

**WORLDWIDE PROTEIN DATA BANK ADVISORY COMMITTEE (WWPDB-AC)
REPORT OF THE VIRTUAL MEETING HELD ON 18 OCTOBER 2024
(HOSTED BY PDB JAPAN)**

Advisory Committee Chair:

Jennifer Martin (Australia, IUCr Representative)

Advisory Committee Co-Chair:

Toshiya Senda (Japan, PDBj Chair)

Advisory Committee Members

Paul Adams (USA, RCSB Chair)

Kirk L. Clark (USA, RCSB)

Juha Huiskonen (Finland, EMDB)

Kyeong Kyu Kim (South Korea, PDBj)

Mehdi Mobli (Australia, BMRB)

Martin Noble (UK, PDBe and EMDB Co-Chair)

Christina Redfield (UK, BMRB)

Corinne Smith (UK, PDBe and EMDB Chair)

Martin Walsh (UK, PDBe)

ISMAR Representative:

Charles Schwieters (USA)

3DEM Representative:

Nieng Yan (China)

wwPDB Principal Investigators:

Stephen K. Burley (RCSB)

Jeffrey C. Hoch (BMRB)

Genji Kurisu (PDBj, Host)

Kyle Morris (EMDB)

Sameer Velankar (PDBe)

wwPDB Regional Representatives

Debasisa Mohanty (PDBi India)

Wenqing Xu (PDBc China)

TABLE OF CONTENTS

1.	Introduction	2
2.	Question 1. Funding issues for wwPDB Core Members	3
3.	Question 2. Future of OneDep Development	6
4.	Question 3. Transition of PDB China to Core Membership	8
5.	Question 4. wwPDB Policies re Obsolete Entries	9
6.	Question 5. Next Advisory Committee meeting	10
7.	Other Issues Raised by the Advisory Committee	10

1. INTRODUCTION

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

wwPDB Mission Statement

- Manage the wwPDB Core Archives as a public good according to 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

The 2024 AC meeting was held virtually over several weeks to provide sufficient time for AC members to discuss agenda items across multiple timezones. PDBj hosted the wwPDB PIs, Genji Kurisu was host PI. The meeting timetable was:

- (i) 27 Sept 2024: wwPDB leadership provided slides, papers and short videos.
- (ii) 4 Oct 2024: 90 min wwPDB-AC-only (Executive Session) virtual meeting held.
- (iii) 10 Oct 2024: AC provided feedback and questions to wwPDB PIs.
- (iv) 17 Oct 2024: wwPDB PIs provided responses to AC questions.
- (iv) 18 Oct 2024, 9pm Japan local time: 3-hr virtual wwPDB-AC meeting
 - Welcome and Introduction, wwPDB PDBj lead, Genji Kurisu (10 min)
 - Joint session on AC questions and comments and PI feedback - ALL (90 min)
 - Executive Session – wwPDB-AC members (45 min)
 - Joint Session to summarise wwPDB-AC feedback – ALL (20 min)
 - Acknowledgements and photo – ALL (10 min)

NOTE: Background information for subsequent sections is in accompanying papers.

2. QUESTION 1. FUNDING ISSUES FOR WWPDB CORE MEMBERS

The PDB celebrates its 53rd birthday on October 20th 2024. The wwPDB leadership reported that the value of open access to wwPDB Core Archives is well understood by respective funders. As evidence, the Core Archive has been funded continuously by the US government since 1971 and the number of FTEs across the wwPDB has increased considerably since the umbrella organisation was founded in 2003. Nevertheless, current funding for core members is constrained to support operations only. Technical debt (refactoring of OneDep) and necessary infrastructure updates remain unfunded or underfunded. The AC was tasked by wwPDB Leadership to develop suggestions on how to address funding issues.

The AC acknowledges the significant continued efforts of wwPDB core members to secure ongoing funding and we share the frustration and concern about challenges for securing essential resources to operate and update this precious worldwide resource. There may be no easy solutions, though we suggest paths for wwPDB leadership to consider.

