

**MS01-1-2 Trends in macromolecular structure data across 50 years of the Protein Data Bank**  
**#MS01-1-2**

D. Harrus<sup>1</sup>, wwPDB Consortium<sup>2</sup>

<sup>1</sup>EMBL-EBI - Cambridge (United Kingdom), <sup>2</sup>wwPDB (United Kingdom)

**Abstract**

The worldwide PDB (wwPDB) is the international consortium responsible for managing the Protein Data Bank (PDB) - the single global repository for three-dimensional structures of biological macromolecules and their complexes. Over the past decade, the size and complexity of macromolecular complexes deposited to the PDB has increased significantly. The PDB archive now contains more than 190,000 experimentally determined structures of biological macromolecules, all of which are publicly accessible without restriction. These structures provide essential information worldwide, to a diverse user community. There are more than 2 billion downloads of data files from the PDB archive each year, with more than 1 million unique IP addresses accessing the archive within the same period.

In 2014, in an effort to fulfil the evolving archive requirements of the scientific community over the coming decades, the wwPDB partners launched OneDep, a global unified system for deposition, biocuration, and validation of macromolecular structures. It replaced legacy pipelines across PDB, EMDB, and BMRB deposition sites and is able to interface with other archival resources. Since then, the goal of the developments of the OneDep system has been to ensure data quality and completeness across all three archives, while supporting growth in the number and complexity of depositions.

In 2021, we celebrated 50 years of the PDB archive, making it one of the longest-running open access scientific databases. Throughout these 50 years, there has been significant evolution of the data in the PDB archive. This evolution has been driven by a number of factors including development in structural determination techniques, adaptation of biocuration practices, and increase in data capture via updated file formats.

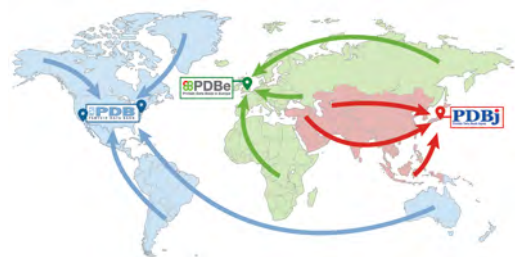
In this poster, we present some of the key trends in data across the PDB archive, highlighting how structural biology data has changed over time and how wwPDB biocuration practices have adapted to handle these changes.

**References**

**The Protein Data Bank Archive.** *Methods in Molecular Biology*. 01 Jan 2021, 2305:3-21 DOI: 10.1007/978-1-0716-1406-8\_1 PMID: 33950382

**Modernized uniform representation of carbohydrate molecules in the Protein Data Bank.** *Glycobiology*. 01 Sep 2021, 31(9):1204-1218 DOI: 10.1093/glycob/cwab039 PMID: 33978738

**Enhanced validation of small-molecule ligands and carbohydrates in the Protein Data Bank.** *Structure*. 02 Mar 2021, 29(4):393-400.e1 DOI: 10.1016/j.str.2021.02.004 PMID: 33657417



WORLDWIDE  
**wwPDB**  
PROTEIN DATA BANK