Imaging Data Management Survey Report and Next steps

Survey Results Summary

An online survey was conducted to collect data on the needs of the UK bioimaging community on imaging data management.
The survey was conducted using google forms and remained open to responses for one month between April to May 2025. The survey received a total of 46 responses from various imaging facilities and institutions across the UK.

The following questions were asked:

1. What type of biological imaging facility do you represent?
2. What imaging modalities does your facility have/use?
3. How is imaging data typically stored at your facility?
4. How is imaging data managed at your facility?
5. What metadata do you record at your facility?
6. If Other or None, please describe
7. How do you capture and record the metadata at your facility?
8. What metadata schema is used at your facility?
9. What file formats do you primarily use for imaging data?
10. Do you use ontologies or any kind of standard vocabulary to describe your data?
11. Does your facility have a dedicated data steward or data manager?
12. What are your biggest challenges in managing data prior to publication?
13. Does your facility deposit data in public repositories like the BioImage Archive and IDR?
14. What challenges do you face in depositing data in public repositories?
15. What improvements would make it easier for your facility to share data with public repositories?
16. What policies govern long-term data storage at your institution?
17. Do you need to share data with collaborators outside of your institution? If so, how do you go about it?
18. What type of support do you provide?
19. What type of training would benefit your facility most?
20. Are there any other challenges you would like to mention?
21. Is it ok to contact you for further information?

For questions 1, 2, 3, 4, 5, 8, 9, 11, 13, 16, 18 and 21, Google Forms created charts.
For other questions requiring text answers, word clouds were created to better see the commonalities between responses.

Word count percentages were calculated by calculating the number of responses that mention the word divided by the total rows in that given column (=question). This would mean a word percentage by respondent unless not all 46 respondents gave an answer to that particular question.

Question 21 contains the email addresses of the respondents who agreed to be contacted further.

The replies to the survey show certain commonalities are shared between the different bioimaging facilities, as summarised below:

- The majority of the respondents are from imaging facilities embedded within universities.
- The three major imaging modalities are light microscopy (76%), electron microscopy (41%), and multi-modal imaging (32%). However, it is evident that a myriad of imaging modalities are used across the facilities, e.g. multi-modal including omics data, multiplexed spatial proteomics, label free microscopy, mass spectrometry, MRI, PET, ultrasound, in vivo imaging, clinical radiology imaging, and light-sheet microscopy.
- Data is typically stored in institutional storage servers (80%), followed by local hard drives (46%), and cloud storage (22%).
- Majority of the metadata recorded is on instrument or user related, like image acquisition, date, PI/researcher name, followed by some sample preparation (recorded by 24% of the respondents) and analysis metadata (35%). 15% of the responses said no metadata is recorded at their facility. Metadata is mostly recorded in some form of file (format varies) and mainly comes from the equipment in itself and some from a booking system (if a booking system was used).
- Hardly any metadata schema is used (only 6.5% of the facilities use REMBI).
- The majority of respondents use proprietary formats (70%), closely followed by tiff (63%). Only 24% use OME-Tiff and 4% OME-Zarr.
- 59% of respondents said they lack a dedicated data steward, and 28% said they don't have a dedicated data steward but they share the role amongst staff.
- Most respondents (24.36%) reported large volumes of data as their biggest challenge, which links to problems with data storage (33%) and data transfer. Majority of the challenges reported are about overall data management: file location, lack of metadata, duplication of files.
- About 41% of the facilities publish the data in public repositories. The main reasons behind this low percentage seem to be lack of time, incentive, awareness, and support.
- Most institutions are governed by institutional or funder long-term storage policies. However, 35% of the respondents said there was no clear policy.
- The majority of respondents need to share data with collaborators outside of their institution. The main ways of sharing are via external hard drives (54%), the cloud (41%), OneDrive (15%), Globus (15%), OMERO (6%) or some form of institutional file transfer system.
- Most of the support provided at the institutions revolves around image acquisition (93%), processing (85%) and analysis (91%).