Commendations:

- 1. The AC commends the efforts, past and ongoing, by partner PIs to identify suitable funding schemes and to secure operational funding for the wwPDB**
- 2. The AC commends the outstanding success, under challenging circumstances, in securing current core funding for this global biodata resource**
- 3. The AC specifically commends the joint grant applications and successes across PDB partners**

We understand that core partners currently run with a head count of 35 FTEs. The cost of infrastructure, maintenance and institutional costs, and the cost of high priority unfunded items was not clarified. The PIs anticipate that technical debt will never be fully retired and that a continual process of improvement will be required. The AC is unclear what base level of funding, beyond the currently funded ops FTEs, would address technical debt/high priority unfunded items. Having this detail would help the AC identify additional creative solutions to address shortfalls, especially for technical debt for which funding mechanisms are limited.

A key challenge for the wwPDB, with the ever-increasing emphasis on open access to scientific data across the sciences, is the growing number of data resources competing for insufficient funding. Moreover, funding calls dedicated to data resources have not expanded the available funding, so wwPDB competes for funding with research proposals. The Global Biodata Coalition (comprising leading funding agencies) has argued strongly that 5% of total research funding be set aside for capturing, preserving and distributing data. Current funding for data resources in US, UK, Europe, and Japan falls far short of this target. In efforts to secure ongoing funding, the wwPDB leadership will need to explore multiple avenues, as outlined here, to ensure that future core funding is not jeopardised when applications are successful but not fully funded.

wwPDB Communications and PR

The wwPDB leadership reported that a broad issue is the perceived long-term value of structural biology. There is a perception by some in the community that AlphaFold/RoseTTAfold

have solved and democratized the protein structure problem leading to questions on the funding priority of experimental structural biology. We stress the continued importance of experimental structural biology and note the impact these developments have had on demonstrating the value to the wider scientific community, who increasingly see provision of structural information as an essential resource. We agree that addressing these broader issues will require joint efforts of wwPDB and other stakeholders across related professional societies (e.g., IUCr, ACA, LACA, ECA, AsCA, AfCA, MSA, EMS, ISMAR, ASBMB, Biophysical Society, FEBS).

It was unclear what **combined** wwPDB comms/PR plans and approaches have been developed to engage the community or high level, non-specialist bodies. Such documentation could emphasize (i) the importance/significance of a curated database, (ii) scientific data integrity (iii) impact for industry and government, (iv) utility in training future machine learning algorithms.

We suggest that wwPDB align its funding needs to current short term funding priorities to capture timely benefit for funding agencies. Topical short-term benefits could include: publisher peer review challenges, fraud detection; AI/ML groups needing annotated high quality data; funding agencies need reproducibility (data dumps won't solve this). It is timely to leverage the direct and fundamental link between the wwPDB archive and the 2024 Nobel prize in Chemistry.

A joint public statement (not publication) describing the importance and significance of wwPDB investment (including annual cost and benefits in quantitative terms - cost-benefit analysis) could help support joint applications and national/local applications for specific funding. We encourage the PIs to communicate to users and potential funders that the vast sums spent on biomedical research are, in part, placed in jeopardy by not properly funding the infrastructure for appropriate stewardship of the output of that funding. The development of AlphaFold is an example of having such infrastructure in place. Sufficient infrastructure funding is required for future epochal developments.

Industry Engagement

Many for-profit industries benefit from accessing the freely available wwPDB archives. For example, the archives serve as a critical research resource for pharma, agrochemical, bioremediation, other biotech and tech companies. We understand that wwPDB discussed the possibility of funding support with a council composed of research and development heads of many of the biggest pharma 4 years ago. At that time, the council decided against supporting a resource that is already supported by corporate taxes and which could set a precedent for agencies defunding important initiatives like the wwPDB when industrial support is provided. We also understand that the open access CC0 license prohibits charging consumers, including industry consumers, for wwPDB archive access, and that there is no appetite to change the license. We acknowledge the efforts made to date to engage with industry, though avenues remain to be pursued with money-generating beneficiaries.

