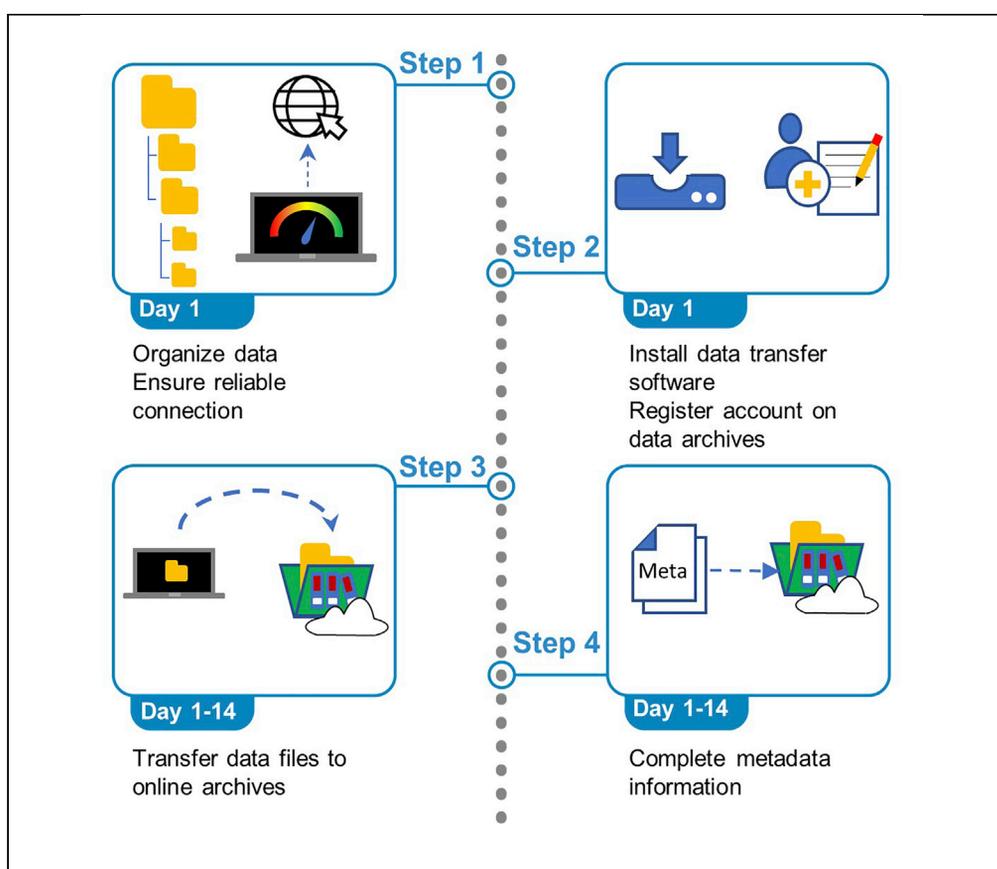


Protocol

Data-deposition protocols for correlative soft X-ray tomography and super-resolution structured illumination microscopy applications



This protocol illustrates the steps necessary to deposit correlated 3D cryo-imaging data from cryo-structured illumination microscopy and cryo-soft X-ray tomography with the BioStudies and EMPIAR deposition databases of the European Bioinformatics Institute. There is currently a real need for a robust method of data deposition to ensure unhindered access to and independent validation of correlative light and X-ray microscopy data to allow use in further comparative studies, educational activities, and data mining.

Andrii Iudin, Ugis Sarkans, Mohamed A. Koronfel, ..., Maria Harkiolaki, Ardan Patwardhan, Gerard J. Kleywegt

maria.harkiolaki@diamond.ac.uk (M.H.)
ardan@ebi.ac.uk (A.P.)

HIGHLIGHTS

Step-by-step guide for CLXM data deposition to EMBL-EBI archives

Establishes technical metadata requirements for efficient image sharing and reuse

Example deposition of CLXM dataset from a study of reovirus infection in U2OS cells

Iudin et al., STAR Protocols 2, 100253
March 19, 2021 © 2020 The Authors.
<https://doi.org/10.1016/j.xpro.2020.100253>



Protocol

Data-deposition protocols for correlative soft X-ray tomography and super-resolution structured illumination microscopy applications

Andrii Iudin,^{1,4} Ugis Sarkans,^{1,4} Mohamed A. Koronfel,^{2,4} Tom Fish,² Ian Dobbie,³ Maria Harkiolaki,^{2,5,*} Ardan Patwardhan,^{1,5,*} and Gerard J. Kleywegt^{1,6}

¹European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK

²Diamond Light Source, Harwell Science and Innovation Campus, Didcot OX11 0DW, UK

³Micron, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK

⁴These authors contributed equally

⁵Technical contact

⁶Lead contact

*Correspondence: maria.harkiolaki@diamond.ac.uk (M.H.), ardan@ebi.ac.uk (A.P.)
<https://doi.org/10.1016/j.xpro.2020.100253>

SUMMARY

This protocol illustrates the steps necessary to deposit correlated 3D cryo-imaging data from cryo-structured illumination microscopy and cryo-soft X-ray tomography with the BioStudies and EMPIAR deposition databases of the European Bioinformatics Institute. There is currently a real need for a robust method of data deposition to ensure unhindered access to and independent validation of correlative light and X-ray microscopy data to allow use in further comparative studies, educational activities, and data mining.

For complete details on the use and execution of this protocol, please refer to Kounatidis et al. (2020).

BEFORE YOU BEGIN

Background

3D cryo-imaging has become an integral method for many advanced biomedical investigations (Verkade and Collinson, 2019). Methodological progress invariably pushes the limits of resolution but also moves data collection in the direction of method integration and correlative approaches. A significant leap was the development of cryo-imaging which enabled nanometer-resolution imaging at near-physiological conditions (Henderson et al., 1990). Although cryo-imaging was initially implemented in electron microscopy (EM), soft X-ray tomography (SXT) is now providing an alternative for imaging of cellular ultrastructure in thicker vitrified samples (on the order of 10 μm compared to < 0.5 μm for EM), without the need for any chemical processing (Duke et al., 2014; Harkiolaki et al., 2018). Recently, a novel correlative imaging scheme between cryo-soft X-ray tomography (cryo-SXT) and cryo-structural illumination microscopy (cryo-SIM) (Figure 1) added the strength of fluorescence localization within the same field of view and at matching resolutions (Kounatidis et al., 2020).

