



## **Implementation of the CoreTrustSeal**

The CoreTrustSeal board hereby confirms that the Trusted Digital repository wwpdb complies with the guidelines version 2017-2019 set by the CoreTrustSeal Board.

The afore-mentioned repository has therefore acquired the CoreTrustSeal of 2016 on December 27, 2017.

The Trusted Digital repository is allowed to place an image of the CoreTrustSeal logo corresponding to the guidelines version date on their website. This image must link to this file which is hosted on the CoreTrustSeal website.

Yours sincerely,

The CoreTrustSeal Board

## Assessment Information

Guidelines Version: 2017-2019 | November 10, 2016  
Guidelines Information Booklet: [DSA-booklet\\_2017-2019.pdf](#)  
All Guidelines Documentation: [Documentation](#)

Repository: wwpdb  
Seal Acquiry Date: Dec. 27, 2017

For the latest version of the awarded DSA for this repository please visit our website: <http://assessment.coretrustseal.org/seals/>

Previously Acquired Seals: None

This repository is owned by: **Worldwide Protein Data Bank**

T +1-848-445-4290  
E [info@wwpdb.org](mailto:info@wwpdb.org)  
W <https://www.wwpdb.org/>

# Assessment

## 0. Context

### Applicant Entry

*Self-assessment statement:*

Repository Type: Domain or subject-based repository

Level of Curation:D. Data-level curation as in C above, but with additional editing of deposited data for accuracy

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 deduce a structure's role in human health and disease, and in 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

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

RCSB Protein Data Bank (RCSB PDB), Protein Data Bank in Europe (PDBe), Protein Data Bank Japan (PDBj), and BioMagResBank (BMRB).

Recent articles (1-2) provide additional general context for the important role of the PDB in the global management of 3D biological structure data.

1. 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. Daughter, M. Jaskolski) Springer New York doi: 10.1007/978-1-4939-7000-1
2. 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

Repository Community

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

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 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.

## Curation

### Biocuration Policies and Practices:

wwPDB data depositions contain comprehensive descriptions of structural models coming 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).

#### CoreTrustSeal Board

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

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 and 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 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. The latter versioning policy was recently announced at (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) 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

Outsource Partners

Not applicable.

Other Relevant Information

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

## Repository Usage:

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.

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).

#### Global Organization and Role:

The PDB archive is managed jointly by the Worldwide Protein Data Bank partnership (wwPDB; [wwpdb.org](http://wwpdb.org))(1) and consists of the RCSB Protein Data Bank,(2, 3) Protein Data Bank Japan (PDBj),(4) the Protein Data Bank in Europe (PDBe),(5) and BioMagResBank (BMRB).(6) The wwPDB organization operates under a formal agreement ([wwpdb.org/about/agreement](http://wwpdb.org/about/agreement)), most recently renewed in 2013. This agreement commits wwPDB partners to standardizing, collecting, validating, annotating, and storing macromolecular structure data as a single global archive for data depositors and disseminating these data via FTP 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. Rose PW, Prlic A, Altunkaya A, Bi C, Bradley AR, Christie CH, Costanzo LD, Duarte JM, Dutta S, Feng Z, Green RK, Goodsell DS, Hudson B, Kalro T, Lowe R, Peisach E, Randle C, Rose AS, Shao C, Tao YP, Valasatava Y, Voigt M, Westbrook JD, Woo J, Yang H, Young JY, Zardecki C, Berman HM, Burley SK. The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.* 2017;45(D1):D271-D81. doi: 10.1093/nar/gkw1000. PubMed PMID: 27794042; PMCID: PMC5210513.
  
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. 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.
  
6. 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

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 1. Mission/Scope

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

## Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

I. Mission/Scope

The mission of the wwPDB is to maintain a single archive of macromolecular structural data that are 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 via FTP to data consumers, all at no charge with no restrictions on data usage.

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

## Reviewer Entry

CoreTrustSeal Board

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 2. Licenses

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

#### 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).

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

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

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

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

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

### 3. Continuity of access

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

#### **Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

III. Continuity of Access

The PDB has a 46-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 long-term stability of the repository. Because the PDB has developed data processing tools and infrastructure which are 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**

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## 4. Confidentiality/Ethics

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

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 which are maintained exclusively for administrative purposes.

PDB provides an optional embargo period of up to one year prior to releasing a data entry. The primary purpose of this embargo 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 prior to data release.