We urge the wwPDB to renew efforts to engage with industry. We encourage strategic consideration of all funding possibilities to ensure that future core funding is not put in jeopardy with successful but insufficient alternative funding. The following ideas provide starting points.

- (i) Identify proposed wwPDB developments (and their cost) that would benefit industry activities (and the broader user base); industry consumers may be more amenable to provide targeted support that benefits them directly rather than undirected funding support. See also point (vi).
- (ii) Consider engaging with industry sectors **beyond pharma** that benefit from access to wwPDB, as outlined above.
- (iii) Highlight to Industry stakeholders the key role wwPDB played in the record-breaking development of COVID vaccines and drugs, and the fundamental role of the archive in the research that led to the 2024 Nobel prizes.
- (iv) Involve commercial beneficiaries in wwPDB advocacy, to lobby government to earmark funds for wwPDB/biodata purposes. High level communications from wealth-generators carry significant weight with government when securing funding.
- (v) Involve industry/tech sector as partners/advocates in wwPDB funding requests. Some specific funding schemes require industry partners, along with their cash or in-kind contribution.
- (vi) Continue to explore possible fee-for-service models that offer added value to customers willing to pay, for example for priority data downloads.
- (vii) Consider engaging with Industry Foundations, through charitable donations if business cases are not sufficient (see also wwPDB Foundation section below).

Sector Engagement

We note that in 2018, feedback from one NSF Program Officer to the RCSB PDB partner suggested that funders may not be positively influenced by endorsements from Nobel prizewinners in grant applications. Documenting value and return on investment are important in funding applications. This could be done through engaging with sector leaders (industry heavyweights, Nobel prize winners, COVID researchers, Presidents of Learned Academies) to support a statement demonstrating the value and ROI of the wwPDB and the importance of continued support of global databases, perhaps referencing the Global Biodata Coalition 5% funding recommendation. Sector leaders could also be invited to endorse or author an open letter to a high-profile journal outlining these same points. Cognate unions and societies could be approached to organise sessions at congresses that demonstrate the value and needs of the wwPDB. [Consider also the UK/Europe annual 'Protein Structure Determination in Industry' meeting. PSDI 2024 will be held in Nov in Paris]. Increasing awareness among the broader scientific community of the importance of the wwPDB and its financial challenges could have a broader benefit, by informing reviewers/decision makers on grant review panels.

wwPDB Foundation

The wwPDB Foundation is chartered as a 501(c)(3) US entity exclusively for scientific, literary, charitable, and educational purposes. Donations to the Foundation are fully tax-deductible in the US. We understand that, in principle, the Foundation's activities are not restricted to outreach, and a decision to expand the scope of activities could be made by the Board of Directors. Currently the wwPDB Foundation Board has representation from RCSB, PDBj, and BMRB.

The wwPDB AC advises the wwPDB to encourage the creation of a culture of consumer giving/charitable donation, from industry/tech consumers, that is proportional to use. A risk

assessment will be needed, to consider *inter alia* the potential return of charitable donations, and the implications for other funders of success or lack of success in charitable fund-raising.

Some options to consider regarding charitable donations:

- (i) Expand membership of the wwPDB Foundation to encompass all core members
- (ii) Establish means to donate (in tax-deductible charitable manner) to eg OneDep development
- (ii) Explore use of wwPDB Foundation as an instrument to engage with Industry Foundations
- (iii) Engage with Industry Foundations for financial support of wwPDB activities that benefit both Industry and the broader community. Major software systems use this model successfully.
- (ii) Consider a Wikipedia-like annual fundraising campaign targeted each year to address a specific OneDep or infrastructure issue that does not fit Government funding criteria, but which is urgent and would address industry needs and be valued by all wwPDB consumers.
- (iii) analogous to the operations of US museums, consider suggesting a voluntary charitable donation on accessing the wwPDB.