Cryo-CLXM data can be acquired at synchrotron facilities such as beamline B24 at Diamond Light Source, UK. Access to this facility is granted based on a peer-reviewed application process (<https://www.diamond.ac.uk/Users/Apply-for-Beamtime.html>).

The life sciences are becoming increasingly data-dependent, and better approaches to facilitating the examination of scientific research data are needed, as exemplified by the FAIR data initiative



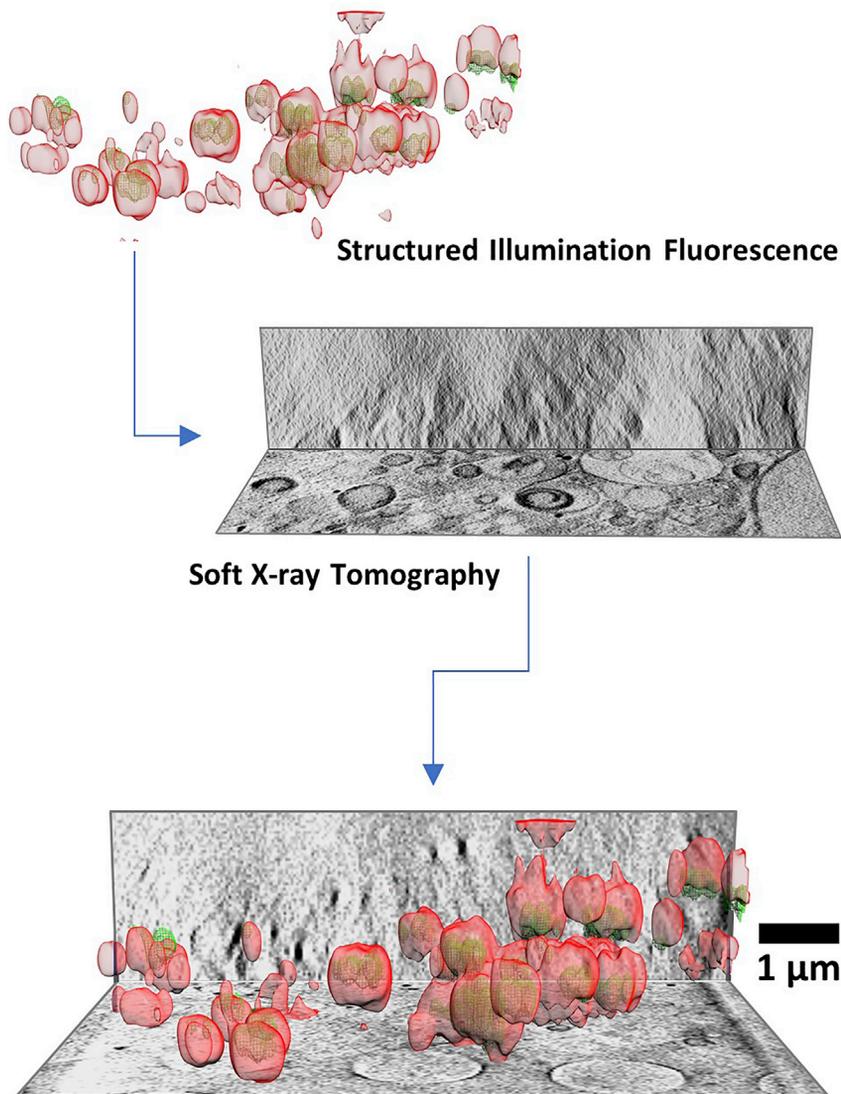


Figure 1. Illustration of correlative imaging combining structured illumination microscopy and soft X-ray tomography

The top panel displays a volume representation of SIM data showing endosomal vesicles (in red) carrying reovirus components (in green), and the SXT volume of the same area denoted by two orthoslices (middle). The bottom panel shows the correlation of SIM and SXT data in 3D. Data adapted from [Kounatidis et al. \(2020\)](#).

([Wilkinson et al., 2016](#)). There is an urgent need for open bio-imaging data archives, using community-supported standards, that provide a platform for large-scale data analysis, data sharing/validation and methods development ([Ellenberg et al., 2018](#)). Recent imaging, computational, and data-storage advancements have led to the establishment of the BioImage Archive at the European Bioinformatics Institute (EMBL-EBI). EMBL-EBI builds and maintains EMPIAR ([Iudin et al., 2016](#)) for storing 2D and 3D electron and X-ray microscopy/tomography data, and BioStudies ([Sarkans et al., 2018](#)) for storing bio-imaging data from all other modalities including light microscopy (and, more generally data associated with a publication for which no dedicated archive exists). EMPIAR and BioStudies together are appropriate resources for storing CLXM data, where related X-ray and light microscopy datasets are linked together through identification codes in each of the two archives. Describing correlative imaging experiments involves a substantial amount of data collection and processing information and parameters. Therefore, a clear protocol for data

and metadata communication is an essential requirement, for both biologists and microscopists, for further advancement of the field.

Note: Data deposition into the EMPIAR archive has to be performed in accordance with the EMPIAR policies and procedures (<https://empiar.org/policies>). More details about the Bio-Studies archive can be found at (<https://www.ebi.ac.uk/biostudies/>).

A number of actions before commencing deposition of data are important to ensure no delays: (a) the necessary data files need to be collated in a sensible directory structure to allow ease of access; (b) metadata for each dataset should be available; (c) appropriate web access and upload speeds should be secured; (d) all necessary file-transfer software should be installed and tested.