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

1. wwPDB Release and Embargo Policies:

[https://www.wwpdb.org/documentation/policy#toc\\_release](https://www.wwpdb.org/documentation/policy#toc_release)

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 5. Organizational infrastructure

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

V. Organizational Infrastructure

The wwPDB Team:

Funding for the operations of the PDB is independently obtained by each of the wwPDB partner organizations. Funding cycles are typically 3-5 years in duration. The RCSB PDB is supported by the National Science Foundation, the National Institutes of Health, and the Department of Energy. 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 National Bioscience Database Center-Japan Science and Technology Agency. The BMRB is supported by the National Institute of General Medical Sciences.

The operations of the PDB are conducted by a highly skilled staff with broad domain expertise in experimental structural biology, life sciences and medical applications, data science, information technology, and software engineering. Team members routinely participate in domain conferences and professional associations. The staff for each of the wwPDB partner sites and their PDB related publications are enumerated in the following links.

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

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://pdj.org/info/staff>

Publications:

<https://pdj.org/info/publications>

BMRB at University of Wisconsin:

Staff:

<http://www.bmrwisc.edu/bmrw/staff.shtml>

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

Staff:

[http://www.rcsb.org/pdb/static.do?p=general\\_information/about\\_pdb/contact/pdb\\_members.html](http://www.rcsb.org/pdb/static.do?p=general_information/about_pdb/contact/pdb_members.html)

Publications:

[https://www.rcsb.org/pdb/static.do?p=general\\_information/news\\_publications/articles/index.html](https://www.rcsb.org/pdb/static.do?p=general_information/news_publications/articles/index.html)

wwPDB joint publications:

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

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

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 6. Expert guidance

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

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 is made up of 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>

<https://pdb-dev.wwpdb.org/contributors/>

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

[http://www.rcsb.org/pdb/static.do?p=general\\_information/about\\_pdb/pdbac.html](http://www.rcsb.org/pdb/static.do?p=general_information/about_pdb/pdbac.html)

PDBe Advisory Committee:

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

BMRB Advisory Committee:

[http://www.bmrwisc.edu/bmrw/aims\\_and\\_policies.shtml](http://www.bmrwisc.edu/bmrw/aims_and_policies.shtml)

The wwPDB Advisory Committee includes membership from each of the wwPDB Partner Advisory Committees. This reciprocal membership provides a communication channel that serves to coordinate and align recommendations for wwPDB organization. Recommendations arising from these oversight bodies are reviewed by the wwPDB leadership. Issues requiring technical assessment are reviewed by the OneDep project team. Informed by advisory and team input, the wwPDB leadership sets priorities, commits resources, and defines the direction of the project.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 7. Data integrity and authenticity

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

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/>).

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

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

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 data files in the repository. 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 which records the changes to the entry at the granularity of data category and data item. (8) A new instance of the wwPDB ftp repository containing explicitly versioned data files will be released in late 2017 (9). The versioned repository is being introduced in parallel with the traditional ftp repository. The latter contains only the most recent version of each data entry. The new versioned organization will permit 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 (10) and are audited within each entry. As described in 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 PDB repository (11). Representative example data files illustrating substantial administrative changes are also provided to users in advance of any changes in the content of the production data repository.

The combination of versioning and revision history provide 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 prior to any significant administrative or remediation repository update.

Depositor Identity:

PDB depositors self-identify during the deposition process. Contact information 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 requested for each depositor.

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. 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](http://mmcif.wwpdb.org/dictionaries/mmcif_pdbx_v50.dic/Groups/index.html)
  
9. wwPDB news item describing delivery of a repository versioning: <http://www.wwpdb.org/news/news?year=2017#5979ef4161fd3d50915a4b00>
  
  
10. wwPDB Deposition and Biocuration policies: <http://www.wwpdb.org/documentation/policy>

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

### **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 8. Appraisal

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

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. All such issues must be addressed by the depositor in order 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 assessment of data quality. This quality assessment must be reviewed by the depositor before finalizing a deposition, and the depositor is strongly encouraged to address any issues arising from this assessment.

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

The depositor is always required to provide data which complies with the wwPDB format standard, satisfies data integrity checks, and satisfies minimum repository content requirements. Experimental limitations may preclude the depositor from addressing all of the issues arising in the scientifically focused data quality assessment. This assessment includes: comparison of observed structure models with community accepted standard values as well as a comparison of observed structures models with their underlying experimental data (e.g., electron density for X-ray or mass density map for 3DEM). Sample imperfections, weak data signal, radiation damage, among others, can contribute to experimental observations that rank less than optimum in this assessment. It is not the policy of the wwPDB archive to reject depositions based on this data quality assessment. The wwPDB provides the detailed data quality assessment in the form of a validation report for each structure in the archive. Providing this information makes it possible for repository users to make informed selections of structure entries according to their data quality requirements.

