RCSB Protein Data Bank Advisory Committee
Report of April 12, 2023 Annual Meeting Videoconference

Chair: Paul Adams

Membership:

Paul Adams, Peter Andolfatto, Bridget Carragher, Wah Chiu, Kirk Clark, Robert B. Darnell, Roland Dunbrack, Paul Falkowski, Thomas Ferrin, Cathy Peishoff, Sue Rhee, Torsten Schwede, Lance Stewart, Kevin H. Gardner, Takita F. Sumter

Present: Paul Adams, Wah Chiu, Kirk Clark,Roland Dunbrack, Thomas Ferrin, Mandë Holford, Cathy Peishoff, Sue Rhee, Torsten Schwede, Lance Stewart, Kevin H. Gardner, Takita F. Sumter

Absent: Peter Andolfatto, Bridget Carragher, Robert B. Darnell, Paul Falkowski, Mandë Holford

RCSB PDB AC E-mail Addresses:

pdadams@lbl.gov, pa2543@columbia.edu, bcarr@nysbc.org, wahc@stanford.edu, kirk.clark@novartis.com, darnelr@rockefeller.edu, roland.dunbrack@gmail.com, falko@marine.rutgers.edu, tef@cgl.ucsf.edu, mholford@hunter.cuny.edu, peishoffc@gmail.com, srhee@carnegiescience.edu, torsten.schwede@unibas.ch, ljs5@uw.edu, kgardner@gc.cuny.edu, sumtert@winthrop.edu

RCSB PDB Leadership: Stephen Burley (Director), Helen Berman (Director Emerita), Andrej Sali (UCSF Site Head)

RCSB PDB Leadership E-mail Addresses: sburley@proteomics.rutgers.edu, berman@rcsb.rutgers.edu, sali@salilab.org

Executive Summary

The Advisory Committee (AC) to the Research Collaboratory for Structural Bioinformatics (RCSB) held a virtual meeting on April 12th, 2023 to review recent progress and provide feedback on specific questions.

Agenda items included:

- Welcome and Introductions
- 2022 Overview and Brief Updates
- Computed Structure Models
- PDB-Dev Unification with PDB archive
- Succession Planning
- Diversity Equity Inclusion and Accountability
- Sustainability

The meeting was opened by Dr. Stephen Burley. Other RCSB PDB participants were Helen M. Berman, Andrej Sali (UCSF), Jose Duarte (Scientific Software Lead and UCSD Manager), Henry Chao (DevOps Lead East), Zukang Feng (Principal Scientific Software Developer), Yana Rose (S3 Lead; Scientific Software Developer & Data Architect), Jasmine Young (S1-2 Lead; RCSB PDB Biocuration Team Lead & wwPDB Global Project Lead), Christine Zardecki (Deputy Director, S4 Lead).

The meeting was divided into 2 halves. The first provided updates on aspects of RCSB activities over the last year and included some questions for the committee. The second half focused on planning for the upcoming funding agency led review of the recently submitted renewal proposal. Appendix 1 provides a summary of the RCSB responses to the 2022 Advisory Panel meeting recommendations. Appendix 2 provides a summary of global PDB deposition and data access statistics in 2022.

Overall Comments from the Advisory Panel

The team is again congratulated on their outstanding work over the last 12 months, which has been dominated by the preparation for the renewal of the RCSB funding. Parallel with these planning activities there has been significant progress in all service components and also the inclusion of a large number of computed models in what is served to users. This last year has been a record for structures deposited, released and served. The committee greatly appreciated the very focused presentation material and the time provided for discussion.

> *Recommendations*
>
> Prepare a concise and compelling presentation for the reverse site visit, with backup slides on topics and questions that might arise during the review. The committee is very willing to review materials prior to the review if that would be helpful.

Detailed Advisory Panel Comments and Feedback

2022 Overview and Brief Updates

The committee heard from Stephen about RCSB performance in 2022, with records being set in most areas, including structure depositions and releases, and web traffic for data and outreach. The team is clearly working well and able to keep up with the growing needs of the structural biology community. However, it is also clear that to meet future needs, an increase in core funding is required.