Data organization

⌚ Timing: 60 min to 1 day

1. Organize the data in directories with appropriate naming and structure for ease of identification and retrieval (Figure 2).
 - a. Define each time point or distinct states as separate samples. For example, if studying a dynamic process through time points, each of the time points imaged would represent a separate sample.
 - b. Identify and collect all data sets associated with each sample.
 - c. Define dataset pairs of multiples that comprise a correlative imaging set.

Note: The maximum file-path length on Windows operating systems is 260 characters.

⚠ **CRITICAL:** The size of a single file should not exceed 1 TB. For larger files, contact the staff of the relevant depository (EMPIAR or BioStudies)

Metadata availability

⌚ Timing: 60 min to 1 day

2. Ensure knowledge of and/or access to the information and parameters which describe each experiment (Table 1). These metadata are essential for data validation and informed reuse. Although currently incomplete metadata will not hinder the submission process, it is strongly recommended that they are submitted along with the experimental data.

Assess available infrastructure

⌚ Timing: 10 min

3. Evaluate your internet connection and speed of upload (e.g., via Ookla: www.speedtest.net); ideally >5 Mbps.

Note: In cases of low upload speed ensure connection stability to accommodate long data-upload times.

File upload software setup and installation

⌚ Timing: 30 min to 2 h

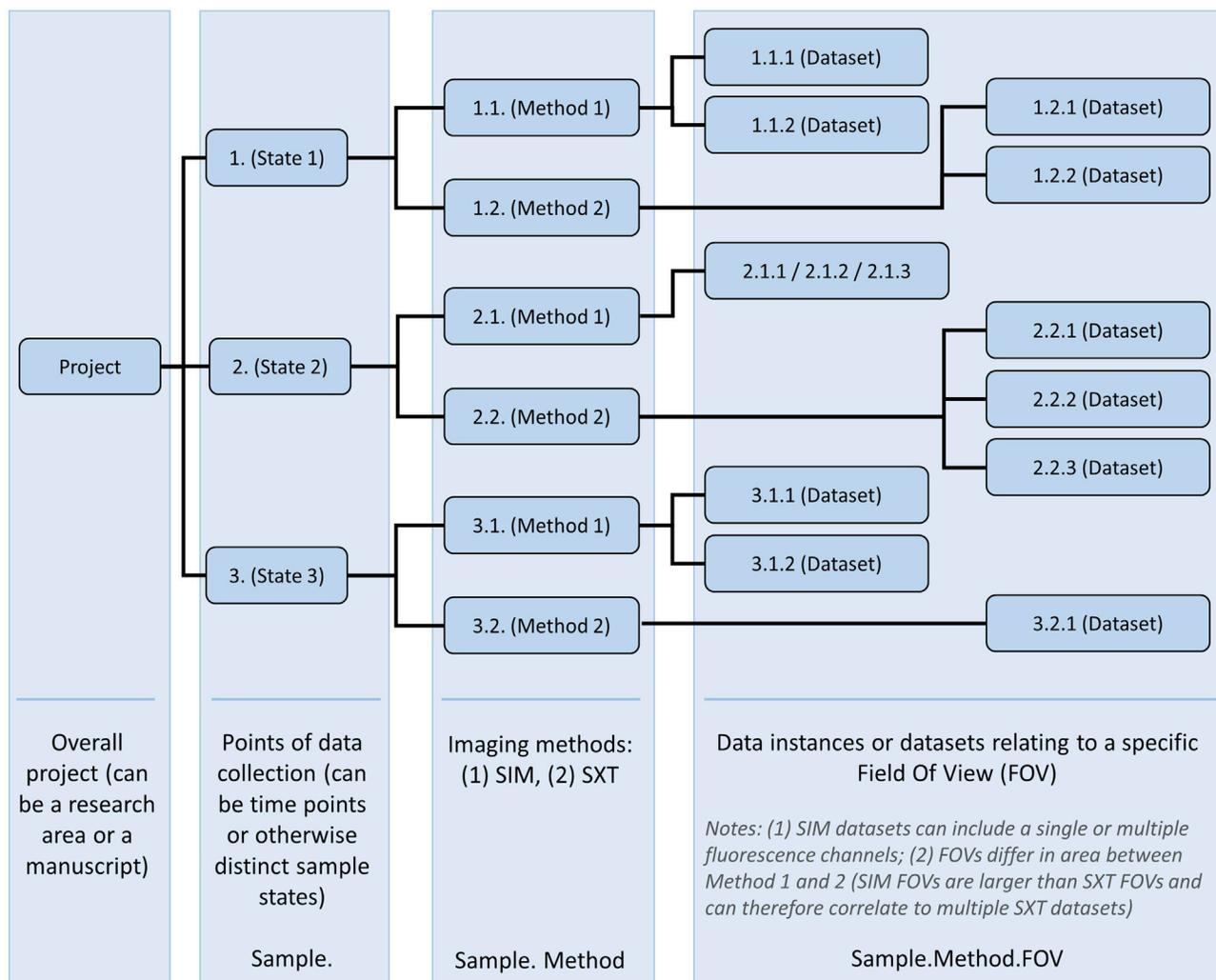


Figure 2. Suggested directory structure for efficient data management

4. Decide on the most appropriate method (software) for the data-transfer process. There are three available methods for uploading SIM data to BioStudies and two methods for uploading data to the EMPIAR database (Table 2).
5. Download and/or set up the chosen method(s) for data transfer as follows (all freely available):
 - a. Globus (<https://www.globus.org>) - requires setup of a personal endpoint (where the data will be uploaded from) and an active account login (the depositor's current organization/institution login details may be used):
 - i. Visit the Globus webpage and click "login." Check if login through your institution is available; otherwise, create a Globus ID or use a Google or ORCID account.
 - ii. Login to your account to reach the "File Manager" interface. Click the "Endpoints" tab from the left-hand-side menu, then click "Create a Personal Endpoint" (Figure S1A).
 - iii. Download and install "Globus Connect Personal" for your operating system (Figure S1B).
 - iv. When installation is complete, login through "Globus Connect Personal" and finish the setup by entering a collection name and description (optional) for this personal endpoint (Figure S1C).
 - b. Aspera (command-line client):

