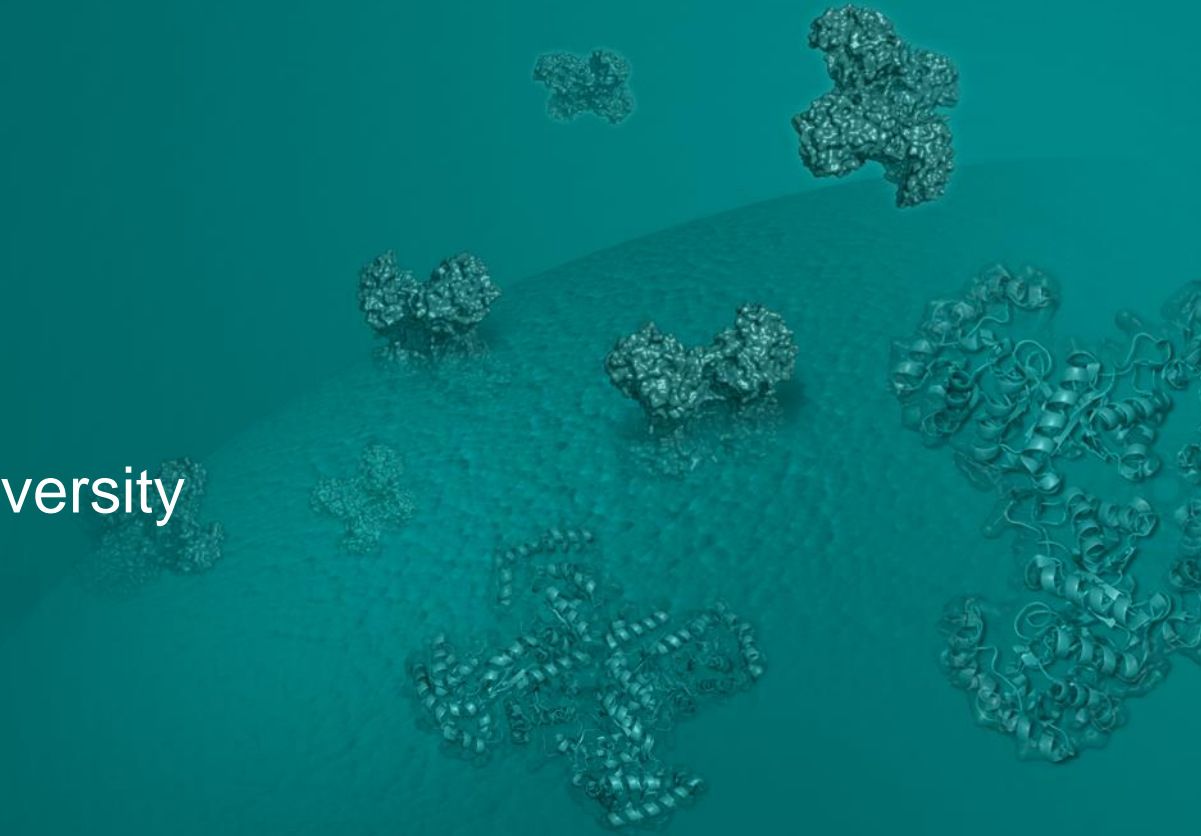


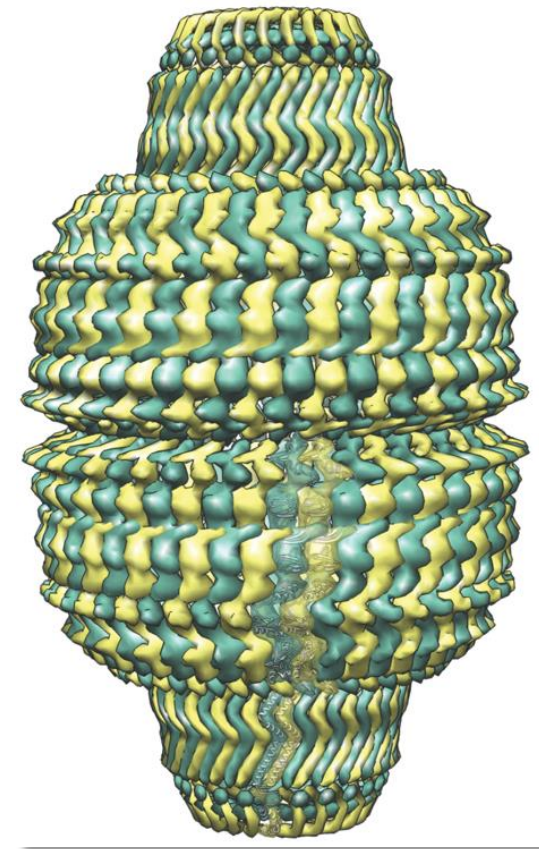
Protein Data Bank

Radka Svobodová
CEITEC Masaryk University



Why archive research data?

