

○常住 規代¹、中根 崇智¹、川端 猛^{1,2}、栗栖 源嗣¹

1. 大阪大学 蛋白質研究所

2. 東北大学 大学院 情報科学研究科

EMPIARとは

- EMPIAR (**E**lectron **M**icroscopy **P**ublic Image **A**rchive) は、3Dバイオイメーシング実験から得られた生の2次元画像データおよび特定の3Dバイオイメーシングデータセットを無償で公開するデータベースです※1
- 電子顕微鏡データベース (EMDB※2) や、生体高分子の立体構造データベース (PDB※3) と相補する役割があります。
- 公開された生データを使って三次元構造を再構成することにより、解析ツールの開発や解析手法の研究に用いられ、分解能向上に役立っています。
- オープンソースの解析ツールを使えば、市販のGPU搭載PCさえあれば、非専門家でも電顕画像の単粒子構造解析手法の学習が可能です。
- 2022-09-05現在、**991** entries & **2.06 PB** のデータが公開されています。

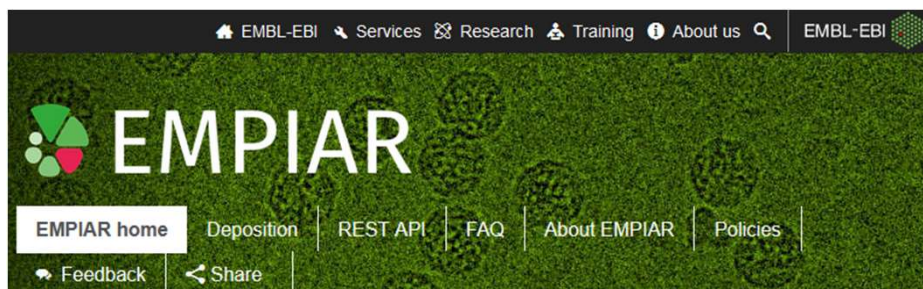
※1: <https://www.ebi.ac.uk/training/online/courses/empiar-quick-tour/what-is-empiar/>

※2: <https://pdj.org/emnavi/?lang=ja>

※3: <https://pdj.org/>

EMPIAR-PDBjとは

- 欧州バイオインフォマティクス研究所（EMBL-EBI）による EMPIAR@EBI の公式ミラーサイトです。
- 本家（EBI）データの効率的なデータの配布を目的に、2018 年に PDBj 内で運用開始しました。
- 2019 年 8 月から、HDD 郵送によるデータ登録支援（2019/08）を開始しています。

