Worldwide Protein Data Bank Advisory Committee (wwPDB-AC)
Report of the 27 October 2023 Meeting
Held virtually (wwPDB PIs hosted by EMDB, UK)

Advisory Committee Chair:
   Jennifer Martin (IUCr Representative)

Advisory Committee Co-Chair:
   Corinne Smith (Chair of the EBI – Molecular and Cellular Structure advisory board for EMDB and PDBe)

Advisory Committee Members
PDB Site Representatives Nominated by wwPDB partners:
   Paul Adams (RCSB-PDB Chair)
   Kirk L. Clark (RCSB-PDB)
   Linda Columbus (BMRB Advisory Board Chair) APOLOGY
   Chwan-Deng (David) Hsiao (PDBj)
   Juha Huiskonen (EMDB)
   Martin Noble (PDBe)
   Anastassis Perrakis (PDBe)
   Christina Redfield (BMRB)
   Toshiya Senda (PDBj)

ISMAR Representative:
   Tatyana Polenova APOLOGY
   Andy Byrd (ad hoc substitute)

3DEM Representative:
   Basil Greber

wwPDB Institutional Representative:
   Gerard Kleywegt (EMBL-EBI) APOLOGY

wwPDB Principal Investigators:
   Stephen K. Burley (RCSB-PDB) Kyle Morris (EMDB)
   Jeffrey C. Hoch (BMRB) Sameer Velankar (PDBe)
   Genji Kurisu (PDBj)

wwPDB Regional Representatives
   Debasisa Mohanty (PDBi India)
The Worldwide Protein Data Bank Advisory Committee (wwPDB-AC) and the leadership (wwPDB-PIs) of the Research Collaboratory for Structural Bioinformatics (RCSB-PDB), the Biological Magnetic Resonance Data Bank (BMRB), the Protein Data Bank in Europe (PDBe), the Electron Microscopy Data Bank (EMDB), and the Protein Data Bank Japan (PDBj) met together in a 3-hour online meeting organized and hosted by EMDB. This meeting was preceded two weeks earlier by an Executive Session, as discussed in the Introduction section of this report.

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

**wwPDB Mission Statement**
- Manage the wwPDB Core Archives as a public good according to the FAIR Principles.
- Provide expert deposition, validation, biocuration and remediation services at no charge to Data Depositors worldwide.
- Ensure universal open access to public domain structural biology data with no limitations on usage.
- Develop and promote community-endorsed data standards for archiving and exchange of global structural biology data

**Table of Contents**

1. Summary of 2023 wwPDB-AC Commendations and Recommendations
2. Introduction
3. Possible threats to wwPDB OA Model (issues outlined in accompanying docs)
4. Future of OneDep Development (issues outlined in accompanying docs)
5. BMRB Funding Update
6. Future of Data Archiving
7. Next Advisory Committee meeting
8. Carry-Forward advice/recommendations from 2022 report
1. SUMMARY OF 2023 wwPDB-AC COMMENDATIONS AND RECOMMENDATIONS

wwPDB-AC COMMENDATIONS

AC members commend the wwPDB teams on:

1. Arranging Archive access through Amazon Web Services with no storage/egress fees
2. Reaching the milestone of 200,000 wwPDB entries in 2023
3. EMDB reaching the milestone of 30,000 wwPDB entries in 2023
4. Delivering PDB NextGen Archive providing enriched annotation from external databases
5. Recognition by Global Biodata Collection (GBC) as a Global Core Biodata Resource
6. Progress made towards joint development projects for OneDep and NextGen PDB archive
7. Appointment of Dr Kyle Morris as EMDB PI
8. Official appointment of PDBc as an Associate wwPDB Member (Acta D72 p792-5)
9. PDBc curators completing 297 new PDB depositions by end of 2022
10. Development of PDBi including annotation of 573 structures in 12 months
11. Successful outreach activities across wwPDB, including strong social media presence
12. Highly impactful wwPDB presentations at the IUCr Melbourne Congress, August 2023
13. Prioritising and championing open access to wwPDB data
14. OneDep achievements made to date with limited resources
15. Working with software developers on their plans to automate OneDep deposition
16. Award of NIGMS grant to BMRB in a highly competitive funding environment
17. Continued international leadership of high quality data-curation.