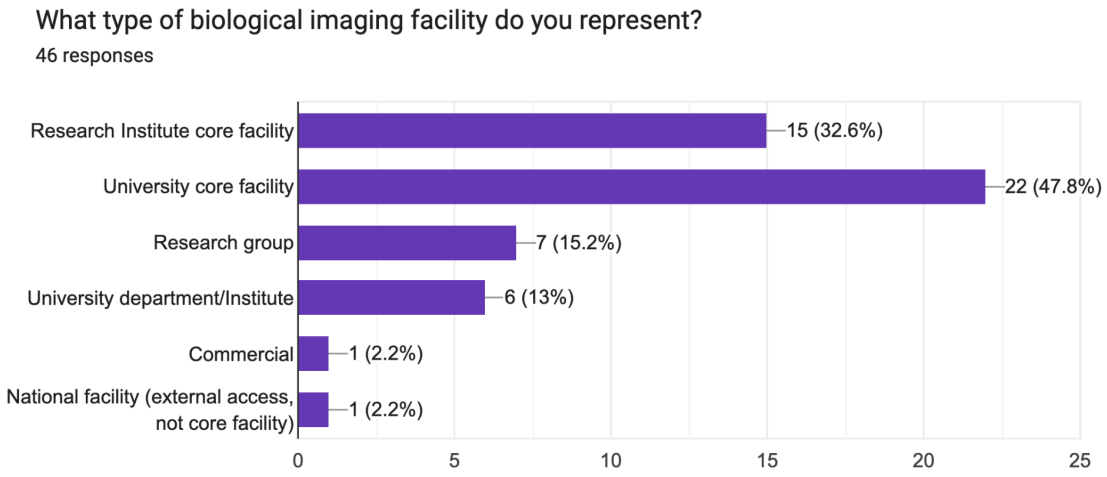
Shared challenges:

- The majority of the respondents (67%) stated they had no standardised approach when asked how they manage imaging data at their facilities. This is highlighted in the follow-up question with regards to problems regarding data location, lack of metadata, file duplication.
- One takeaway point from the metadata question is that the metadata that is recorded and the level of description largely depends on the user. The facility will record mainly metadata related to its usage. As a matter of fact, the main form of metadata records comes from equipment-generated files.
- Another major issue is the lack of use of metadata schemas (85%) and ontologies. It bears significance to say that only 15 out of 46 respondents answered the ontologies question.
- Most commonly used file format is some form of proprietary one, although it is closely followed by .tiff (63%). Very few use OME formats.
- Regarding the data steward question, if we summarise the 2 major responses, we can say that 86.9% lack dedicated support. This could have great bearing on some of the responses we had regarding the lack of a standardised way when recording metadata or the challenges faced when trying to deposit data in public repositories prior to publication.
- Handling large volumes of data is a shared issue. This connects to data storage and data transfer.
- Less than half use public repositories to deposit their imaging data (41%).
- 35% of respondents said there was no clear policy governing long-term data storage.
- Other challenges not listed above were: data ownership, different requirements between repositories and also funders, limited staff and recognition, technical difficulties with formats, cost of data storage, lack of studies available for commercial re-use.

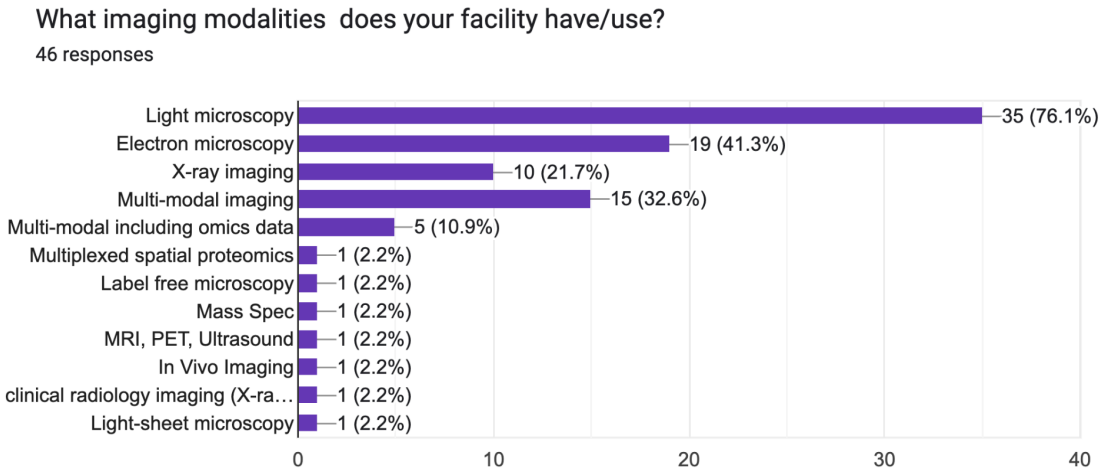
Shared needs:

- Only 40% of facilities provide any support on data management. This was one of the main training needs voiced by the respondents, alongside image acquisition, processing/conversion and analysis.
- Clear guidelines, policies, and universal standards were mentioned by a high proportion of respondents as being necessary to improve depositing data in public repositories. The lack of clear guidelines include those that come from the funders, as well as potentially the repositories themselves (though not explicit in all responses). There were 2 respondents mentioning some sort of monetary compensation would be needed to encourage researchers to share. Some respondents added training is needed to help with data deposition.