Recommendations:

Regarding funding Issues faced by the wwPDB, the AC recommends that wwPDB PIs:

- 1. Clarify funding required to address technical debt and high priority infrastructure**
- 2. Develop and implement a combined communications/PR plan to communicate the value and ROI of investing in the wwPDB.**
- 3. Develop a joint public statement re importance & significance of wwPDB investment.**
- 4. Consider, in funding applications, aligning the benefits of PDB funding to shorter term goals of stakeholders**
- 5. Redouble efforts to engage with industry, as outlined in detail in this report.**
- 6. Engage with sector leaders to communicate to the broader scientific community the importance of supporting the wwPDB into the future.**
- 7. Engage with Professional Societies and Unions (through AC reps where possible) to arrange Congress sessions that communicate value, benefit & needs of the wwPDB.**
- 8. Consider expanding the wwPDB Foundation remit to encompass charitable giving eg for OneDep, and to engage with Industry Foundations.**
- 9. Prioritise the long list of wwPDB activities, in the event that all funding possibilities are exhausted.**

3. QUESTION 2. FUTURE OF ONEDEP DEVELOPMENT

An easy-to-use data deposition interface is critical for user engagement. Users report that data deposition is not improving, that new and non-expert users find OneDep to be complicated, and that it is difficult to know what parameters must be included; ~80% of fields are not mandatory.

The wwPDB is faced with at least two significant challenges around deposition - increased complexity of submissions and a code base that needs to be significantly overhauled. Addressing either issue will likely require new funding. We understand that the OneDep team is

currently assessing technical debt, determining the impact and cost of remediation. This cost/benefit analysis is a necessary prerequisite to forging a plan for retiring technical debt.

The increased volume and complexity of depositions prompted questions from the AC of how strategic decisions could be made to triage the amount of detail required for biocuration. For example, should every deposition receive the same amount of attention, could AI/ML be used, how are models unsupported by experimental data (e.g. models used for interpretation of low resolution EM volumes) treated in comparison to those supported by data? Could different types of deposition be dealt with in ways that use resources more efficiently? The wwPDB PIs responded that every deposition should receive the same amount of attention during curation, though with the distinction as to whether processes require human curation or automatic curation. The use of AI/ML is under active consideration by wwPDB PIs, with a caveat that natural language recognition and other AI/ML technologies are not yet sufficiently accurate for automating curation (as per multiple NIH supported data resources, NIH Data Science Meeting November 27-28 2023, Bethesda, MD).

AC members welcomed the move to engage with the structure-solving community to promote pre-curation of data. The structure-solving community also face resource challenges so it will be important for wwPDB to work with data generators (e.g. synchrotrons), structure solvers (e.g. CCP4), and data harvesters (e.g. OneDep team) to make such an approach work. The wwPDB leadership reported that lowering the burden on depositors and biocurators is an overarching goal of the wwPDB. The OneDep team is pursuing multiple paths to achieve this goal, including sharing metadata among linked depositions, context-dependent user interfaces, and coordination with third-party software developers to capture more metadata. BMRB and PDBe are coordinating with INSTRUCT facilities to archive fragment screening datasets. A 6-month digital research infrastructure (DRI) software grant awarded by UKRI will drive collaboration between PDBe/EMDB and Diamond/eBIC/CCP's to capture metadata from experiment and data analysis to create deposition ready files for depositors to use via depUI and deposition API. These developments are important initial steps towards supporting depositors to automatically populate depositions more completely and robustly than by human input. This project will be further leveraged now that the BBSRC-NSF/BIO proposal submitted in Feb 2024 is successful, by providing resources to continue development of the OneDep deposition API.

