



Protein Data Bank in Europe Knowledge Base

# Data Deposition Guide

*Last updated 7th July 2020*

This document describes the process of contributing annotations as a member of the PDBe-KB consortium.

A separate Consortium Guideline document with information on the agreed terms of collaborations is available at <https://www.ebi.ac.uk/pdbe/pdbe-kb/guidelines>

## 1. Data Deposition

The deposition infrastructure consists of 3 main parts:

1. Data Exchange Schema
2. Format Validation Tools
3. File Transfer

### 1.1 Data Exchange Schema

Collaborating partners contribute annotations by converting their data to JSON (<https://json.org>) formatted files based on the agreed JSON specification schema. This schema and an example JSON can be found at <https://gitlab.ebi.ac.uk/pdbe-kb/funpdbe/funpdbe-schema>

### 1.2 Format Validation Tools

The JSON files can be validated using a Python package available at <https://gitlab.ebi.ac.uk/pdbe-kb/funpdbe/funpdbe-validator>. This lightweight Python package performs every check performed during deposition:

- Checks against the schema
- Data sanity checks
- Cross-matching the residue indices in JSON to those available in the PDB

**Note for the last point:** The residue numbering in the JSON files is required to be the same as the author residue numbering in PDB (i.e. `auth_seq_id`).



## 1.3 File Transfer

Each partner resource is provided with a private FTP area by the PDBe-KB team. The partner resources transfer the validated JSON files to the assigned private FTP areas provided by EMBL-EBI. The folder structure should follow the PDB FTP convention, i.e. the JSON file for the PDB entries '1a00' and '1a0n' should both be in the sub-folder 'a0' as shown on the example below:



Remote site:	/upload/a0		
	4g		
	5g		
	6g		
	7g		
	8g		
	9g		
	a0		
	a1		
	a2		
	a3		
	a4		

  

Filename	Filesize	Filetype	Last modified
..			
1a00.cath-funsites.json	499,684	JSON File	14/02/2020 14:...
1a01.cath-funsites.json	499,684	JSON File	14/02/2020 14:...
1a02.cath-funsites.json	223,967	JSON File	14/02/2020 14:...
1a07.cath-funsites.json	94,656	JSON File	14/02/2020 14:...
1a08.cath-funsites.json	94,656	JSON File	14/02/2020 14:...
1a09.cath-funsites.json	94,656	JSON File	14/02/2020 14:...
1a0l.cath-funsites.json	369,860	JSON File	14/02/2020 14:...
1a0n.cath-funsites.json	18,245	JSON File	14/02/2020 14:...

The transferred files are then processed, validated and integrated with the core PDB data in the PDBe graph database (<https://pdbe.org/graph-schema>).

**Please note:** The PDBe will request this FTP account once the data format is finalised and reviewed.

## 2. Downstream Processes

The integrated annotations are exposed via the PDBe graph API endpoint (<https://pdbe.org/graph-api>). These API endpoints power the PDBe-KB aggregated views, such as the Aggregated Views for Proteins (<https://pdbe-kb.org/proteins>).

Importantly, the annotations are always displayed together with links that use the URLs provided by the partner resource so that the users can directly reach the original data resource.



**Please note:** Not every annotation is displayed on all the PDBe and PDBe-KB pages, i.e. some annotations may only be displayed on specific pages.