#### Data Formats:

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

1. wwPDB Deposition and Biocuration Policies:  
<http://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:  
<http://www.wwpdb.org/documentation/file-format>

## Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 9. Documented storage procedures

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

IX. Documented Storage Procedures

OAIS Archive Reference Model Support:

Conformance with the OAIS Archive Reference Model (1) requires support of the OAIS information model concepts in OIAS Section 2.2. The concepts in the OAIS Archive Reference Model have the following correspondences in the context of the PDB archive. The implementation of the responsibilities enumerated in OIAS Section 3.1 describing compliance with the reference model are 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 specialist and non-specialist, are considered part of the Designated Community.

**OAIS Archive:** The archival data content stored and delivered on wwPDB ftp data repository.

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

**Representation Information:** PDBx/mmCIF metadata dictionary describing the semantics of each data item 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 ftp repository. Data objects provide internal linkage to a particular version 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 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 ftp 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 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 wwPDB OneDep Biocuration platform (2).

#### Data Security Requirements:

PDB data deposition is performed through a OneDep web user interface. Each deposition session is password protected and a secure web protocol (HTTP over TLS/SSL) is used. Prior to 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 that is 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 ftp repository. Each of these partner sites also maintains redundant on-line copies of the wwPDB ftp repository to support performant and highly available user access.

Access to 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 ftp 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 exchange of deposition sessions between wwPDB partner sites. This capability allows for failover of in-progress deposition services between wwPDB sites in the event of a loss of service. This functionality is regularly used to maintain availability and accommodate any required periods of data center maintenance.

#### Risk Management:

Principal objectives of the wwPDB organization are providing for the security, availability, and long-term preservation for 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 the problem of 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 loss of data acquisition services.

#### Data Consistency:

The wwPDB releases of data on a coordinated weekly schedule. The BMRB, PDBe, PDBj and RCSB PDB wwPDB partners maintain synchronized copies of the ftp 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 ftp repository copies. Additional indices of content that are 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](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](http://mmcif.wwpdb.org/dictionaries/mmcif_pdbx_v50.dic/Categories/database_2.html)
  
5. wwPDB Repository Advisory Notice: <ftp://ftp.wwpdb.org/pub/pdb/advisory.txt>.
  
6. wwPDB Download Instructions: <http://www.wwpdb.org/download/downloads>
  
7. wwPDB Deposition and Biocuration Policies: <https://www.wwpdb.org/documentation/policy>

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 10. Preservation plan

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

X. Preservation Plan

Preservation Plan:

The documentation of 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 Biocuration processing performed by the repository 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 for 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 depositor and the repository are clearly 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) for the purpose of maintaining 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: <http://www.wwpdb.org/documentation/policy>.
2. For Biocuration policies, see: <http://www.wwpdb.org/documentation/procedure>
3. For versioning policies, see: <http://www.wwpdb.org/news/news?year=2017#5979ef4161fd3d50915a4b00>
4. For retrospective Biocuration/remediation see: <http://www.wwpdb.org/documentation/remediation>

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*



## 11. Data quality

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

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. All such issues must be addressed by the depositor in order to finalized 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:

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

In addition to verifying basic data integrity, completeness, and compliance with the PDBx/mmCIF metadata standard, data are subjected to a rigorous assessment of data quality. This assessment is informed by recommendations from the wwPDB Task Forces and Working Groups described in Section VI. Expert Guidance. These recommendations include both data quality content as well as software applications available to the community to compute the target quality metrics. The particular content of the assessment 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 as well as in a software accessible data file. Validation reports are provided in the repository for each structure entry in the PDB archive. A complete description of report content and a selection of example validation reports are available at [wwpdb.org](http://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. A number of key journals which publish macromolecular structural data require that wwPDB validation reports accompany 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 have a positive impact on improving 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 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 primary citation which is the case for more than 80% of the archive entries. The metadata for the primary citation includes a PubMed identifier and Digital Object Identifier (DOI) if these are available. Each PDB entry is individually assigned a DOI which provides a linkage to the structure entry in the PDB archive.

1. wwPDB validation reports:  
<http://www.wwpdb.org/validation/validation-reports>
2. See Section I. Quality at Glance in the example validation report:  
[http://www.wwpdb.org/validation/samples/1cbs/1cbs\\_validation.pdf](http://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:  
<http://www.wwpdb.org/validation/validation-servers>

5. OneDep validation API:

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

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 12. Workflows

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

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 web site and handling of data as described in these procedures must be acknowledged by depositors prior to 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 set of standard workflows executed by the wwPDB OneDep platform (3). These workflows guarantee that all data processing operations are performed in a uniform manner. 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 so that they may flexibly adapt to changing scientific and technical requirements.

