Repository Usage and Impact:

Global usage of repository holdings, including aggregate and entry-level statistics, is updated regularly at (1). Analysis of the impact of services provided by the European Bioinformatics Institute, including the Protein Data Bank in Europe (PDBe), is presented in (2). The primary citation (3) for the wwPDB US partner (RCSB) is one of the top-cited scientific publications of all time. A bibliometric analysis of this primary citation (2) performed by Clarivate Analytics (4) in 2017 shows the PDB motivated high-quality research throughout the world. Papers citing the PDB had a citation-based impact exceeding the world average in 16 scientific fields, including Biology & Biochemistry, Computer Science, Plant & Animal Sciences, Physics, Environment/Ecology, Mathematics, and Geosciences. The PDB structure data contributions to recent FDA drug approvals (5) and anti-cancer drug approvals (6) have been recently reviewed.

1. wwPDB usage statistics:

<https://www.wwpdb.org/stats/download>

2. EBI/PDBe impact Analysis:

<https://beagrie.com/static/resource/EBI-impact-report.pdf>

3. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. (2000) The Protein Data Bank *Nucleic Acids Research*, 28: 235-242.

4. Basner, J. Impact Analysis of "Berman HM et al., (2000), The Protein Data Bank. (Clarivate Analytics, DOI: 10.2210/rcsb_pdb/pdb-cit-anal-2017, 2017).

5. Westbrook JD, Burley SK. How Structural Biologists and the Protein Data Bank Contributed to Recent FDA New Drug Approvals. *Structure*. 2019; 27: 211-7. DOI: 10.1016/j.str.2018.11.007. PubMed PMID: 30595456

6. Westbrook JD, Soskind R, Hudson BP, Burley SK. Impact of Protein Data Bank on Anti-neoplastic Approvals. *Drug Discov Today*. 2020;25:837-50. DOI: 10.1016/j.drudis.2020.02.002.

Global Organization and Role:

The PDB archive is managed jointly by the Worldwide Protein Data Bank partnership (wwPDB; wwpdb.org)(1) and consists of the RCSB Protein Data Bank (2, 3), Protein Data Bank Japan (PDBj) (4-5), the Protein Data Bank in Europe (PDBe)(6-7), and the BioMagResBank (BMRB)(6). The wwPDB organization operates under a formal agreement (wwpdb.org/about/agreement), most recently renewed in 2013. This agreement commits wwPDB partners to collect, standardize, validate, annotate, and store macromolecular structure data in a single global archive for data depositors and to disseminate these data via a web repository to data consumers, all at no charge with no restrictions on data usage.

1. Berman HM, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. *Nat Struct Biol*. 2003;10(12):980. DOI: 10.1038/nsb1203-980. PubMed PMID: 14634627; PMCID: 1669775.

2. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235-42. Epub 1999/12/11. DOI: 10.1093/nar/28.1.235. PubMed PMID: 10592235; PMCID: 102472.

3. Burley SK, Bhikadiya C, Bi C, Bittrich S, Chen L, Crichlow G, Christie CH, Dalenberg K, Costanzo LD, Duarte JM, Dutta S, Feng Z, Ganesan S, Goodsell DS, Ghosh S, Green RK, Guranovic V, Guzenko D, Hudson BP, Lawson CL, Liang Y, Lowe R, Namkoong H, Peisach E, Persikova I, Randle C, Rose A, Rose Y, Sali A, Segura J, Sekharan M, Shao C, Tao

- Y-P, Voigt M, Westbrook JD, Young JY, Zardecki C, Zhuravleva M. RCSB Protein Data Bank: Powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering, and energy sciences. *Nucleic Acid Research*. 2020:gkaa1038. DOI: 10.1093/nar/gkaa1038.
4. Kinjo AR, Bekker GJ, Suzuki H, Tsuchiya Y, Kawabata T, Ikegawa Y, Nakamura H. Protein Data Bank Japan (PDBj): updated user interfaces, resource description framework, analysis tools for large structures. *Nucleic Acids Res*. 2017;45(D1): D282-D8. DOI: 10.1093/nar/gkw962. PubMed PMID: 27789697; PMCID: PMC5210648.
5. Kinjo, A.R., Bekker, G.-J., Wako, H., Endo, S., Tsuchiya, Y., Sato, H., Nishi, H., Kinoshita, K., Suzuki, H., Kawabata, T., Yokochi, M., Iwata, T., Kobayashi, N., Fujiwara, T., Kurisu, G. and Nakamura, H. New tools and functions in Data-out activities at Protein Data Bank Japan (PDBj). *Protein Science* 27 (1): 95-102 (2018) [PubMed] [doi:10.1002/pro.3273]
6. Armstrong DR, Berrisford JM, Conroy MJ, Gutmanas A, Anyango S, Choudhary P, Clark AR, Dana JM, Deshpande M, Dunlop R, Gane P, Gáborová R, Gupta D, Haslam P, Koča J, Mak L, Mir S, Mukhopadhyay A, Nadzirin N, Nair S, Paysan-Lafosse T, Pravda L, Sehnal D, Salih O, Smart O, Tolchard J, Varadi M, Svobodova-Vařeková R, Zaki H, Kleywegt GJ, Velankar S., PDBe: improved findability of macromolecular structure data in the PDB. *Nucleic acids research* Volume 48 (2020) p.D335-D343 DOI: 10.1093/nar/gkz990
7. Velankar S, van Ginkel G, Alhroub Y, Battle GM, Berrisford JM, Conroy MJ, Dana JM, Gore SP, Gutmanas A, Haslam P, Hendrickx PM, Lagerstedt I, Mir S, Fernandez Montecelo MA, Mukhopadhyay A, Oldfield TJ, Patwardhan A, Sanz-Garcia E, Sen S, Slowley RA, Wainwright ME, Deshpande MS, Iudin A, Sahni G, Salavert Torres J, Hirshberg M, Mak L, Nadzirin N, Armstrong DR, Clark AR, Smart OS, Korir PK, Kleywegt GJ. PDBe: improved accessibility of macromolecular structure data from PDB and EMD. *Nucleic Acids Res*. 2016;44(D1): D385-95. DOI: 10.1093/nar/gkv1047. PubMed PMID: 26476444; PMCID: PMC4702783.
8. Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Kent Wenger R, Yao H, Markley JL. BioMagResBank. *Nucleic Acids Res*. 2008; 36(Database issue):D402-8. Epub 2007/11/07. DOI: 10.1093/nar/gkm957. PubMed PMID: 17984079; PMCID: 2238925.

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept.

ORGANIZATIONAL INFRASTRUCTURE

1. Mission/Scope

R1. The repository has an explicit mission to provide access to and preserve data in its domain.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

I. Mission/Scope

The wwPDB maintains a single archive of macromolecular structural data that is freely and publicly available to the global community. The wwPDB organization operates under a public formal agreement (1), most recently renewed in 2013. This agreement wwPDB partner commitment to standardizing, collecting, validating, annotating, and storing macromolecular structure data as a single global archive for data depositors and disseminating these data in a common repository to data consumers, all at no charge with no restrictions on data usage.

The wwPDB Mission statement (<http://www.wwpdb.org>):

- + 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.

1. wwPDB Agreement: <https://www.wwpdb.org/about/agreement>

Reviewer Entry

Reviewer 1

Comments:

Accept.

Reviewer 2

Comments:

Accept

2. Licenses

R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

II. License

Data files contained in the PDB archive (<ftp://ftp.wwpdb.org>) are free of all copyright restrictions and made fully and freely available for both non-commercial and commercial use. Users of the data should attribute the original authors of that structural data. By using the materials available in the PDB archive, the user agrees to abide by the conditions described in the PDB Advisory Notice (1). This advisory notice defines the current license and terms of use for the repository,

The commitment to providing open access to PDB data is further described in the wwPDB Agreement document (2).

1. wwPDB Repository Advisory Notice: <https://ftp.wwpdb.org/pub/pdb/advisory.txt>

2. wwPDB Agreement: <https://www.wwpdb.org/about/agreement>

The wwPDB organization has recently agreed to replace the current license based on the PDB Advisory Notice with a more standard open-source license. the Creative Commons CC0 license (<https://creativecommons.org/share-your-work/public-domain/cc0/>). The terms of this license are more permissive than the PDB Advisory Notice as there is no formal acknowledgment requirement. The acknowledgment requirement has proved incompatible with restrictions on the number of references allowed by some journal reporting results analyses on larger numbers of PDB data entries. The organization plans to officially revise the licensing policy by the end of 2021.

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:
Accept

3. Continuity of access

R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

III. Continuity of Access

The PDB has a 49-year track record for providing open data access with continuity of supporting repository services. Since 2003, the archive has been managed by a global partnership, which further strengthens the repository's long-term stability. Because the PDB has developed data processing tools and infrastructure shared by all of the wwPDB partners, an orderly transfer of responsibility is possible should any of the partners withdraw from the global agreement. Such an eventuality is addressed by the wwPDB Agreement document (1); moreover, this Agreement also provides for extending the wwPDB partnership to address future growth in structure data production.

1. wwPDB Agreement: <https://www.wwpdb.org/about/agreement>

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept.

4. Confidentiality/Ethics

R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

IV. Confidentiality/Ethics

The data collected and distributed by the PDB are considered public data and do not present ethical disclosure risks. As a matter of policy, the wwPDB does not publicly distribute depositor contact details maintained exclusively for administrative purposes.

PDB provides an optional embargo period of up to one year before releasing a data entry. This embargo's primary purpose is to allow for the coordinated release of PDB entries with the publication of their associated primary citations. PDB detects publications through routine scanning of publication repositories or direct notification from depositors or publishers. In all cases, depositors are notified for confirmation before data release.

Complete details of the release embargo policies are described in (1).

1. wwPDB Release and Embargo Policies: https://www.wwpdb.org/documentation/policy#toc_release

Reviewer Entry

Reviewer 1

Comments:

Accept.

Reviewer 2

Comments:

Accept.