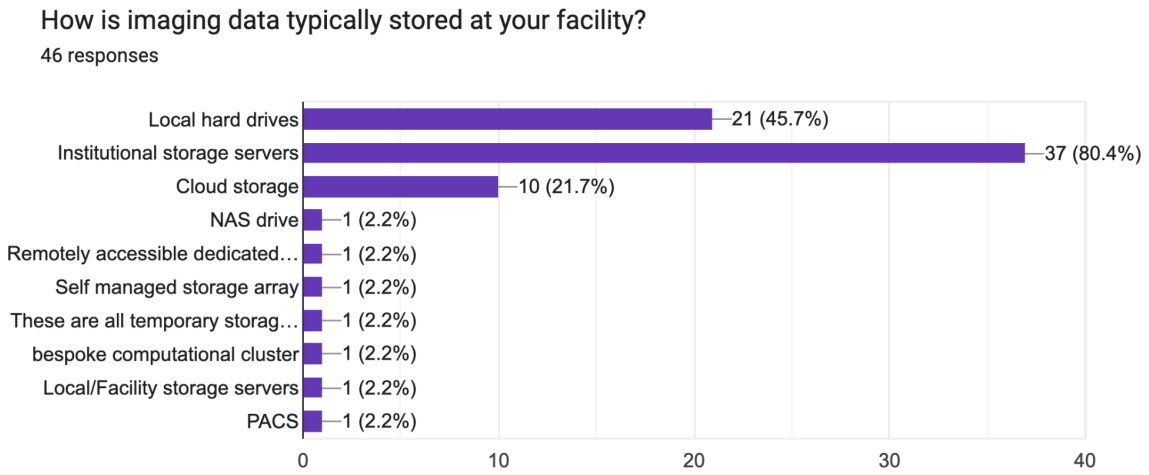
What type of biological imaging facility do you represent?



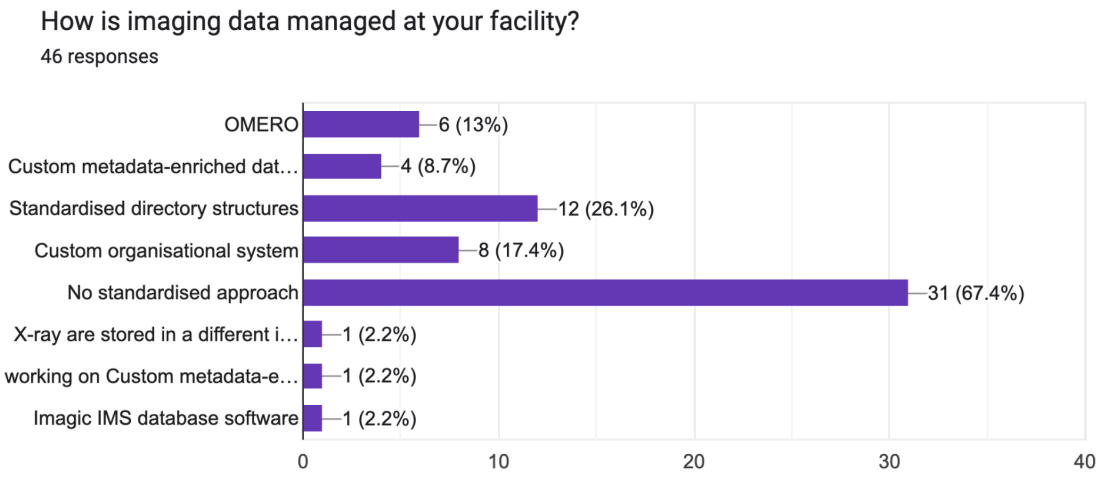
What imaging modalities does your facility have/use?



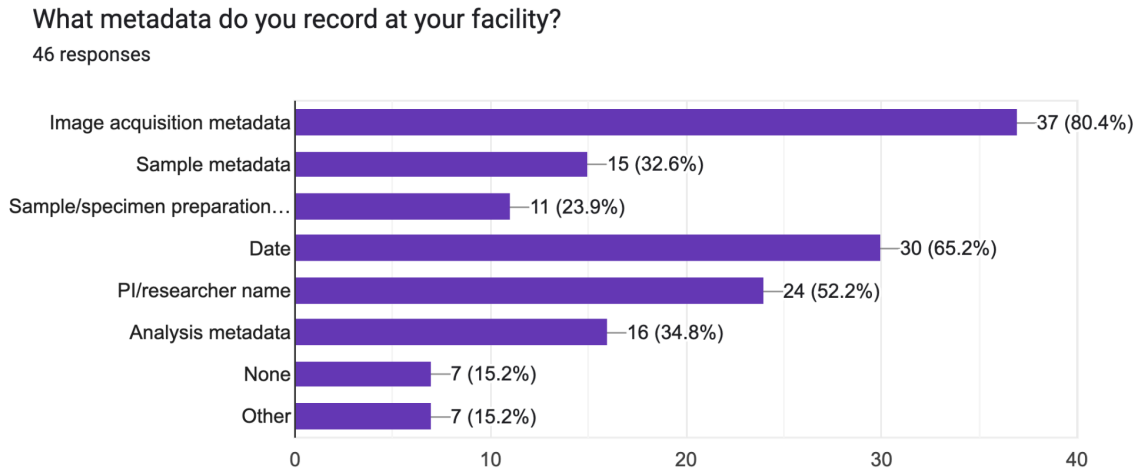
How is imaging data typically stored at your facility?