Table 1. Metadata fields necessary for submitting data to BioStudies and EMPIAR

	Cryo-SIM datasets (BioStudies)	X-ray microscopy datasets (EMPIAR)
General information	Dataset title Data collection date Data collection location (facility and RRID if applicable) Instrument used (beamline and microscope) Researchers and technical staff (names, affiliations, ORCIDs) Funding (beamline proposal or associated) Publication DOI (if applicable)	Dataset title Data collection date Data collection location (facility and RRID if applicable) Instrument used (beamline and microscope) Researchers and technical staff (names, affiliations, ORCIDs) Funding (beamline proposal or associated) Publication DOI (if applicable)
Sample specification	Sample ID (origin publication, ATCC number, etc.) Fluorophores used and administration route (endogenously expressed or otherwise) Fiducialization regime (e.g., nanoparticles-size-chemical composition-geometry) Sample carrier and substrate	Sample ID (origin publication, ATCC number, etc.) Fluorophores used and administration route (endogenously expressed or otherwise) Fiducialization regime (e.g., nanoparticles-size-chemical composition-geometry) Sample carrier and substrate
Microscope setup	Objective* (NA and serial number) Lasers used Detector type* Emission filters FOV* Pixel size	Energy of illumination (eV or Å; normally in the region of 500 eV for water-window imaging) Zone plate used* Magnification* FOV Pixel size*
Acquisition parameters	Number of excitation wavelengths Excitation wavelength, exposure, and power Number of focal planes (size Z) Stack height ($((\text{size Z} \div 15) - 1) \times 0.125$) Diffraction angle Linewidth per k_0 per wavelength* Acquisition data order ($S_{I_{\text{acquired}}}$) Reported data order ($S_{I_{\text{reported}}}$) Detector settings* (gain, binning, temperature, fan settings) Attenuation filter setting* Software (version) used 2D mosaic of grid area (optional)**	Exposure time and type Attenuation* Tilt-series parameters <ul style="list-style-type: none"> o Angle range (start/end)* o Angle step* o Number of frames Overlaps (with adjacent FOVs)* Notes on radiation damage if present* 2D mosaic of grid area (optional)**
Raw data	Format type (.dv/.mrc) Image dimensions (width and height in pixels and microns) Pixel type (e.g., unsigned byte, 32-bit float, etc.) Max/min intensities recorded Associated brightfield Z stack file**	Type (background subtracted/referenced)* Format type (.txrm/.tiff) Image dimensions (width and height in pixels and microns)* Pixel type (e.g., unsigned byte, 32-bit float, etc.) Non-referenced raw data file (optional)**
Data processing	Software (version) used Associated OTF files (optional)** Reconstruction summary values per channel: <ul style="list-style-type: none"> o Best fit o Optimized k_0 o Calculated linespacing o Amplitude 	Software (version) used Associated background reference image** Processing type (manual/automated) Reconstruction protocol (WBP, SIRT, Patch) Reconstruction summary values***: <ul style="list-style-type: none"> o Residual error mean/residual error standard deviation o Weighted mean residual error o Correlation coefficients o Fourier shell/ring correlation cut off
Processed data	3D image dimensions Voxel size Resolution Color depth (bits per pixel)	Image format Image dimensions Pixel type (e.g., unsigned byte, 32-bit float, etc.)
Chromatic shift	Software (version) used ChA-ChB transformation matrix: $t_z, t_y, t_x, r_z, m_z, m_y, m_x$ (t, translation; r, rotation; m, magnification)	
Correlation	Transformation matrices, and/or other means to define correlation	Related dataset IDs (these are "accession codes" which are received upon successful deposition of data to one of EBI's archives; see Expected outcomes section). If available, provide the accession code of a related dataset which was uploaded to a different archive (e.g., for CLXM datasets, these are the accession codes for associated SXT and SIM datasets respectively). Transformation matrices, and/or other means to define correlation

Metadata that may be supplied by the facility are denoted with * (all other data can be mined or calculated from file-header metadata; header metadata can be accessed, as text XML format, by opening 2D or 3D image files in specialized freeware such as Fiji (<https://imagej.net/Fiji/Downloads>) using its Bio-Formats importer). Individual files relating to data processing and acquisition are denoted with **. Values obtained post-data processing through respective reconstruction software (IMOD or equivalent) are denoted with ***.

- i. Download the appropriate Aspera command-line client for your operating system (<https://www.ibm.com/aspera/connect/>).
- ii. Extract the contents of the downloaded compressed file into a directory of your choice.
- c. Aspera (web interface):
 - i. Download and run the IBM Aspera Connect installation file (<https://www.ibm.com/aspera/connect/>).
 - ii. Install the IBM Aspera Connect web-browser add-on/extension.
- d. FTP/sFTP: use any freely available FTP client (such as FileZilla <https://filezilla-project.org/download.php>) which is available for your operating system.
- e. HTTP: No specific setup is required.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Soft X-ray tomography data of U2OS cells	EMPIAR archive	EMPIAR-10416
Structured illumination microscopy data of U2OS cells	BioStudies archive	S-BIAD-19
Software and algorithms		
Globus	https://www.globus.org	n/a
Aspera	https://www.ibm.com/aspera/connect/	n/a
Other		
Cryo-structured illumination microscope (Diamond Light Source, Beamline B24)	Custom built	n/a
Cryo-transmission X-ray microscope (Diamond Light Source, Beamline B24)	Zeiss	n/a

STEP-BY-STEP METHOD DETAILS

This section details the steps required for depositing SIM data to BioStudies and SXT data to EMPIAR. An overview of the whole process is provided in [Figure 3](#).