5. Organizational infrastructure

R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

V. Organizational Infrastructure

Global Organization and Role:

The PDB archive is managed jointly by the Worldwide Protein Data Bank partnership (wwPDB; wwpdb.org)(1) and consists of the RCSB Protein Data Bank (2, 3), Protein Data Bank Japan (PDBj) (4-5), the Protein Data Bank in Europe (PDBe)(6-7), and the BioMagResBank (BMRB)(6). The wwPDB organization operates under a formal agreement (wwpdb.org/about/agreement), most recently renewed in 2013. This agreement commits wwPDB partners to collect, standardize, validate, annotate, and store macromolecular structure data in a single global archive for data depositors and to disseminate these data via a web repository to data consumers, all at no charge with no restrictions on data usage.

Each of the wwPDB partner organizations independently obtains funding for the operations of the PDB. Funding cycles are typically 3-5 years in duration. RCSB PDB is supported by grants from the National Science Foundation, the US Department of Energy, and the National Cancer Institute, the National Institute of Allergy and Infectious Diseases, and the National Institute of General Medical Sciences of the National Institutes of Health. PDBe is supported by the European

Molecular Biology Laboratory, Wellcome Trust, Biotechnology and Biological Sciences Research Council, the National Institutes of Health, and the European Union. PDBj is supported by the National Bioscience Database Center-Japan Science and Technology Agency. The BMRB is supported by the National Institute of General Medical Sciences.

PDB operations are conducted by highly skilled staff with broad domain expertise in experimental structural biology, life sciences and medical applications, data science, information technology, and software engineering. The Principal Investigators (PIs) for each wwPDB partner site form provide overall leadership for the wwPDB organization. The PIs meet regularly to set direction and priorities for the wwPDB team. A wwPDB OneDep Project Manager leads operational and production activities of the wwPDB and provides routine progress reports to the wwPDB PIs. Developers and Biocurators from all project sites meet virtually on a bi-weekly schedule to review operational and development activities. Current staffing commitments from the partner sites for wwPDB biocuration and development include 12.6 FTEs from RCSB, 8.3 FTEs from PDBe, 5.8 FTEs from PDBj, and 1.85 FTEs from BMRB. These commitments are adjusted annually based on the current rates of data deposition and planned development activities.

The full staff for each of the wwPDB partner sites and their PDB-related publications are enumerated in the following links.

PDBe Team at the European Molecular Biology Laboratory/European Bioinformatics Institute:

Staff: <http://www.ebi.ac.uk/about/people/sameer-velankar>

Publications:

<https://www.ebi.ac.uk/pdbe/about/publications>

<https://www.ebi.ac.uk/pdbe/about/events>

PDBj and BMRBj Teams at Osaka University:

Staff: <https://pdbj.org/info/staff>

Publications: <https://pdbj.org/info/publications>

BMRB at University of Connecticut:

Staff: <https://bmr.io/bmr/staff.shtml>

RCSB PDB Team at Rutgers University and University of California at San Diego (UCSD):

Staff: <https://www.rcsb.org/pages/team>

Publications: <https://www.rcsb.org/pages/publications>

wwPDB joint publications:

<https://www wwpdb.org/about/publications>

<https://www wwpdb.org/about/outreach>

1. Berman HM, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. *Nat Struct Biol.* 2003;10(12):980. DOI: 10.1038/nsb1203-980. PubMed PMID: 14634627; PMCID: 1669775.
2. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res.* 2000;28(1):235-42. Epub 1999/12/11. DOI: 10.1093/nar/28.1.235. PubMed PMID: 10592235; PMCID: 102472.
3. Burley SK, Bhikadiya C, Bi C, Bittrich S, Chen L, Crichlow G, Christie CH, Dalenberg K, Costanzo LD, Duarte JM, Dutta S, Feng Z, Ganesan S, Goodsell DS, Ghosh S, Green RK, Guranovic V, Guzenko D, Hudson BP, Lawson CL, Liang Y, Lowe R, Namkoong H, Peisach E, Persikova I, Randle C, Rose A, Rose Y, Sali A, Segura J, Sekharan M, Shao C, Tao Y-P, Voigt M, Westbrook JD, Young JY, Zardecki C, Zhuravleva M. RCSB Protein Data Bank: Powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering, and energy sciences. *Nucleic Acid Research.* 2020:gkaa1038. DOI: 10.1093/nar/gkaa1038.
4. Kinjo AR, Bekker GJ, Suzuki H, Tsuchiya Y, Kawabata T, Ikegawa Y, Nakamura H. Protein Data Bank Japan (PDBj): updated user interfaces, resource description framework, analysis tools for large structures. *Nucleic Acids Res.* 2017;45(D1): D282-D8. DOI: 10.1093/nar/gkw962. PubMed PMID: 27789697; PMCID: PMC5210648.
5. Kinjo, A.R., Bekker, G.-J., Wako, H., Endo, S., Tsuchiya, Y., Sato, H., Nishi, H., Kinoshita, K., Suzuki, H., Kawabata, T., Yokochi, M., Iwata, T., Kobayashi, N., Fujiwara, T., Kurisu, G. and Nakamura, H. New tools and functions in Data-out activities at Protein Data Bank Japan (PDBj). *Protein Science* 27 (1): 95-102 (2018) [PubMed] [doi:10.1002/pro.3273]
6. Armstrong DR, Berrisford JM, Conroy MJ, Gutmanas A, Anyango S, Choudhary P, Clark AR, Dana JM, Deshpande M, Dunlop R, Gane P, Gáborová R, Gupta D, Haslam P, Koča J, Mak L, Mir S, Mukhopadhyay A, Nadzirin N, Nair S, Paysan-Lafosse T, Pravda L, Sehnal D, Salih O, Smart O, Tolchard J, Varadi M, Svobodova-Vařeková R, Zaki H, Kleywegt GJ, Velankar S., PDBe: improved findability of macromolecular structure data in the PDB. *Nucleic acids research* Volume 48 (2020) p.D335-D343 DOI: 10.1093/nar/gkz990
7. Velankar S, van Ginkel G, Alhroub Y, Battle GM, Berrisford JM, Conroy MJ, Dana JM, Gore SP, Gutmanas A, Haslam P, Hendrickx PM, Lagerstedt I, Mir S, Fernandez Montecelo MA, Mukhopadhyay A, Oldfield TJ, Patwardhan A, Sanz-Garcia E, Sen S, Slowley RA, Wainwright ME, Deshpande MS, Iudin A, Sahni G, Salavert Torres J, Hirshberg M, Mak L, Nadzirin N, Armstrong DR, Clark AR, Smart OS, Korir PK, Kleywegt GJ. PDBe: improved accessibility of macromolecular structure data from PDB and EMDB. *Nucleic Acids Res.* 2016;44(D1): D385-95. DOI: 10.1093/nar/gkv1047. PubMed PMID: 26476444; PMCID: PMC4702783.
8. Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Kent Wenger R, Yao H, Markley JL. BioMagResBank. *Nucleic Acids Res.* 2008; 36(Database issue):D402-8. Epub 2007/11/07. DOI: 10.1093/nar/gkm957. PubMed PMID: 17984079; PMCID: 2238925.

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept

6. Expert guidance

R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either inhouse or external, including scientific guidance, if relevant).

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

VI. Expert Guidance

The wwPDB and each of the partner organizations have external advisories that provide both general and scientific guidance.

wwPDB Advisory:

The wwPDB Advisory Committee comprises an international team of experts in X-ray crystallography, cryoEM, NMR, and bioinformatics. The advisory meets annually. Advisory membership and meeting details can be found at <https://www.wwpdb.org/about/advisory>.

wwPDB Community Task Forces:

The wwPDB has also established method-specific Task Forces and Working Groups to provide recommendations for best

practices for data content and data quality. These groups include leading experts and application developers in each method area. The following links provide further details of the activities of these groups.

X-ray Validation Task Force:

<https://www.wwpdb.org/task/xray>

NMR Validation Task Force:

<https://www.wwpdb.org/task/nmr>

EM Validation Task Force:

<https://www.wwpdb.org/task/em>

Small Angle Scattering Task Force:

<https://www.wwpdb.org/task/sas>

Ligand Validation:

<https://www.wwpdb.org/task/ligand-validation-workshop>

Integrative and Hybrid Methods Task Forces:

<https://www.wwpdb.org/task/hybrid>

Publications describing the recommendations from these Task Forces and Working Groups are listed on the wwPDB

Publications page: <https://www.wwpdb.org/about/publications>

wwPDB Partner Advisory Committees:

The wwPDB partner projects also sponsor their own external Scientific Advisory Committees. Each partner project receives further oversight from their respective funding bodies.

RCSB PDB Advisory Committee: <https://www.rcsb.org/pages/advisory-committee>

PDBe Advisory Committee: <https://www.ebi.ac.uk/pdbe/about/sac>

BMRB Advisory Committee: https://bmr.io/bmr/aims_and_policies.shtml

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept.

DIGITAL OBJECT MANAGEMENT

7. Data integrity and authenticity

R7. The repository guarantees the integrity and authenticity of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

Digital Object Management

VII. Data integrity and authenticity

Data Integrity

The wwPDB uses the macromolecular Crystallographic Information Framework (mmCIF) (1-6) as a metadata standard. mmCIF was originally developed by the International Union of Crystallography (IUCr) (7). Since 2014, the wwPDB, together with the PDBx/mmCIF Working Group (<https://www.wwpdb.org/task/mmcif>) oversee the evolution of the standard. The details of the PDBx/mmCIF metadata specifications, tutorial information, and links to supporting software tools are maintained at PDBx/mmCIF Resource Site (<http://mmcif.wwpdb.org/>). In 2019, PDBx/mmCIF became the official deposition standard for the wwPDB (8).

The PDBx/mmCIF framework provides a rich collection of software accessible metadata that allows PDB data processing tools to assess compliance with a particular version of the data standard. This includes tests for required data fields, conformance with controlled vocabularies and boundary values, referential integrity within and between data sections. Changes in the metadata specification are versioned and audited by a revision history within the specification. Data files record this version information to permit compliance to be evaluated. Revisions are managed such that changes within a major version of the PDBx/mmCIF metadata standard are backward-compatible with all data files in the repository. Major version changes in the metadata standard are accompanied by an administrative update to all of the repository data files. These administrative updates are typically focused on improving consistency (e.g., nomenclature) or organization of the archive entries and do not change the primary data values (e.g., atomic positions, structure factor amplitudes, or chemical shifts).