Concise summaries of recent progress in the 4 service components of RCSB highlighted significant advances. Jasmine Young continues her excellent leadership of Service 1 and 2. The sustained improvements to the OneDep system are recognized by the committee, as is the work to include computed structure models. We are also very pleased to see that PDBc is now an Associate Member and actively processing depositions from Asia. Yana Rose is leading Service 3, with significant efforts on delivery of computed models to PDB users, combined with new visualization tools. The committee was particularly impressed with the User Experience Design review, which has already led to improvements in content delivery. Christine Zardecki continues her excellent leadership of Service 4, where there has been a website redesign to emphasize training, and holding boot camps for undergraduate and graduate students. The committee was

very pleased to see the creation of a Training, Outreach, and Education Working Group to provide feedback on RCSB resources and help promote RCSB-based training.

Recommendations

The RCSB is encouraged to increase the adoption of community-based feedback where it will help refine services and the RCSB user experience.

Several questions were posed to the committee, with responses below:

Any concern on our responses to last year's AC report?

The committee has no concerns.

Any concerns regarding 2022 wwPDB AC activities?

The committee has no concerns but concurs with that committee's concerns about the slow pace of implementing new validation metrics for cryo-EM. However, we recognize that this is not under RCSB control and applaud the RCSB for implementing the Q-score as a new metric for validation of cryo-EM structures.

Any suggestions on venues to communicate extended PDB/CCD IDs?

The committee recommends leveraging the larger community organizations and their associated meetings, such as IUCr and ACA. The RCSB did receive feedback from the ChimeraX team in response to the proposed plan. As much of the work to respond to this change will be by the software developers, continued and timely engagement with the mmCIF/PDBx working group is recommended. Some members of the panel explained that making the necessary code changes to support extended PDB/CCD IDs is a very large project and could take considerable time.

Thoughts on approaches to CSM integration?

The committee appreciates the RCSB efforts to integrate computed structure models (CSMs) and view this as critical to the future success of the organization. However, this is not without its challenges and concerns. For example, there was significant discussion about whether CSMs should be provided by default for user searches, or whether the user should be required to opt-in. The committee had differing opinions, with some supporting opt-out and some supporting opt-in. For some there is the concern that users will be unable to differentiate between CSMs and experimental structures and treat them as equivalent quality and utility. What was universally agreed was that the current method of notifying users that they can opt-in is not prominent enough and likely overlooked by many users. One proposed solution was to keep the opt-in approach but provide prominent messages to the users as a result of a search that indicates if CSMs are available, accompanied by some caveats about quality and use.

Recommendations

The committee suggests engaging with the User Experience Design group to get feedback on what would be reasonable approaches to providing CSMs to the community. The committee also recommends more user engagement to inform about CSMs, how they are made available and their strengths and weaknesses.

What would be best approach for making new RCSB PDB features more known/accessible?

The committee recognizes that informing the community about new features can be challenging, although the RCSB has a strong track record of outreach in various outlets to keep users up to date. However, it is probably worth crafting some information outreach to be consistent with current best practices – e.g. short video based guides.

Recommendations

The committee recommends creating a series of web-based short video guides to inform users about new features, which can then be a lasting source of information for anyone new to the RCSB. A YouTube channel provides an easy mechanism for users to find content. In addition, it is probably timely to think about a redesign of the RCSB web site, which has had the same look and feel for several years and could benefit from adopting current best practices for website design, including links to the short video guides on feature use.

Topics for future crash courses/training?

The RCSB crash courses are an effective way to reach the user community and a place to continue to invest effort if resources permit. The committee felt this was an opportunity to inform and train users on a variety of topics, including advanced searches, programmatic interfaces to access RCSB resources, computed structure models, and graphical tools such as Mol*.

Current strategies for 2023 Review

The committee heard from Stephen about the logistics and planning for the reverse site visit in June, to determine the next round of funding for the RCSB. We congratulate the team on putting together a strong proposal for consideration by the funding agencies and peer reviewers. The committee is also very encouraged to hear that the funders are entertaining a funding request that provides a much-needed increase in support for core operations. We support the approach of focusing on capacity building and sustainability, and pivoting outreach to training. The goal of embracing AI/ML technologies more broadly is strongly supported by the committee, and they could have a significant impact on biocurator efficiency and data query interfaces at the RCSB. There were discussions on specific topics, that would be part of the limited material presented at the reverse site visit:

Computed Structure Models

RCSB plans for expanding the library of CSMs were presented and discussed. The committee agrees that this is essential for the long-term viability of the RCSB and the benefit to the research community is very significant. It is recognized that this has implications for how core services will need to operate, and the RCSB plans are appropriate, and should simplify operations in the future.

