PDBe-KB

Protein Data Bank in Europe - Knowledge Base

Collaboration Guidelines

Last updated 26th Oct 2022

This document describes the PDBe-KB consortium guidelines and terms of collaboration.

1. Introduction

The Protein Data Bank in Europe - Knowledge Base (PDBe-KB) is a community-driven resource managed by the PDBe team, collating functional annotations and predictions for structure data in the PDB archive. PDBe-KB is a collaborative effort between PDBe and a diverse group of bioinformatics resources and research teams. This document outlines the framework for contributing to the PDBe-KB collaborative resource, the services PDBe-KB agrees to provide and the conditions PDBe-KB consortium members accept.

2. Terms of Collaboration

2.1 PDBe-KB

1. PDBe-KB will maintain the infrastructure for data deposition and retrieval.
2. PDBe-KB will support data exchange format schema(s).

3. The schema will evolve in consultation with collaborating partners.
4. PDBe-KB will provide programmatic access to expose contributed annotations.
5. PDBe-KB will link back to the collaborating partners' resources, clearly attributing credit for their contributions.
6. PDBe-KB will maintain an open-access library of reusable data visualisation components.

2.2 Consortium members

1. The data contributed to PDBe-KB by consortium members will be made available under CC-BY 4.0 license.
2. Members are responsible for the quality of the data they contribute.
3. Members describe their data, software or database by filling in a Google Form.
4. Protocols for data generation must be published in peer-reviewed publications.
5. In the case of predicted/calculated annotations, the contributing partner commits to depositing data at least once a year: e.g., to provide annotations for newer PDB entries or update the existing annotations when the underlying algorithms change significantly.
    a. Manually curated annotations may be exempt from this condition on a case-by-case basis.
    b. Depositors can change/update/delete their entries at any time.

3. General Data Protection Regulation (GDPR) notice

Collaborating partners must agree to the PDBe-KB GDPR notice before registering a data deposition account.

4. Technical appendix (data collection)

Residue-level annotations and their metadata will be deposited using the PDBe-KB deposition system. The deposition system consists of the components described below.

4.1 PDBe-KB JSON specification

A JSON schema is defined to capture residue-level annotations. The schema is available at https://github.com/PDBe-KB/funpdbe-schema. Contributing partners agree to comply with the JSON specification to deposit data via the PDBe-KB deposition system.

4.2 Validation tool

A validator tool (Python3.x) is provided for the contributors to parse input JSON files, validate them against the PDBe-KB schema, and perform all the data checks the local pipeline would perform. The tool is available for download from https://github.com/PDBe-KB/funpdbe-validator.

4.3 Deposition

All PDBe-KB partners are provisioned with a private FTP area provided by EMBL-EBI, where consortium members can transfer their JSON files containing the annotations. The PDBe-KB validation and processing pipeline picks up these files and integrates the data with the core PDB data in a graph database.