PDB data entries include an internal version number and a revision history that records the entries' changes at the granularity of data category and data item (9). An instance of the wwPDB repository containing explicitly versioned data files was released in late 2017 (10). The versioned repository was introduced in parallel with the traditional ftp repository. The latter contains only the most recent version of each data entry. The versioned organization permits retaining all major versions (i.e., latest minor version) of each entry in the active repository.

Data Authenticity:

PDB entries are revised either at the request of the depositor of record or through an administrative update performed by the Biocuration staff. Change requests from depositors are managed following published policies (11) and are audited within each entry. As described in the previous section, each data entry contains a version and a detailed revision history which identifies specific entry modifications at the level of individual data items. Entries also record the version of metadata to which they comply. Revisions made by the Biocuration staff targeted at improving data consistency or organization are similarly audited within affected entries. Advance notification of 60 days is provided to repository users of planned administrative or remediation changes in the PDB repository (12). Representative example data files illustrating substantial administrative changes are also provided to users in advance of any changes in the production data repository content.

The combination of versioning and revision history provides software accessible tracking of the provenance and specific details of any changes to PDB data entries. This makes it possible for a user to appreciate how any entry has changed since its initial release in the repository.

The wwPDB creates and maintains annual snapshots of the state of the full repository (<ftp://snapshots.wwpdb.org>). Snapshots are also created before any significant administrative or remediation repository update.

Depositor Identity:

PDB depositors self-identify during the deposition process. Contact details for both conventional mail and e-mail are collected. A principal investigator must be identified for each deposition session which may differ from the contact author. While there is currently no universal digital identifier used in this domain, ORCIDs (<https://orcid.org/>) are required information for each depositor.

Data Access:

The wwPDB partners provide a wide variety of search, analysis, and reporting features. These search services all share the common structure and experimental data delivered in the wwPDB repository, synchronized weekly in concert with the wwPDB update schedule. Search services are delivered by resources hosted in each of the wwPDB partner institutions. The entry points for these services are described and presented on the wwPDB website (wwpdb.org). These web entry points are enumerated in section R13. PDB data are distributed in a structured form conforming to a versioned underlying schema and/or data dictionary. These data specifications provide the foundation for validating the essential data integrity of wwPDB data products. A further level of scientific validation (Section 11) is performed and these results accompany each released PDB entry. Access to wwPDB schema and dictionaries is provided in both searchable and browsable form at mmcif.wwpdb.org.

1. Fitzgerald PMD, Westbrook JD, Bourne PE, McMahon B, Watenpaugh KD, Berman HM. 4.5 Macromolecular dictionary (mmCIF). In: Hall SR, McMahon B, editors. *International Tables for Crystallography G Definition and exchange of crystallographic data*. Dordrecht, The Netherlands: Springer; 2005. p. 295-443.
2. Westbrook J, Yang H, Feng Z, Berman HM. 5.5 The use of mmCIF architecture for PDB data management. In: Hall SR, McMahon B, editors. *International Tables for Crystallography*. Dordrecht, The Netherlands: Springer; 2005. p. 539-43.
3. Westbrook JD, Berman HM, Hall SR. 2.6 Specification of a relational Dictionary Definition Language (DDL2). In: Hall SR, McMahon B, editors. *International Tables for Crystallography*. Dordrecht, The Netherlands: Springer; 2005. p. 61-72.
4. Westbrook JD, Hall SR. 4.10 DDL2 dictionary. In: Hall SR, McMahon B, editors. *International Tables for Crystallography*. Dordrecht, The Netherlands: Springer; 2005. p. 473-81.
5. Westbrook JD, Bourne PE. STAR/mmCIF: an ontology for macromolecular structure. *Bioinformatics*. 2000;16(2):159-68. PubMed PMID: 10842738.
6. Westbrook J, Yang H, Feng Z, Berman H, editors. *The use of mmCIF architecture for PDB data management*, Eds., pp. Dordrecht, The Netherlands: Springer; 2005.
7. IUCr CIF Specification: <http://www.iucr.org/resources/cif/spec>
8. Paul D. Adams, Pavel V. Afonine, Kumaran Baskaran, Helen M. Berman, John Berrisford, Gerard Bricogne, David G. Brown, Stephen K. Burley, Minyu Chen, Zukang Feng, Claus Flensburg, Aleksandras Gutmanas, Jeffrey C. Hoch, Yasuyo Ikegawa, Yumiko Kengaku, Eugene Krissinel, I Genji Kurisu, Yuhe Liang, Dorothee Liebschner, Lora Mak, John L. Markley, Nigel W. Moriarty, Garib N. Murshudov, Martin Noble, Ezra Peisach, Irina Persikova, Billy K. Poon, Oleg V. Sobolev, Eldon L. Ulrich, Sameer Velankar, Clemens Vonrhein, John Westbrook, Marcin Wojdyr, Masashi Yokochij and Jasmine Y. Young. Announcing mandatory submission of PDBx/mmCIF format files for crystallographic depositions to the

Protein Data Bank (PDB) (2019) Volume 75, Pages 451-454 doi:10.1107/S2059798319004522

9. See for instance, the PDBx/mmCIF metadata categories: `pdbx_audit_revision_history`, `pdbx_audit_revision_category`, and `pdbx_audit_revision_item` at http://mmcif.wwpdb.org/dictionaries/mmcif_pdbx_v50.dic/Groups/index.html

10. wwPDB news item describing the delivery of a repository versioning:

<http://www.wwpdb.org/news/news?year=2017#5979ef4161fd3d50915a4b00>

11. wwPDB Deposition and Biocuration policies: <http://www.wwpdb.org/documentation/policy>

12. wwPDB Remediation activities: <https://www.wwpdb.org/documentation/remediation>

Reviewer Entry

Reviewer 1

Comments:

Accept

Reviewer 2

Comments:

Accept.

8. Appraisal

R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

VIII. Appraisal

Deposition Standards:

Content requirements for PDB depositions are fully described in public project documentation (1). Tutorials, FAQs, and

video guides complement this documentation with specific deposition examples. The scope of data collected reflects community recommendations (See Section VI) to provide for data quality assessment and to broadly enable data reuse.

The PDB deposition system (OneDep) (2) manages the interactions with depositors through a web-based user interface. This system guides the depositor through the deposition process, ensuring that all required data items are provided and are compliant with the community PDBx/mmCIF data standard (<http://mmcif.wwpdb.org>). Any uploaded data files in a supported format are checked for format compliance and data integrity. Any anomalies are reported to the depositing user for corrective action. The depositor must address all such issues to finalize a deposition session and receive a PDB accession code. In addition to verifying format compliance, basic data integrity, and completeness, data are subjected to a rigorous data quality assessment. The depositor must review this quality assessment before finalizing a deposition, and the depositor is encouraged to address any issues arising from this assessment. Because it may not be possible for the depositor to address all of these data quality issues, failure to address all of these issues does not block completing the deposition process.

Data Formats:

A small number of mature data formats are commonly used in structural biology to store atomic coordinates and supporting experimental data (3). These formats are directly supported by the OneDep deposition system (2). In 2019, PDBx/mmCIF became the deposition standard for crystallographic depositions (4). PDB also provides data assembly and harvesting tools (<https://pdb-extract.wwpdb.org>) to enable depositors to prepare compliant and complete data files for deposition extracted from content stored in intermediate data and application log files.

1. wwPDB Deposition and Biocuration Policies:

<https://www.wwpdb.org/documentation/policy>

2. J. Young, J. D. Westbrook, Z. Feng, R. Sala, E. Peisach, T. J. Oldfield, S. Sen, A. Gutmanas, D. R. Armstrong, J. M. Berrisford, L. Chen, M. Chen, L. Di Costanzo, D. Dimitropoulos, G. Gao, S. Ghosh, S. Gore, V. Guranovic, P. M. S. Hendrickx, B. P. Hudson, R. Igarashi, Y. Ikegawa, N. Kobayashi, C. L. Lawson, Y. Liang, S. Mading, L. Mak, M. S. Mir, A. Mukhopadhyay, A. Patwardhan, I. Persikova, L. Rinaldi, E. Sanz-Garcia, M. R. Sekharan, C. Shao, G. J. Swaminathan, L. Tan, E. L. Ulrich, G. Van Ginkel, R. Yamashita, H. Yang, M. A. Zhuravleva, M. Quesada, G. J. Kleywegt, H. M. Berman, J. L. Markley, H. Nakamura, S. Velankar, S. K. Burley. (2017) OneDep: Unified wwPDB System for Deposition, Biocuration, and Validation of Macromolecular Structures in the Protein Data Bank (PDB) Archive. *Structure* 25: 536-545 DOI: 10.1016/j.str.2017.01.004

3. wwPDB File Format Documentation:

<https://www.wwpdb.org/documentation/file-format>

4. Paul D. Adams, Pavel V. Afonine, Kumaran Baskaran, Helen M. Berman, John Berrisford, Gerard Bricogne, David G. Brown, Stephen K. Burley, Minyu Chen, Zukang Feng, Claus Flensburg, Aleksandras Gutmanas, Jeffrey C. Hoch, Yasuyo Ikegawa, Yumiko Kengaku, Eugene Krissinel, I Genji Kurisu, Yuhe Liang, Dorothee Liebschner, Lora Mak, John L. Markley, Nigel W. Moriarty, Garib N. Murshudov, Martin Noble, Ezra Peisach, Irina Persikova, Billy K. Poon, Oleg V. Sobolev, Eldon L. Ulrich, Sameer Velankar, Clemens Vonrhein, John Westbrook, Marcin Wojdyr, Masashi Yokochij and Jasmine Y. Young. Announcing mandatory submission of PDBx/mmCIF format files for crystallographic depositions to the

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept.