- **Accessibility**
 - One-stop shop
 - Uniform data representation/annotation
- **Persistence**
 - Typical websites have a half-life of 2 years
 - Professional management
- **Context**
 - Comparisons against all other entries
 - Validation
 - Integration with other resources
- **Facilitate further analysis**
 - Database-wide studies and data-mining

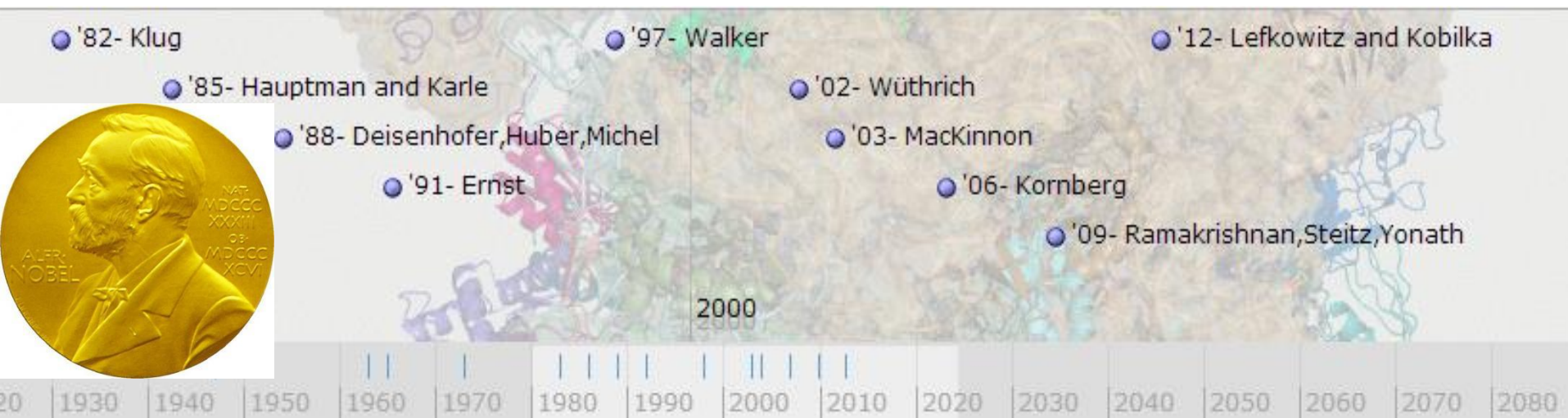


Macromolecular structures (1950-1980)

- Myoglobin (1958) – Kendrew *et.al.* - The most striking features of the molecule were its irregularity and its total lack of symmetry
- Lysozyme (1965) – Phillips *et.al.*
- Ribonuclease (1967) – Kartha, Bello & Harker
- Papain (1968) – Drenth *et.al.*
- Haemoglobin (1968) – Perutz *et.al.* – *Stereochemistry of cooperative effects in Haemoglobin* – M.F. Perutz (1971)
- Insulin (1971) – Blundell *et.al.*
- **PDB was established in 1971 with 7 structures**

