## Witness Statement: R. Andrew Byrd, Economics and Impact of the Protein Data Bank (PDB)

**1. What is the main sources of income (who pays and how) and how are these presented as a business model?**

The PDB Archive is a high-value, primary archival database. This resource is the single global repository for experimentally determined 3D structures of biological macromolecules. It complements GenBank, the US data repository for experimentally determined DNA and RNA sequences. Since 2003, the PDB Archive has been managed by the Worldwide Protein Data Bank or wwPDB (wwpdb.org) organization. Independently-funded regional Data Centers located in the US, United Kingdom, and Japan collaborate on data curation [deposition, annotation, validation, and remediation] while maintaining separate resources for visualization and analysis of PDB data. wwPDB partners include the RCSB Protein Data Bank (US; supported by NSF, NIH, and DOE), the Protein Data Bank in Europe or PDBe (UK; EMBL-EBI, Wellcome Trust, MRC, BBSRC, and the EU), Protein Data Bank Japan or PDBj (Japan; JST-NBDC) and BioMagResBank or BMRB (US; NIH). Formation of the wwPDB was an important step towards the long-term stability of access to the PDB Archive.

**2. What are the pros and cons of that business model?**

In early 2014, wwPDB partners began staged deployment of a common deposition, annotation, and validation system (OneDep) that enables data curation load balancing among all regional Data Centers. Each Data Center now takes responsibility for data depositions coming from its assigned catchment area. Thus, the wwPDB partners use regionally contributed funds to process depositions with the common system originating from within their respective geographies. This model is subject to the continued funding of current members, but can also be expanded to new locations.

**3. Why is this appropriate to the repository's mission and the services offered?**

Given the central role of the PDB as a high-value, primary archival database central to basic and applied biological and biomedical research and education in the life sciences, the ideal funding model would be coordinated support by national and international funding agencies, both public and private.

**4. How do you rate the level of acceptance from the 'designated community' / stakeholders?**

Deposition to the PDB archive is required by virtually all journals that publish experimentally determined structure models. Many funding agencies also require PDB depositions for structural biology research.
The usage of PDB data is very high: 500 million downloads of coordinates per year; more than 1,000,000 unique visitors to the websites per year. The PDB is among the top two cited biological databases.

**5. What options, if any, are being considered for additional income streams, costs limitation?**

The popularity and use structure of the PDB Archive begs the obvious question: Can the wwPDB be funded by mechanisms that balances costs across our user communities? Ember and Hanisch (2013) reviewed several models in terms of needs for long-term stability, support for Open Access to data, equity for depositors, and equity for institutions.
*Membership* models do not provide public access and biases toward institutions with more resources.
*Submission Fees* would transfer the burden of supporting the archive to the structural biology community, which may not be able to afford these costs. Only research subject to the policies of funding agencies and journals would be archived, assuming that the funders and the journals have the resources to ensure compliance. Research not subject to funder regulations, such as from industry, would be irretrievably lost.
*User Fees* would be unlikely to cover the costs of data curation and archiving, and limit data usage to organizations with more resources. This mechanism also runs counter to the the wwPDB mission and core values.
Another model not reviewed by Ember and Hanisch would be *Charging the Journals a Validation Fee for PDB Archive Deposits*. While not specifically precluded by the Charter of the Worldwide Protein Data Bank (wwPDB), this mechanism would also go against the spirit of the wwPDB agreement. Implementing such a mechanism would also require negotiating individual contracts between the wwPDB and each of the scientific journals and then arriving at a revenue sharing arrangement amongst the wwPDB Regional Data Centers.