9. Documented storage procedures

R9. The repository applies documented processes and procedures in managing archival storage of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

IX. Documented Storage Procedures

OAIS Archive Reference Model Support:

Conformance with the OAIS Archive Reference Model (1) requires the support of the OAIS information model concepts in OAIS Section 2.2. The OAIS Archive Reference Model concepts have the following correspondences in the context of the PDB archive. Implementing the responsibilities enumerated in OAIS Section 3.1 describing compliance with the reference model is provided in Sections I to XVI of this application for certification.

Producers: The depositor community for the wwPDB includes researchers primarily in the field of structural biology.

Consumers: The wwPDB has an international community of users, including biologists (in fields such as structural biology, biochemistry, genetics, pharmacology); other scientists (in fields such as bioinformatics, software developers for data analysis and visualization); student and educators (all levels); media writers, illustrators, textbook authors; and the general public. Data producers are also data consumers in the context of the PDB.

Designated Community: In the context of the PDB, all consumers, both specialists, and non-specialists are considered part of the Designated Community.

OAIS Archive: The archival data are stored and delivered in the wwPDB data repository.

Management: The wwPDB organization is responsible for managing the data deposition, biocuration, and archiving of data in the wwPDB repository.

Representation Information: PDBx/mmCIF metadata dictionary describing each data item's semantics in the PDB data repository.

Data Object: Individual data files containing the 3D structure data, descriptive metadata, and supporting experimental data are stored in the wwPDB repository. Data objects provide internal linkage to a particular version of the PDBx/mmCIF metadata dictionary containing the required representation information.

Preservation Descriptive Information: PDB data objects contain provenance information, including detailed revision history (3), reference information in the form of accession codes (4), and contextual information provided by descriptive metadata (PDBx/mmCIF). The latter contextual information may include the biological role, relationships with other data resources (e.g., reference sequence databases, taxonomy, function ontologies), and other experimental method details. Fixity details (e.g., data file checksums) are maintained implicitly by data transfer protocols but are not exposed in the repository. Access rights for the repository are described in the PDB Advisory Notice (5).

Packaging Information:

The packaging information for the wwPDB repository is described in the PDB download instructions (6).

IX. Documented storage procedures (additional questions)

Data Storage and Data Management Workflows:

The full life cycle of PDB data is documented in the wwPDB Processing Procedures and Policies Document (7). The deposition, biocuration, and archiving tasks in the PDB data life cycle are implemented by the standard workflows performed by the wwPDB OneDep Biocuration platform (2).

Data Security Requirements:

PDB data deposition is performed through the OneDep web user interface. Each deposition session is password-protected, and a secure web protocol (HTTP over TLS/SSL) is used. Before data release into the public archive, deposition session data access is limited to the depositing user and PDB Biocuration staff. Communication between the Depositor and the Biocurator regarding the content of a deposition session conducted through the password-protected secure web channel.

Data Preservation Policy:

The life cycle policy documentation includes the long-term storage of data sets in the single common data archive replicated and delivered globally at the BMRB, PDBe, PDBj, and RCSB PDB partner sites.

Repository Replication:

The wwPDB releases data into the PDB repository on a coordinated weekly schedule. Globally, BMRB, PDBe, PDBj, and RCSB PDB deliver synchronized copies of the wwPDB repository. Each of these partner sites also maintains redundant on-line copies of the wwPDB repository to support performant and highly available user access.

Access to the global wwPDB ftp archive access site is described in detail (6). Annual snapshots and milestone copies of the wwPDB ftp archive are also maintained at <ftp://snapshots.wwpdb.org/> and <ftp://snapshots.pdbj.org/>.

Data Recovery and Data Availability:

Recovery of current and historical data in the public archive is provided through local and global replication of the wwPDB repository. In addition to the weekly replication of public archive data to each of the wwPDB partner sites, the wwPDB OneDep Biocuration platform provides additional support for the exchange of deposition sessions between wwPDB partner sites. This capability allows for failover of in-progress deposition services between wwPDB sites in a loss of service. This functionality is regularly used to maintain availability and accommodate any required periods of data center maintenance.

Risk Management:

The principal objectives of the wwPDB organization are providing for the security, availability, and long-term preservation of the PDB data repository and supporting deposition and biocuration services. As an organization, this is achieved by adopting common practices for data management and providing robust infrastructure to support the hosting and delivery of repository services and data. The multi-site capability established by the wwPDB organization avoids a single point of failure, thereby reducing the risk of adverse impacts on continuity of access or long-term stability of the archive. This

partnership further provides global redundancy in deposition and biocuration services, reducing the risk of losing data acquisition services.

Data Consistency:

The wwPDB releases data on a coordinated weekly schedule. The BMRB, PDBe, PDBj, and RCSB PDB wwPDB partners maintain synchronized copies of the repository. The RSYNC protocol used for data synchronization provides an internal checksum mechanism as part of the data transfer operation to ensure consistency among the repository copies. Additional indices of content provided within the repository allow software tools at each site to further verify the completeness and correctness of any update.

Media Refresh:

wwPDB relies primarily on multiple on-line copies of archival data files hosted on Enterprise quality storage hardware. This hardware auto-detects and corrects for media inconsistencies or failures. Where tape media is used for off-line archiving, tape copies are refreshed periodically by the backup system software (e.g., IBM Tivoli Storage Manager Scheduler), and physical tape media is replaced on a regular schedule.

1. Reference Model for an Open Archival Information System (OAIS). Magenta Book CCSDS 6500-M-2. Washington: Consultative Committee for Space Data Systems, NASA; 2012.
2. J. Young, J. D. Westbrook, Z. Feng, R. Sala, E. Peisach, T. J. Oldfield, S. Sen, A. Gutmanas, D. R. Armstrong, J. M. Berrisford, L. Chen, M. Chen, L. Di Costanzo, D. Dimitropoulos, G. Gao, S. Ghosh, S. Gore, V. Guranovic, P. M. S. Hendrickx, B. P. Hudson, R. Igarashi, Y. Ikegawa, N. Kobayashi, C. L. Lawson, Y. Liang, S. Mading, L. Mak, M. S. Mir, A. Mukhopadhyay, A. Patwardhan, I. Persikova, L. Rinaldi, E. Sanz-Garcia, M. R. Sekharan, C. Shao, G. J. Swaminathan, L. Tan, E. L. Ulrich, G. Van Ginkel, R. Yamashita, H. Yang, M. A. Zhuravleva, M. Quesada, G. J. Kleywegt, H. M. Berman, J. L. Markley, H. Nakamura, S. Velankar, S. K. Burley. (2017) OneDep: Unified wwPDB System for Deposition, Biocuration, and Validation of Macromolecular Structures in the Protein Data Bank (PDB) Archive. *Structure* 25: 536-545 DOI: 10.1016/j.str.2017.01.004
3. See the PDBx/mmCIF metadata categories: `pdbx_audit_revision_history`, `pdbx_audit_revision_category`, and `pdbx_audit_revision_item` at http://mmcif.wwpdb.org/dictionaries/mmcif_pdbx_v50.dic/Groups/index.html
4. See the PDBx/mmCIF metadata category: http://mmcif.wwpdb.org/dictionaries/mmcif_pdbx_v50.dic/Categories/database_2.html
5. wwPDB Repository Advisory Notice: <https://ftp.wwpdb.org/pub/pdb/advisory.txt>.
6. wwPDB Download Instructions: <https://www.wwpdb.org/download/downloads>
7. wwPDB Deposition and Biocuration Policies: <https://www.wwpdb.org/documentation/policy>

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept

10. Preservation plan

R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

X. Preservation Plan

Preservation Plan:

Documentation of the PDB data life cycle is described in (1). This document describes the deposition requirements, policies for assigning accession codes, release procedures including embargo provisions, archiving/preservation of data products in the PDB repository, and post-release change management policies for both depositor and the repository. The transformations of deposited data that may accompany the repository's Biocuration data processing are separately documented (2).

The policies regarding entry change by depositors following data release in the PDB repository are sensitive to the type of change. Substantial post-release changes to an entry impacting primary data have historically been accompanied by re-accessioning the entry. wwPDB policies have recently been extended to provide a more flexible versioning scheme for selected depositor-initiated changes (3). In all cases where an entry is substantially changed after release, both the prior and revised entry remain available in the archive.

Deposition, Biocuration, and Archiving Policies:

The policy documents described in the previous section (1,2) document deposition requirements (e.g., data content and data file formats), the transformations performed by the repository during Biocuration, and the packaging of data products for release in the PDB repository are well documented (1-2). The requirements and responsibilities of the depositor and the repository are enumerated in these documents. The long-term preservation of the data products in the PDB public repository and the change management policies by both the depositor and repository are described. These change management policies include the retrospective Biocuration of the repository conducted by the wwPDB (4) to maintain a high-level consistency across the archive.

Compliance:

As a condition for receiving a PDB accession code at the time of submission, depositors must accept the terms described in the wwPDB policy documents (1-2). PDB deposition is a pre-condition for publication in all major scientific journals when publishing a new structure determination study. PDB deposition is also a requirement of many public research funding agencies.

1. For general deposition and release policies, see <https://www.wwpdb.org/documentation/policy>.
2. For Biocuration policies, see: <https://www.wwpdb.org/documentation/procedure>
3. For versioning policies, see: [https://www.wwpdb.org/news/news?year=2017 - 5979ef4161fd3d50915a4b00](https://www.wwpdb.org/news/news?year=2017-5979ef4161fd3d50915a4b00)
4. For retrospective Biocuration/remediation see: <https://www.wwpdb.org/documentation/remediation>

Data Recovery and Data Availability:

Recovery of current and historical data in the public archive is provided through local and global replication of the wwPDB repository. In addition to the weekly replication of public archive data to each of the wwPDB partner sites, the wwPDB OneDep Biocuration platform provides additional support for the exchange of deposition sessions between wwPDB partner sites. This capability allows for failover of in-progress deposition services between wwPDB sites in a loss of service. This functionality is regularly used to maintain availability and accommodate any required periods of data center maintenance. Annual snapshots and milestone copies of the wwPDB ftp archive are also maintained at <ftp://snapshots.wwpdb.org/> and <ftp://snapshots.pdbj.org/>.

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept.