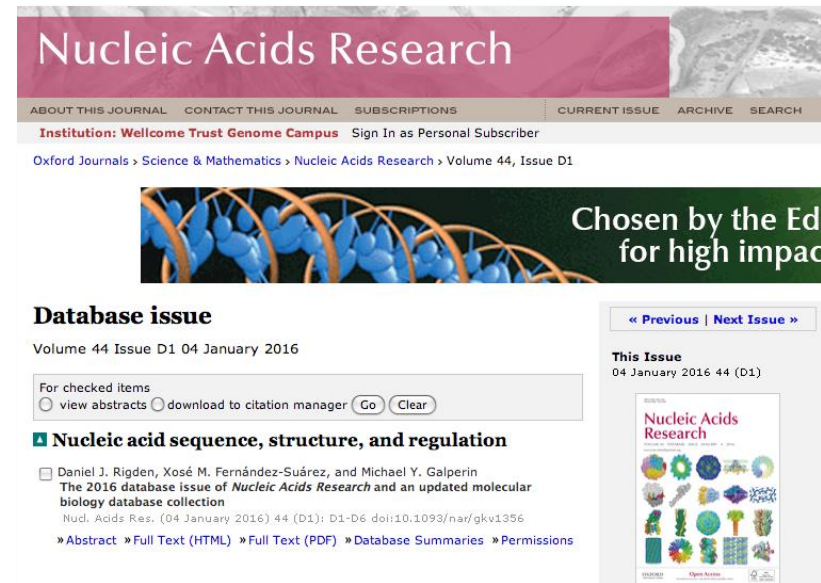
Immeasurable scientific value

- 1946 Sumner “Enzymes can be crystallised”
- 1962 Crick, Watson, Wilkins DNA
- 1962 Perutz & Kendrew Haemoglobin & myoglobin
- 1972 Anfinsen, Moore & Stein Ribonuclease
- 1982 Klug Nucleic acid-protein complexes (TMV)
- 1988 Deisenhofer, Huber & Michel Photosynthetic RC



Many PDB-derived databases/resources!

- Since 2011, >25% of new databases described in annual NAR Database issues used PDB data (119 of 452)
- In total, >200 databases (of 1685 in Jan-2016 NAR Database collection) use PDB data, including:
 - 123 structure databases
 - 49 sequence databases
 - 22 metabolic and signalling pathways databases



Nucleic Acids Research

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURRENT ISSUE ARCHIVE SEARCH

Institution: Wellcome Trust Genome Campus Sign In as Personal Subscriber

Oxford Journals > Science & Mathematics > Nucleic Acids Research > Volume 44, Issue D1

Chosen by the Editor for high impact

Database issue

Volume 44 Issue D1 04 January 2016

For checked items
 view abstracts download to citation manager

Nucleic acid sequence, structure, and regulation

Daniel J. Rigden, José M. Fernández-Suárez, and Michael Y. Galperin
The 2016 database issue of *Nucleic Acids Research* and an updated molecular biology database collection
Nucl. Acids Res. (04 January 2016) 44 (D1): D1-D6 doi:10.1093/nar/gkv1356
» Abstract » Full Text (HTML) » Full Text (PDF) » Database Summaries » Permissions

« Previous | Next Issue »

This Issue
04 January 2016 44 (D1)

Cost of relocating the PDB archive

- Current PDB holdings exceed 123,456 experimentally determined 3D structures of biological macromolecules
- Estimated cost of replicating a PDB entry ranges between US\$50,000 to US\$250,000
- Conservative cost of replicating the PDB archive (assuming average unit cost of US\$100,000) equals