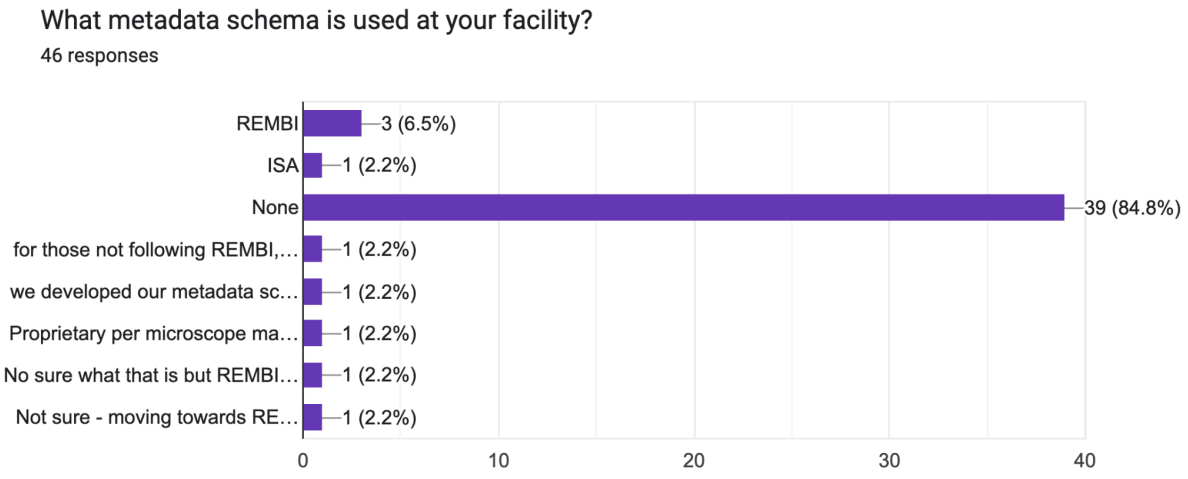
How is imaging data managed at your facility?



What metadata do you record at your facility?



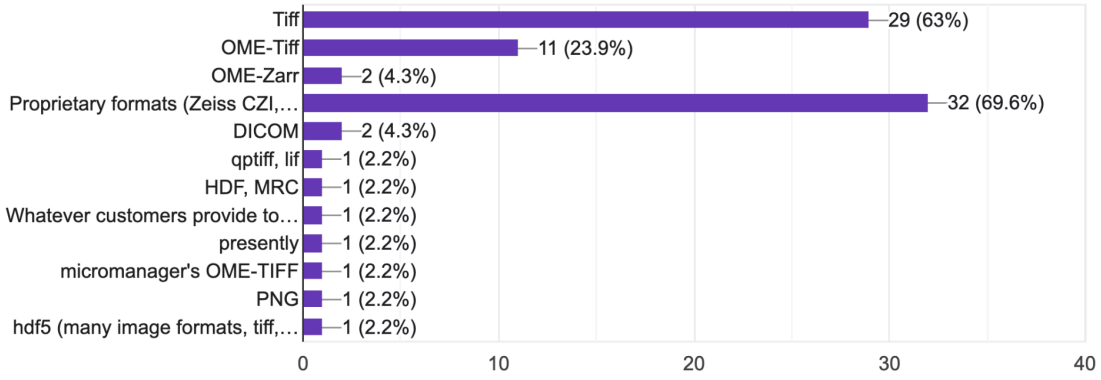
What metadata schema is used at your facility?



What file formats do you primarily use for imaging data?

What file formats do you primarily use for imaging data?