△ CRITICAL: At present SIM data should be deposited in BioStudies first to generate a BioStudies accession code, which is needed for depositing SXT data in EMPIAR.

Deposition of light microscopy data (BioStudies)

⌚ **Timing:** 30 min to 1 week depending on number of files and overall data size

The deposition of light microscopy data to the BioStudies archive comprises three main steps: (a) registering an account with BioStudies, (b) uploading the data, and (c) uploading the metadata (as per [Table 1](#)). These steps are described in detail below.

1. Register at the BioStudies database
 - a. Open the BioStudies submission page: <https://www.ebi.ac.uk/biostudies/submissions>.
 - b. Click on "Register."
 - c. Complete the form providing the following personal information: name, email address, ORCID (optional), and a password.

Note: The data will be stored, processed, and shared in a GDPR-compliant fashion according to EMBL-EBI and BioStudies GDPR privacy notices.

- d. Activate your account by opening the link sent to the email address you provided.

Table 2. Comparison of the file-transfer methods available for uploading data to the BioStudies and EMPIAR databases

Method	Application	Pros	Cons
Globus (a web application for file transfers)	EMPIAR only	Requires no software download	Requires setup of local endpoint
Aspera (command-line-based transfer client)	EMPIAR and BioStudies	Fastest (more suitable for long-distance and/or large data transfers)	Might require additional firewall settings setup (dependent on local setup) Requires familiarity with command-line operations
Aspera (web interface)	EMPIAR only	Does not require knowledge of command-line operations	Might require additional firewall settings setup (dependent on local setup) A maximum limit on files to upload at once, dependent on the memory available to the web-browser
File Transfer Protocol, FTP/sFTP	EMPIAR and BioStudies	Multitude software available on any operating system	Slower compared to Aspera FTP poses security risks (use software that supports sFTP)
Hypertext Transfer Protocol, HTTPS	BioStudies only	Least demanding technically No software downloads are required	Slowest of all methods

2. Login to the BioStudies submission system using the same link as in step 1a. In the home portal you can find a history of previous data submissions, as well as those still in progress under the “Drafts” tab.

3. To begin the data-upload process click on the “Files” tab in the upper right corner (Figure S2A). Depending on your chosen method for upload follow the instructions below:

a. Aspera:

- i. Click the “FTP/Aspera” button on the upper right corner (Figure S2A).
- ii. In the pop-up window, note the code for the “secret” directory as shown in Figure S2B (this will be the destination directory to which the data are to be uploaded).
- iii. In the command prompt, type the path to the ascp file in the directory where you installed the Aspera client (refer to step 2b in the “File upload software setup and installation” section above), followed by the connection settings, the path of the source directory to be uploaded and, the path of the destination “secret” directory. This information is provided in the FTP/Aspera pop-up window (Figure S2B).

For example, to begin the upload, first enter the path to the Aspera executable (this varies by operating system):

Windows:

```
C:\Users\\AppData\Local\Programs\Aspera\Aspera Connect\bin\ascp.exe
```

Mac:

```
~/Applications/Aspera\ Connect.app/Contents/Resources/ascp
```

Linux:

```
~/aspera/connect/bin/ascp
```

Note: Default paths are used in the above examples. The path to the Aspera executable (ascp) will depend on where you installed the Aspera client

Followed by the arguments:

```
-P33001 -i ..\etc\asperaweb_id_dsa.openssh -d <directory to upload> bsaspera_w@hx-fasp-1.ebi.ac.uk:.<;secret directory>
```

Note: Refer to [Troubleshooting 1](#) if Aspera transfer is blocked by your local firewall.

b. Transfer via FTP: you can use any freely available FTP client such as FileZilla (<https://filezilla-project.org/>) or use the command line to execute FTP commands as follows:

- i. Click the “FTP/Aspera” button on the upper right corner (Figure S2A).
- ii. In the pop-up window note the code for the “secret” directory as shown in Figure S2B (this will be the destination directory where the data are to be uploaded to).