US\$12 billion

Molecular and cellular structures

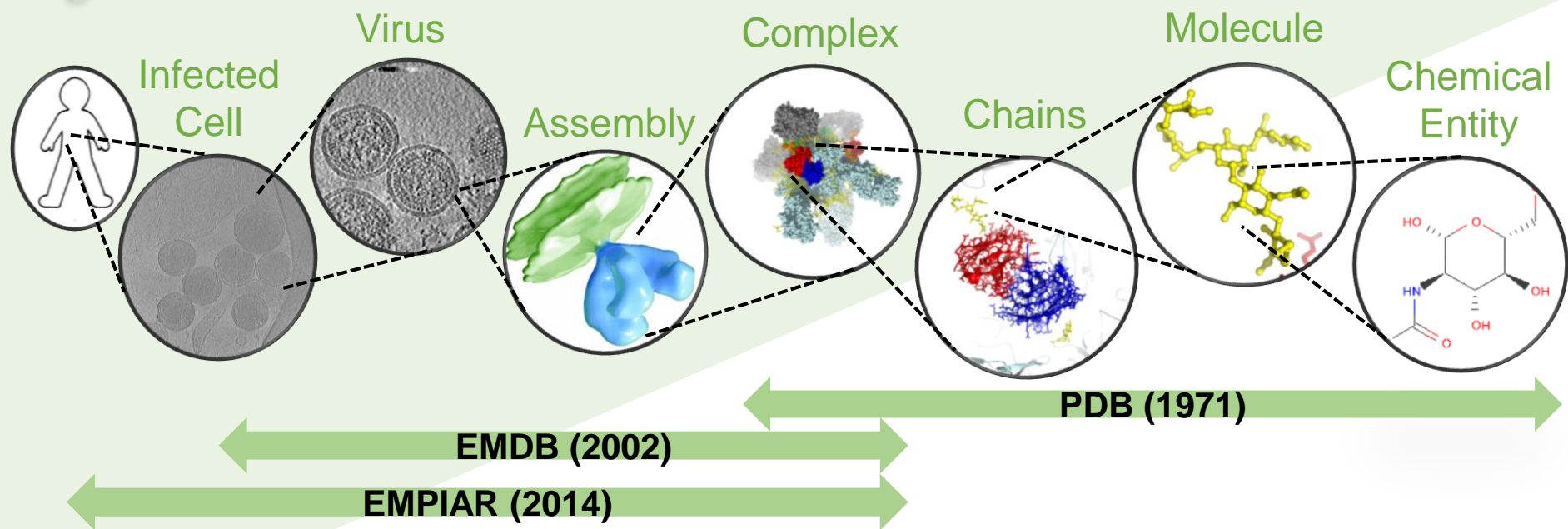
- CLEM
- 3DSEM
- SXT

- ET
- SAXS
- EM

- X-ray
- NMR
- EM

Organism

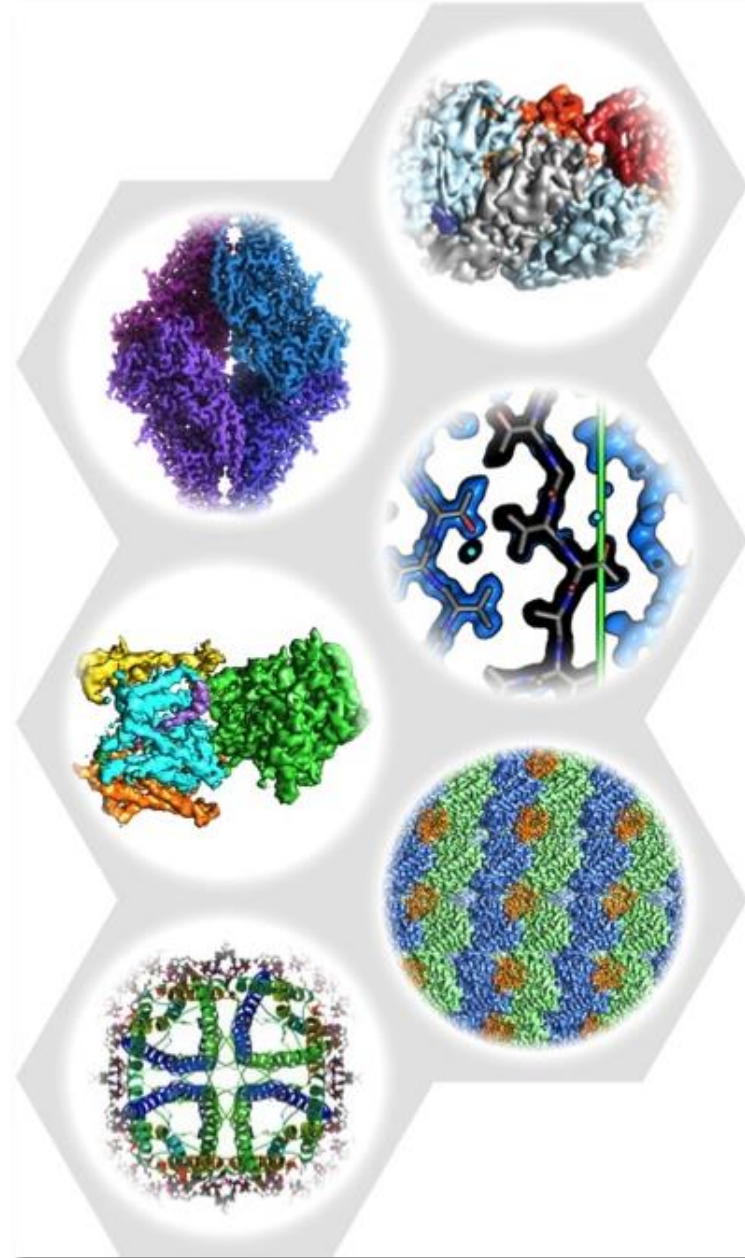
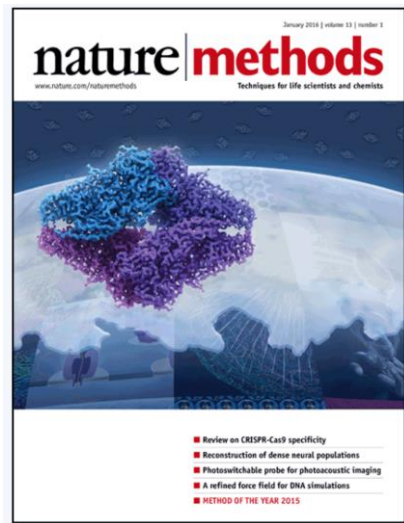
Atom