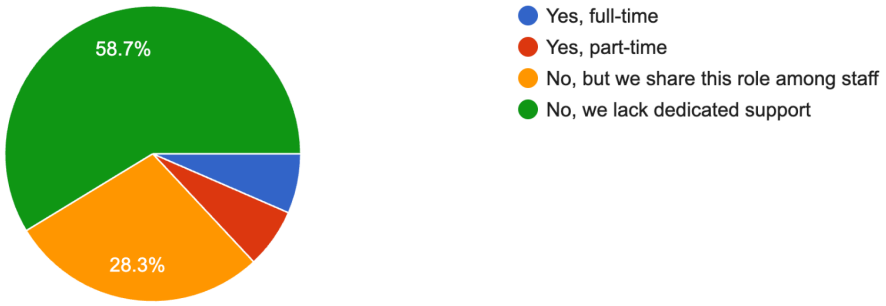
46 responses



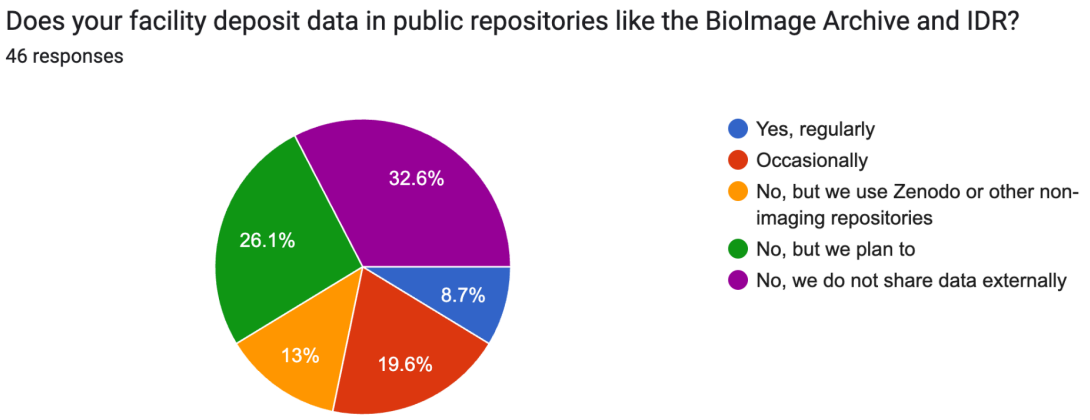
Does your facility have a dedicated data steward or data manager?

Does your facility have a dedicated data steward or data manager?

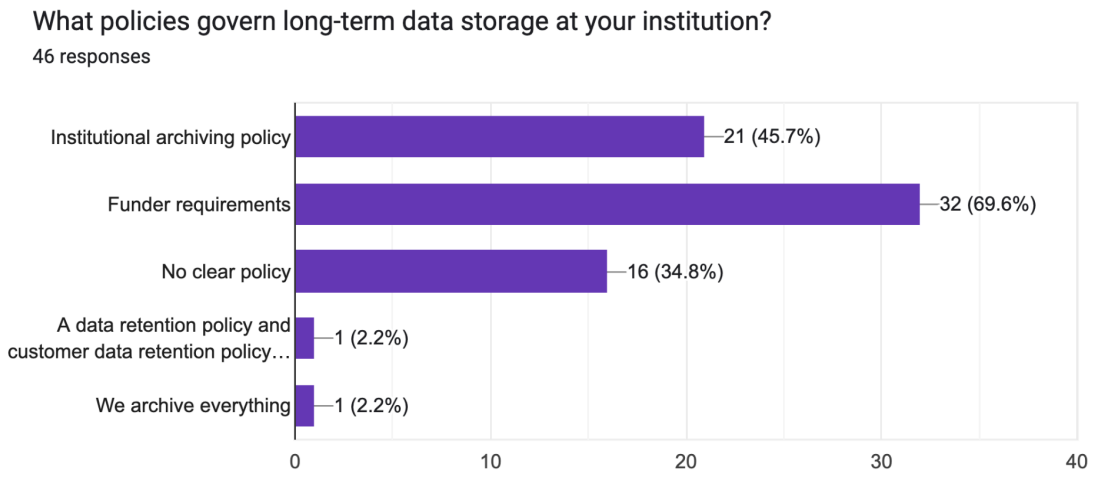
46 responses



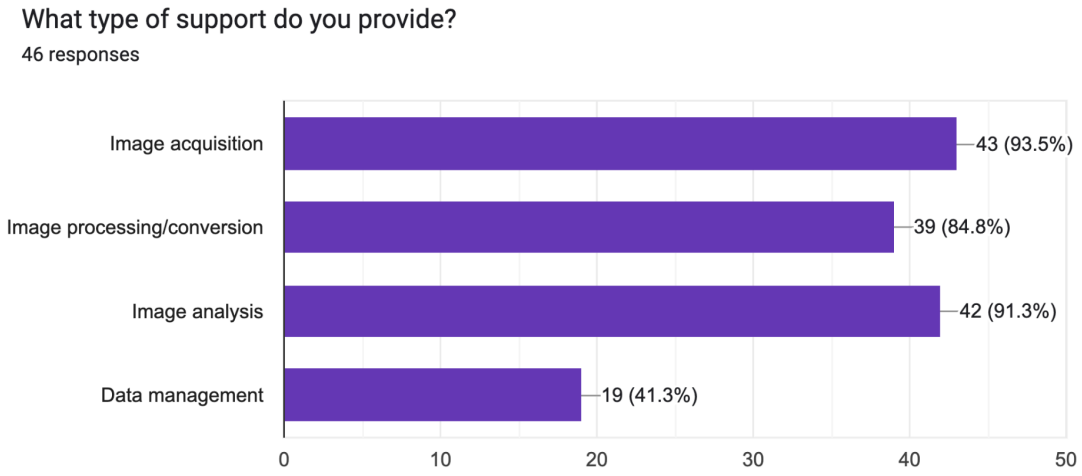
Does your facility deposit data in public repositories like the BioImage Archive and IDR?