11. Data quality

R11. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

XI. Data Quality

Compliance with Data Standards:

The wwPDB OneDep system guides the depositor through the deposition process, ensuring that all required data items are provided and are compliant with the community PDBx/mmCIF data standard (<http://mmcif.wwpdb.org>). Any uploaded data files in a supported format are checked for format compliance and data integrity. Any anomalies are reported to the depositing user for corrective action. The depositor must address all such issues to finalize a deposition session and receive a PDB accession code. During Biocuration, processing data are similarly managed using PDBx/mmCIF. Tools within the OneDep software system perform test conformance with the data standard at each processing step and ensure that final data products of Biocuration are compliant with the data standard.

Data Quality Assessment:

In addition to verifying basic data integrity, completeness, and compliance with the PDBx/mmCIF metadata standard, data are subjected to a rigorous data quality assessment. This assessment is informed by recommendations from the wwPDB Task Forces and Working Groups described in Section VI. Expert Guidance. These recommendations include data quality content and software applications available to the community to compute the target quality metrics. The assessment's particular content depends on the structure determination but typically includes checks of molecular geometry, stereochemistry, and structure model fit to experimental data. The data quality assessment for each entry is compiled in a

textual validation report and a software accessible data file. Validation reports are provided in the repository for each structure entry in the PDB archive. A complete description of the report content and a selection of example validation reports are available at wwpdb.org (1).

Presentations of data quality assessment have been developed for both specialist and non-specialist users. For depositors and editorial reviewers, the full details of all assessment criteria are tabulated. Many key journals that publish macromolecular structural data require that wwPDB validation reports accompany a manuscript submission. For non-specialist users, a simple graphical depiction is provided that highlights a small number of essential quality metrics (2). These metrics are presented in both absolute and relative terms, where the latter provides a percentile ranking among entries of a similar class within the archive.

Access to data quality assessments in wwPDB validation reports has been demonstrated to positively impact the overall data quality in the PDB archive (3). To enable the depositors to take advantage of this information before deposition, wwPDB provides validation reports through an anonymous web server and a web API (4-5).

The wwPDB provides access to a wide variety of community vetted validation metrics and comparative ranking information. The wwPDB does not collect or distribute ranking or commentary on individual entries by 3rd parties.

Related Citations:

PDB entries are linked to a primary citation which is the case for more than 80% of the archive entries. The primary citation's metadata includes a PubMed identifier and Digital Object Identifier (DOI) if these are available. Each PDB entry is individually assigned a DOI, which links the structure entry in the PDB archive. Any entry, its metadata, and its associated data quality assessments can be reached from the landing page connected to the DOI such as dx.doi.org/10.2210/pdb1KIP/pdb/. All of these data products are similarly accessible from the search tools provided by each of the wwPDB partners.

1. wwPDB validation reports:

<https://www.wwpdb.org/validation/validation-reports>

2. See Section I. Quality at Glance in the example validation report:

https://www.wwpdb.org/validation/samples/1cbs/1cbs_validation.pdf

3. Shao C, Yang H, Westbrook JD, Young JY, Zardecki C, Burley SK. Multivariate Analyses of Quality Metrics for Crystal Structures in the Protein Data Bank Archive. *Structure*. 2017;25:458-68. DOI: 10.1016/j.str.2017.01.013.

4. OneDep validation servers:

<https://www.wwpdb.org/validation/validation-servers>

5. OneDep validation API:

<https://www.wwpdb.org/validation/onedep-validation-web-service-interface>

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept.

12. Workflows

R12. Archiving takes place according to defined workflows from ingest to dissemination.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

XII. Workflows

Deposition, Biocuration, and Archiving Workflows:

The policy documents described in the previous section (1,2) document deposition requirements and steps (e.g., data content and deposition tasks), the transformations performed by the repository during Biocuration, and the packaging of data products for release in the PDB repository are documented here (1-2). These procedures are published on the wwPDB Portal website, and depositors must acknowledge handling of data as described in these procedures before receiving a PDB accession code. Various aspects of these procedures have been described in previous sections VIII. Appraisal, IX. Storage Procedures, X. Preservation Plan, and XI. Data Quality.

The data processing operations performed by the wwPDB are technically implemented as a set of standard workflows executed by the wwPDB OneDep platform (3). These workflows guarantee that all data processing operations are performed uniformly. Workflows enforce standard conventions for naming and versioning data objects, reporting diagnostic information, and recording auditing details. Workflows are used to instrument computational and repetitive

tasks and then deliver their outputs in a standard manner to a Depositor or Biocurator for review. Biocurators interact with OneDep workflows through an administrative web application. This application allows the Biocurator to access, track, and control the annotation activities for multiple entries.

OneDep workflows are described in simple declarative syntax to adapt to changing scientific and technical requirements flexibly.

Workflows also automate data processing operations associated with release processing in the life cycle of PDB data entries. This automation includes additional status verification, which reinforces PDB security protocols preventing pre-disclosure of embargoed data sets.

Data Scope and Data Extensibility:

The scope of PDB depositions includes atomic coordinates that are substantially determined by experimental measurements on actual sample specimens containing biological macromolecules (4). Currently, coordinate sets produced by X-ray crystallography, NMR, electron microscopy, neutron diffraction, powder diffraction, and fiber diffraction can be deposited to the PDB, provided the molecule studied meets the minimum size requirement. Other data repositories (e.g., Cambridge Crystallographic Data Center) are responsible for archiving experimental data for smaller molecular structures. Supplementary experiment datasets are similarly managed by their designated community domain archival repository. For example, small-angle scattering datasets that may accompany a PDB entry are curated and managed by the Small Angle Scattering Biological Data Bank (SASBDB, www.sasbdb.org).

Within this broad scope, there is a continuing need to extend data content to keep pace with the rapidly evolving landscape methods and technologies in experimental structural biology. In this regard, the PDB data architecture benefits from the extensibility of the PDBx/mmCIF data architecture (mmcif.wwpdb.org). PDBx/mmCIF is designed to support facile content extension, and software tools supporting the OneDep platform take advantage of this feature. Because OneDep software tools are designed to adapt to extensions in the PDBx/mmCIF metadata dictionary, OneDep workflows naturally accommodate new data content.

Workflow change management:

Workflows are an integral part of the OneDep software platform that is shared by all wwPDB partner sites. The global OneDep project team manages changes in workflows and other aspects of the OneDep software. This team meets bi-weekly to discuss and address operational, production, and development issues associated with the OneDep platform. Changes required in the platform are scheduled for development, and after appropriate testing, are released and deployed at all partner sites. The deliberations of the project team, development timelines, and related project management documentation are stored in a project collaboration system, Confluence. Project development is tracked using the issue management tool, Jira. Software changes are managed using the Git version control software. Software testing and deployment workflows are orchestrated using tools provided by GitHub and GitLab.

Advance notification of 60 days is provided to depositors and repository users of substantial changes in wwPDB deposition, biocuration, or archiving procedures.

1. For deposition and archiving procedures, see <https://www.wwpdb.org/documentation/policy>.
2. For Biocuration procedures, see: <https://www.wwpdb.org/documentation/procedure>
3. J. Young, J. D. Westbrook, Z. Feng, R. Sala, E. Peisach, T. J. Oldfield, S. Sen, A. Gutmanas, D. R. Armstrong, J. M. Berrisford, L. Chen, M. Chen, L. Di Costanzo, D. Dimitropoulos, G. Gao, S. Ghosh, S. Gore, V. Guranovic, P. M. S. Hendrickx, B. P. Hudson, R. Igarashi, Y. Ikegawa, N. Kobayashi, C. L. Lawson, Y. Liang, S. Mading, L. Mak, M. S. Mir, A. Mukhopadhyay, A. Patwardhan, I. Persikova, L. Rinaldi, E. Sanz-Garcia, M. R. Sekharan, C. Shao, G. J. Swaminathan, L. Tan, E. L. Ulrich, G. Van Ginkel, R. Yamashita, H. Yang, M. A. Zhuravleva, M. Quesada, G. J. Kleywegt, H. M. Berman, J. L. Markley, H. Nakamura, S. Velankar, S. K. Burley. (2017) OneDep: Unified wwPDB System for Deposition, Biocuration, and Validation of Macromolecular Structures in the Protein Data Bank (PDB) Archive. Structure 25: 536-545 DOI: 10.1016/j.str.2017.01.004
4. H.M. Berman, S.K. Burley, W. Chiu, A. Sali, A. Adzhubei, P.E. Bourne, S.H. Bryant, J. Roland L. Dunbrack, K. Fidelis, J. Frank, A. Godzik, K. Henrick, A. Joachimiak, B. Heymann, D. Jones, J.L. Markley, J. Moulton, G.T. Montelione, C. Orengo, M.G. Rossmann, B. Rost, H. Saibil, T. Schwede, D.M. Standley, and J.D. Westbrook (2006) Outcome of a workshop on archiving structural models of biological macromolecules. Structure. 14: 1211-1217

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept.

13. Data discovery and identification

R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

XIII. Data Discovery and Identification

Data and Metadata Search Facilities:

The wwPDB partners provide a wide variety of search, analysis, and reporting features. These search services all share the common structure and experimental data delivered in the wwPDB repository, synchronized weekly in concert with the PDB update schedule. Search services are delivered by resources hosted in each of the wwPDB partner institutions. The entry points for these services are described and presented on the wwPDB website (wwpdb.org). These web entry points are enumerated in the following section, along with links to selected user search documentation and information describing API-level search features.

PDBj:

Web Search: <https://pdbj.org/mine>

Selected Search Documentation:

<https://pdbj.org/help/search-menu>

Search API:

<https://pdbj.org/help/rest-interface>

PDBe:

Web Search:

<https://www.ebi.ac.uk/pdbe/>

<https://www.ebi.ac.uk/pdbe/pdbe-kb>

Selected Search Documentation:

<https://www.ebi.ac.uk/pdbe/training/tutorials>

Search API:

<https://www.ebi.ac.uk/pdbe/pdbe-rest-api>

RCSB PDB:

Web Search:

<https://www.rcsb.org/#Category-search>

<https://www.rcsb.org/search/advanced>

Selected Search Documentation:

<https://www.rcsb.org/search/help/overview>

Search API:

<https://www.rcsb.org/pages/webservices>

BMRB:

Web Search:

<https://bmrbl.io/search>

Search API:

<https://bmrbl.io/bmrbl/linking.shtml>

Metadata Documentation Search Facilities:

The details of the PDBx/mmCIF metadata specifications, tutorial information, and links to supporting software tools and metadata search services are maintained at <http://mmcif.wwpdb.org/>. Metadata dictionaries are maintained at the GitHub repository <https://github.com/wwpdb-dictionaries>.

