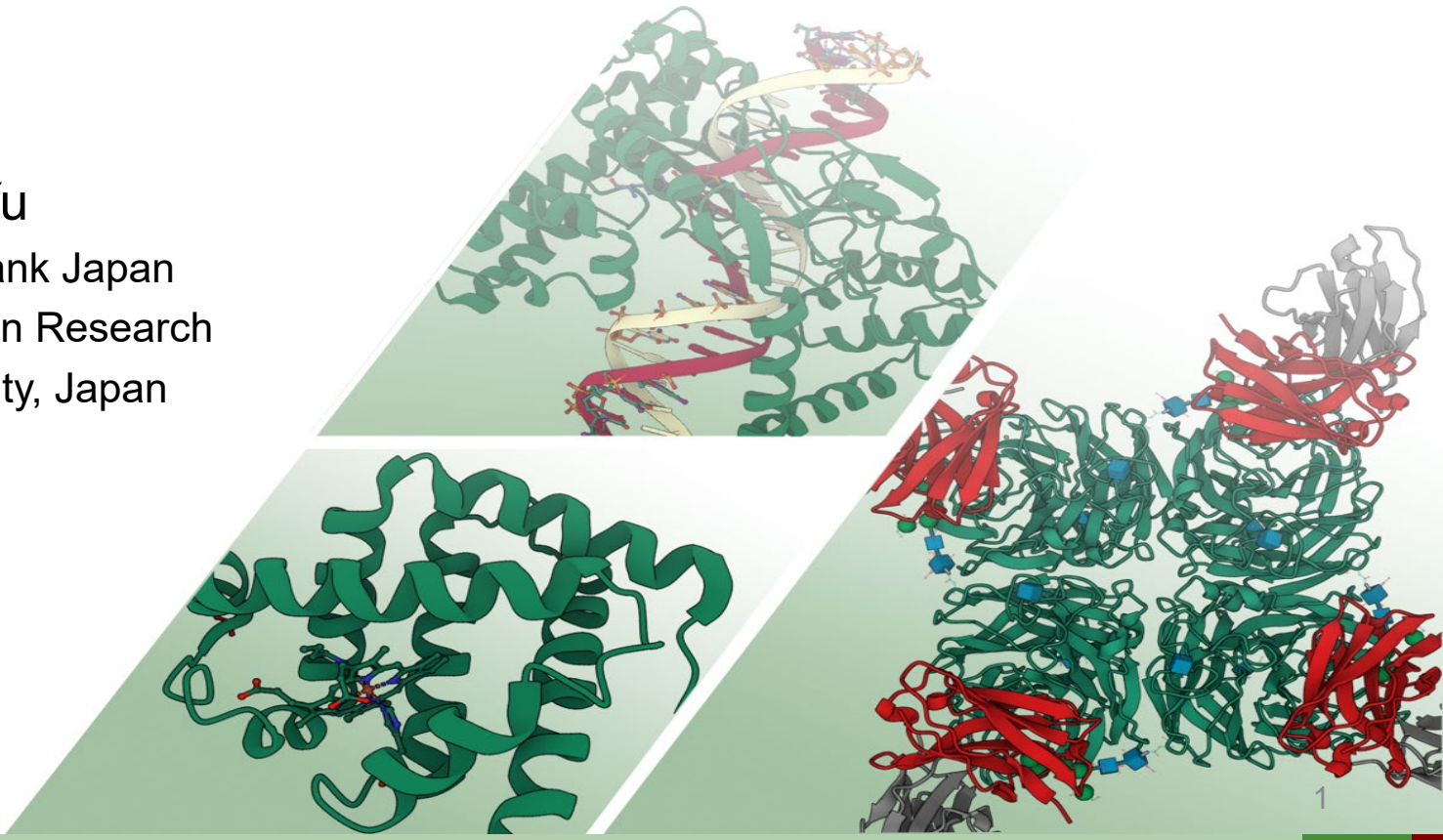
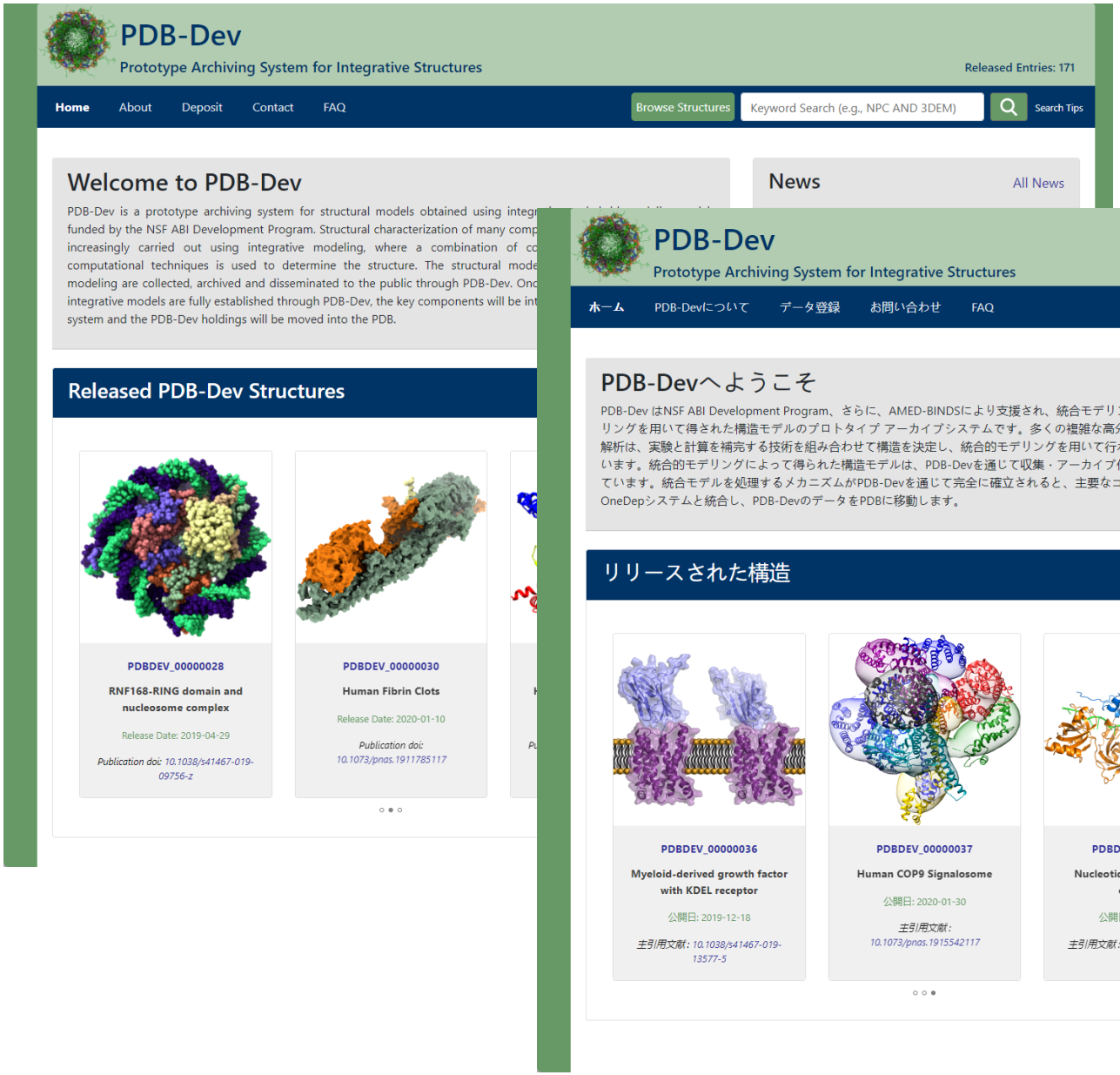
Introduction to Data Deposition for PDB-Dev