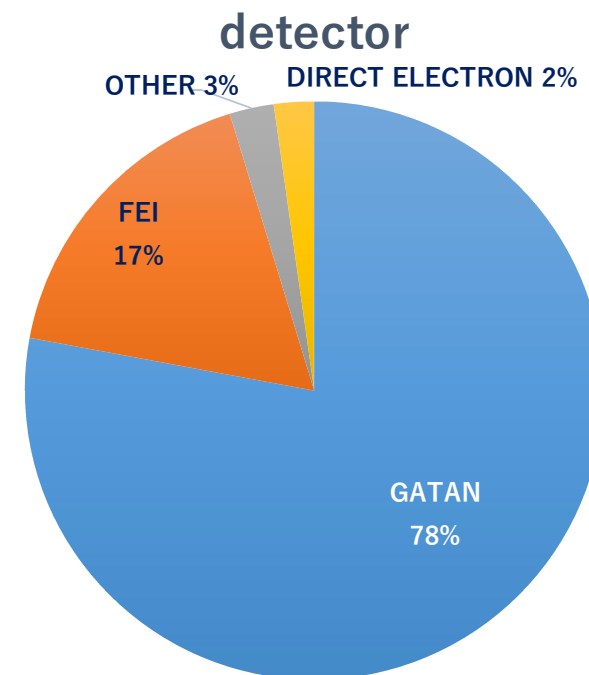
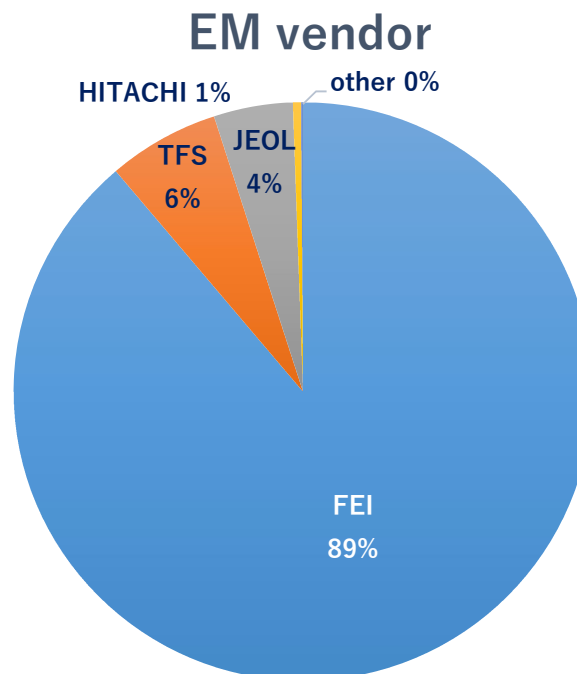
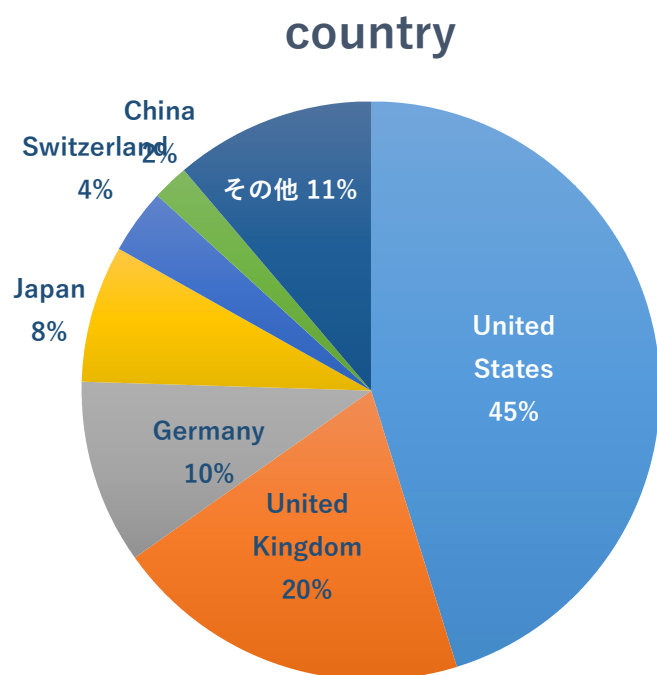