Workflows also automate data processing operations associated with release processing in 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 include 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.

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 data architecture of the PDB benefits from the extensibility of the PDBx/mmCIF data architecture ([mmcif.wwpdb.org](http://mmcif.wwpdb.org)). PDBx/mmCIF is designed to support facile content extension, and software tools supporting 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:

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: <http://www.wwpdb.org/documentation/policy>.
2. For Biocuration procedures, see: <http://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**

*Accept or send back to applicant for modification:*

Accept

*Comments:*



### 13. Data discovery and identification

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

#### **Applicant Entry**

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

XIII. Data Discovery and Identification

Data and Metadata Search Facilities:

A wide variety of search, analysis and reporting features are provided by the wwPDB partners. These search services all share the common structure and experimental data delivered by in the wwPDB repository which are 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](http://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

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

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

Selected Search Documentation:

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

Search API:

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

PDBe:

Web Search:

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

Selected Search Documentation:

[https://www.ebi.ac.uk/pdbe/sites/ebi.ac.uk.pdbe/files/documents/service\\_tutorials/PDBe-search\\_0.pdf](https://www.ebi.ac.uk/pdbe/sites/ebi.ac.uk.pdbe/files/documents/service_tutorials/PDBe-search_0.pdf)

Search API:

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

RCSB PDB:

Web Search:

<http://www.rcsb.org/>

Selected Search Documentation:

[https://www.rcsb.org/pdb/secondary.do?p=v2/secondary/search.jsp - search\\_advanced-search](https://www.rcsb.org/pdb/secondary.do?p=v2/secondary/search.jsp - search_advanced-search)

Search API:

<https://www.rcsb.org/pdb/software/rest.do>

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

BMRB:

Web Search:

<http://www.bmrwisc.edu/>

Search API:

<http://www.bmrwisc.edu/bmrw/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/>.

Repository Registries:

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

Registry of Research Data Repositories:

<http://www.re3data.org/>

Minimal Information Required In the Annotation of Models (MIRIAM):

<http://www.ebi.ac.uk/miriam>

Directory of Open Access Repositories (OpenDOAR):

<http://www.opendoar.org/>

FAIRSharing:

<https://fairsharing.org/>

DataMed/BioCADDIE:

<https://datamed.org/>

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:  
<http://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

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 14. Data reuse

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

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:

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)



A small number of mature data formats are commonly used in structural biology to store atomic coordinate and supporting experimental data (2). These formats are directly supported by the wwPDB OneDep deposition system (3). The primary format for the PDB archive 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 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:

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

The details of the PDBx/mmCIF metadata specifications, tutorial information, and links to supporting software tools are maintained at 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 (14-18) describing the PDB data content along with supporting access and analysis tools.

1. wwPDB Deposition and Biocuration Policies: <http://www.wwpdb.org/documentation/policy>
2. wwPDB File Format Documentation: <http://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:  
<http://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:  
<http://www.wwpdb.org/news/news?year=2017#5963997661fd3d50915a4af7>
  
  
11. wwPDB PDBx/mmCIF Resource Site: <http://mmcif.wwpdb.org>
  
  
12. Chemical Reference Data documentation: <http://www.wwpdb.org/data/ccd>
  
  
13. Molecular Reference Data documentation: <http://www.wwpdb.org/data/bird>

14. BMRB Educational Materials: <http://www.bmrw.wisc.edu/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. PDBej Educational Services: <https://pdbej.org/help/educational-services-menu>
19. RCSB PDB PDB-101 Educational Site: <https://pdb101.rcsb.org/>

## Reviewer Entry

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 15. Technical infrastructure

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

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) embraces 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 has 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 bi-annual schedule to develop longer-term detailed technical plans.

#### CoreTrustSeal Board

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)



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 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 the 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 the 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 lease 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 (Subversion (SVN) and Git). External software dependencies are managed by the project build system. 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 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 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 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.