Jian Yu
Protein Data Bank Japan
Institute for Protein Research
Osaka University, Japan



PDB-Dev: https://pdb-dev.wwpdb.org/

PDBj Protein Data Bank Japan



日本語化

Deposition and Data harvesting system

PDBj
Protein Data Bank Japan

PDB-Dev
Prototype Archiving System for Integrative Structures
Released Entries: 171

Home | About | Deposit | Contact | FAQ | Browse Structures | Keyword Search (e.g., NPC AND 3DEM) | Search Tips

Welcome to PDB-Dev

PDB-Dev is a prototype archiving system for structural models obtained using integrative or hybrid modeling and is funded by the NSF ABI Development Program. Structural characterization of many complex macromolecular assemblies is increasingly carried out using integrative modeling, where a combination of complementary experimental and computational techniques is used to determine the structure. The structural models obtained through integrative modeling are collected, archived and disseminated to the public through PDB. integrative models are fully established through PDB-Dev, the key components system and the PDB-Dev holdings will be moved into the PDB.

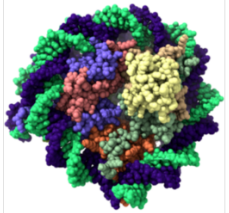
News

All News

New Deposition API

Starting January 22nd 2024, PDB-Dev accepts

Released PDB-Dev Structures

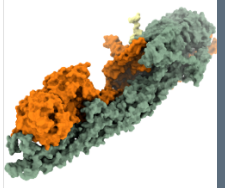


PDBDEV_00000028

RNF168-RING domain and nucleosome complex

Release Date: 2019-04-29

Publication doi: 10.1038/s41467-019-09756-z



PDBDEV_00000030

Human Fibrin Clots

Release Date: 2020-01-10

Publication doi: 10.1073/pnas.1911785117

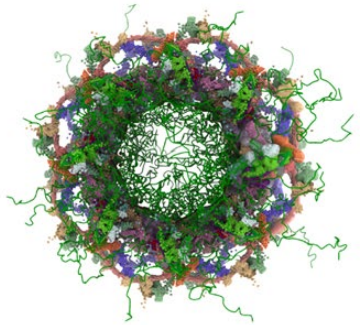
PDB-Dev | Entry | Data Categories | Templates | Documentation | Go to RID | Sign Up | Log In

PDB-Dev Data Harvesting System

Welcome to the **PDB-Dev Data Harvesting System**.

The PDB-Dev Data Harvesting System provides a web interface for depositors to assemble all the information required for archiving integrative structures in PDB-Dev and to create a compliant mmCIF file. This includes the submission of integrative structures, associated spatial restraints and starting models used, modeling protocols and metadata information (citations, authors, software, data in external repositories, reference sequence information etc.).

The Data Harvesting System provides an automated mechanism for collecting diverse and heterogeneous types of integrative structures and restraint data. The system has been developed using the DERIVA scientific asset management platform and is supported by the PDBx/mmCIF dictionary, the IHM-dictionary and the python-ihm software library.



Use the links below to start exploring the system.

