Economics of Biodata Archiving Viewed Through the Lens of the RCSB Protein Data Bank

Professor Stephen K Burley[1]
[1]RCSB Protein Data Bank
stephen.burley@rcsb.org

The Protein Data Bank (PDB) was established in 1971 as the first open-access digital data resource in biology with just seven X-ray crystallographic structures of proteins. Today, the single global PDB archive houses more than 200,000 experimentally-determined three- dimensional (3D) structures of biological macromolecules that are made freely available to many millions of users worldwide with no limitations on usage. 3D biostructure information facilitates basic and applied research and education across the sciences, impacting fundamental biology, biomedicine, biotechnology, and energy sciences. The Worldwide Protein Data Bank partnership (wwPDB, wwpdb.org) currently includes five Full Members (RCSB PDB, PDBe, PDBj, BMRB, and EMDB) and one Associate Member (PDBc), which together manage the PDB, EMDB, and BMRB Core Archives. wwPDB Members are committed to ensuring that structural biology data are Findable, Accessible, Interoperable, and Reusable (FAIR).

This talk will explore the economics of biodata archiving viewed through the Lens of the RCSB Protein Data Bank. Topics to be discussed will include impact of the PDB archive on (i) local, regional, and global economies; (ii) small-molecule and biologics drug discovery and development; (iii) technological innovation more broadly; and (iv) federal funding agencies (NSF, NIH, and DOE). The failure of the free market to adequately prepare for the COVID-19 pandemic will also be discussed.

RCSB PDB is funded by the National Science Foundation (DBI-1832184), the US Department of Energy (DE-SC0019749), and the National Cancer Institute, National Institute of Allergy and Infectious Diseases, and National Institute of General Medical Sciences of the National Institutes of Health under grant R01GM133198.