1. IUCr CIF Format Specification:  
<http://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: <http://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: <http://www.chem.qmul.ac.uk/iupac/AtWt/index.html>
  
9. IUPAC Nomenclature and Symbolism for Amino Acids and Peptides (one-letter codes) including selenocysteine (Sec; U) and pyrrolysine (Pyl; O): <http://www.chem.qmul.ac.uk/iupac/AminoAcid/A2021.html>

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 data bases 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](https://doi.org/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](https://doi.org/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. (<http://www.chem.qmul.ac.uk/iubmb/enzyme/>)

16. Gene Ontology (GO): Ashburner et al. Gene ontology: 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, Mizrahi 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

<http://www.bmrwisc.edu/standards/>

25. wwPDB Legacy PDB Format Standard V3:

<http://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  
<http://www.bmrwisc.edu/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/nsmb.3041

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

## Reviewer Entry

CoreTrustSeal Board

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)



*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 16. Security

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

### Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

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 for 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 are described in the preceding Section XV. Technology, and the mitigation of any potential security risks is described in the following sub-sections (Institutional IT Security Procedures and Application Security).

Institutional IT Security Procedures:

**CoreTrustSeal Board**

W [www.coretrustseal.org](http://www.coretrustseal.org)

E [info@coretrustseal.org](mailto:info@coretrustseal.org)

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 the regional privacy requirements of this sensitive clinical data. 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 the collective security requirements of all deployments. 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. The policies and the environment of each site 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 and monitoring, power conditioning, live and/or virtual surveillance. Access to data center areas is restricted 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 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 cypher 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 light-weight 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 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 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 resilient 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 code base. 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)

(9-11). Web services and data exchange operations are implemented using secure protocols (SSL/TLS).

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:  
<http://www.sdsc.edu/services/it/cybersecurity.html>
  
4. EMBL/EBI Wide Impact Cyber Security Risk Framework:  
<https://www.cyberwiser.eu/>
  
5. University Wisconsin Madison Office of Cyber Security:  
<https://it.wisc.edu/about/office-of-the-cio/cybersecurity/>
  
6. Osaka University Cyber Security Research Alliance:  
[http://cybersecurity.ist.osaka-u.ac.jp/index\\_e.html](http://cybersecurity.ist.osaka-u.ac.jp/index_e.html)
  
7. Computer Emergency Response Team (CERT):  
<https://www.us-cert.gov/ncas/alerts>  
<https://cert.europa.eu/>

[https://www.jpcert.or.jp/english/menu\\_alertsandadvisories.html](https://www.jpcert.or.jp/english/menu_alertsandadvisories.html)

8. National Vulnerability Database (NVD):  
<https://nvd.nist.gov/vuln/full-listing>

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

10. OWASP Top Ten Vulnerabilities:  
[https://www.owasp.org/index.php/Category:OWASP\\_Top\\_Ten\\_Project](https://www.owasp.org/index.php/Category:OWASP_Top_Ten_Project)

11. OWASP Secure coding guidelines:  
[https://www.owasp.org/index.php/Secure\\_Coding\\_Cheat\\_Sheet](https://www.owasp.org/index.php/Secure_Coding_Cheat_Sheet)

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

## 17. Comments/feedback

*Minimum Required Statement of Compliance:*

0. N/A: Not Applicable.

## Applicant Entry

*Statement of Compliance:*

4. Implemented: This guideline has been fully implemented for the needs of our repository.

*Self-assessment statement:*

Some additional items in the PDF template not in the HTML application forms.

WDS V. Periodic Assessment

### **M5. The repository has defined processes for responding to changing scientific requirements and to evolving technologies.**

As described in Sections XII. Workflows and XV. Technical Infrastructure PDB data and technical systems have been designed to address the challenges of the rapidly evolving landscape methods and technologies in experimental structural biology. In this regard, the data architecture of the PDB enjoys the extensibility of the PDBx/mmCIF data architecture ([mmcif.wwpdb.org](http://mmcif.wwpdb.org)). PDBx/mmCIF is designed to support facile content extension, and software tools supporting 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.

The process by which support for new methods and technologies are identified and prioritized is informed by the recommendations from our communities of data producers and consumers as described in Section VI. Expert Guidance.



The delivery of project features supporting new methods and technologies is managed by the PDB OneDep project team as described In Section XV. Technical Infrastructure.

**Please describe any formal, periodic assessment the organization undergoes to ensure its responsiveness to new scientific and technological developments.**

Project oversight and expert advisory have been described in Section VI. Expert Guidance. Additionally, each wwPDB partner project is subject to regular reviews from their respective funding agencies which review how well projects are serving the contemporary needs of their scientific communities. The leading edge of new methods and technologies are often first observed by PDB Biocurators. As such, adapting to these changes in requirements is a routine Biocurator responsibility. PDB Biocuration staff meet regularly to review and address these new challenges.

## **Reviewer Entry**

*Accept or send back to applicant for modification:*

Accept

*Comments:*

Thanks for providing additional information.