Repository Registries:

The Protein Data Bank is indexed in the following public repository registries:

Registry of Research Data Repositories:

<http://www.re3data.org/> (<https://www.re3data.org/repository/r3d100011104>)

Identifiers.org (EBI/EMBL):

<https://identifiers.org/> (<https://registry.identifiers.org/registry/pdb>)

Directory of Open Access Repositories (OpenDOAR):

<http://www.opendoar.org/> (<https://v2.sherpa.ac.uk/id/repository/2960>)

FAIRSharing:

<https://fairsharing.org/> (<https://fairsharing.org/FAIRsharing.mckkb4>)

Recommended Data Citations:

The recommended citation for the wwPDB repository is (1).

Citing recommendations (2) for authors and journals includes further details describing citations for individual data sets. For example, an individual PDB entry can be identified by PDB accession code or by the Digital Object Identifier (DOI)

assigned to the entry.

1. H. M. Berman, K. Henrick, H. Nakamura. (2003) Announcing the worldwide Protein Data Bank. Nat Struct Biol 10: 980.

DOI: 10.1038/nsb1203-980

2. wwPDB Citing and Journal information:

<https://www.wwpdb.org/documentation/journals>

Persistent Identifiers:

A Digital Object Identifier (DOI) is assigned to each PDB entry at release time (e.g., for PDB entry 4HHB, 10.2210/pdb4hhb/pdb).

Reviewer Entry

Reviewer 1

Comments:

Accept.

Reviewer 2

Comments:

Accept.

14. Data reuse

R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:

4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:

4 – The guideline has been fully implemented in the repository

Response:

XIV. Data Reuse

Data and Metadata Requirements at Deposition:

As described in Section VIII. Appraisal, content requirements for PDB depositions are described in public project documentation (1). Data requirements are defined in terms of the PDBx/mmCIF data representation (<http://mmcif.wwpdb.org/>), which is an international standard for this community.

Content and Format Extensibility:

A small number of mature data formats are commonly used in structural biology to store atomic coordinates and supporting experimental data (2). These formats are directly supported by the wwPDB OneDep deposition system (3). The primary format for the PDB archive is PDBx/mmCIF. The PDBx/mmCIF content is also serialized and delivered in XML/PDBML format (4-5) and RDF formats (6-7). For backward compatibility, archival data content continues to be delivered in the legacy PDB Format (8).

The PDBx/mmCIF data representation is designed to support facile content extension, and software tools supporting the OneDep platform take advantage of this feature. The PDBx/mmCIF data format inherits the extensibility of the metadata representation. PDBML and RDF transliterations of the PDBx/mmCIF data similarly inherit this data content extensibility.

Maintaining Repository Content and Format Consistency:

wwPDB policies and practices for retrospective biocuration and archive-wide remediation have been described in Sections VII. Data Integrity and Authenticity and X. Preservation Plan. Examples of archive-wide remediation efforts aimed at improving repository consistency are described here (9). The 2017 archive-wide update of data files to conform with the PDBx/mmCIF V5 dictionary is a notable example (10).

Repository Content Documentation:

The details of the PDBx/mmCIF metadata specifications, tutorial information, and links to supporting software tools are maintained at the PDBx/mmCIF Resource Site (11). Chemical and molecular reference data documentation is maintained here (12-13).

The wwPDB partners host a rich collection of documentation, tutorial, and education materials (14-19) describing the PDB data content along with supporting access and analysis tools.

Data and Metadata Access:

Data are distributed in a structured format that includes a reference to the underlying data dictionary and/or schema. The dictionary and schema provide documentation for each item of data in a PDB entry. Conformance to the

dictionary/schema standard is a requirement of data release. Access to supporting data dictionaries and schema is provided at the wwPDB dictionary and schema sites (mmcif.wwpdb.org/pdbml.wwpdb.org).

1. wwPDB Deposition and Biocuration Policies: <https://www.wwpdb.org/documentation/policy>
2. wwPDB File Format Documentation:
<https://www.wwpdb.org/documentation/file-format>
3. J. Young, J. D. Westbrook, Z. Feng, R. Sala, E. Peisach, T. J. Oldfield, S. Sen, A. Gutmanas, D. R. Armstrong, J. M. Berrisford, L. Chen, M. Chen, L. Di Costanzo, D. Dimitropoulos, G. Gao, S. Ghosh, S. Gore, V. Guranovic, P. M. S. Hendrickx, B. P. Hudson, R. Igarashi, Y. Ikegawa, N. Kobayashi, C. L. Lawson, Y. Liang, S. Mading, L. Mak, M. S. Mir, A. Mukhopadhyay, A. Patwardhan, I. Persikova, L. Rinaldi, E. Sanz-Garcia, M. R. Sekharan, C. Shao, G. J. Swaminathan, L. Tan, E. L. Ulrich, G. Van Ginkel, R. Yamashita, H. Yang, M. A. Zhuravleva, M. Quesada, G. J. Kleywegt, H. M. Berman, J. L. Markley, H. Nakamura, S. Velankar, S. K. Burley. (2017) OneDep: Unified wwPDB System for Deposition, Biocuration, and Validation of Macromolecular Structures in the Protein Data Bank (PDB) Archive. *Structure* 25: 536-545 DOI: 10.1016/j.str.2017.01.004
4. Westbrook J, Ito N, Nakamura H, Henrick K, Berman HM. PDBML: the representation of archival macromolecular structure data in XML. *Bioinformatics*. 2005;21(7):988-92. DOI: 10.1093/bioinformatics/bti082. PubMed PMID: 15509603
5. PDBML Schema Resources: <http://pdbml.wwpdb.org/>
6. Kinjo AR, Suzuki H, Yamashita R, Ikegawa Y, Kudou T, Igarashi R, Kengaku Y, Cho H, Standley DM, Nakagawa A, Nakamura H. Protein Data Bank Japan (PDBj): maintaining a structural data archive and resource description framework format. *Nucleic Acids Res*. 2012;40(Database issue):D453-60. Epub 2011/10/07. DOI: 10.1093/nar/gkr811. PubMed PMID: 21976737; PMCID: 3245181.
7. PDB/RDF Mapping and Deliver Details: <https://pdbj.org/help/rdf>
8. PDB Legacy File Format:
<https://www.wwpdb.org/documentation/file-format-content/format33/v3.3.html>
9. wwPDB Data Remediation Activities:
<https://www.wwpdb.org/documentation/remediation>
10. PDB V5 repository update:
[https://www.wwpdb.org/news/news?year=2017 - 5963997661fd3d50915a4af7](https://www.wwpdb.org/news/news?year=2017-5963997661fd3d50915a4af7)
11. wwPDB PDBx/mmCIF Resource Site: <http://mmcif.wwpdb.org>
12. Chemical Reference Data documentation: <https://www.wwpdb.org/data/ccd>
13. Molecular Reference Data documentation: <https://www.wwpdb.org/data/bird>
14. BMRB Educational Materials: <https://bmrbl.io/education/>
15. PDBe Teaching Materials: <https://www.ebi.ac.uk/pdbe/training/teaching-materials>
16. PDBe Tutorials: <https://www.ebi.ac.uk/pdbe/training/tutorials>
17. PDBe Featured PDB Entries: <https://www.ebi.ac.uk/pdbe/quips>
18. PDBj Educational Services: <https://pdbj.org/help/educational-services-menu>
19. RCSB PDB PDB-101 Educational Site: <https://pdb101.rcsb.org/>

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept.

TECHNOLOGY

15. Technical infrastructure

R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

Technology

XV. Technical Infrastructure

Reference Standards and Implementations:

PDBx/mmCIF is the archival format (1) and content standard for the wwPDB (2-5).

The wwPDB PDBx/mmCIF metadata standard has been described in Section VII. Data integrity and authenticity and the implementation in the wwPDB Deposition and Biocuration system in Sections VIII. Appraisal and XII Workflows.

The wwPDB deposition and biocuration (6) embrace standards for:

- IUPAC atomic symbols (7)
- IUPAC atomic weights (8)
- IUPAC one and three-letter abbreviations for standard amino acids (9)
- IUPAC nomenclature atom names in standard amino acids and nucleotides including protonated forms (10)
- wwPDB chemical reference data for constituent molecules including residue and atom nomenclature (11-13)
- IUCr Crystallographic Space-Group Symmetry (14)
- IUPAC-IUBMB Enzyme Nomenclature (15)
- Gene Ontology Consortium - Annotation Terms (16-17)
- NCBI Aggregation of Taxonomy Nomenclature (18-19)
- IUPAC Chemical Identifiers (InChI) (20)
- Open Simplified Molecular-Input Line-Entry System (SMILES) (21)
- Document Object Identifier (DOI) (22)
- Open Researcher and Contributor Id (ORCID) (23)
- Collected standards supporting NMR restraint and chemical shifts data (24)

Standards supporting data formats in the wwPDB repository include:

- IUCr CIF/mmCIF Format Standard (1)
- PDB Legacy Format Standard V3 (25)
- PDBML/XML (26-29)
- PDBML XML/XSD Schema (30-31)
- PDB RDF/OWL (32-38)
- NMRSTAR - BMRB archival format standard for NMR restraints and chemical shifts data (39-40)
- NMR Exchange Format Standard– (NEF) (41)

Infrastructure Development Plan

Application Platform Development:

The development of the OneDep platform is a collaborative effort of the wwPDB partners. The platform was first released in 2013, and both maintenance and feature development have continued since that point. The membership of the OneDep project team includes both Developers and Biocurators from each of the partner sites. A global project manager oversees the activities of the team, which meets virtually on a weekly schedule to discuss operational and development planning matters. Face-to-face meetings for the team are organized on a bi-annual schedule to develop longer-term detailed technical plans.