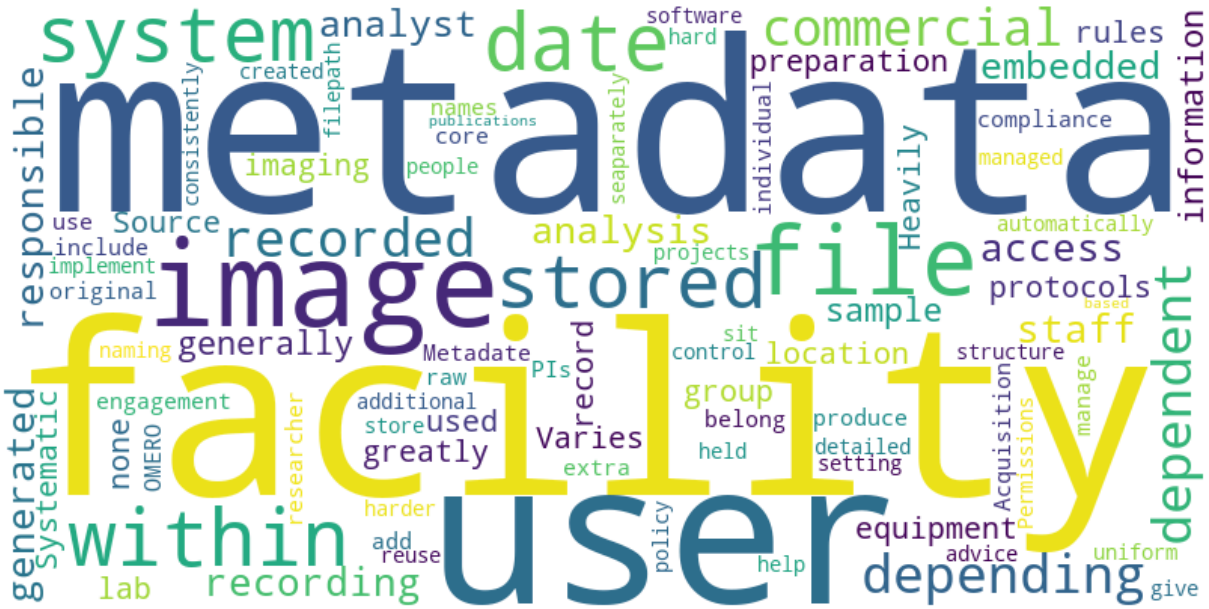
What policies govern long-term data storage at your institution?



What type of support do you provide?



What metadata do you record at your facility? If Other or None, please describe

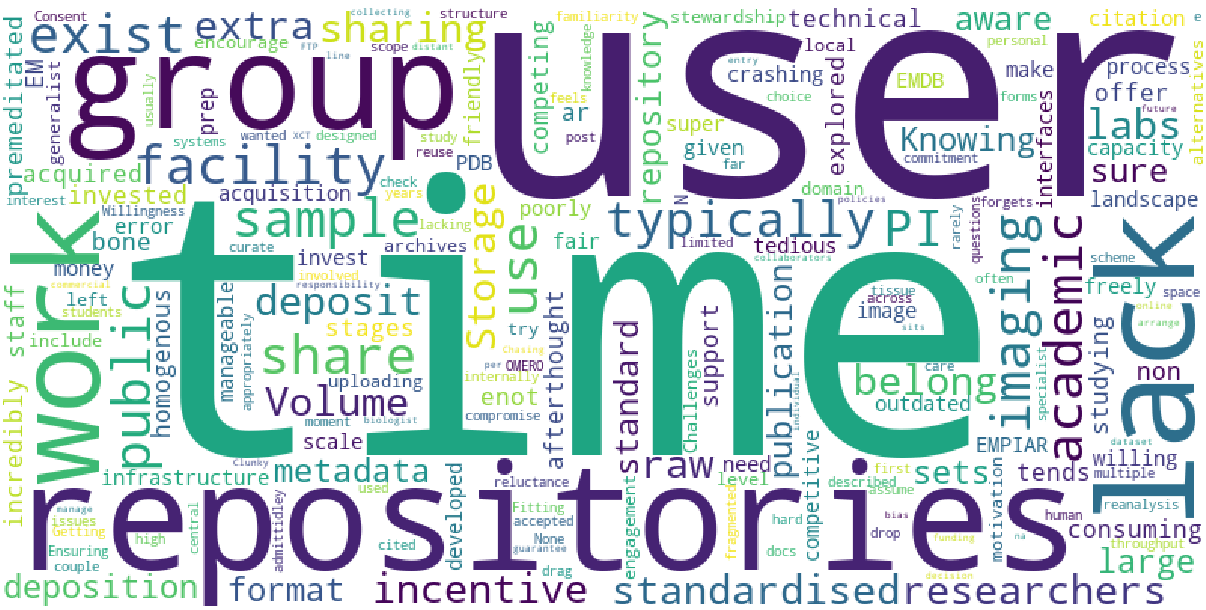


How do you capture and record metadata at your facility?