At the April 25th 2024 RCSB AC meeting high-level plans were presented to integrate the PDBDev archive into the main archive. A question arose at the wwPDB AC meeting as to how this would impact, if at all, OneDep. The AC thought it important to ensure that depositors see a single deposition interface. In response, the wwPDB PIs reported that Incorporation of PDB-Dev structures (as PDB-IHM structures) is a work in progress. For the immediate future, there will be two deposition interfaces both of which will be accessible from the OneDep home page. The AC understands that rapid progress towards a single deposition interface is not possible at this time, given current resource limitations, but reiterates that the wwPDB should remain committed to the implementation of a single deposition interface in the longer term.

The AC considers that OneDep is a process both as an application, and as a means of having a data-in capability for each site and between sites. The AC considered it would be enlightening for the wwPDB PIs to undertake a formal Process Excellence analysis of OneDep. This analysis would help prioritize steps to focus on, steps to measure differently (KPI standards), and steps that are no longer relevant. Similar to the RCSB approach to UX, it may be possible to engage a University class to take on this task as a project rather than hiring a consulting firm.

wwPDB PIs recognised the potential benefit to the community of working closely with the developer community on validation tools as part of the perceived technical debt embedded in OneDep. In turn, for developers there would be a benefit in having validation tools used during structure determination closely match those used during data deposition.

Commendations:

- 4. We recognise and deeply appreciate the hard work and collaborative effort that goes into maintaining and developing OneDep across multiple partners**
- 5. We commend the positive impact that OneDep has had on user ability to deposit structures and for annotators to more efficiently process depositions. This has led to a measurable increase in deposition throughput across the wwPDB sites.**
- 6. Development of “investigation” multi-structure depositions appears on track and will be a great example of collaborative implementation and transformative impact.**

Recommendations:

- 10. Provide a full report on OneDep (including information on addressing ease of use; pre-curation of data, engagement with data generators/structure solvers/data harvesters/software developers), prior to the 2025 AC meeting, and Include OneDep as a deep dive agenda item in 2025**
- 11. Engage with users/depositors to communicate that simplifying OneDep is a key priority and that this goal is subject to funding.**
- 12. Implement a Process Excellence analysis of OneDep (e.g. Lean Six Sigma), through engagement with a University class.**
- 13. Consider what other options may be available to support further streamlining OneDep should the collaborative UK/US funding not be awarded.**

4. QUESTION 3. TRANSITION OF PDBC TO CORE MEMBERSHIP

The AC acknowledges and celebrates the outstanding efforts of wwPDB partners, particularly PDBj, in supporting and training PDBc as an associate wwPDB member. We recognise that this process began during the very difficult, travel-restricted COVID era. The proposed transition plan is well-considered and the proposed timeline and milestones for core membership are appropriate. If all milestones are met, PDBc could become a full partner by 2027.

Depositions from China have sharply increased (400-500 per month) with a ~2:1 CryoEM:X-ray deposition ratio. This ratio is higher than that of wwPDB core partners, and EM depositions take more time to curate. It is clear that PDBj could not handle the load if not for PDBc.

Given the success of PDBc's associate membership of wwPDB, we encourage wwPDB leadership to fast-track its efforts to involve new partners to further spread the financial and operational load of this important global biodata resource. We are especially keen to learn more about development of a BMRBe site, and a South American PDB site. We welcome the news that PDBj is working with colleagues in South Korea to establish local Data Out facilities and backup secure storage facilities. This will provide geographical redundancy and enhance access, visibility, and data security across regions.