wwPDB-AC RECOMMENDATIONS

POSSIBLE THREATS TO OPEN ACCESS MODEL

1. Recommend caution in tackling issues of control of access, at this time.
2. Recommend against proceeding with a joint paper (proposal iii) at this time.
3. Recommend a curiosity-driven approach to gain insight from funding agencies about the risks they see, and to assist PIs and AC to formulate appropriate counter-arguments. This approach could determine whether adopting (proposal iii) in the future might have negative consequences, and whether a statement of impact to specific agencies might be more effective.
4. Recommend that PIs provide immediate updates to AC of any change in the situation, especially of any impact on wwwPDB members’ ability to provide open access to wwPDB data.
5. Recommend that PIs liaise with other international providers of open access data, and the Global Biodata Coalition, to understand the broader situation and likely changes and impact.

FUTURE OF ONEDEP DEVELOPMENT

6. Support the planned development of OneDep, and recommend that wwPDB ensure that improvements do not disenfranchise contributions of new software developers.
7. Recommend and encourage development and reporting of realistic OneDep development goals and development of a timeline with associated milestones and deliverables.
8. Recommend wwPDB PIs intensify their efforts, working with community-based software developers, to apply for funds and secure additional resources. It may be beneficial to highlight the leverage that community-based developers provide to investments made in this area.

BMRB FUNDING

9. Agree that diversifying BMRB funding sources is sensible, and recommend exploring the following possibilities (while ensuring BMRB remains fully open access):
a. Instrument manufacturers (including in-kind contributions);
b. Other Federal funders, such as NSF;
c. European Commission funding through collaboration with European laboratories, for example through the INFRATECH calls of the Horizon Europe program, and developing a relationship with European Research Infrastructure Consortiums (ERICs) such as Instruct-ERIC and the 3D-BiolInfo community of ELIXIR.) It is specifically encouraged to continue engagement with the Director of Instruct-ERIC, Professor Harald Schwalbe, for exploring joint funding opportunities on “investigations” deposition.
d. Devolving some maintenance and development through leverage of data modeling and other tools at other wwPDB resources

e. Plan together as PIs how some BMRB/PDB activities might be integrated into the broader wwPDB consortium.

FUTURE OF DATA ARCHIVING

10. Support the proposal to prepare a position paper highlighting the importance of curated data repositories and the limitations of generalist repositories.

11. Recommend that, in addition to inviting AC members as authors, PIs approach potential authors from user groups beyond structural biology. eg. virology, biology, medical research, vaccine industry, pharmaceutical industry, computational biology, the AlphaFold, DeepMind and Rosetta teams, etc.

12. Recommend wwPDB leverage the proposed development of a more automated data deposition system to articulate a long term vision for data archiving that supports wwPDB mission, FAIR principles and a sustainable system consistent with federal funder requirements.

13. Recommend engaging with NIH, to help NIH develop a more effective approach to data management that supports their own research mission.

14. Recommend wwPDB PIs continue to explore possibilities to leverage funding through the European Open Science Cloud (EOSC) and EOSC-Life initiatives, and to increase presence in related consortia, if possible.

15. Recommend wwPDB PIs develop a holistic communication plan for a data archiving future that includes the proposed publication, stakeholders, and other communication channels.

16. Recommend joining with other databases, and through the Global Biodata Coalition, to emphasize the importance of and demonstrate the value of expert curated resources.

17. Recommend wwPDB PIs develop a broader, coordinated PR plan to address the multiple challenges on the horizon for wwPDB; such a plan could explore the possibility for more emphasis on dissemination through social media platforms.

NEXT ADVISORY COMMITTEE MEETING

18. Agreed that the next wwPDB-AC meeting be held on Friday October 18th 2024, with a similar format to 2023, including an AC only pre-meeting on Oct 4th.

CARRY FORWARD ADVICE/RECOMMENDATIONS FROM 2022 wwPDB-AC REPORT

19. Recommend that the EM-validation white paper be submitted ASAP and no later than end of November 2023, to ensure that its publication remains relevant.

20. Recommend that wwPDB PIs provide the wwPDB AC with a timeline with qualitative and quantitative measures for how PDBc (and PDBi in future) will be assessed as “meeting the global standard to attain full wwPDB membership”.