Do you use ontologies or any kind of standard vocabulary to describe your data?

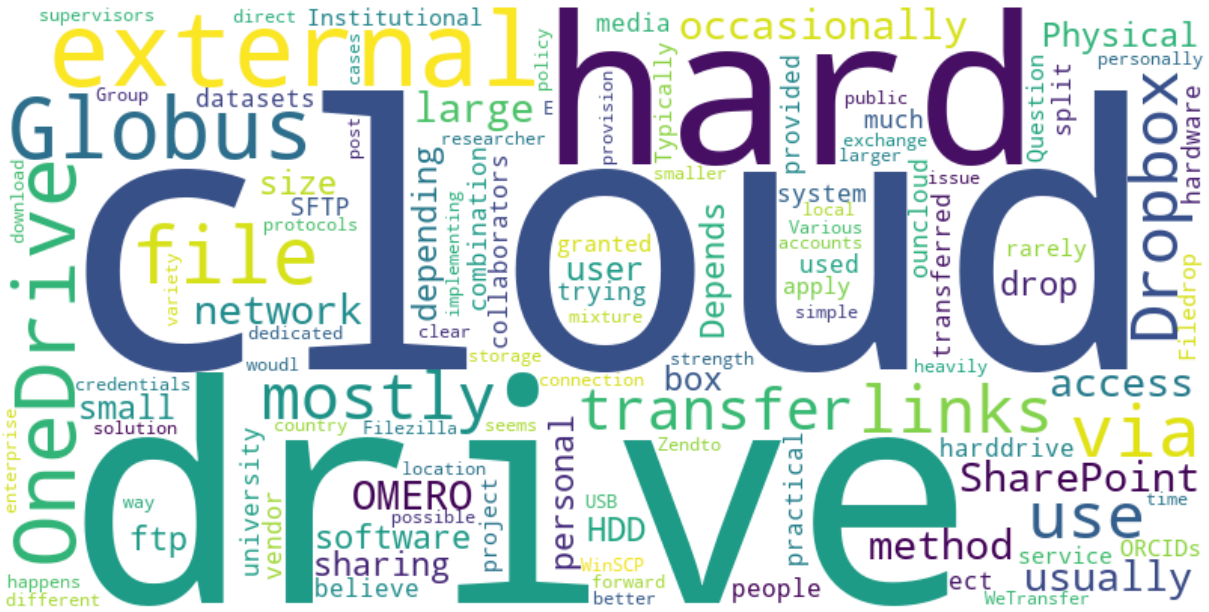
What are your biggest challenges in managing data prior to publication?

What challenges do you face in depositing data in public repositories?

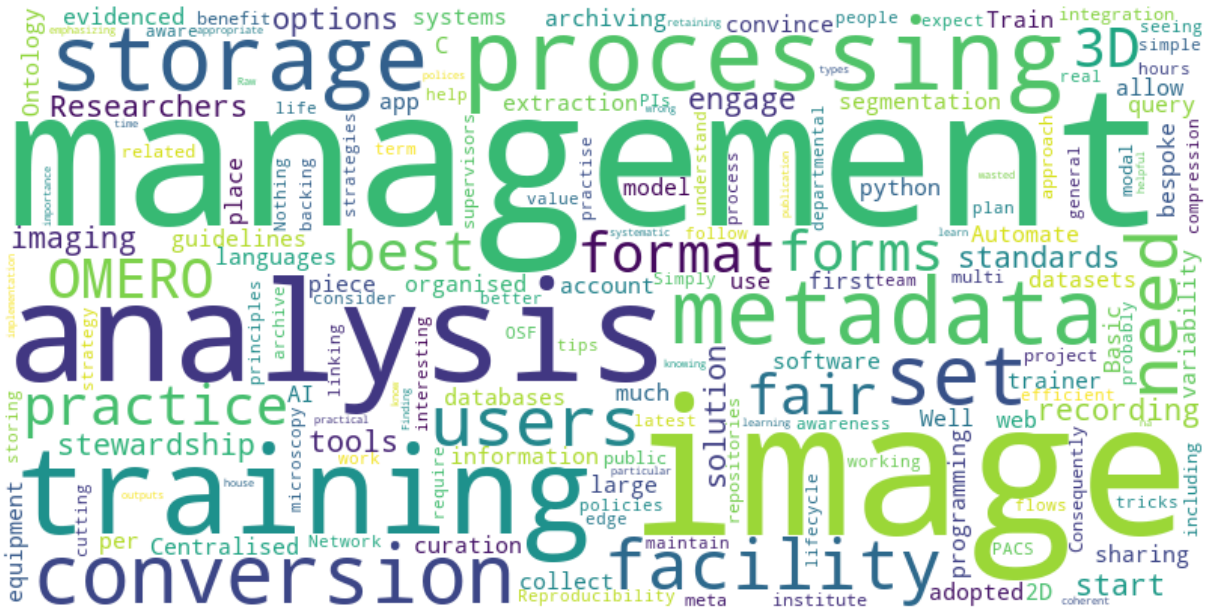


What improvements would make it easier for your facility to share data with public repositories?