Software products of the OneDep team are deployed on a regular schedule following a standard testing protocol. Each project site reserves and maintains separate physical resources to support the development and testing/staging activities.

Establishing and Evolving Supporting Physical Resource Requirements:

The OneDep deposition and biocuration platform runs on physical or virtual compute servers and local storage resources hosted in the data centers of the wwPDB regional partner institutions. Based on the experience in developing and using the OneDep platform, estimates for the resource requirements for typical deposition and biocuration data processing workloads have been established. The capability of scaling to address exceptionally demanding workflows arising from very large and complex structure depositions is part of the system design. OneDep delegates computationally intensive services to its Workflow Management System. This system provides capacity adjustment by allowing additional workflow engine servers to be dynamically provisioned to extend the pool of available compute resources.

As described in Section X. Documented Storage Procedures, each wwPDB site hosts Enterprise-quality storage resources to support their local data processing operations. Additional on-line and/or off-line tape copies of production file systems may be maintained to support high availability and to provide for disaster recovery. This local redundancy is provided in addition to the global replication of the corpus of data in the public wwPDB repository. Storage growth requirements are estimated by monitoring deposition statistics (42), funding trends, through careful tracking of emerging technologies. These growth estimates inform requirements for both data file storage and the networking infrastructure to support required data exchange operations.

Hardware resources are locally hosted and supported by regional partner projects. As such, the refresh of hardware resources across sites is not entirely uniform. Typically, replacement of production hardware occurs on a 3 to 5-year cycle. This duration corresponds to common warranty or leasing agreement lifetimes. Hardware components with longer lifetimes are maintained by extended service agreements providing a guaranteed quality of service for replacement. Data centers also provide environmental and power conditioning to protect hardware resources.

Software Inventory and Documentation:

Locally developed software components of the OneDep Platform are maintained in project software version control systems (Git and Subversion (SVN)). The project build system manages external software dependencies. The build system defines these dependencies (e.g. software artifact, version, access details) and their specific installation steps in a modular form that can be executed to produce reproducible deployments. Versioned software dependencies are cached in a project CDN so they can be efficiently accessed by a project build process at any wwPDB project site. Project builds bind a collection of external dependencies that are also individually versioned. Complete project installation and maintenance documentation is maintained in conjunction with the project build system. This documentation describes all aspects of the OneDep setup starting from a base Linux OS installation and including all supporting services and maintenance operations. This documentation is used at all of the project partner installation sites and is regularly revised.

Community Software:

The wwPDB OneDep Platform leverages community and open-source software tools as a design objective. An important example of this approach has been the development of the wwPDB Validation Pipeline where each data quality assessment software module has been chosen based on its standing as representing a community consensus and best-practice.

Managing Deposition Load:

Distributing the deposition and biocuration responsibilities geographically among regional partner sites allows the wwPDB to respond to the changing demands of data producers. The OneDep platform stores progress tracking statistics such as deposition load (42) and biocuration data processing time. The system also provides a UI for depositors to provide feedback. Collectively, this allows monitoring how well the deposition and biocuration processes address both the data load and feature expectations of data producers. wwPDB project team members and project regional directors meet regularly to evaluate project status and adjust resource commitments to respond to regional demands.

Compliance:

The wwPDB partners are signatories of the wwPDB Agreement (43) which lays out essential expectations for services to be delivered by each partner site. Each partner site is continuously evaluated by Community and funding organizations, other advisories, and stakeholders for the requisite expertise and infrastructure to provide good stewardship for the PDB repository data. On a very concrete level, each site is required to host the common OneDep software platform, which by its standard workflows and tracking, guarantees that agreed data processing and data management practices are followed at all partner sites. Partner sites work collaboratively to proactively identify operational problems and rapidly troubleshoot and resolve issues.

1. IUCr CIF Format Specification:

<https://www.iucr.org/resources/cif/spec>

2. Westbrook J, Yang H, Feng Z, Berman HM. 5.5 The use of mmCIF architecture for PDB data management. In: Hall SR, McMahon B, editors. *International Tables for Crystallography*. Dordrecht, The Netherlands: Springer; 2005. p. 539-43.

3. Westbrook JD, Berman HM, Hall SR. 2.6 Specification of a relational Dictionary Definition Language (DDL2). In: Hall SR, McMahon B, editors. *International Tables for Crystallography*. Dordrecht, The Netherlands: Springer; 2005. p. 61-72.

4. Westbrook JD, Hall SR. 4.10 DDL2 dictionary. In: Hall SR, McMahon B, editors. *International Tables for Crystallography*. Dordrecht, The Netherlands: Springer; 2005. p. 473-81.

5. wwPDB PDBx/mmCIF Resource Site: <http://mmcif.wwpdb.org>

6. wwPDB Deposition and Biocuration policies:

<https://www.wwpdb.org/documentation/procedure>

7. IUPAC Periodic Table of Elements (2016), Atomic Symbols:

<https://iupac.org/what-we-do/periodic-table-of-elements/>

8. IUPAC Commission on Isotopic Abundances and Atomic Weights:

https://iupac.org/who-we-are/committees/committee-details/?body_code=210#subcommittees_and_commissions

9. IUPAC Nomenclature and Symbolism for Amino Acids and Peptides (one-letter codes) including selenocysteine (Sec; U) and pyrrolysine (Pyl; O): <https://www.qmul.ac.uk/sbcs/iupac/AminoAcid/>

10. Protein and Nucleic Acid Nomenclature:

Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, Sykes BD, Wright PE, Wüthrich K. Recommendations for the presentation of NMR structures of proteins and nucleic acids. IUPAC-IUBMB-IUPAB Inter-Union Task Group on the standardization of databases of protein and nucleic acid structures determined by NMR spectroscopy, *J. Biomol. NMR*, 1998,12,1-23.

11. wwPDB Chemical Component Dictionary: J.D. Westbrook, C. Shao, Z. Feng, M. Zhuravleva, S. Velankar, J. Young (2014) The chemical component dictionary: complete descriptions of constituent molecules in experimentally determined 3D macromolecules in the Protein Data Bank Bioinformatics DOI: 10.1093/bioinformatics/btu789

12. wwPDB Chemical Component Dictionary S. Sen, J. Young, J. M. Berrisford, M. Chen, M. J. Conroy, S. Dutta, L. Di Costanzo, G. Gao, S. Ghosh, B. P. Hudson, R. Igarashi, Y. Kengaku, Y. Liang, E. Peisach, I. Persikova, A. Mukhopadhyay, B. C. Narayanan, G. Sahni, J. Sato, M. Sekharan, C. Shao, L. Tan, M. A. Zhuravleva (2014) Small molecule annotation for the Protein Data Bank DOI: 10.1093/database/bau116

13. S. Dutta, D. Dimitropoulos, Z. Feng, I. Persikova, S. Sen, C. Shao, J. Westbrook, J. Young, M.A. Zhuravleva, G.J. Kleywegt, H.M. Berman (2014) Improving the representation of peptide-like inhibitor and antibiotic molecules in the Protein Data Bank Biopolymers 101:659-668 DOI: 10.1002/bip.22434

14. Crystallographic Space Group Symmetry:

International Tables for Crystallography (2016). Volume A, Space-group symmetry. doi:10.1107/97809553602060000114, <http://it.iucr.org/A/>

15. Enzyme Nomenclature (EC): Moss GP. "Recommendations of the Nomenclature Committee". International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes by the Reactions they Catalyze. (<https://www.qmul.ac.uk/sbcs/iubmb/enzyme/>)

16. Gene Ontology (GO): Ashburner et al. Gene ontology: a tool for the unification of biology (2000) *Nat Genet* 25(1):25-9

17. Gene Ontology (GO): The Gene Ontology Consortium. Gene Ontology Consortium: going forward. (2015) *Nucl Acids Res* 43 Database issue D1049–D1056

18. NCBI Taxonomy Resources:

<https://www.ncbi.nlm.nih.gov/books/NBK54428/>

19. NCBI Taxonomy Database:

Sayers EW, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Landsman D, Lipman DJ, Madden TL, Maglott DR, Miller V, Mizrachi I, Ostell J, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Yaschenko E, Ye J (2009). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2009 Jan;37(Database issue): D5-15. Epub 2008 Oct 21.

20. IUPAC International Chemical Identifier (InChI)

Stephen E. Stein, Stephen R. Heller, and Dmitrii Tchekhovskoi, An Open Standard for Chemical Structure

Representation: The IUPAC Chemical Identifier, in Proceedings of the 2003 International Chemical Information

Conference (Nimes), Infonortics, pp. 131-143.

21. Open SMILES

<http://opensmiles.org/opensmiles.html>

22. DOI - ISO 26324:2012 Information and Documentation, Digital Object Identifier System:

<https://www.iso.org/standard/43506.html>

23. ORCID, ISO Standard (ISO 27729) International Standard Name Identifier (ISNI):

<https://www.iso.org/standard/44292.html>

24. Compilation of BMRB NMR Experimental Standards <https://bmrbl.io/standards/>

25. wwPDB Legacy PDB Format Standard V3:

<https://www.wwpdb.org/documentation/file-format-content/format33/v3.3.html>

26. Westbrook J, Ito N, Nakamura H, Henrick K, Berman HM. PDBML: the representation of archival macromolecular structure data in XML. *Bioinformatics*. 2005;21(7):988-92. DOI: 10.1093/bioinformatics/bti082. PubMed PMID: 15509603

27. PDBML Schema Resources: <http://pdbml.wwpdb.org/>

28. Extensible Markup Language (XML) 1.1 (Second Edition):

<https://www.w3.org/TR/2006/REC-xml11-20060816/>

29. Namespaces in XML 1.1 (Second Edition):

<https://www.w3.org/TR/2006/REC-xml-names11-20060816/>

30. W3C XML Schema Definition Language (XSD) 1.1 Part 1: Structures:

<https://www.w3.org/TR/2012/REC-xmlschema11-1-20120405/>

31. W3C XML Schema Definition Language (XSD) 1.1 Part 2: Datatypes Types:

<https://www.w3.org/TR/2012/REC-xmlschema11-2-20120405/>

32. Kinjo AR, Suzuki H, Yamashita R, Ikegawa Y, Kudou T, Igarashi R, Kengaku Y, Cho H, Standley DM, Nakagawa A, Nakamura H. Protein Data Bank Japan (PDBj): maintaining a structural data archive and resource description framework format. *Nucleic Acids Res*. 2012;40(Database issue):D453-60. Epub 2011/10/07. DOI: 10.1093/nar/gkr811. PubMed PMID: 21976737; PMCID: 3245181.