<https://www.ebi.ac.uk/empiar/>



<https://empiar.pdbj.org/>

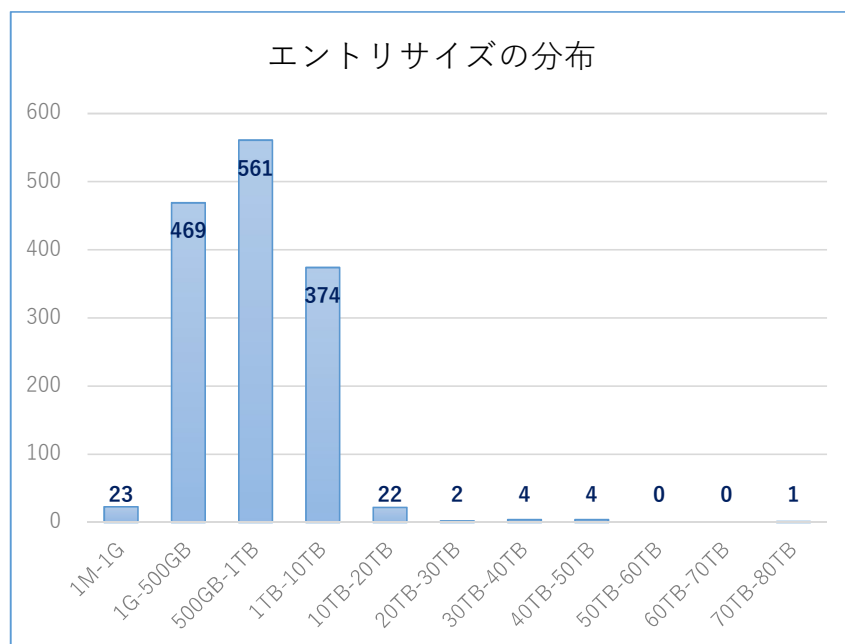
各国から多彩なデータが登録されている



※1: 筆頭 corresponding author の国籍で分類

※2: <https://pdbj.org/emnavi/>

1 エントリあたりサイズは頭打ち傾向



エントリサイズTOP10

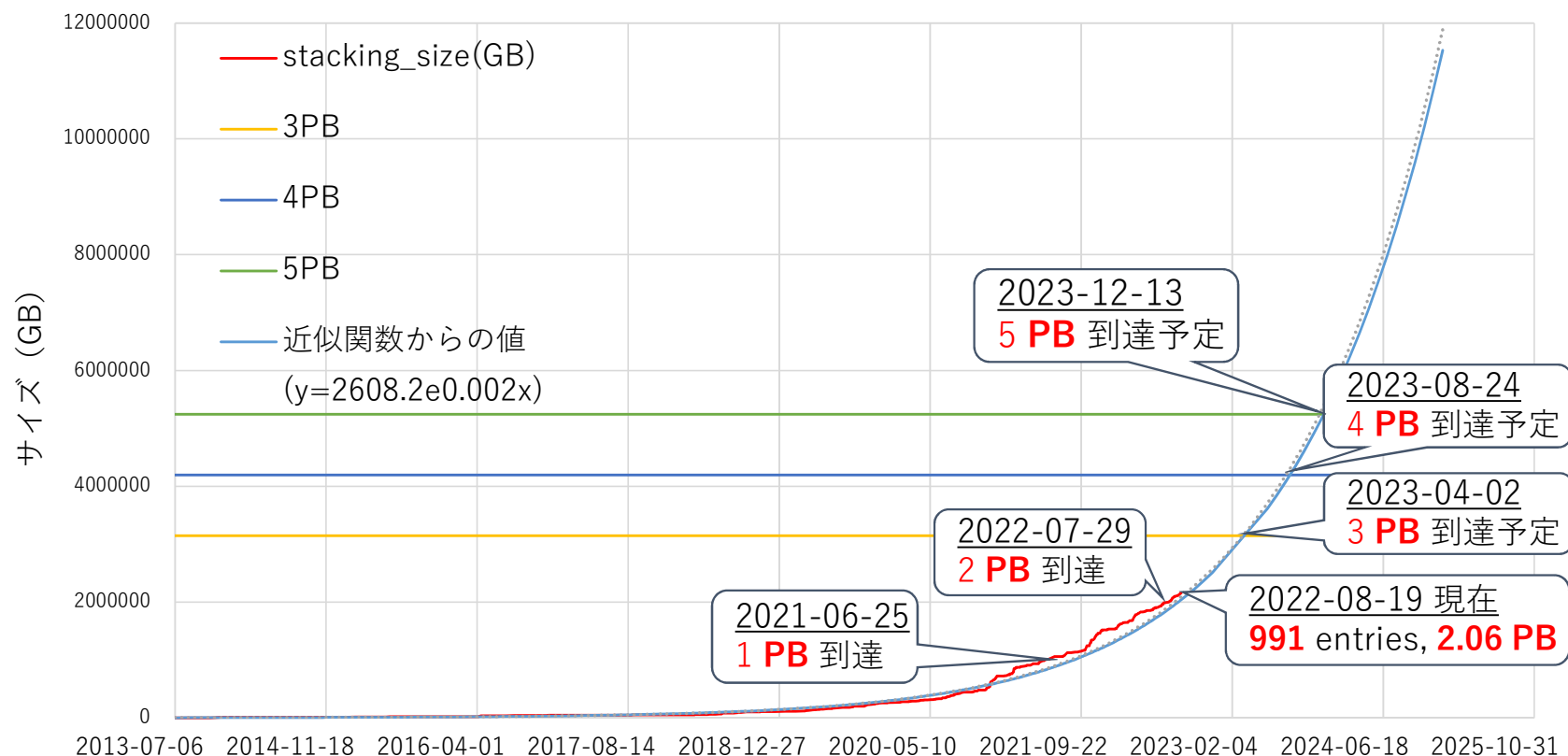
empiar_id	release_date	size(TB)
10781	2021-10-11	74.64
10517	2020-11-27	46.86
10809	2021-10-22	45.74
10375	2020-12-18	43.81
10591	2021-02-12	41.66
10735	2021-08-06	35.78
10437	2021-11-08	33.99
10473	2020-11-27	33.93
10999	2022-07-22	32.36
10638	2021-05-07	24.41