PDB-Dev Unification with PDB archive

The committee supports the proposal to unify PDB-Dev with the core archive. Adequate resources will need to be provided as part of the renewal to support the unification and long-term support of this more complex infrastructure. The committee believes that this is very

important for the research community, as structural biology moves more towards hybrid models driven in part by the availability of predicted structures.

Diversity Equity Inclusion and Access

The committee applauds the RCSB attention to diversity, in multiple dimensions. Beyond the excellent activities already proposed or underway, we suggest leveraging any institutional strategic plans that highlight diversity efforts and perform targeted outreach to institutions in IDEA states (as defined by NIH).

Sustainability

The RCSB participation in the Global Biodata Coalition is a good development and provides strong evidence of the team's commitment to a sustainable resource.

Recommendations

Provide more material to demonstrate the credibility of the Global Biodata Coalition, as funders and reviewers are unlikely to be familiar with it.

Discussion based on the questions posed to the committee (removed)

Next Advisory Committee Meeting: scheduling, location

A doodle poll will be conducted to gather possible dates for an in-person meeting at Rutgers in 2024.

Appendix 1: Responses to Report of March 15th, 2022 Annual Meeting
(AC commentary in italics and has been truncated for brevity)

2023 Overall Comments from the Advisory Panel

The team is again congratulated on their outstanding work over the last 12 months, as the field

of structural biology undergoes significant changes with the wider availability of highly accurate structure prediction methods. The team gave an excellent set of focused presentations, and their collective leadership of the RCSB and their individual projects is applauded. It was good to see that new people have joined the team despite the pandemic, and that interviews are underway for additional hires. The committee recognizes that the major activity for the team in the coming months is preparation for the renewal of the RCSB funding.

2022 AC Recommendations

- The committee encourages the RCSB team to make use of the committee's expertise to assist in preparing for the renewal.

2023 Response: We appreciate the feedback that the advisory committee provided at the October 20, 2022 pre-proposal AC meeting and during the proposal writing process, and look forward to your comments on the materials prepared for the Site Visit.

Deposition/Validation/Biocuration, Remediation, and Archive Management

The committee heard from Jasmine Young about deposition and biocuration activities…again congratulate Jasmine for her leadership of OneDep and creating a strong collaboration across the wwPDB.

2022 Recommendations

- At a future committee meeting the team should present their strategy for increasing automation of deposition of more complex structures, including any implications for developer resources needed. This might be timely given the renewal of funding activities in the coming months.

2023 Response: Several plans have been put in place for more automation to support deposition, validation, and biocuration of larger or more complex structures. Some are on the 2023-2024 roadmap, while others are described in the RCSB PDB renewal proposal (2024-2028).

Improvements on automation:

- Improve validation run time by providing modular and parallel validation calculations (RCSB PDB)
- Enhance file re-upload infrastructure to make the process more efficient for depositors (PDBe)
- Better handling of composite maps for 3DEM structures (EMDB)
- Provide more sanity checking of cross-field metadata, when possible (PDBe)
- Extend pdb_extract tool on data harvesting for Cryo-EM/Cryo-ET structures to provide more complete data files for PDB deposition (RCSB PDB)
- Collaborate with refinement software developers to provide more complete data file for PDB deposition (All)
- Continue providing more automated annotation and checking such as outlier detection and presenting to depositors and biocurators (RCSB PDB/PDBe)
- Incorporate 3DEM image recognition to detect unusual 3DEM maps (EMDB)
- Scaling for performance improvements (All sites)
- Deploy GroupDep features within OneDep, e.g., simultaneous deposition of multiple related structures such as investigation study (RCSB PDB)

Data Exploration (RCSB.org)

The committee heard from Yana Rose on the improvements in data delivery through the RCSB.org web site…

Looking to the future of the RCSB.org web site, the committee is very enthusiastic about the recently initiated Rutgers user experience design review and the involvement of Paul Craig and the BioMolViz community. We expect that this will further enhance the usability of the site in the future...

2022 Recommendations

- Develop a plan for user interaction with computed structure models in the RCSB.org interface. This should leverage community input and be considered as part of the UXD review if possible. Attention will need to be paid to the likely increase in the number of non-structure aware users interested in computed structure models.

2023 Response: On September 27, 2021, the RCSB PDB AC endorsed the plan to incorporate CSMs at RCSB.org. During 2022, the infrastructure and software stack behind RCSB.org was updated to accommodate the new data, which represented a 6-fold increase in data volume. Interoperation of CSMs with all tools and features at RCSB.org was enabled by the extension of the PDBx/mmCIF data standard with the new ModelCIF data standard developed for CSMs. This expansion of the purview of RCSB.org allows us to continue serving as a one-stop shop for studying the 3D structures of biomolecules by providing PDB data consumers with access to CSMs covering the entire human proteome as well as that of many model organisms, selected pathogens, organisms relevant to bioenergy research, and protein complexes from select studies.