21. Recommend that meeting “global standard” for wwPDB membership should include a diverse National Advisory Committee.
22. Recommend that pre-meeting papers provided by wwPDB PIs include responses to recommendations from the previous year's wwPDB-AC report.

23. Recommend that AC reports from individual wwPDB members and associate members be provided to wwPDB-AC members at the same time as other papers for AC meetings.
2. INTRODUCTION

In 2023, the AC meeting was held virtually over several weeks to provide sufficient time to discuss the agenda across multiple timezones. EMDB hosted the wwPDB PIs, with Kyle Morris (newly appointed lead of EMDB) the host PI. The meeting proceeded with the following timetable:

1. Three weeks prior to the scheduled virtual wwPDB-AC meeting, each wwPDB PI provided a 5-10 min pre-recorded video to accompany their pre-meeting slides and information papers.
2. Two weeks prior to the scheduled virtual wwPDB-AC meeting, a 1-hour wwPDB-AC-only (Executive Session) virtual meeting was held to discuss the papers.
3. One week prior to the scheduled virtual wwPDB-AC meeting, the wwPDB-AC provided a consolidated list of feedback and questions to wwPDB PIs. The wwPDB PIs provided responses to the questions raised by the wwPDB-AC three days prior to the scheduled meeting.
4. A 3-hour virtual wwPDB-AC meeting was held on Friday 27th October, with the following agenda:
   • Welcome and Introduction, wwPDB EMDB lead, Kyle Morris (10 min)
   • Discuss PI questions and other issues raised by AC - ALL (90 min)
   • Executive Session – wwPDB-AC members and ISMAR and 3DEM representatives (45 min)
   • Summarise wwPDB-AC feedback and further discussion – ALL (20 min)
   • Acknowledgements and photo – ALL (5 min)

AC members agreed that this virtual multi-week/multi-meeting format was an improvement on the 2022 virtual arrangement, providing more opportunity to discuss issues and request additional information.

Background details for this and subsequent sections are provided in accompanying papers.

COMMENDATIONS

AC members commend the wwPDB teams on:

1. Arranging Archive access through Amazon Web Services with no storage/egress fees
2. Reaching the milestone of 200,000 wwPDB entries in 2023
3. EMDB reaching the milestone of 30,000 wwPDB entries in 2023
4. Delivering PDB NextGen Archive providing enriched annotation from external databases
5. Recognition by Global Biodata Coalition (GBC) as a Global Core Biodata Resource
6. Progress made towards joint development projects for OneDep and NextGen PDB archive
7. Appointment of Dr Kyle Morris as EMDB PI
8. Official appointment of PDBc as an Associate wwPDB Member (Acta D72 p792-5)
9. PDBc curators completing 297 new PDB depositions by end of 2022
10. Development of PDBi including annotation of 573 structures in 12 months
11. Successful outreach activities across wwPDB, including strong social media presence
12. Highly impactful wwPDB presentations at the IUCr Melbourne Congress, August 2023
3. POSSIBLE THREATS TO wwPDB OA MODEL (OUTLINED IN ACCOMPANYING DOCS)
wwPDB provided three proposed opportunities to strengthen open access:
(i) wwPDB Leader engagement of national/regional funders to explain the potential impact of open access limitations on global research efforts across fundamental biology, biomedicine, and energy sciences.
(ii) wwPDB Leadership engagement with the Global Biodata Coalition.
(iii) wwPDB Leadership and the wwPDB AC could together contribute to the debate by publishing a “position paper” in a high-profile journal.

wwPDB-AC members strongly agreed that wwPDB data must be kept open access, and that wwPDB PIs should continue to ensure that open access remains at the heart of wwPDB activities. However questions were raised by AC members about the nature of the threat, the timing, and the proposed approach.

AC members noted from the discussions that, under a US federal government directive, a US host university of a wwPDB PI required that access be blocked to certain non-wwPDB resources by investigators based in Russia. The issue had not yet impacted wwPDB activities per se, but wwPDB-PIs were concerned that the situation could become more significant and more widespread, as international tensions grow.