Do you need to share data with collaborators outside of your institution? If so, how do you go about it?



What type of training would benefit your facility most?



Are there any other challenges you would like to mention?

Next steps based on imaging data management needs survey

Based on the survey results, we propose to provide support and training for the following:

1. Imaging Data Management & Metadata in imaging
2. Big Imaging Data Storage, Transfer and Sharing
3. Public Repository Submission
4. Train-the-Trainer Programme

1. Imaging Data Management & Metadata in imaging

Aim: to improve the consistency and quality of imaging data management, and to support the adoption of metadata standards and FAIR principles.

Important concepts: FAIR principles, use of metadata schemas, controlled vocabularies and ontologies to standardise metadata capture, organising and naming data, backup and version control, different levels of metadata, i.e. project including researcher/user, sample, image acquisition, image, analysis etc., data life cycle, Data Management Plans (DMPs and DMP tooling like Data Stewardship Wizard, RDMkit) and data management tools (OMERO, databases, ro-crate etc.)

Selling points: avoid data loss and duplication, easier transfer of projects, answering to reviewers, provenance of data tied to research integrity and reproducibility (by documenting every step of the way), good data management saves time in the long run, standardised data can be programmatically handled and analysed (AI-ready data), easier management of big or multi-modal data

Resources: develop a workshop to cover this, recent webinar series, docs on ontologies and metadata schema for imaging data (from e.g. FoundingGIDE), games?, consultations

Think about:

(i) How, what and when to capture metadata
(ii) Standardised metadata: use of ontologies, and schemas like REMBI and MIFA
(iii) How to store, manage and share metadata
(iv) Best practices and tools (incl. OMERO) in image data management
(v) practical examples of FAIR

2. Image Data Transfer, Storage and Sharing

Aim: Learn about standardised file formats and conversion strategies, storage and sharing options, and transfer tools like Aspera or Globus

Important concepts: data flow, open and cloud friendly file formats, file conversion, handling large files, data storage and sharing solutions particular to imaging data, preserving raw data as well as processed data and analysis.

Selling points: Easier to handle and transfer large files and submit to repositories as long-term storage of large files. Easy viewing and analysis of large image data.

Resources: training materials, workshops and consultations

Think about:

(i) Data flow

(ii) Standardised data formats: open source cloud-friendly formats like OME-Zarr and conversion strategies

(iii) Infrastructure (involvement of the IT department)

(iv) Remote visualisation and analysis of data

3. Public Repository Submission and Data Re-use

Aim: Convey the importance of sharing data and examples of data re-use

Important concepts: open data vs FAIR open data, metadata collection for BIA and IDR submission, licensing, embargo of data sets

Selling points: Meet funder policies on data sharing, increase potential collaborations and citations, alternative to long-term data storage, scientific collaboration

Resources: online tutorials for step-by-step guidelines on how to submit data (existing and new). Success stories

Think about:

(i) FAIR open data

(ii) Advantages of using public repositories, i.e. funder policies, long-term storage, increased recognition and collaboration

(iii) Thinking ahead: submission requirements

(iv) Submission tutorials

(v) Success stories and data re-use examples

4. Train-the-Trainer Programme

Aim: to train facilities staff so they can communicate the benefits of data management and data sharing but keeping in mind that this should feed into the researchers/PIs.

Important concepts: encourage the development of internal data policies, and to clarify points like data ownership and access, but link it to institutional and funder policies on data management and release, data provenance tied to research integrity and reproducibility, data embargo, exemplify the use of templates, DMPs

Selling points: good data management practices save time in the long run (transfer of projects, answering to reviewers), and meets funders requirements, and you can avoid data loss and duplication

Resources: Material developed for points 1 and 3 primarily.

Think about:

(i) identify why researchers resist good data management practices and data sharing (time, difficult to do, lack of clear benefit)
(ii) identify what is important to them (publications, impact, reproducibility, saving time in the long run)
(iii) Identify early adopters and data champions to showcase successful stories, and what could happen when you don't think about this
(iv) Useful metrics

Image Analysis (within the data life cycle)

There are already existing courses and resources available on this topic. Therefore this project isn't planning to address image analysis individually but its place in the data life cycle and integration of it will be addressed in the other modules.