New Account: Register using Globus

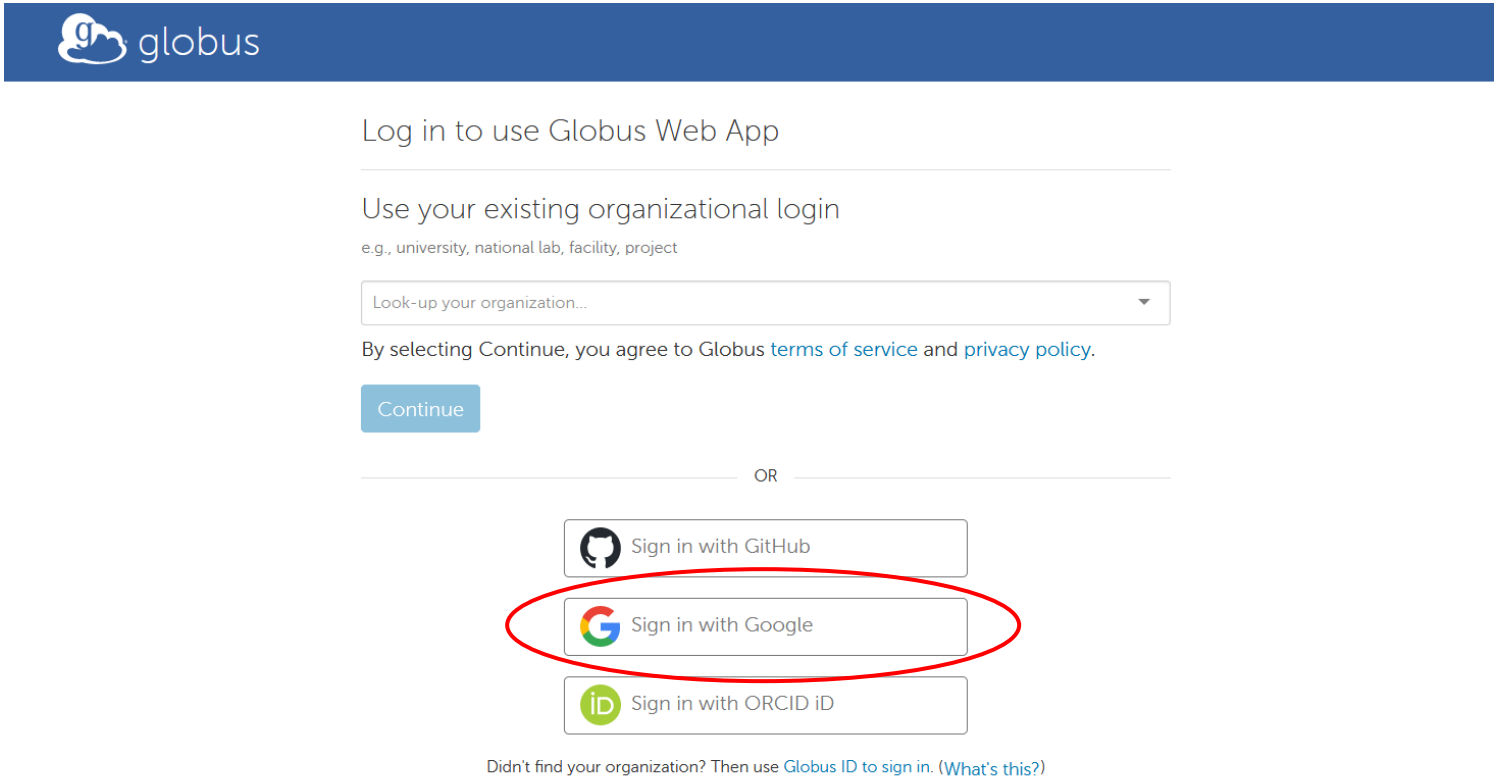
Submit Data: Create a New Entry

Documentation: User Guide

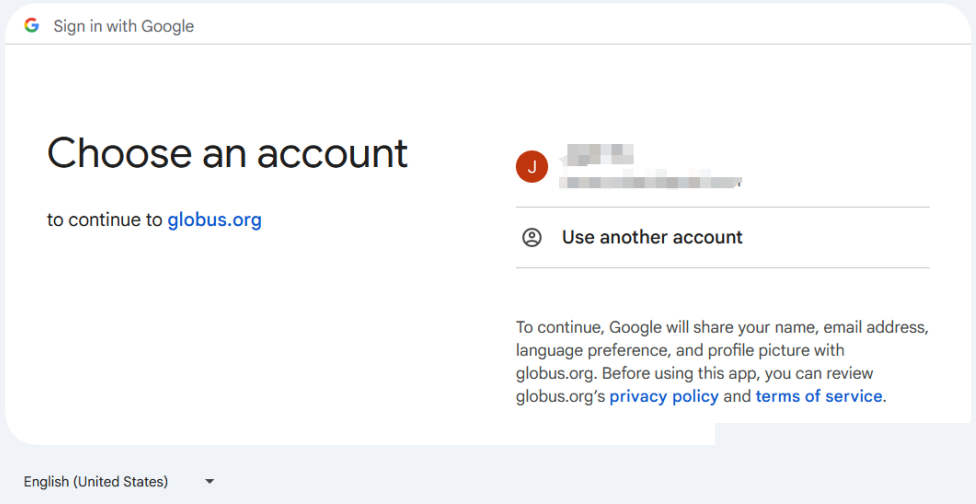
Please report problems to pdb-dev@mail.wwpdb.org.

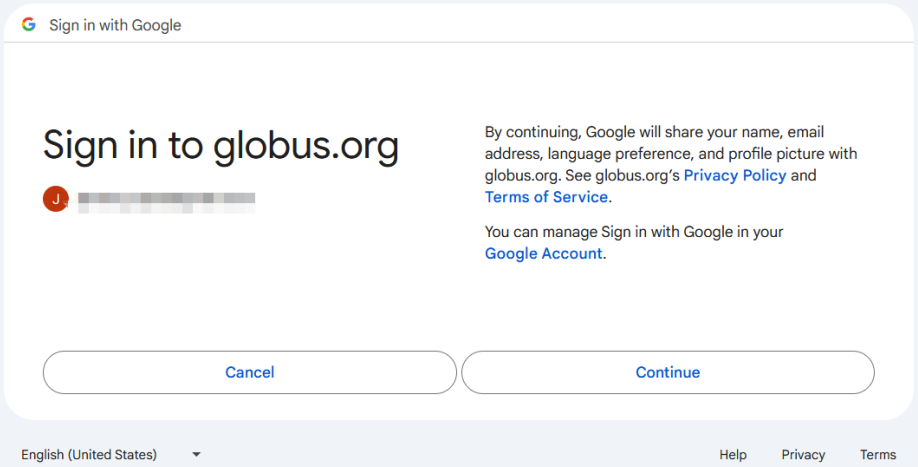
RUTGERS THE STATE UNIVERSITY OF NEW JERSEY | UCSF University of California San Francisco | USC Viterbi School of Engineering | deriva

➢ **Choose how to login to Globus (using organizational account, GitHub, google account or ORCID ID).**