Not in the picture: Biological context, Time

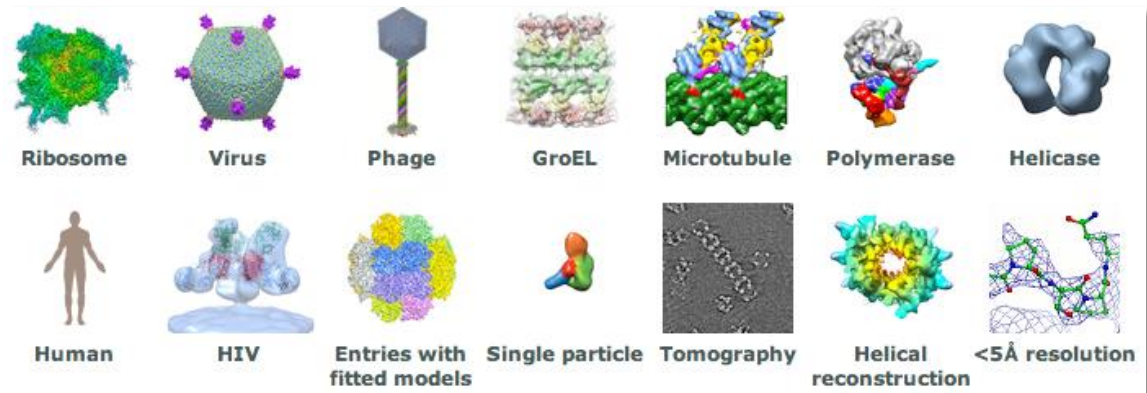
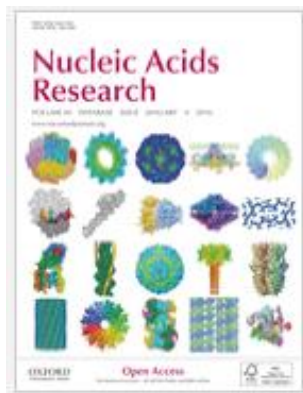
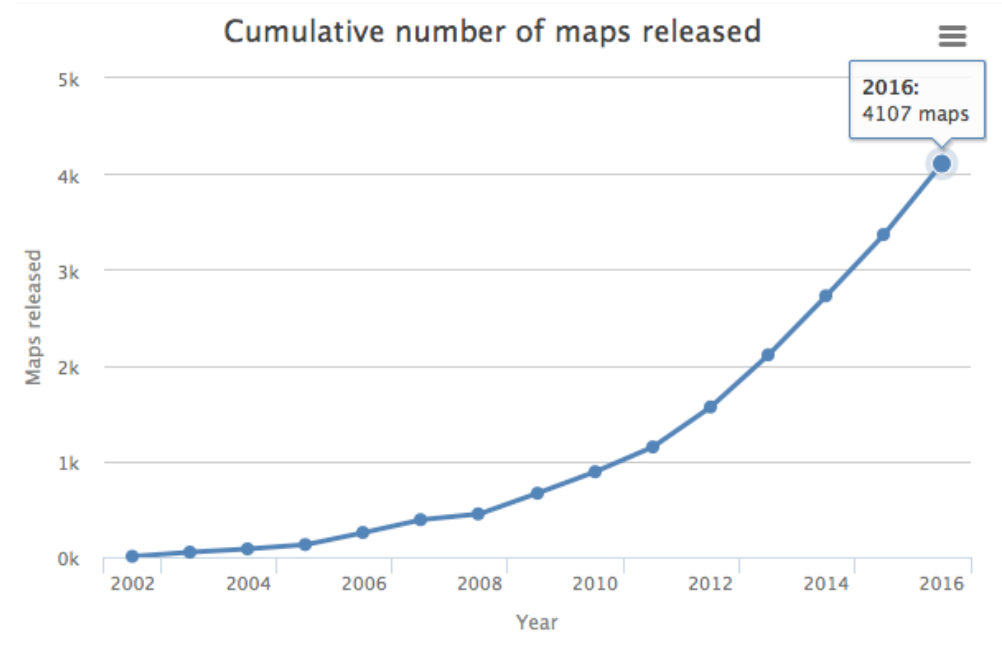