数十TBのエントリが頻出していた時期もありました。

自動撮影により画像ファイル数は増加していますが、TIFF圧縮の普及により1ファイルあたりのサイズは1GB未満のものが多くなり、1エントリのサイズの伸びは鈍化しています。

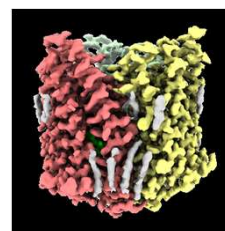
データサイズは指数関数的に増加中

リリース開始（2013-07-06）からのデータサイズの推移と予測

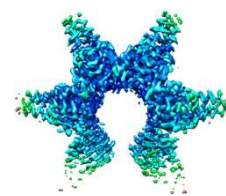


EMPIAR-PDBjにより登録代行されたエントリー

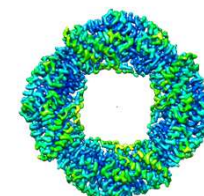
年	件数
2019	2
2020	6
2021	6
2022	15 (公開済み) + 8 (公開待ち)
合計	37 (公開待ち含む)



10926



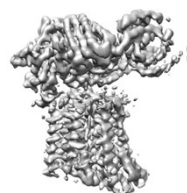
11072



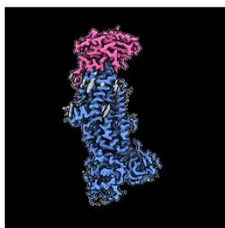
11079



10733



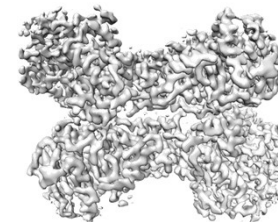
11046



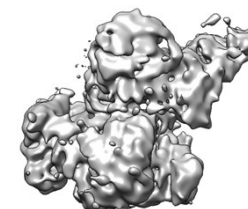
11057



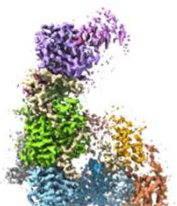
11056



11059



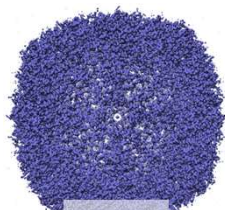
11061



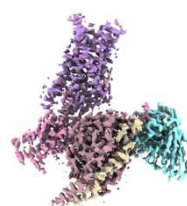
11099



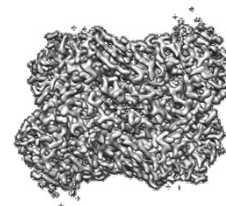
11091



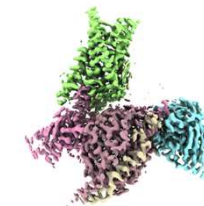
11101



11074



11024



11073

EMPIARデータの登録

EMPIARデータの登録には大きく 2 つのステップがあります。

1. メタデータの作成
2. データの転送

EMPIARの登録手順※¹は公開を希望するすべて人に開かれています。

しかし、数TBのデータ転送に必要な、“[IBM Aspera Connect](#)” や “[Globus Connect Personal](#)” の導入や、ネットワーク環境の制約※²により登録を躊躇している方が多いのが現状です。

EMPIAR-PDBj ではそういった方々に対して、登録代行サービス（Deposition代行）を提供しています。

※1: マニュアル “[EMPIAR deposition manual](#)” を参照

※2: PDBjの1Gbpsネットワークで1TBのデータセットを転送するのに約8時間以上必要

EMPIARデータ登録代行サービス

EMPIAR-PDBj では、データ登録を希望する方に以下のようなサービスを行っています。

- メタデータ作成補助
- データの入ったメディアを大阪大学 蛋白質研究所に送付、蛋白研ネットワークからEBIへデータ転送
- 非圧縮 MRC 動画ファイルの TIFF 形式への圧縮

代行作業の内容は人それぞれです。登録をお考えの方は一度[ご相談](#)ください。

EMPIARにエントリを作成する

登録データを作成するには以下の2つの方法があります。

1. JSONファイル

- 登録用JSONファイルを作成 ([sample](#))
- [メール](#)に添付してPDBjへ送信 (PDBjでEBIに送信)

2. Webインターフェース

- EMPIAR-EBI Depositionサイトでエントリを作成
- EMPIAR-PDBj担当者に権限 (View, Edit & Submit) を追加
- ブラウザ上でエントリ項目を入力

初めての場合は“2.Webインターフェースによる方法”をおすすめします。
いずれの場合でも、登録をお考えの際には事前に[ご相談](#)ください。

JSONファイルによる登録情報の作成

- [EMPIAR-EBI Deposition](#) のログインアカウントが不要
- 入力データの管理が容易
 - 次回以降、登録用JSONファイルを使い回すことができます。
 - テキスト形式なのでバージョン管理システムなどでの管理が可能です。
- オフラインでも編集可能
 - EBIへ送信するのはPDBjで行いますので、EBIのサーバーの状態に依存しません。

以下のようなデメリットがあります。

- JSONファイル（テキスト形式）を編集するのが面倒
 - [VSCode](#)などJSON形式に対応したエディタを使えば楽に編集することができます。
 - 文法や[入力規則 \(JSON schema\)](#)に従った修正はこちらで行います。
- Citation, Author 情報が手入力

Webインターフェースによる登録情報の作成

- [EMPIAR-EBI Deposition](#) のログインアカウントが必要
 - 氏名、メールアドレス、所属機関名があれば作成可能です。
 - Google, ORCIDのログイン情報を使用することもできます。
- エントリを共有し複数人で編集が可能
- Citation情報を自動で入力可能
 - DOI, PubMed, EMDB, ORCID などから citation, author の情報を入力補完可能です。

以下のようなデメリットがあります。

- Webサイトの状態（メンテナンスや不具合など）により作業ができない場合があります。
- 複数人が同時に編集作業を行うことはできません。
- データをローカルに保存することができない
 - テキストファイルへのエクスポート機能がなく、スクリーンショットやPDFへの印刷で保存する方法しかありません。

データ転送代行サービスのご紹介

EMPIAR-PDBj独自のサービスとして、[データ転送代行サービス](#)を行っています。

データメディアを大阪大学に送っていただくことにより、大阪大学のEMPIARスタッフが代わりにEBIにデータをアップロードします。データ転送のみのご依頼も可能です。

1. [Depositionサイト](#)でエントリを作成してアップロード用のトークンを取得（代行可）

ID ▼	EMPIAR ID (accession code) ▼	Title ▼	Upload Token ▼	Status ▼	Permissions ▼
47	N/A		ここに表示される17桁英数字		View, Edit and Submit

2. データの入ったメディアを 大阪大学 宛に送付（着払い可）
 3. 大阪大学からEMPIAR-EBIの所定の場所にデータをアップロード。前出のアップロードトークンを使用します。
 4. アップロード完了後HDDを返送します
- ※ データメディアの貸出も行っています。事前にご相談ください。

EMPIARデータ公開までの流れ

1. メタデータを登録
2. データを転送
3. DepositionデータSubmit
 - EBIに登録データを引き渡す。以後、編集はWebサイトからは行えなくなり、データを修正するにはEBIにリクエストを出して個別対応して貰う必要があるため、Submit前に登録データを再度確認してもらう必要がある。
 - この時点でEMPIAR-IDが付与される（Submitボタンを押してから数分程度）
4. 公開準備完了通知
 - EBI側で公開準備が完了した旨の連絡が来る（Submit完了後数日～数週間、EBIのキュレーターの稼働状況次第）。
 - 公開データの場合はこの時点で即公開リクエストに応じる旨のアクションを送信可能。放置していた場合でも所定の期間で自動公開される。
 - 論文公開待ちの場合は登録の際に指定したアクション "Release instruction" に従って処理される
5. 公開
 - 数日～数週間

新 EMPIAR-PDBj Web が新しくなります


EMPIAR PDBj

Electron Microscopy Public Image Archive

[EMPIAR home](#)
[Deposition](#)
[REST API](#)
[FAQ](#)
[About EMPIAR](#)
[Policies](#)
[Workshop](#)

Feedback [Language: Japanese](#)

大阪大学のEMPIAR-PDBjチームでは、アジアの電顕研究者が、EMPIARデータベースに大きな電顕画像を転送するお手伝いをしています。ネットワークでイギリスに転送する代わりに、ハードディスク自体を郵送・宅配便で大阪大学に送付、あるいは阪大のサーバまでネットワーク転送していただければ、こちらで仲介して登録サイトへの転送を代行します。EMPIARへの登録を希望する方で、阪大へのデータ送付・転送をご希望の方は、まず、電子メールで、登録したい電顕画像データについて、ご相談ください。

[Read more](#)






Quick Links

Talks and Tutorials

10

Release date	Imageset	Title	Authors and references	Size
2022-09-06	EMPIAR-11013	Cryo-EM images of equine apo-	Brown HG	127.4
		ferritin recorded as demonstration of MeasureIce software [626 multi-frame micrographs composed of 8	Subscriber: 5555271] [DOI: 10.1038/s42003-022-03698-x]	GB

Advanced Filter

Filter

Enter keywords separated by space

Filter Condition

☐ AND
 ☐ OR

Microscope

Detector

Author

※1: [Relational Database Management System](#)

2022年10月公開予定

新 EMPIAR-PDBj Webの特徴

- データベースの独立
 - EBIで提供されているTEXT形式 (json,xml) のデータをRDMS (MariaDB) にインポートすることにより、SQLクエリーを用いた詳細な検索が可能に。
 - PDBj独自項目を追加することにより、独自仕様への柔軟な対応が可能に。
- 保守性向上
 - EMPIAR-PDBjへのローカライズでブラックボックス化していたソフトウェア内部を、フルスクラッチで書き直してプログラム全体構造が可視化
 - ライブラリを、可能な限りオープンソースで向こう数年の保守が確約されているものに置き換えてメンテナンス性を向上
 - デバッグ環境の整備 (IDEによるデバッグ対応、ソースコードのバージョン管理)
- 動作速度の向上
 - RDMSにより必要なデータのみを取得することによりトップページ表示の高速化
 - ファイルツリーのデータをテキストに書き出すことにより表示の高速化
 - DBのチューニングは随時実施予定

新 EMPIAR-PDBj Web ソフトウェア概要

- Web（主要なもののみ）
 - Python 3.9.x
 - Django 3.2.x
 - MariaDB 10.x (MySQL clone)
 - Bootstrap 4.x
 - jQuery 3.6.x
- Storage（稼働中）
 - GlusterFS 7.2
 - Aspera High-Speed Transfer Server 4.3.1