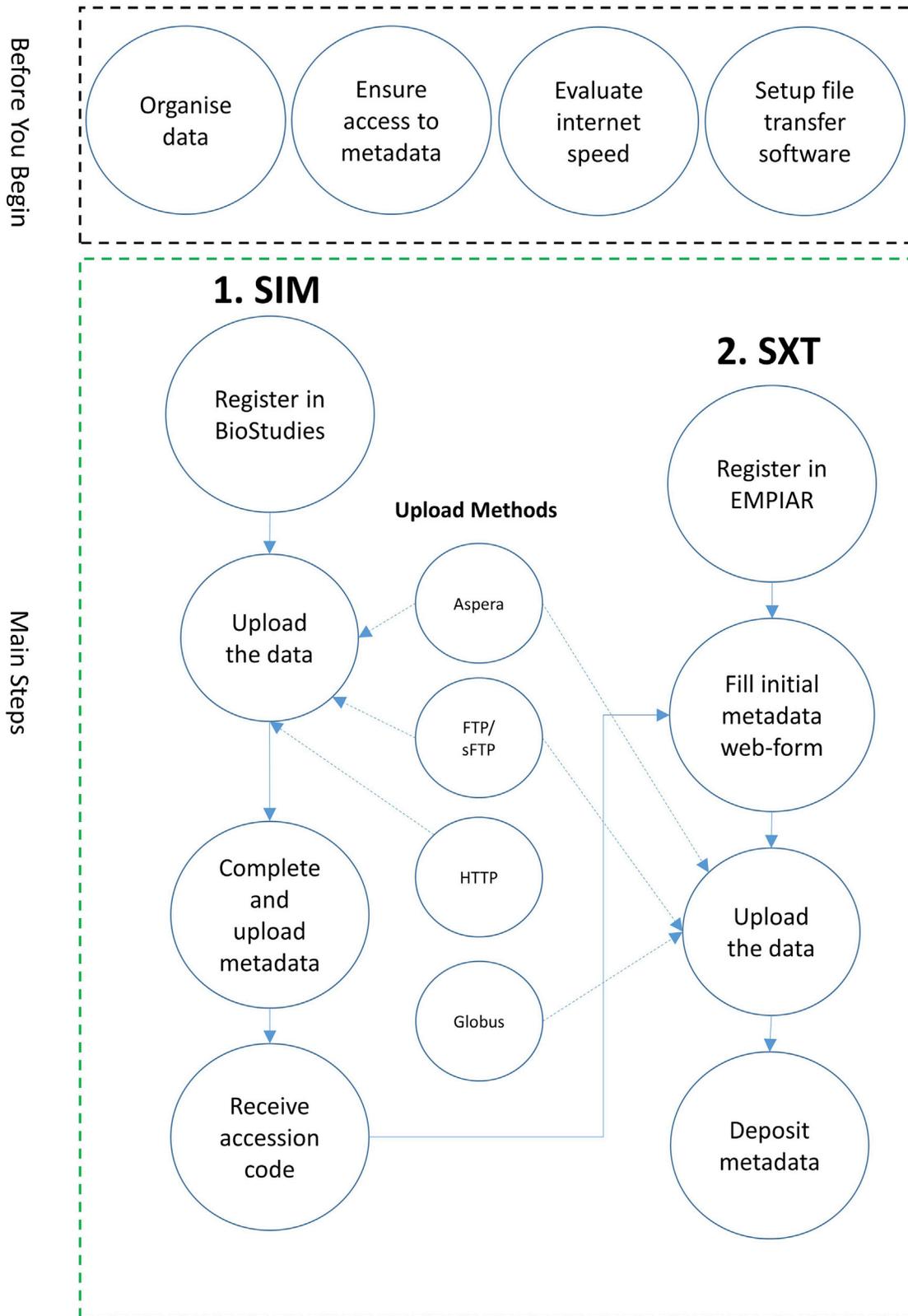


Figure 3. Overview of the workflow for uploading CLXM data to the appropriate EMBL-EBI archives

- iii. Open a command prompt and type *ftp* then at the *ftp>* prompt type *open* followed by the link to your chosen FTP server.
- iv. When prompted type your username and password.
- v. Define the source and destination directories as follows: type *cd* followed by the path of the source directory, then type *put* followed by the destination directory which will be your secret directory noted from step (ii).

Note: Further help on how to transfer files using FTP or Aspera can be found at <https://www.ebi.ac.uk/ega/about/ftp-aspera>.

- c. Transfer via HTTP: click "File upload" or "Folder upload," select from your local system the files or folder to be uploaded and confirm. The upload progress for each of the files and directories selected will be shown in the "Progress" column of the "Files" table (Figure S2A).

▮▮ **Pause point:** 1 min – 1 week depending on number of files and overall data size.

4. Upload the relevant metadata describing the dataset (based on Table 1). This can be done in one of two ways:
 - a. Web form (under development):
 - i. Fill all mandatory fields.
 - ii. Click the "Submit" button which appears after all relevant fields are completed.
 - b. Excel template:
 - i. Download template from <ftp://ftp.biostudies.ebi.ac.uk/biostudies/pub/Biolmaging/templates/> (file "SIM_template.xlsx").
 - ii. Complete the template.
 - iii. Send the completed file to biostudies@ebi.ac.uk for processing.

Note: You will now receive your BioStudies accession code which can be cited in your publication. This code will also be required to complete the deposition of the corresponding SXT data to EMPIAR.

Deposition of X-ray microscopy data (EMPIAR)

⌚ **Timing:** 30 min to 1 week depending on number of files and overall data size

5. Register an account with the EMPIAR database
 - a. Open the EMPIAR deposition webpage: <https://empiar.org/deposition/login>.
 - b. Click on "Register as a new user."
 - c. Complete the web form providing general information, such as your name, contact details, and a password.

Note: The data will be stored, processed, and shared in a GDPR-compliant fashion, according to EMBL-EBI (<https://www.ebi.ac.uk/data-protection/privacy-notice/embl-ebi-public-website>) and EMPIAR Deposition System Data Protection (<https://empiar.org/deposition/gdpr>) schemes.

- d. Activate the account by opening the web link sent to the email address that you provided.

Optional: you can use your Google, Facebook, or ORCID account to sign in.

6. Login to the EMPIAR deposition system using the same link as in step 5a to reach the deposition home portal; then click on "Begin/Continue an EMPIAR deposition" (Figure S3A).
7. Click on "Create a new deposition" (Figure S3B) to reach the "Deposition overview" page.

Table 3. Description of the available release instructions which define the date for the public release of the data

Release Instruction	Description
REL	As soon as the processing has finished and the entry approved by the depositor, the release procedure will be initiated. For data sizes in the TB range this process could take several days
EMDBPUB	Wait until the associated EMDB entry has been released before releasing the EMPIAR entry (only applicable to EMPIAR entries related to one or more EMDB entries)
HPUB	Wait until either the primary citation has been published or one year has elapsed since the date of deposition (whichever comes first)
HOLD	Wait until 1 year has passed from deposition

8. Complete the web form providing some of the initial metadata associated with your dataset as follows:

- a. *Deposition image*: click “choose a file” to upload a figure (not subject to copyright, e.g., by a journal) that illustrates a facet of your data or study (see [Figure S4A](#)).

Note: This image will be used on the EMPIAR website to illustrate your entry.

- b. *Citations*: complete the details of the publication where the data are described (e.g., DOI, journal name, authors etc.).

Note: Publication details can be automatically filled if you provide a DOI or PubMed ID.

Note: More than one publication can be added by pressing the “Add Citation” button. The first publication provided will be the primary one.

Note: If the data are not published yet, then some of the fields can be omitted or filled by pressing the associated “N/A” button. Publication details can be updated by curators at a later stage but remains the depositor’s responsibility to communicate publications associated with this data as and when they are released.