Commendations:

- 7. PDBc is commended on the outstanding achievement of processing 75% of deposits from within China. This achievement in such a short time is very impressive!**
- 8. The wwPDB, particularly PDBj, is commended for the support and training provided to bring in PDBc as an associate wwPDB member during very challenging times.**

Recommendations:

- 14. That prior to being considered for core membership, PDBc demonstrate logistical success in contributing to global advancement of code (eg OneDep) in a bidirectional manner.**
- 15. That prior to being considered for core membership in 2027, that PDBc provide evidence of renewal of funding beyond that date.**
- 16. That wwPDB leadership advance plans for other associate members.**

5. QUESTION 4. WWPDB POLICIES RE OBSOLETING ENTRIES

The AC agrees with the wwPDB leadership that the wwPDB Obsolete Entries policy needs updating and harmonising, and we support the path that the wwPDB is implementing. The incidence of fraudulent entries in the wwPDB is thought to be very small, though in future there may be a larger number of entries that could be re-refined by non-authors, to correct errors.

A policy for Obsolete Entries due to fraud should include standards for documentation to ensure an appropriate paper trail. For example, what are the sources of information that trigger the obsolescence, the timelines, etc. Will the information that triggers obsolescence be made public? Will this same information be distributed to all sites (this could have implications regarding discoverability from a legal perspective). The wwPDB should limit case-by-case liability (damages and legal costs) arising from decisions to obsolete fraudulent entries. This may be addressed during the legal consultation phase.

Research ethics training is key for all wwPDB members. As the wwPDB grows and expands its workforce, with both associate and core members increasingly involved in wwPDB activities, ensuring harmonised research ethics training for everyone will be crucial. This initiative will also provide an advantage, or possibly a prerequisite, in securing funding from government sources. It may be timely to review all current ethics and integrity policies and procedures for wwPDB curators and wwPDB members to ensure these are up to date and aligned across sites.

Commendation:

9. We commend the wwPDB for identifying the need to update and harmonise the Policy and for including public consultation in its development.

Recommendations:

17. That the policy for obsoleting entries include standards for documentation to ensure an appropriate paper trail.

18. Consider whether the information that triggers obsolescence of a fraudulent entry should be (a) distributed to all PDB sites, or (b) made public. Both actions could have legal implications.

19. Consider whether the Policy should encompass obsolescence of entries by a non-author for reasons other than integrity/fraud (eg re-refinement to correct errors).

20. Undertake a broad policy review and institute a review timeline to ensure that all policies are harmonised and reviewed regularly (perhaps every 3 years).

6. QUESTION 5. ADVISORY COMMITTEE MEETING 2025

The 2025 meeting will be virtual, with the same format as this year. The date of the AC meeting is confirmed as Friday October 17, with the AC-only meeting scheduled for Friday Oct 3. Several wwPDB AC members and the AC chair will rotate off the committee this year.

Recommendation:

21. Arrange replacements for AC members and AC Chair, as soon as possible in 2024 to ensure availability of new members for 2025 meetings.

22. Ensure new AC members are aware of the full AND the AC-only 2025 dates, and that calendar invites are issued as soon as possible to hold dates and times.

7. OTHER ISSUES RAISED BY THE AC

Commendation:

10. The wwPDB AC warmly appreciates the time taken and the detail provided in the response to the recommendations of the 2023 AC report.

A 2024 RCSB AC report recommendation was “cryo-EM validation tools provided by the wwPDB for depositions should be discussed at an upcoming wwPDB AC meeting”. A major concern is that validation tools for cryo-EM depositions are lagging and need to be implemented on a faster timeline. We also see uncertainty in the roadmap to support cryo-ET studies at the wwPDB. This is becoming pressing as the method is rapidly developing and gaining wider use.

Recommendation:

23. The wwPDB validation pipeline and timeline for cryo-EM and cryo-ET be included as an agenda item for a deep dive at the 2025 meeting.

Finally, we wish to highlight the Scientific Comment authored by Bond & Sussman, published in Acta D Structural Biology 1 December 2024 (after the AC meeting). The PDB-AC agrees with Bond & Sussman's suggested two-step process and think that the wwPDB should play a central role in any such discussion.

<https://journals.iucr.org/d/issues/2024/12/00/me6288/index.html>

Recommendation:

24: wwPDB Pls consider the 2024 Bond & Sussman two-step process.