The initial set of Computed Structure Models (CSMs) were released at RCSB.org on August 31, 2022 (announcement, user guide documentation).

The initial response was mixed: users were interested in exploring CSMs using RCSB PDB resources, but did not like that CSMs were included in search results by default. We quickly changed the user interface so that users need to "opt in" explicitly for conclusion. This change and the immediacy of the response was very well received.

To further promote these features, "RCSB Protein Data Bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning" as also published in the 2023 *Nucleic Acids Research* Database Issue (51: D488–D508 doi: 10.1093/nar/gkac1077).

An additional set of CSMs was released on February 1, 2023 (announcement).

2022 Recommendations

- Report out to the committee on the results of the UXD review at the next meeting, and seek our input prior to implementing the proposed changes.

2023 Response: The Spring 2022 UXD Report (PDF) and our response (PDF) was shared with Advisors in May 2022. Some of the features that have been incorporated include footer rearrangement to conform to the standard column look, standardized design of the "Contact Us" button, and improvements to better communicate search results. Other recommendations will be implemented in 2023.

.

Over the summer, the UXD students worked with former RCSB PDB AC member Paul Craig

during his sabbatical at Rutgers on a review of Mol*. Recommendations are available (PDF) and will be reviewed and implemented in 2023.

An additional UXD review is scheduled for the summer of 2023 with a focus on Structure Summary pages.

Delivering Computed Structure Models Alongside PDB Data

The committee heard from Stephen Burley on the RCSB plans for serving computed structure models alongside the experimental PDB data. The plan is to serve ~800 thousand models from key proteomes first, and then propose scaling this to 100 million models as part of the renewal of funding…This will clearly be an area for further developments as the prediction methods evolve, and the RCSB should keep a close eye on community development of new validation approaches in addition to their own developments.

2022 Recommendations

- *At the next committee meeting present statistics on the access of computed models by RCSB users, and any insights on their use by the community.*

  2023 Response**:** Our initial communication plan included wide distribution of the initial announcement, release of user guide documentation, and an overview feature on PDB-101. A Virtual Crash Course was held September 22, 2022 (training materials are published at PDB-101) where initial feedback was collected. Types of requests made in this feedback included interest in being able to more directly compare CSMs alongside experimental PDB structures to being able to offer more CSMs for other organisms, as well as allowing users to upload their own model files (e.g., of their own structure predictions) for performing structure similarity search. We are pleased to state that we have addressed these requests, including the release of the Groups Alignment Viewer, inclusion of additional CSMs from ModelArchive, and the release of support for user-uploaded coordinates in structure similarity search.

Following the initial delivery of CSMs on RCSB.org (September 2022), our RCSB PDB Usage Analytics system demonstrates an increase in the average monthly website bandwidth and number of unique visitors (based on client IP addresses) relative to before the delivery date (when comparing the six months prior to September 2022 versus the six months afterwards). Additionally, we have been monitoring access to CSMs (via downloads and visits to Structure Summary pages) as well as how often users are including CSMs in their searches. Although the relative usage of experimental PDB structures still far exceeds that of CSMs, we are observing consistent access to CSMs via each of these metrics.

Outreach/Education (PDB-101); PDB50

Christine Zardecki presented many of the great efforts in outreach and education over the last year. A highlight was the activities celebrating 50 years of the PDB archive...The committee wondered what else can be done to extend the reach to underserved communities, appreciating that some of those activities are better in person than virtual.

2022 Recommendations

- *At the next committee meeting present progress and future plans for outreach to underrepresented minorities and underserved communities.*

2023 Response: “Diversity and Inclusion” are organizational Core Values (See Vision & Mission and DEIA at RCSB.org). RCSB PDB works to increase participation and access through advocating for our community members and emphasizing the importance of promoting diversity in learning and development. Rutgers, UCSD, and UCSF, and, in particular, Rutgers Institute for Quantitative Biomedicine (IQB) institutional priorities guide our ongoing DEIA strategic planning. RCSB PDB has been a leader of DEIA within IQB; a major milestone was Director Emerita Berman mentoring students who launched the first Rutgers SACNAS (Society for Advancing Chicanos/Hispanics & Native Americans in Science) Chapter. Four team members served on the Inaugural Planning Committee for Diversity and Inclusion at IQB and helped develop the IQB DEIA strategic plan. RCSB PDB Leadership continue to serve on the Committee, informing efforts across the RCSB PDB.