33. PDB/RDF Mapping and Deliver Details: <https://pdj.org/help/rdf>

34. RDF 1.1 XML Syntax:

<https://www.w3.org/TR/2014/REC-rdf-syntax-grammar-20140225/>

35. RDF Schema 1.1:

<https://www.w3.org/TR/2014/REC-rdf-schema-20140225/>

36. OWL 2 Web Ontology Language XML Serialization (Second Edition):

<https://www.w3.org/TR/2012/REC-owl2-xml-serialization-20121211/>

37. OWL 2 Web Ontology Language RDF-Based Semantics (Second Edition):

<https://www.w3.org/TR/2012/REC-owl2-rdf-based-semantics-20121211/>

38. OWL 2 Web Ontology Language XML Serialization (Second Edition):

<https://www.w3.org/TR/2012/REC-owl2-xml-serialization-20121211/>

39. BMRB data standard, NMRSTAR

<https://bmrbl.io/dictionary/>

40. NMR Restraint Grid – Remediation of NMR

Doreleijers JF, Vranken WF, Schulte C, Lin J, Wedell JR, Penkett CJ, Vuister GW, Vriend G, Markley JL, Ulrich EL. The

NMR restraints grid at BMRB for 5,266 protein and nucleic acid PDB entries. J Biomol NMR. 2009;45(4):389-96. DOI: 10.1007/s10858-009-9378-z. PubMed PMID: 19809795; PMCID: 2777234.

41. NMR Exchange Format Standard (NEF):

A. Gutmanas, P.D. Adams, B. Bardiaux, H.M. Berman, D.A. Case, R.H. Fogh, P. Güntert, P.M.S. Hendrickx, T. Herrmann, G.J. Kleywegt, N. Kobayashi, O.F. Lange, J.L. Markley, G.T. Montelione, M. Nilges, T.J. Ragan, C.D. Schwieters, R. Tejero, E.L. Ulrich, S. Velankar, W.F. Vranken, J.R. Wedell, J. Westbrook, D.S. Wishart, G.W. Vuister (2015) NMR Exchange Format: a unified and open standard for representation of NMR restraint data Nature Structural & Molecular Biology 22: 433-434 DOI: 10.1038/.3041

42. wwPDB deposition statistics: <https://www.wwpdb.org/stats/deposition>

43. wwPDB Agreement and Appendix (<http://www.wwpdb.org/about/agreement> and https://cdn.rcsb.org/wwpdb/docs/about/charter_appendix.pdf)

Reviewer Entry

Reviewer 1

Comments:
Accept.

Reviewer 2

Comments:
Accept.

16. Security

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

XVI. Security

Service Availability:

The wwPDB organization is founded on the principle of providing highly available deposition, biocuration, validation, and archiving services. Each wwPDB partner provides regional hosting for these wwPDB services. This alone significantly reduces the risk of loss of wwPDB services globally. The further backup and recovery capabilities of the repository and OneDep system have been described in detail in Section IX. Documented Storage Procedures. The technical infrastructure to support these systems is described in the preceding Section XV. Technology and the mitigation of any potential security risks are described in the following sub-sections (Institutional IT Security Procedures and Application Security).

Institutional IT Security Procedures:

wwPDB hosting institutions each support large IT infrastructures and devote significant staff and resources to Information Security Management Systems and Cyber Security (1-5). As each institutional site either hosts a medical school or hosts patient research data, institutions have developed additional security infrastructure to enforce this sensitive clinical data's regional privacy requirements. While the wwPDB does not manage any clinically sensitive data, it benefits from the special security scrutiny required to host these data. Because the wwPDB maintains common services in all site deployments, each deployment must satisfy all deployments' collective security requirements. Consequently, the wwPDB benefits from the sum of the security protocols implemented by all its hosting sites.

Each hosting institution maintains a campus security policy (1-5) which provides the security environment in which the wwPDB partner projects operate. Each site's policies and environment have evolved in response to observed and anticipated threats against shared infrastructure and services. There is a demonstrated commitment on the part of each hosting institution to aggressively protect its information infrastructure and rapidly adjust security policies and procedures to meet any incident challenges.

Campus data centers supported by our hosting institutions provide physical security for equipment and cabling, data center entry access controls, environmental conditioning, environmental monitoring, power conditioning, live and/or virtual surveillance. Access to data center areas is restricted to staff with system administration roles. Project office facilities are equipped with surveillance and entry access controls.

Hosting institutions also manage the security of human resources across the employment life cycle. Accepting institutional security policies and acknowledging individual staff responsibilities for IT security are pre-conditions for obtaining access credentials for the campus network and IT resources. wwPDB specific security protocols at each project site reinforce the campus security practices. These project protocols limit the physical locations or networks from which sensitive project services and data may be accessed. At employment termination, both institutional and project access credentials are revoked.

Some key network services and infrastructure provided by our institutions include security risk assessment, regular vulnerability detection scanning and reporting, network firewall protection, active threat detection and mitigation (e.g., intrusion, DDOS), partitioned public and local network architecture, abuse, and security incident management.

Network vulnerability detection includes regular monitoring of network-accessible services, which identify known security issues (6-7), insecure services, and vulnerable cipher suites. Additional reports are provided of known vulnerabilities in applications in common Linux distributions. The OneDep software build and install tools are updated regularly in response to vulnerability reports.

Network partitioning and firewall protection allow OneDep deployments to limit the footprint of public network exposure. Only lightweight web user interfaces and web APIs for OneDep deposition and validation functions are exposed as public services.

These OneDep public-facing components are implemented with a mature and uniform software stack (e.g., Linux, Apache, Python, Django) with limited external dependencies (e.g., MySQL).

Beyond these well-supported infrastructure and system tools, potential vulnerabilities in public-facing applications are isolated to a subset of the OneDep platform that is entirely under project control. Managing the risks associated with internally developed project software is described in the following section (Application Security). The OneDep biocuration and workflow system components (Section XII. Workflows) are deployed within networks with restricted access controls. These components carry dependencies for a wide range of community software tools (Sections XI. Data Quality and XV. Technical Infrastructure). The OneDep workflow and web API architecture allow these community applications to be executed within network and server environments with sufficient isolation to reduce the impact of any vulnerabilities in these components on other OneDep operations.

wwPDB also engages an external domain name service (DNS) provider, NS1 (<https://ns1.com/>). NS1 provides global network points of presence, allowing global service health monitoring, load balancing, and failover. NS1 has also provided resilience against increasingly common denial of service and spoofing attacks against public DNS infrastructure.

Application Security:

In addition to addressing external security risks, the wwPDB takes steps to avoid introducing security vulnerabilities in developing applications and deploying services. The OneDep development team has adopted uniform development practices aimed at creating and maintaining a high-quality and secure codebase. This includes a mutually agreed limited set of programming technologies, dependencies, standards for code organization, version control, and multi-stage testing. Maintaining shared and well-defined standard procedures reduces the risk of sloppy coding practices that may impact reliability or introduce code vulnerabilities. The OneDep development team also benefits from larger development teams of the wwPDB partner projects and their hosting institutions.

For OneDep Platform, public-facing web applications benefit from built-in security protections from the Django web

framework (e.g., injection and cross-site scripting) (8), and development patterns are informed by security best practice recommendations from public forums such as the Open Web Application Security Project (OWASP) (9-10) and automated software scanning services (11-12). Web services and data exchange operations are implemented using secure protocols (SSL/TLS).

Compliance:

The wwPDB partners are signatories of the wwPDB Agreement (43) which lays out essential expectations for services to be delivered by each partner site. Each partner site is continuously evaluated by Community and funding organizations, other advisories, and stakeholders for the requisite expertise and infrastructure to provide good stewardship for the PDB repository data. The OneDep works collaboratively to assess potential security threats to the platform. The global team works together to ensure that all software components are appropriately updated with security patches and network services are maintained and configured following secure practices.

Security issues that do arise are discussed in regular meetings to raise awareness of all team members and achieve rapid appropriate actions.

1. Rutgers Information Protection and Security:

<https://rusecure.rutgers.edu/policies-standards-and-compliance>

2. UCSD IT Security:

<http://blink.ucsd.edu/technology/security/services/index.html>

3. SDSC Cyber Security and Security Consulting:

<https://www.sdsc.edu/services/it/cybersecurity.html>

<https://security.sdsc.edu/>

4. EMBL/EBI Wide Impact Cyber Security Risk Framework:

<https://www.cyberwiser.eu/>

5. University Connecticut Vulnerability Assessment Center:

<https://its.uconn.edu/service/vulnerability-assessment/>

6. Computer Emergency Response Team (CERT):

<https://www.us-cert.gov/ncas/alerts>

<https://cert.europa.eu/>

https://www.jpccert.or.jp/english/menu_alertsandadvisories.html

7. National Vulnerability Database (NVD):

<https://nvd.nist.gov/vuln/full-listing>

8. Django Web Framework: <https://www.djangoproject.com/>

9. OWASP Top Ten Vulnerabilities:

https://www.owasp.org/index.php/Category:OWASP_Top_Ten_Project

10. OWASP Secure coding guidelines:

https://www.owasp.org/index.php/Secure_Coding_Cheat_Sheet

11. GitGuardian: <https://www.gitguardian.com/>

12. Dependabot: <https://dependabot.com/>

Reviewer Entry

Reviewer 1

Comments:

Accept.

Reviewer 2

Comments:

Accept.

APPLICANT FEEDBACK

Comments/feedback

These Requirements are not seen as final, and we value your input to improve the CoreTrustSeal certification procedure. Any comments on the quality of the Requirements, their relevance to your organization, or any other contribution, will be considered as part of future iterations.

Response:

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments: