EMPIAR – the Electron Microscopy Pilot Image Archive (pdbe.org/empiar)

CRYO-EM 3D Image Analysis Symposium 2016

Ardan Patwardhan

EMPIAR

MRC Medical Research Council

BBSRC bioscience for the future

PDBe

EMBL-EBI

Why did we set up EMPIAR?

- Community consultations
  - Outcome of Data Management Challenges in 3D Electron Microscopy workshop organized by PDBe in 2011
  - Discussion at GRC 2012

- Usage: validation, development, testing, teaching and…
  - Safe storage of your data!

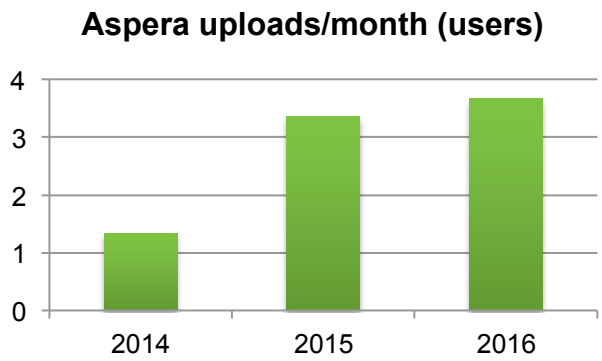
- 1 big image data set can easily exceed size of EMDB
  - EMDB simply not setup for regular 100GB+ transfers and storage
  - Need to try and test technologies – not easy on established EMDB production process
  - Needs additional funding

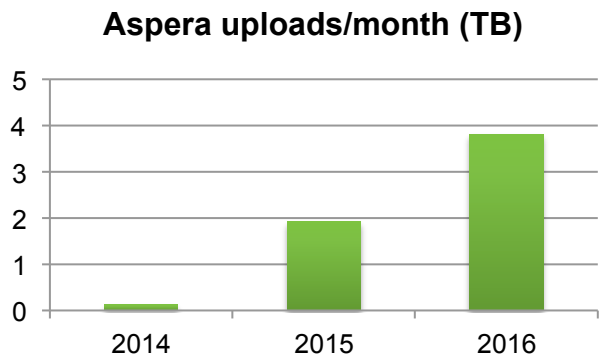
Setting up EMPIAR

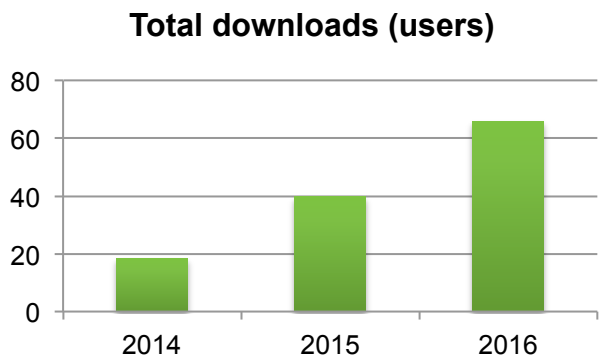
- Manpower – funding from MRC/BBSRC for 3 years
- Hardware infrastructure – EBI
- Need fast/robust transfer technologies
  - EBI used Aspera + Globus for PB+ transfers of sequencing data
  - Cost only on server side (EBI)
- Storage - plans for how we could scale up to PB range
- Manage growth
  - EMPIAR deposition requires EMDB deposition
  - Encourage but do not mandate
  - Lower barrier to deposition
    - Don’t be too prescriptive and provide tools to make it easier

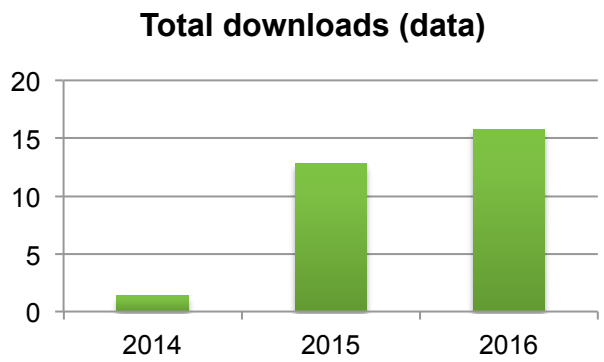
EMPIAR metrics

- Number or entries: 47 (22TB; average size ~ 470GB)
- 5 TB+ sets
- Transfer speed: uploads 1-2 TB/24h (Europe, US, Australia)
- “empiar” cited 20 times
- Nature methods publication: doi:10.1038/nmeth.3806









EMPIAR – resources

Welcome to EMPIAR

EMPIAR, the Electron Microscopy Pilot Image Archive, is a public resource for raw, 2D electron microscopy images. Here, you can browse, upload, and download and reprocess the thousands of raw, 2D images used to build a 3D structure. More ...