Reaching communities through free, public websites (RCSB.org and PDB-101) is our most powerful means of advancing our DEIA goals. As a matter of policy, we do not collect identifying or demographic data from PDB Data Consumers. We do, however, conduct periodic User Surveys and Exit Surveys from Training events. Baseline data were established with a 2021 survey of RCSB.org visitors wherein ~35% of respondents identified as non-White/Caucasian. We aim to broaden user demographics by expanding Training efforts and advertising these materials widely and targeting MSIs. We are also committed to improving RCSB PDB web portal accessibility through UXD reviews and improvements.

We promote RCSB PDB and recruit for personnel at the Annual Biomedical Research Conference for Minoritized Scientists (ABRCMS) and the National Diversity in STEM conference, and at targeted institutions in concert with the Rutgers School of Graduate Studies. We maintain an alumni database for networking and assessment purposes. RCSB PDB follows inclusive hiring practices (*e.g.,* advertising positions in targeted venues like SACNAS, defining searches broadly, considering candidates with non-traditional backgrounds) and interviewing all qualified underrepresented minority applicants.

RCSB PDB proactively invites scholars from underrepresented and underserved communities to Seminars, Crash Courses, Symposia, Training Events, and Boot Camps, drawing participants from community colleges, universities, and industry. These events are attended by thousands annually, with some events (such as PDB50 Symposia) continuing to find audiences through online recordings. Since January 2021, the annual IQB Winter Boot Camp on Science Communication (virtual) has been promoted to and avidly attended by students from Minority Serving Institutions (MSIs, 18 of 25 participants in 2022). Each attendee receives Rutgers college course credit plus authorship on the resulting publication. We plan to continue the IQB Winter Boot Camp annually, monitoring registrants, course feedback, and *Molecule of the Month* article page views for assessment. The IQB Business Office collects and monitors demographic data for all RCSB PDB-sponsored events.

Since 2015, RCSB PDB has partnered with the Research Intensive Summer Experience at Rutgers program (RISE; rise.rutgers.edu) to host summer undergraduate research interns, with a focus on students from underrepresented, disadvantaged, non-traditional or first-generation college backgrounds, or those with limited research opportunities (in partnership with Dr. Evelyn Erenrich, Chief Diversity Officer and Associate Dean, Rutgers School of Graduate Studies).

Students are paired with faculty mentors and receive additional professional development focused on graduate school admissions. As of January 2023, RCSB PDB team members had mentored 24 RISE students, yielding poster awards and peer-reviewed publications. Going forward, we plan to recruit RISE students for scientific software projects. The RISE-RCSB PDB collaboration has created a community network extending also to students who were merely applicants; we recently hired two individuals whom we met as result of their interest in the program. Team members also mentor students at host institutions on projects implemented within RCSB.org and PDB-101 resources.

We have additionally established a "Training, Outreach, and Education Working Group" chaired by former RCSB PDB Advisor Paul Craig. This working group will also provide guidance on DEIA activities.

Operations, Funding and Strategic Initiatives

The committee heard about recent operational improvements, recruiting activities and diversity, equity, and inclusion…

Two strategic initiatives were highlighted, both very important. The migration of backend RCSB.org services to cloud computing is of great significance, and essential for the feasibility of RCSB operations in the future…

The second initiative was the PDB-Dev archive and improved validation information for integrative/hybrid methods structures. This is an important growth area, which will likely be accelerated by the availability of the computed structure models...

We heard about the succession planning efforts for a new RCSB director in the coming year….

Finally, the funding activities since the last renewal were reviewed. The committee notes the huge effort that has gone into raising additional federal grant monies to make up the difference between the funding requested and the funding received at the last renewal. While these efforts were necessary, we do wonder if the funding agencies are best served by requiring this very successful and important community resource to devote so much time and effort to these additional fundraising activities. Maybe a more holistic view of the RCSB funding by the agencies in the upcoming renewal process would serve them and the RCSB better.