EMDB – Electron Microscopy Data Bank



EMDB - Electron Microscopy Data Bank

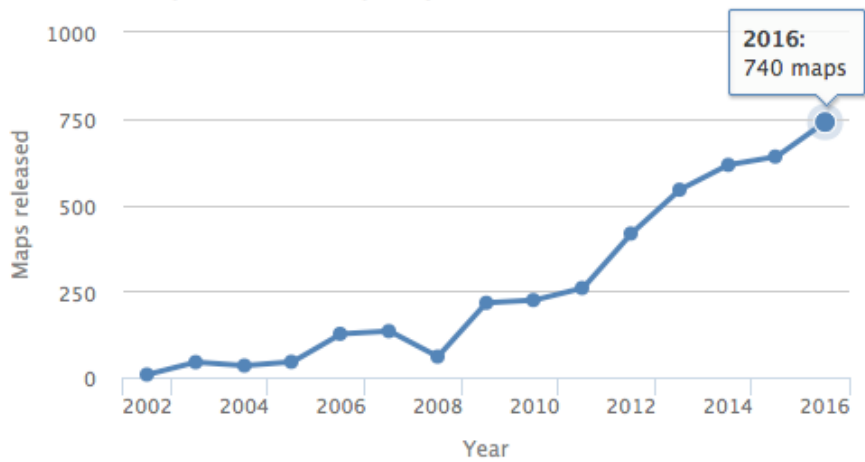
- Archives 3DEM and electron tomography volume data
- Founded at EMBL-EBI in 2002
- Operated jointly by PDBe and RCSB since 2007 (PDBj annotates since 2013)
- >4100 entries (Oct-16)



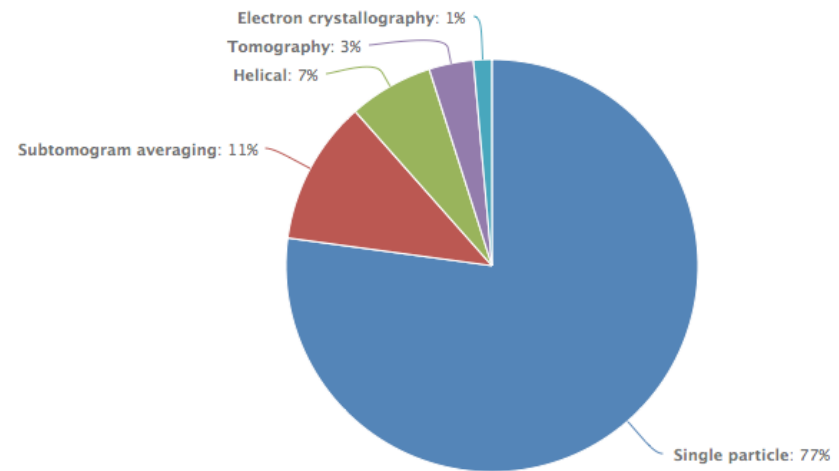
EMDB holdings (Sep-2016)



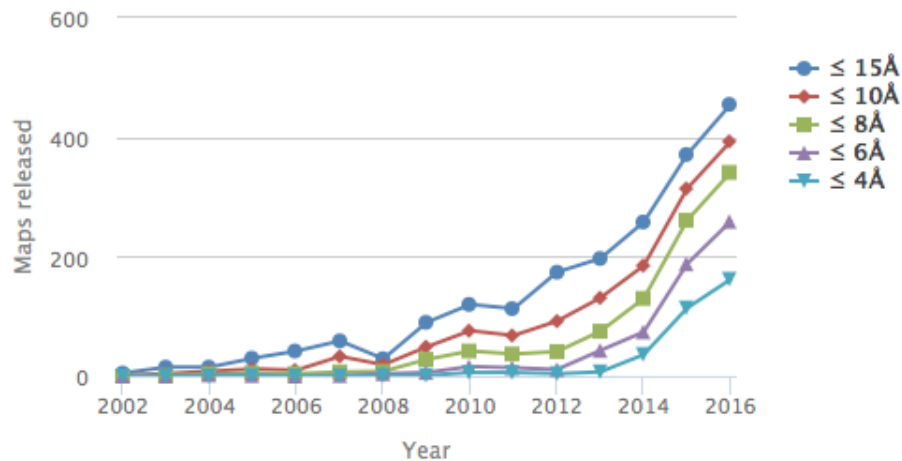
Maps released per year (4107 in total)



Distribution of released maps (4107 in total) as a function of technique used



Maps achieving given resolution levels



Resolution distribution for released maps