- c. *Entry details*: enter a title for the dataset, followed by “Entry type” and “Release instructions” as follows:
 - i. *Entry type*: select the type or modality used to acquire the dataset from the available options. For the scope of this protocol, select the “CLXM” option to upload the SXT data here (while the corresponding SIM data should be uploaded in the BioStudies archive as explained above).

Note: You will be required to provide the BioStudies accession code (refer to footnote 1) where the corresponding SIM data are stored (see step 8d).

- ii. *Release instructions*: choose an option (listed in [Table 3](#)) to indicate when the entry should be released (made available to the public) once it has been submitted and processed successfully.
- d. *Cross-references*: enter the accession codes of related datasets. For CLXM data, the BioStudies code for the corresponding light microscopy data is required. Click “N/A” for other unrelated codes such as EMDB (see [Figure S4B](#)).
- e. *Entry authors, Corresponding author and Principal Investigator(s)*: provide the relevant information of the research team that conducted the study.

Note: Entry authors may differ from the publication(s) authors specified previously.

Note: All authors must be listed under “Entry authors” including those listed as corresponding author and principal investigator(s).

9. Press the “Save & Validate” button on the “Deposition overview” page. If all fields have been successfully completed, you will be transferred to the “Upload data” page.
10. Upload the data using one of the three available methods:
 - a. Transfer via Globus:
 - i. In the “Upload data” page, click “Globus” under “Options for upload” and note your username and password for the EMBL-EBI endpoint (Figure S5A).
 - ii. Ensure you have set up your Globus account and personal endpoint (refer to step 2a in “File upload software setup and installation”).
 - iii. Login at the Globus webpage www.globus.org to reach the “File Manager” interface.
 - iv. In the “Collection” field, search for and select “EMBL-EBI Private endpoint,” then enter your authentication username and password noted from step (i). Enter your EMPIAR directory by selecting the link provided back in the “Upload data” page under the option “Globus” (Figure S5B).
 - v. In the Globus “File Manager,” click “Transfer or Sync...” then search for and select your personal endpoint in the second “Collection” field (Figure S5C).
 - vi. Under your personal endpoint, navigate to the directory to be uploaded and click “Start” to begin the transfer to the EMPIAR directory. You can follow the status of your transfer under the “Activity” tab (Figure S5C).
 - b. Transfer via Aspera (command-line client):
 - i. In the command prompt, type the path to the ascp file in the directory where you installed the Aspera client (refer to step 2b in the “File upload software setup and installation” section above), followed by the connection settings, the path of the source directory to be uploaded, and the path of the destination EMPIAR directory. This information is provided by selecting “Aspera Command Line” under “Options to upload.”
For example, to begin the upload, first enter the path to the Aspera executable (this varies by operating system):
Windows:

```
C:\Users\\AppData\Local\Programs\Aspera\Aspera Connect\bin\ascp.exe
```

Mac:

```
~/Applications/Aspera\ Connect.app/Contents/Resources/ascp
```

Linux:

```
~/aspera/connect/bin/ascp
```


Note: Default paths are used in the above examples. The path to the Aspera executable (ascp) will depend on where you installed the Aspera client
Followed by the arguments:

```
QT -P 33001 -l 200M -L- -k3 <directory to upload> emp_dep@hx-fasp-1.ebi.ac.uk:upload/<your_unique_directory_name>/data
```
 - ii. When prompted enter the provided password for your EMPIAR upload.
 - c. Transfer via Aspera Connect (web interface):
 - i. Click on “Aspera web interface” under “Options to upload.”
 - ii. Select either the “Upload files” or “Upload directories” option.
 - iii. Select the files or directories to upload from the pop-up dialog (a maximum limit of files/directories selected will be set by your web-browser and operating system).
 - d. Transfer via FTP: you can use any freely available FTP client such as FileZilla (<https://filezilla-project.org/>) or use the command line to execute FTP commands following similar steps as in step 3b of the “Deposition of Light Microscopy Data” section. You will need to contact EMPIAR team (empdep-help@ebi.ac.uk) to obtain your destination directory.

Note: Refer to [Troubleshooting 1](#) if Aspera transfer is blocked by your local firewall.

Note: Refer to [Troubleshooting 2](#) if problems are encountered using the above upload methods.

▮▮ **Pause point:** 1 min – 1 week depending on number of files and overall data size

11. In the “Upload data” page, click the “Associate image sets with the data” link from the left-hand-side menu ([Figure S5A](#)).
12. Complete the web form of the requested metadata to describe at least one image set. Additional metadata (in accordance with [Table 1](#)) should be entered in the “Details” field of the web form.

Note: This is for bioimage data only. Any textual and other supplementary data should not be created as a separate image set, but rather described in the free text “Details” field. Some of the fields can be automatically filled from the header of an image file by clicking the “Populate Form” button.

Optional: check the data for consistency by:

- a. Downloading a script that compares checksums of files on your machine and those stored in EMPIAR using the MD5 message-digest algorithm.
- b. Downloading a listing of files present on the EMPIAR side.

13. Press the “Save + Validation” button on the “Associate image sets with the data” page.

Optional: describe EMDDB-SFF (<https://emdb-empiar.org/emschema/segmentation.html/>) segmentation files. Segmentation is the decomposition of 3D volumes into regions that can be associated with defined (biological) objects. You can also provide the original segmentation files from software such as Amira HyperSurface (.surf), Segger (.seg), EMDDB Map masks (.map), IMOD (.mod), etc.

Note: Click the “Add segmentations” button at the bottom of the “Associate image sets with the data” page. A page “Associate segmentations with the data” will be opened where you can select one of your image sets and associate it with a segmentation file. This can be done for multiple files / image sets. If you change your mind you can click “Remove all segmentations.” You will not be able to submit unless you either remove all segmentations or save them with validation.

14. Click the “Submit entry” button.

Note: You will be redirected to the main deposition page where you can see your accession code, which can be cited in your publication.

Note: Once the entry has been checked by the EMPIAR curation team, you will receive a request to approve the entry.