Deposit your data in EMPIAR to share it with the structural biology community.

Browse and download EMPIAR datasets using the table below.

Show 50 entries

Dataset	Title	Authors
EMPIAR-10058	Cryo-EM dataset of T20S proteasome [293 multi-frame micrographs composed of 12 frames each in MRC format]	Danev R, Baumeister W [Pubmed: 26949259] [DOI: 10.7554/eLife.13046
EMPIAR-10057	Volta phase plate in-focus dataset of T20S proteasome [158 multi-frame micrographs composed of 12 frames each in MRC format]	Danev R, Baumeister W [Pubmed: 26949259] [DOI: 10.7554/eLife.13046

Website

EMPIAR-10028

Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine

Publication: Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine
Wong W, Bai XC, Brown A, Fernandez IS, Hanssen E, Condron M, Tan YH, Baum J, Scheres SHW
Elife **3** (2014)
PMID: 24913268
DOI: 10.7554/ELIFE.03080

Related PDB entries: 3j79, 3j7a
Related EMDB entry: 2660
Deposition date: 29 May 2015
Release date: 17 Jun 2015
Dataset size: 1203 GB
Dataset DOI: 10.6019/EMPIAR-10028

Contains:
micrographs
picked particles

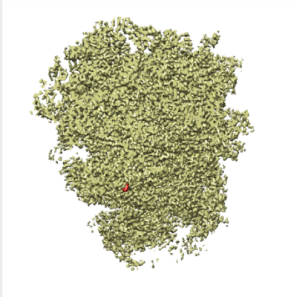


Image sets

Unaligned multi-frame micrographs

Category: micrographs - multiframe
Image format: MRC
No. of images or tilt series: 1081
Frames per image: 16
Image size: (4096, 4096)
Pixel type: 32 BIT FLOAT
Pixel spacing: (1.34 Å, 1.34 Å)

Micrographs 2 TB

Download

Entry pages

EMPIAR – resources

Web-based deposition

EMPIAR deposition system

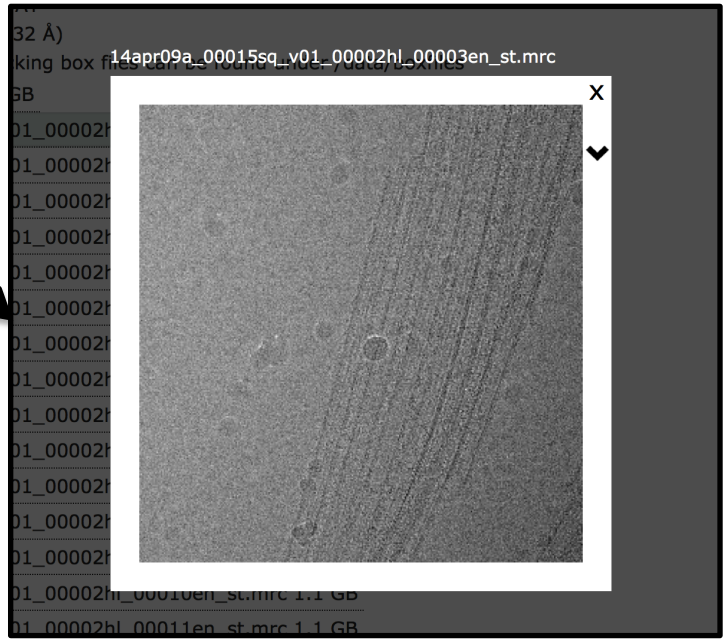
You are logged in as **testuser1**
- Edit profile
- Helpdesk
- Deposition manual
- Log out

Create a new deposition

Show 50 depositions Search:

ID	Lock	Title	Upload Token	Status	Owner	Permissions
56 View	Unlocked	Cryo electron tomography of herpes simplex virus during axonal transport and secondary envelopment in primary neurons	[redacted]	SUBMITTED	False	View, Edit and Submit
55 View	Unlocked			NOT SUBMITTED	False	View
54 View	Unlocked	Volta phase plate cryo-EM of the small protein complex Prx3	[redacted]	SUBMITTED	False	View

Thumbnail images



EMPIAR - plans

- Better support for 3D SEM and soft X-ray tomography data
- Functionality to enable sharing of data with anonymous referees prior to public release of data

Challenges and opportunities

- Grant funded
  - ends next year
  - define scope for renewal
  - community support vital – justify need and demonstrate usage
- Riding high on good-will and enthusiasm
  - Example for future initiatives for archiving of bio-imaging data
- Work with CCP-EM on data harvesting pipelines
- Frame averaged images or movie data
  - we are starting to get 10TB+ data sets
  - compression of full frame data?