The directive from the host university may in part be due to US export control regulations. The specific US resource affected is one that provides access to computing platforms and software. Such access requires users to register and login. In that case, identifying and blocking access (controlling “export”) to specific users may be possible. For international databases, like the wwPDB, anonymous access to data is the norm and - at least at the moment - not controlled. Data deposition to wwPDB does require identifying information, but deposition of data would be unlikely to fall under the category of export.

Taking the current situation into account, the AC members were concerned about implementing actions when there has as yet been no direct impact on the wwPDB. AC members questioned whether initiating wider debate at this time - when there is no clearly established threat - might make the wwPDB more vulnerable, though AC members appreciate the complexity and significance of the possible threats to Open Access.

COMMENDATION
AC members commend wwPDB teams for:
13. Prioritising and championing open access to wwPDB data

RECOMMENDATIONS
AC members:
1. Recommend caution in tackling issues of control of access, at this time.
2. Recommend against proceeding with a joint paper (proposal iii) at this time.
3. Recommend a curiosity-driven approach to gain insight from funding agencies about the risks they see, and to assist PIs and AC to formulate appropriate counter-arguments. This approach could determine whether adopting (proposal iii) in the future might have negative consequences, and whether a statement of impact to specific agencies might be more effective.
4. Recommend that PIs provide immediate updates to AC of any change in the situation, especially of any impact on wwwPDB members’ ability to provide open access to wwPDB data.
5. Recommend that PIs liaise with other international providers of open access data, and the Global Coalition for Data, to understand the broader situation and likely changes and impact.
4. FUTURE OF ONEDEP DEVELOPMENT (OUTLINED IN ACCOMPANYING DOCS)

Does the AC concur with the plans for automating the deposition process as the future direction for OneDep development?

SAC members strongly agreed that these are excellent and timely plans and that automation would be welcomed by users. Increasing the relevance of wwPDB for the future will also depend, to some extent, on capturing massive datasets being generated by modern methodologies which are currently largely undeposited (e.g. high throughput diffraction experiments supporting drug discovery, and especially hit-finding) and which will be a high priority amongst stakeholders.

AC members noted that the track record for meeting OneDep development timelines is poor. In response, the PIs outlined three major reasons for delays, these being (i) insufficient resources, (ii) unanticipated changes to deposition requirements (new data files, additional metadata requirements), and (iii) the requirement to implement essential operating system updates and maintenance tasks, which divert limited resources. With rapid advances in experimental techniques and software packages, OneDep requires constant updates to accommodate new requirements. The pace of change has accelerated in the last 5-10 years, whereas the resources available for OneDep development have remained constant at best or declined. Additionally, RCSB PDB, PDBe, PDBj, EMDB and BMRB depend upon securing competitive grants to carry out this work, which makes planning and allocation of resources challenging.

OneDep is more than 15 years old and some technology requires changes to keep the system running when operating systems or libraries are updated. PIs have taken steps to address some issues by refactoring processes and continuing to address the technical debt as and when possible. Ultimately, the only way to minimise delays is to make additional resources available for the OneDep team.

OneDep automation is one of the most significant developments in recent years and will require a complex, highly collaborative software development effort. To be done properly, additional resources and close coordination will be needed across the wwPDB and the software development community.

AC members queried how wwPDB will work with software developers to ensure appropriate resources and consultations are available for a smooth integration of this major new OneDep development. PIs responded that the wwPDB OneDep team organises hackathons with software developers to integrate the deposition API into individual software packages. This model will continue to be used to ensure efficient interaction with developers. Also, RCSB PDB and PDBe have submitted a letter of intent to submit a full grant proposal to BBSRC and NSF in collaboration with software developers (CCP4, Phenix, Global phasing and CCP-EM) to develop the deposition API and automated deposition. If this EOI is successful and a grant funded, it will provide a mechanism to fund and coordinate future developments in the automation of deposition and annotation processes.

A grant application to fund OneDep implementation of “investigations” as groupings of depositions, was unsuccessful. wwPDB AC and PIs discussed alternative strategies for resourcing of these crucial developments. It was noted that software development consortia such as CCP4 and Phenix recognise the urgency of automating OneDep, but are equally resource-constrained. wwPDB and community software development teams will continue to work together on grant applications.

The AC members shared the view of wwPDB PIs that streamlining structure deposition through automation is a major priority and that additional resources will be required to deliver these developments.
COMMENDATIONS
AC members commend the wwPDB teams on:
14. OneDep achievements made to date with limited resources
15. Working with software developers on their plans to automate OneDep

RECOMMENDATIONS
AC members:
6. Support the planned development of OneDep, and recommend that wwPDB ensure that improvements do not disenfranchise contributions of new software developers.
7. Recommend and encourage development and reporting of realistic OneDep development goals and development of a timeline with associated milestones and deliverables.
8. Recommend wwPDB PIs intensify their efforts, working with community-based software developers, to apply for funds and secure additional resources. It may be beneficial to highlight the leverage that community-based developers provide to investments made in this area.

5. BMRB FUNDING UPDATE (OUTLINED IN ACCOMPANYING DOCS)
Does the AC have suggestions for alternative funding mechanisms to sustain BMRB?
The BMRB NIGMS funding application was approved but with a 44% reduction on the requested amount. This level of funding is insufficient for the necessary remediation of BMRB data, expanding curated data types, retiring technical debt, enhancing user experience, and improving sustainability. The 44% reduction in the proposed budget came with no guidance from NIGMS on how it should be allocated. BMRB has been invited by NIGMS to submit revised aims and a corresponding budget reflective of the award amount. In consultation with the BMRB AC, BMRB will prioritize quotidian operation (“keeping the lights on”). It appears that BMRB has reached the limits of NIGMS support, and will need to seek other sources of support in order to modernize BMRB.

COMMENDATIONS
AC members commend the wwPDB on the:
16. Award of NIGMS grant to BMRB in a highly competitive funding environment

RECOMMENDATIONS
9. Agree that diversifying BMRB funding sources is sensible, and recommend exploring the following possibilities (whilst ensuring BMRB remains fully open access):
   a. Instrument manufacturers (including in-kind contributions);
   b. Other Federal funders, such as NSF;
   c. European Commission funding through collaboration with European laboratories, for example through the INFRATECH calls of the Horizon Europe program, and developing a relationship with European Research Infrastructure Consortiums (ERICs) such as Instruct-ERIC and the 3D-BioInfo community of ELIXIR.) It is specifically encouraged to continue engagement with the Director of Instruct-ERIC, Professor Harald Schwalbe, for exploring joint funding opportunities on “investigations” deposition.
   d. Devolving some maintenance and development through leverage of data modeling and other tools at other wwPDB resources
   e. Plan together how some BMRB/PDB activities might be integrated into the broader wwPDB consortium.
6. FUTURE OF DATA ARCHIVING (OUTLINED IN ACCOMPANYING DOCS)
Creating and maintaining trusted repositories is expensive and labor-intensive. Costs are growing.

• Short term: Additional funding is needed to sustain operations
• Long term: Research is needed to develop technologies that reduce skilled labor costs (AI/ML)

1. How can we engage funders and the community to inform the discussion on the well-intentioned mandates from funders and publishers for archiving data in generalist repositories, but which will have an unintended detrimental impact on funding of specialist repositories and data FAIRness?
2. Would the AC be willing to join a wwPDB effort publishing a “position paper” in a high-profile journal?

wwPDB members described significant differences of opinion across US funding agencies regarding the value of expert data validation/curation/archiving and allocation of financial responsibility to ensure this happens [versus the less expensive (for the agency) option of simply asking every funded investigator to deposit their data into a public repository with no data standards, etc]. The AC used an analogy of data dumps as a ‘swamp’ of data, in comparison with well-curated data as a ‘free flowing river’. The wwPDB PIs believe that additional public campaigning is required to resolve these fundamental differences of opinion among funding agencies. The AC members note that NIGMS funding of RCSB PDB has not increased in a decade, while both NSF and DOE have increased funding during the same period. AC members strongly support the notion that emphasizing the significance and impact of well-curated data resources is critical, and commend the wwPDB on their leadership of this issue.

COMMENDATIONS
AC members commend the wwPDB teams on their:
17. Continued international leadership of high quality data-curation.

RECOMMENDATIONS
AC members:
10. Support the proposal to prepare a position paper highlighting the importance of curated data repositories and the limitations of generalist repositories.
11. Recommend that, in addition to inviting AC members as authors, PIs approach potential authors from user groups beyond structural biology. eg. virology, biology, medical research, vaccine industry, pharmaceutical industry, computational biology, and AlphaFold, DeepMind and Rosetta teams, etc.
12. Recommend wwPDB leverage the proposed development of a more automated data deposition system to articulate a long term vision for data archiving that supports wwPDB mission, FAIR principles and a sustainable system consistent with federal funder requirements.
13. Recommend engaging with NIH, to help NIH develop a more effective approach to data management that supports their own research mission.
14. Recommend wwPDB PIs continue to explore possibilities to leverage funding through the European Open Science Cloud (EOSC) and EOSC-Life initiatives, and to increase presence in related consortia, if possible.
15. Recommend wwPDB PIs develop a holistic communication plan for a data archiving future that includes the proposed publication, stakeholders, and other communication channels.
16. Recommend joining with other databases, and through the Global Biodata Collection, to emphasize the importance of and demonstrate the value of expert curated resources.
17. Recommend wwPDB PIs develop a broader, coordinated PR plan to address the multiple challenges on the horizon for wwPDB; such a plan could explore the possibility for more emphasis on dissemination through social media platforms.
7. NEXT ADVISORY COMMITTEE MEETING

A 2024 virtual AC meeting was agreed requiring a similar approach to 2023, where PI presentations and information (including individual AC reports, videos, and responses to recommendations from 2023 report) are provided at least 3 weeks ahead of the full meeting, and a 1-hour AC only meeting is held two weeks before the full meeting. For this PDBj-hosted meeting, wwPDB PIs will be located in Osaka, and the AC meeting will be virtual. To cover the global contributions to wwPDB, the same time slots as in 2023 will be used: 9pm-midnight Japan. wwPDB PIs will join the AC meeting from individual hotel rooms.

RECOMMENDATION

The AC members:

18. Agreed that the next wwPDB-AC meeting be held on Friday October 18th 2024, with a similar format to 2023, including an AC only pre-meeting on Oct 4th.

6. CARRY FORWARD ADVICE/RECOMMENDATIONS FROM 2022 wwPDB-AC REPORT

In our 2022 report, AC members asked that the 2023 pre-meeting papers include responses to recommendations/questions raised in the wwPDB-AC 2022 report. We note that this hasn’t happened, so we include relevant 2022 recommendations again in the 2023 report.

The 2022 AC report included the following: The wwPDB-AC recommended that the EM-validation white paper (already delayed in 2021) be submitted as soon as possible so that its publication would support changes already implemented by the wwPDB, and demonstrate community endorsement of these changes. The wwPDB-AC also requested an update on plans to establish a working group focused on 3DEM content in both the PDB and EMDB archives (keeping in mind diversity and inclusion).

This topic remains critical and timely. Validation standards for cryo-EM structures still need to be fully adopted across the cryo-EM community. At this point in time, the white paper is increasingly out of date and so needs to be published ASAP or dropped. Many of the points made in the paper may have already been implemented, or new approaches may now be more relevant. The emergence of AI/machine learning-based tools, e.g. for map manipulation, will result in further challenges and require additional standards and/or guidelines in the future. One option is to hold a second validation meeting and start over. However, concerns were raised that the history of the first workshop and delays in publishing the white paper greatly reduced enthusiasm by former workshop members for engaging in future workshops. The wwPDB PIs undertook to ensure positive action and engagement with the community going forward and respond to workshops deliverables in a reasonable timeframe.

RECOMMENDATIONS

AC members:

19. Recommend that the EM-validation white paper be submitted ASAP and no later than end of November 2023, to ensure that its publication remains relevant. The EM validation white paper was submitted 29 Nov 2023 to arXiv http://arxiv.org/abs/2311.17640.

20. Recommend that wwPDB PIs provide the wwPDB AC with a timeline with qualitative and quantitative measures for how PDBc (and PDBi in future) will be assessed as “meeting the global standard to attain full wwPDB membership”.

21. Recommend that meeting “global standard” for wwPDB membership should include a diverse National Advisory Committee.

22. Recommend that pre-meeting papers provided by wwPDB PIs include responses to recommendations from the previous year’s wwPDB-AC report.

23. Recommend that AC reports from individual wwPDB members and associate members be provided to wwPDB-AC members at the same time as other papers for AC meetings.