EXPECTED OUTCOMES

After successful deposition of datasets into the BioStudies and EMPIAR archives, the user will receive the following information (via their email address):

1. Accession codes: these are codes that can be cited in publications to allow users to access the dataset. Accession codes are in the form of S-BIAD##### and EMPIAR-##### for the BioStudies and EMPIAR archives, respectively.

The recommended method to quote a dataset in a publication is by including the following statement “The dataset is available in the BioStudies database (<https://www.ebi.ac.uk/biostudies>) under accession number S-BIAD##### and in the EMPIAR database (<https://empiar.org>) under accession number EMPIAR-#####”.

2. A “secret” link that can be used to access the dataset before the public release date. Access through this link does not require a login and it can, for example, be passed to journal editors to share the data with peer reviewers. However, this link should not be included in publications.
3. When the conditions for public release are met, a new email notification will be received to indicate that the data have become publicly available.
4. Partial/Full validation of the submitted data (under development).

The EMBL-EBI holds over 300 petabytes of data and serves millions of requests per day. Delivering secure, searchable data resources that facilitate research data accessibility and reuse (see <https://empiar.org/reuse>) is one of the core missions of EMBL-EBI. We expect that the BioImage Archive will fuel new scientific breakthroughs similar to those seen in genome sequencing. To achieve this, it is essential to establish the technical and biological metadata required for efficient image sharing and informed reuse, and ensure adoption of those guidelines in the community. Example of submitted data and deposition entries can be seen in [Figure S6](#) and accessed in the BioStudies and EMPIAR archives using the following accession codes S-BIAD19 and EMPIAR-10416, respectively.

LIMITATIONS

It is conceivable that further metadata fields could be required in the future to allow independent validation of raw and reconstructed data. In addition, the metadata fields highlighted ([Table 1](#)) are specific for the two methods presented here (SXT and SIM) as these are implemented at the correlative cryo-imaging beamline B24 at the UK synchrotron Diamond Light Source and may be renamed in the future to become part of consensus metadata fields for related imaging techniques.

TROUBLESHOOTING

Problem 1

Firewall issues with home server.

Potential solution

Contact your systems administrator to enable an access point for your file transfer. You might need to modify the rules to accommodate the following:

Permit TCP outbound to 193.62.197.71 on port 33001

Permit UDP outbound to 193.62.197.71 on port 33001

Permit UDP inbound from 193.62.197.71 on port 33001

Problem 2

None of the transfer methods works

Potential solution

Email biostudies@ebi.ac.uk and/or empdep-help@ebi.ac.uk and provide as much information as possible: logs, time and date, command used, operating system, location.

RESOURCE AVAILABILITY

Lead contact

Gerard J. Kleywegt, gerard@ebi.ac.uk

Materials availability

This study did not generate new unique reagents.

Data and code availability

Datasets included can be found in EMPIAR and BioStudies with accession codes EMPIAR-10416 and S-BIAD-19 respectively.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xpro.2020.100253>.

ACKNOWLEDGMENTS

We would like to thank the staff and associates of beamline B24 for technical support and constructive feedback. Special thanks to Ilias Kounatidis, Chidinma Okolo, Nina Vyas, and Kamal Nahas. The work on EMPIAR is supported by MRC/BBSRC (grant MR/P019544/1) and by EMBL-EBI. The work on BioStudies is supported by BBSRC (grant BB/R015082/1) and by EMBL-EBI. The UK Government via UKRI continues to support EMBL-EBI's technical hardware infrastructure with EMBL providing supporting operational costs.

AUTHOR CONTRIBUTIONS

A.I., U.S., and M.A.K. wrote the manuscript. T.M. and I.D. helped with metadata format and evaluation. A.P., M.H., and G.J.K. reviewed and coordinated the development of the protocol.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

- Duke, E.M.H., Razi, M., Weston, A., Guttmann, P., Werner, S., Henzler, K., Schneider, G., Tooze, S.A., and Collinson, L.M. (2014). Imaging endosomes and autophagosomes in whole mammalian cells using correlative cryo-fluorescence and cryo-soft X-ray microscopy (cryo-CLXM). *Ultramicroscopy* *143*, 77–87.
- Ellenberg, J., Swedlow, J.R., Barlow, M., Cook, C.E., Sarkans, U., Patwardhan, A., Brazma, A., and Birney, E. (2018). A call for public archives for biological image data. *Nat. Methods* *15*, 849–854.
- Harkiolaki, M., Darrow, M.C., Spink, M.C., Kosior, E., Dent, K., and Duke, E. (2018). Cryo-soft X-ray tomography: using soft X-rays to explore the ultrastructure of whole cells. *Emerg. Top. Life Sci.* *2*, 81–92.
- Henderson, R., Baldwin, J.M., Ceska, T.A., Zemlin, F., Beckmann, E., and Downing, K.H. (1990). Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* *213*, 899–929.
- Iudin, A., Korir, P.K., Salavert-Torres, J., Kleywegt, G.J., and Patwardhan, A. (2016). EMPIAR: a public archive for raw electron microscopy image data. *Nat. Methods* *13*, 387–388.
- Kounatidis, I., Stanifer, M.L., Phillips, M.A., Paul-Gilloteaux, P., Heiligenstein, X., Wang, H., Okolo, C.A., Fish, T.M., Spink, M.C., Stuart, D.I., et al. (2020). 3D correlative cryo-structured illumination fluorescence and soft X-ray microscopy elucidates reovirus intracellular release pathway. *Cell* *182*, 515–530.e17.
- Sarkans, U., Gostev, M., Athar, A., Behrang, E., Melnichuk, O., Ali, A., Minguet, J., Rada, J.C., Snow, C., Tikhonov, A., et al. (2018). The BioStudies database—one stop shop for all data supporting a life sciences study. *Nucleic Acids Res.* *46* (D1), D1266–D1270.
- P. Verkade, and L. Collinson, eds. (2019). *Correlative Imaging* (Wiley).
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., et al. (2016). Comment: The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* *3*, 160018.