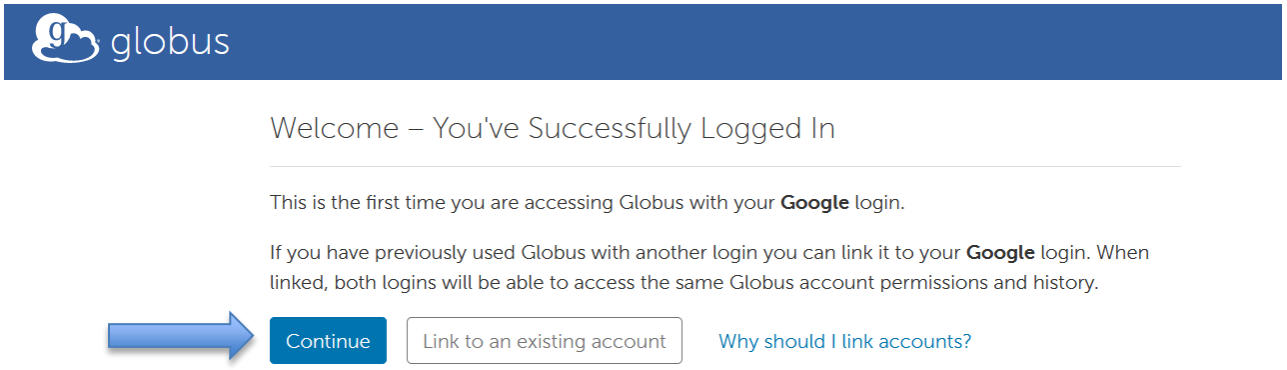


Sign in with Google account

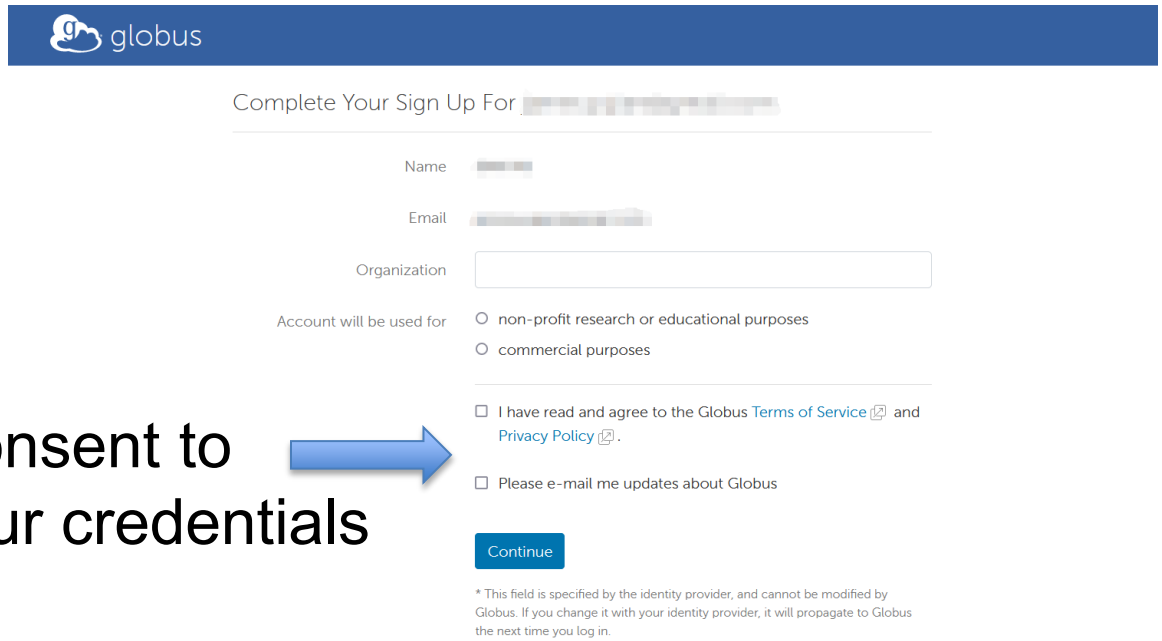




Successfully log in

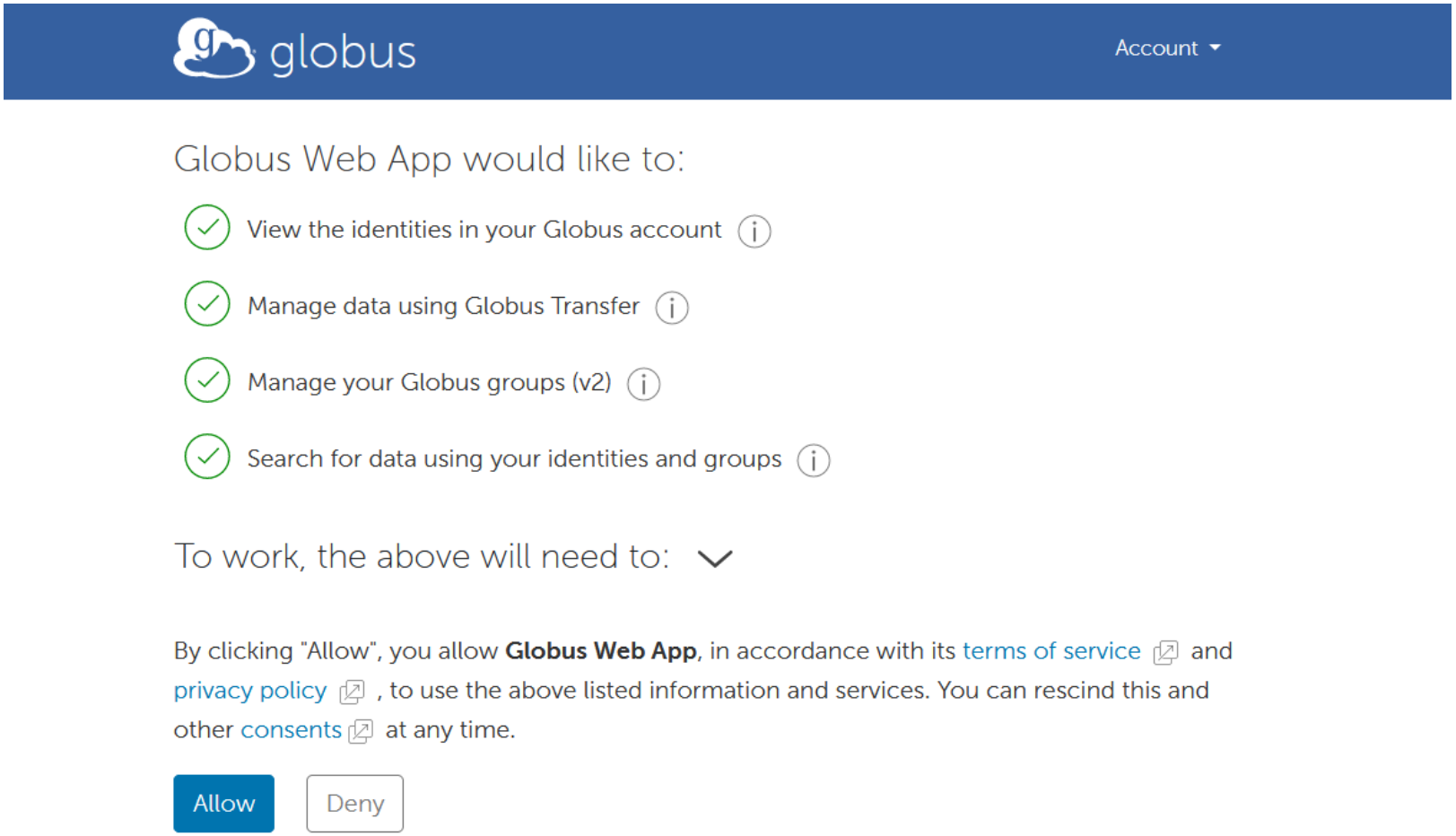


Complete a short signup form

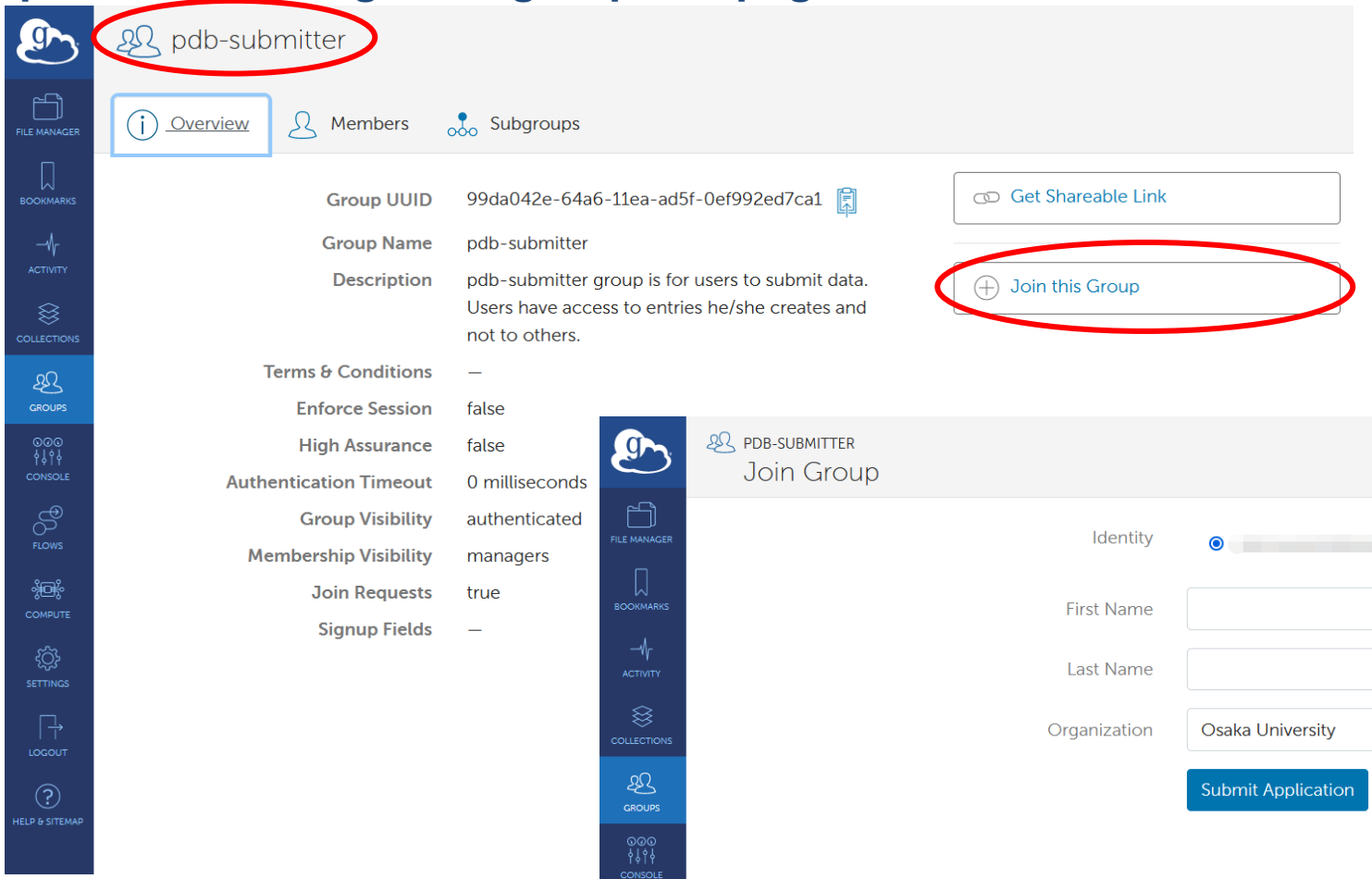


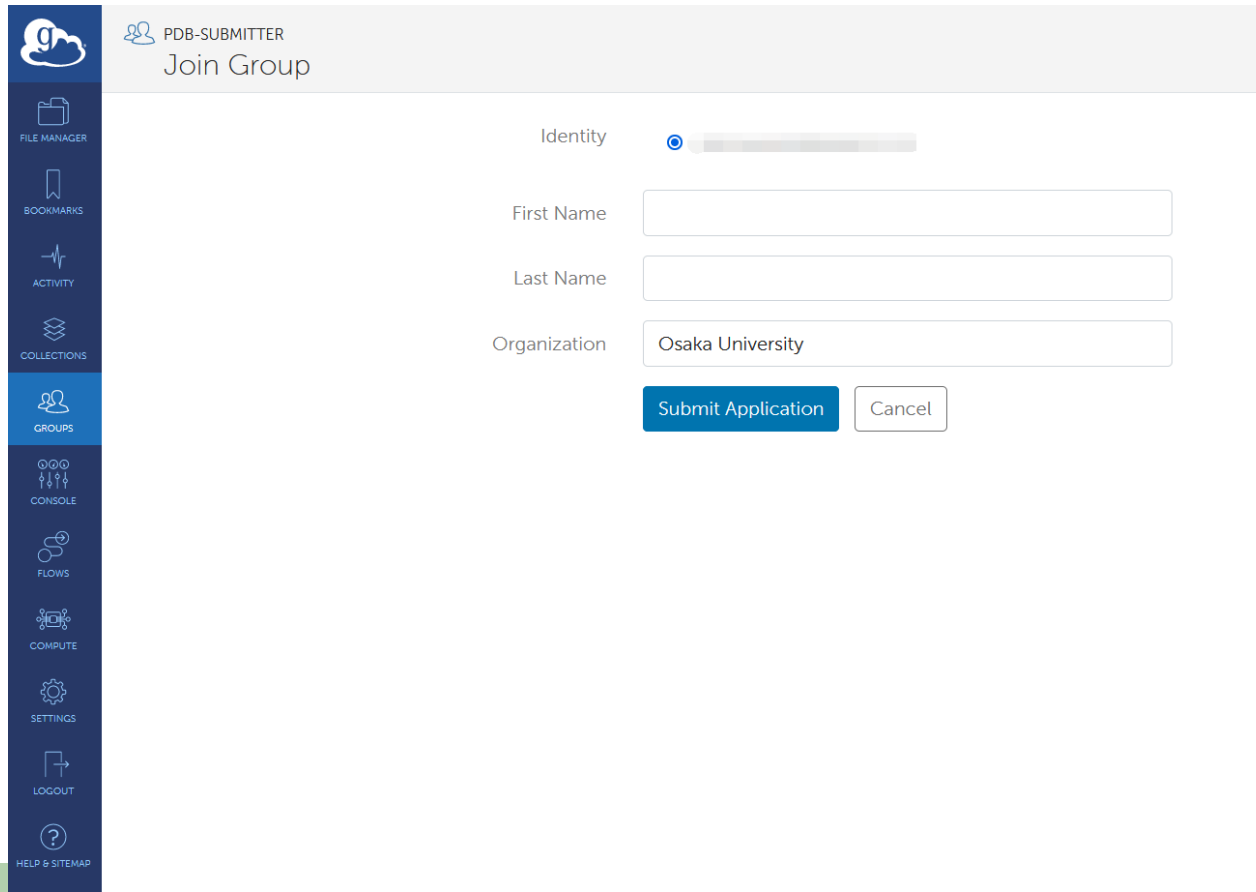
Provide consent to access your credentials