2022 Recommendation

- *Provide an update on succession planning progress before the next regular committee meeting, both for RCSB and its home Institute the Rutgers Institute for Quantitative Biomedicine.*

  2023 Response

  The Director continues to work with Rutgers to repeat the process used for his recruitment, and are starting to schedule candidate seminars starting in Fall 2023. A contingency plan is in place in case of any emergency. Andrej Sali would serve as interim lead, and the Operations Team are well-positioned to provide continuity as needed. We are confident that the current organizational structure and controls will ensure the PDB endure for decades.

2022 Recommendation

- *Provide an update on the outcome of the upcoming NSF/NIH/DOE mid-funding cycle review site visit before the next regular committee meeting.*

  2023 Response: The report from the funding agencies is available for review; the agencies completed a SWOT analysis of the RCSB PDB project and "Although the panelists recognized the value of PDB's approach to open data, the essential prohibition on use charges is a severe threat to sustainability. The charter itself may not be sustainable without a change in the funding strategy. Alternate funding strategies for the future should be evaluated."

Discussion

There was a discussion about what the team should focus on in preparation for the upcoming renewal proposal, to be submitted around February 2023 (actual date TBD, but reasonably extrapolated on the timing of the last renewal). The committee feels that the RCSB is well placed to be successful with a renewal proposal, especially given the stable technical platform, great advances in the structure deposition system, and the increases in biocurator efficiency. The increasing number of cryo-EM derived models presents the RCSB team with a need for increased deposition/curation capacity and data delivery systems to better handle more complex models. This emphasizes the importance of a strong, stable and expanded team. One of the most exciting opportunities for the RCSB in their renewal planning is articulating how they will become the nation's centralized resource for providing experimental and computed models to the research community and beyond. We believe that they are uniquely qualified to do this, and if provided the necessary resources the impact will be immense. It is incontrovertible that the public availability of the protein structures in the PDB enabled computational prediction methods to reach their current level of success. There are still many prediction problems to be solved, including but not limited to, nucleic acid structures, protein/ligand complexes, and post translational modifications. The RCSB team can make an excellent case that they need to be able to expand their activities to ensure that they can provide the data for the community to solve these prediction problems in the future. They are also uniquely positioned to help researchers make use of the combination of experimental and computed structures to accelerate experimental structure determination. This will undoubtedly expand into integrative/hybrid methods structure determination, which will be practiced much more widely with the availability of computed models. The RCSB needs to be prepared to help the community obtain the best models, understand their strengths and weaknesses, and ultimately archive them for future use by others.

Recommendations

- *Engage the Advisory Committee (AC) in reviewing the renewal plans early on. Specifically, a virtual meeting should be held with the committee 4-6 months prior to the renewal submission.*

2023 Response*:* We thank the advisors for their very constructive feedback that enabled our team to submit a stronger proposal on February 28, 2023. We plan to have a high-level discussion with advisors on April 12 about general plans for the site visit, and will share the site visit materials that will be generated as soon as we receive any guidance from the NSF.

Appendix 2: PDB 2022 Metrics

In aggregate, 16,344 depositions were received and processed between January 1st and December 31st, 2022, with an average turnaround of two weeks by the wwPDB. This represents an increase from the 14,571 entries deposited in 2021.

Breakdown of depositions by discipline was as follows:

X-ray: 10,624 (65% of entries deposited, up from 9,937 in 2021)
NMR: 287 (1.8%, down from 364)
EM: 5407 (33%, up from 4,254)
Other: 26 (.2%, up from 16)

Breakdown of depositions by wwPDB processing site was as follows:

RCSB PDB: 7022 (43%)
PDBj: 4236 (26%)
PDBe-EBI: 4811 (29%)
PDBc: 275 (<2%)

Breakdown of depositors by location was as follows:

North America 5630 (34.4%)
Europe 5471 (33.5%)
Asia 4707 (28.8%)
South America 87 (<1%)
Oceania 439 (2.7%)
Africa 10 (<1%)

For 2022, Google Analytics (GA) reported that RCSB.org hosted more than 4.7 million unique users, ~11.9 million sessions, and ~45 million page views. In tandem with GA-based web analytics, RCSB PDB Usage Analytics reported ~7.2 million unique clients (unique IP addresses), 61 million sessions (~37 million of which are estimated to be human users), and ~2.8 billion requests/interactions (e.g., data downloads, service usage, web page content views).

In 2022, 2.2 billion data files in various file formats, including structure files, experimental data files, chemical and molecular reference data files, and validation reports, were downloaded and/or viewed from RCSB PDB-hosted FTP and websites.

Additional data were downloaded from wwPDB partners PDBe and PDBj, for nearly 3 billion data file downloads.