Provide consent to access your credentials



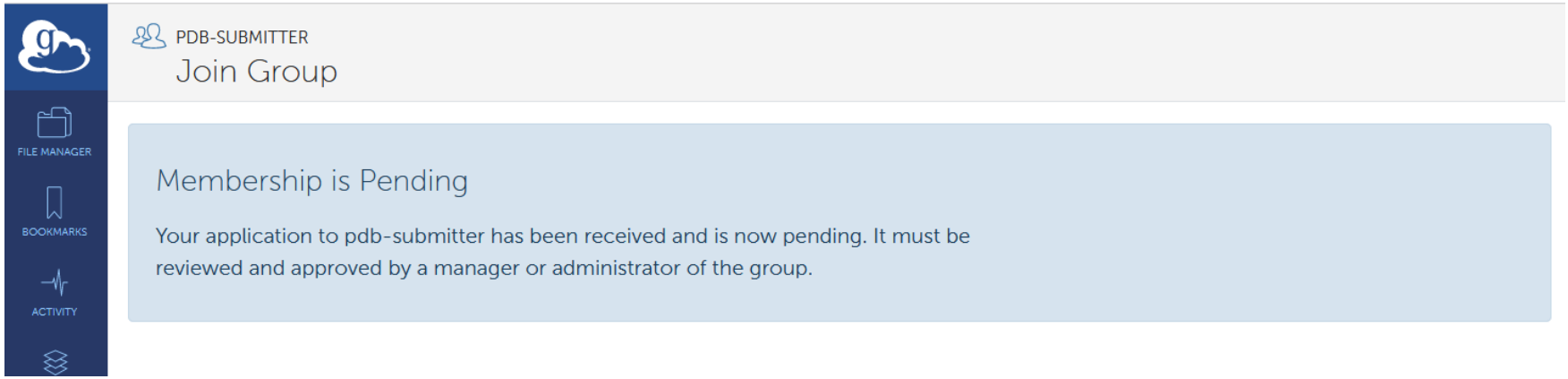
“pdb-submitter” globus group webpage





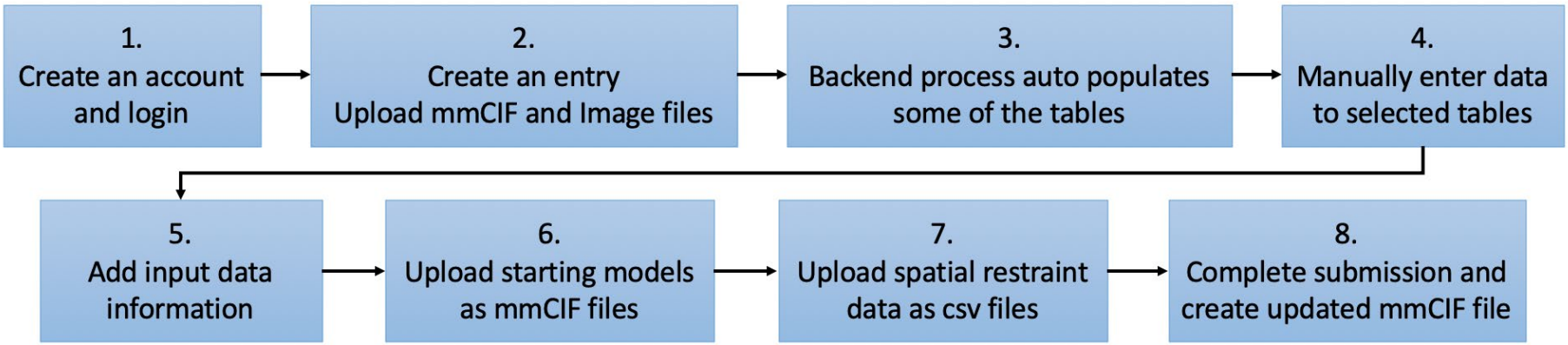
Wait for the approvement

Once the administrator has approved your join request, go to *https://data.pdb-dev.org* and login to set up the globus account.



Workflow outline for data submission

Steps involved in creating a complete mmCIF file for PDB-Dev



Step 1: Create new entry and upload mmCIF and image files

PDB-Dev | Entry | Data Categories | Templates | Documentation | Go to RID

PDB-Dev Data Harvesting System

Welcome to the **PDB-Dev Data Harvesting System**.

The PDB-Dev Data Harvesting System provides a web interface for depositors to assemble all the information required for archiving integrative structures in PDB-Dev and to create a compliant mmCIF file. This includes the submission of integrative structures, associated spatial restraints and starting models used, modeling protocols and metadata information (citations, authors, software, data in external repositories, reference sequence information etc.).

The Data Harvesting System provides an automated mechanism for collecting diverse and heterogeneous types of integrative structures and restraint data. The system has been developed using the DERIVA scientific asset management platform and is supported by the PDBx/mmCIF dictionary, the IHM-dictionary and the python-ihm software library.

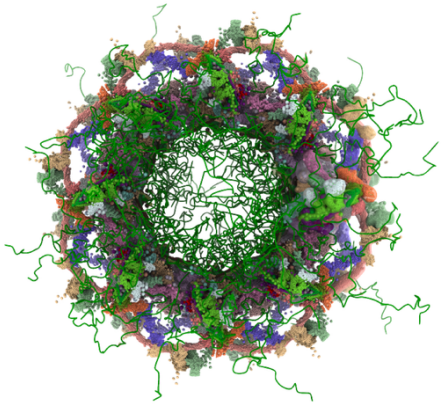
Use the links below to start exploring the system.

New Account: Register using Globus

Submit Data: Create a New Entry

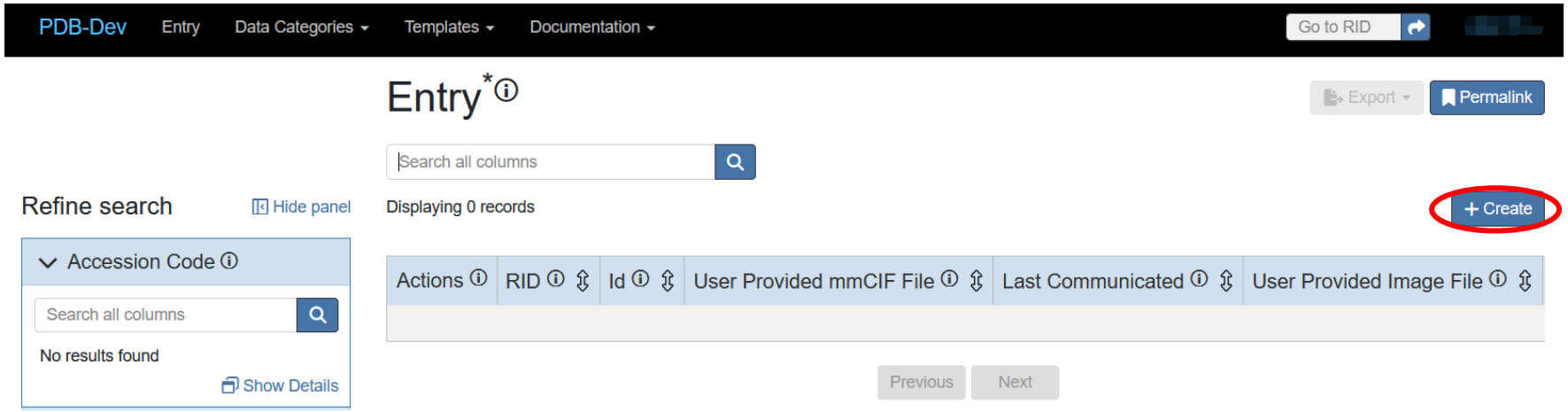
Documentation: User Guide

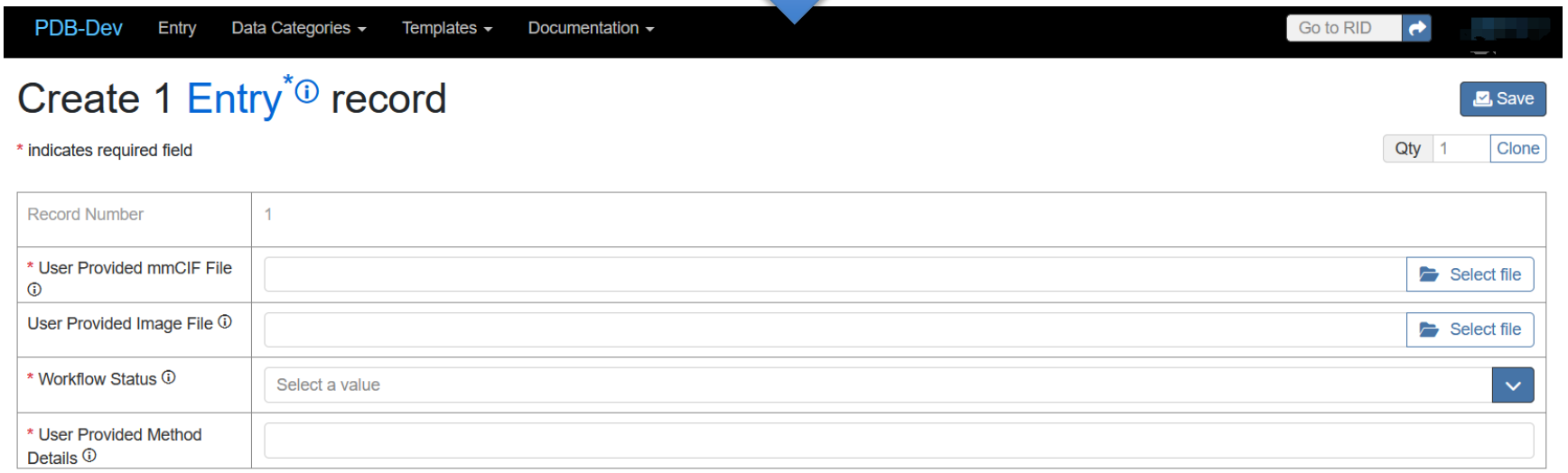
Please report problems to pdb-dev@mail.wwpdb.org.



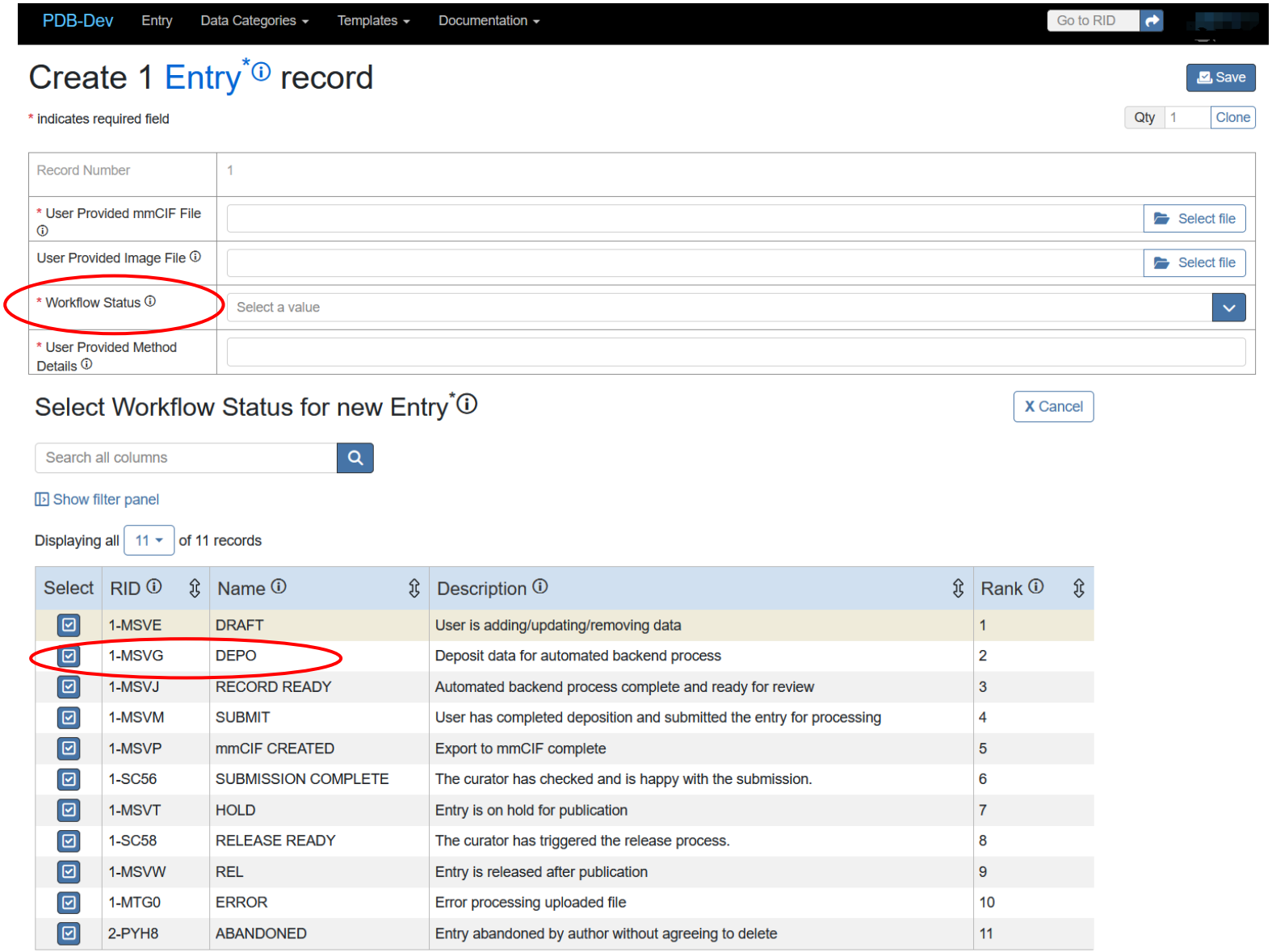
RUTGERS THE STATE UNIVERSITY OF NEW JERSEY | UCSF University of California San Francisco | USC Viterbi School of Engineering Information Sciences Institute | deriva

Create entry and input the data at entry page





Input the data at entry page



Select	RID	Name	Description	Rank
☑	1-MSVE	DRAFT	User is adding/updating/removing data	1
☑	1-MSVG	DEPO	Deposit data for automated backend process	2
☑	1-MSVJ	RECORD READY	Automated backend process complete and ready for review	3
☑	1-MSVM	SUBMIT	User has completed deposition and submitted the entry for processing	4
☑	1-MSVP	mmCIF CREATED	Export to mmCIF complete	5
☑	1-SC56	SUBMISSION COMPLETE	The curator has checked and is happy with the submission.	6
☑	1-MSVT	HOLD	Entry is on hold for publication	7
☑	1-SC58	RELEASE READY	The curator has triggered the release process.	8
☑	1-MSVW	REL	Entry is released after publication	9
☑	1-MTG0	ERROR	Error processing uploaded file	10
☑	2-PYH8	ABANDONED	Entry abandoned by author without agreeing to delete	11

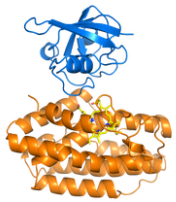
“Workflow Status” for the entry updated to “Record Ready”

Sections | Hide panel

- Summary
- System Generated Files (0)
- Description of the Structure* (1)
- Authors* (0)
- Citations* (0)
- Authors in Citations* (0)
- Software* (0)
- Chemical Components* (22)
- Molecular Entities* (2)
- Common Names of Entities (0)
- Systematic Names of Entities (0)
- Source of Genetically Manipulated Entities (0)
- Reference sequence information* (0)
- Alignment information with the reference sequence* (0)

Entry*: D_3-7B8W

+ Create | Copy | Edit | Delete

RID	3-7B8W
Id	D_3-7B8W
RCB	
RMB	
User Provided mmCIF File	cluster1_1.cif
mmCIF File Size (Bytes)	354 kB
User Provided Image File	
3D Visualization	Not Available
Image File Size (Bytes)	1.72 MB
Workflow Status	RECORD READY
Process Status	Success
Record Status Detail	
Deposit Date	2024-03-12
User Provided Method Details	X-ray, NMR, HADDOCK

Step 2: Add/modify data manually

Sections | Hide panel

- Summary
- System Generated Files (0)
- Description of the Structure* (1)
- Authors* (0)
- Citations* (0)
- Authors in Citations* (0)
- Software* (0)
- Chemical Components* (22)
- Molecular Entities* (2)
- Common Names of Entities (0)
- Systematic Names of Entities (0)
- Source of Genetically Manipulated Entities (0)
- Reference sequence information* (0)
- Alignment information with the reference sequence* (0)
- Point differences in the alignment with the reference sequence (0)
- Polymeric Entities* (2)
- Non-polymeric Entities

Entry*: D_3-7B8W

+ Create | Copy | Edit | Delete

Authors*

+ Add records | Bulk Edit | Explore

Authors of the structure; mmCIF category: audit_author; Example of author name format: Jones, T.J.

Displaying 0 records

Actions	RID	Ordinal	Name
No Results Found			

Previous | Next

Citations*

+ Add records | Bulk Edit | Explore

Citation information for the primary publication as well as other relevant software and methodology; mmCIF category: citation

Displaying 0 records

Actions	RID	Id	Title	Year	Abbreviated Journal Name (CASSI)	Journal Volume	Journal Issue

Previous | Next

Authors in Citations*

+ Add records | Bulk Edit | Explore

Authors associated with publications in the citation list; mmCIF category: citation_author; Example of author name format: Jones, T.J.

Displaying 0 records

Actions	RID	Citation Id	Ordinal	Name
No Results Found				

Previous | Next

Software*

+ Add records | Bulk Edit | Explore

Input the starting model

PDBj Protein Data Bank Japan

Starting Structural Models*

+ Add records | Bulk Edit | Explore

Information regarding starting structural models used in the integrative modeling study; mmCIF category: ihm_starting_model_details

Displaying 0 records

Actions	RID	Starting Model Id	Entity Id	Entity Description	Asym Id	Entity Poly Segment Id

Create 1 Starting Structural Models* record

Save

* indicates required field

Qty 1 Clone

Record Number	1
* Structure Id	D_3-7B8W
* Starting Model Id	
* Entity Id	Select a value
Entity Description	
* Asym Id	Select a value
Entity Poly Segment Id	Select a value
* Starting Model Source	Select a value
* Author Provided Chain Id for the Starting Model	
Starting Model Sequence Offset	
Description	
* Dataset List Id	Select a value
mmCIF File for the Starting Model	Select file

Experimental data

PDBj Protein Data Bank Japan

Input Datasets*

+ Add records | Bulk Edit | Explore

List of all input datasets used in the integrative modeling study, including data from different types of experiments and starting models from different sources; mmCIF category: ihm_dataset_list

Displaying all 1 records

Actions	RID	Id	Data Type	Database Hosted	Details
	3-7BZC	1	NMR data	NO	

Edit Input Datasets*: 1

Save | Delete | Reset

* indicates required field

Record Number	1
* Structure Id	D_3-7B8W
* Id	1
* Data Type	NMR data
* Database Hosted	NO
Details	

Datasets Archived in Other Repositories

+ Add records | Bulk Edit | Explore

Details of input datasets archived in other repositories, including experimental data and starting structural models; mmCIF category: ihm_dataset_related_db_reference

Displaying all 1 records

Actions	RID	Id	Dataset List Id	Database Name	Accession Code	Version	Details
	3-7BZE	1	1	BMRB	BMR00000		

Edit Datasets Archived in Other Repositories: 3-7BZE

Save | Delete | Reset

* indicates required field

Record Number	1
* Structure Id	D_3-7B8W
* Id	1
* Dataset List Id	1
* Database Name	BMRB
* Accession Code	BMR00000

Other restraint files

PDBj Protein Data Bank Japan

Uploaded Restraint Files

+ Add records | Bulk Edit | Explore

Uploaded restraint data files (in csv/tsv format) related to the entry. Crosslinking restraints, predicted contacts, generic distance restraints etc. can be uploaded as CSV/TSV files in the correct format. Only one file of a particular type can be uploaded. The appropriate tables will be populated based on the uploaded data. Templates for the CSV files can be downloaded from https://data.pdb-dev.org/chaise/recordset/#1/PDB:Entry_Related_File_Templates

Create 1 Uploaded Restraint Files record

Save

* indicates required field

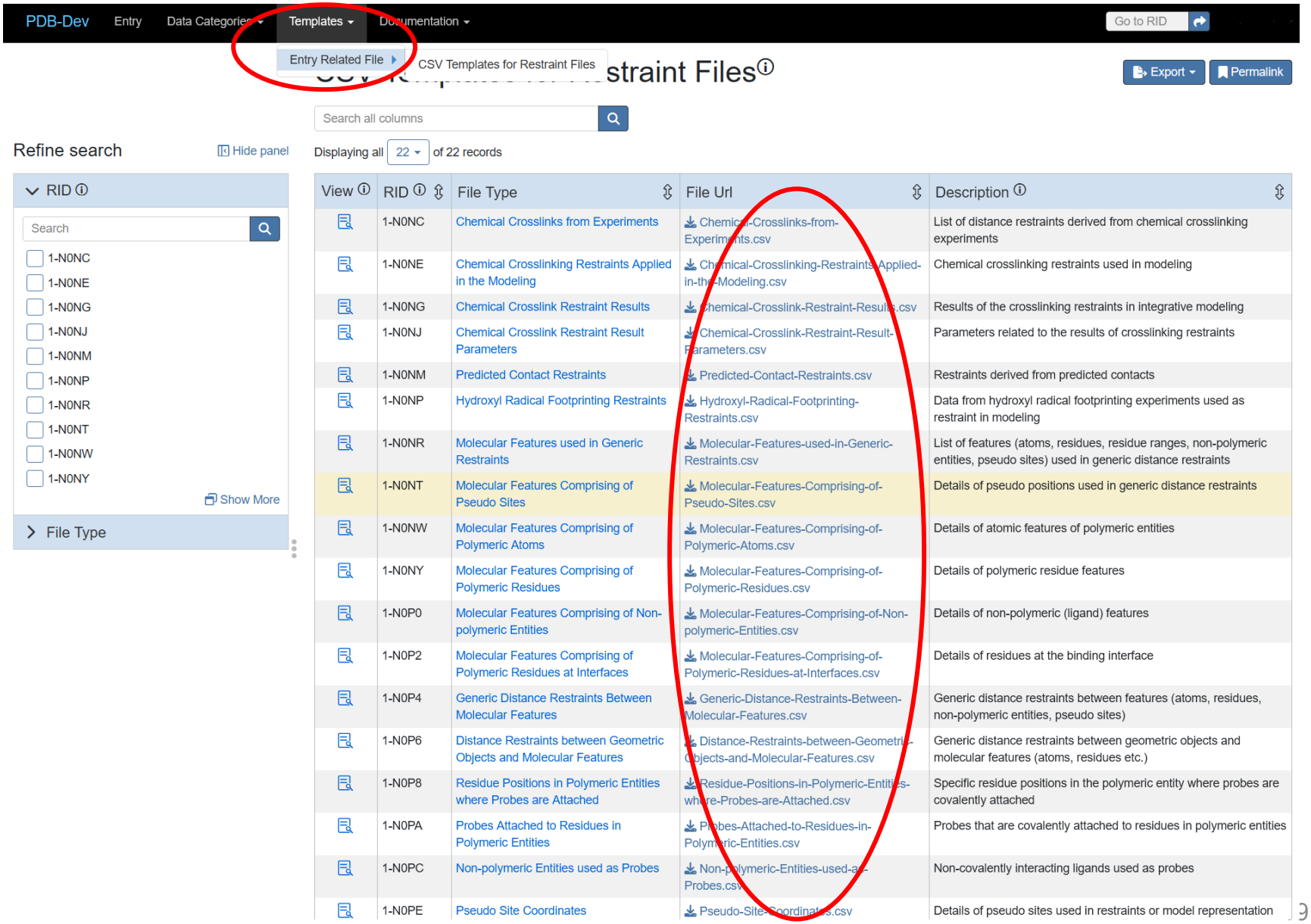
Qty 1 Clone

Record Number	1
* Structure Id	D_3-7B8W
* File Type	Select a value
* File Format	Select a value
* Uploaded File	Select file
Description	
* Restraint Workflow Status	DRAFT

Displaying all 22 of 22 records

Select	RID	Name	Table Name	Description	RCB	RMB	RCT	RMT
☑	1-MT0J	Chemical Crosslinks from Experiments	ihm_cross_link_list	List of distance restraints derived from chemical crosslinking experiments			2020-04-15 04:02:37	2021-06-15 05:29:38
☑	1-MT0M	Chemical Crosslinking Restraints Applied in the Modeling	ihm_cross_link_restraint	Chemical crosslinking restraints used in modeling			2020-04-15 04:02:37	2021-06-15 05:29:38
☑	1-MT0P	Chemical Crosslink Restraint Results	ihm_cross_link_result	Results of the crosslinking restraints in integrative modeling			2020-04-15 04:02:37	2021-06-15 05:29:38
☑	1-MT0R	Chemical Crosslink Restraint Result Parameters	ihm_cross_link_result_parameters	Parameters related to the results of crosslinking restraints			2020-04-15 04:02:37	2021-06-15 05:29:38
☑	1-MT0T	Predicted Contact Restraints	ihm_predicted_contact_restraint	Restraints derived from predicted contacts			2020-04-15 04:02:37	2021-06-15 05:29:38
☑	1-MT0W	Hydroxyl Radical Footprinting Restraints	ihm_hydroxyl_radical_fp_restraint	Data from hydroxyl radical footprinting experiments used as restraint in modeling			2020-04-15 04:02:37	2021-06-15 05:29:38

CSV Templates for Restraint Files



Submission step

Submit entry and wait for mmCIF generation. Once all data are entered/uploaded, the “Workflow Status” column in the “Entry” table needs to be changed to “Submit”.

Accession codes

Accession codes will be provided by PDB-Dev after the submission has been processed to ensure compliance with the IHMCIF dictionary and all the necessary information has been provided. The accession code will be updated in the “entry” table under the “accession code” column.

Thanks for your attention !