

MMC 2025 - BioImagingUK session

Findings

Velasquez, S.M.¹, Fletcher, G.², Burel, J-M.³, Parson, M.⁴, Swedlow, J.³, Hartley, M.¹, and Yoldas Kupcu, A.¹.

¹ European Molecular Laboratory, European Bioinformatics Institute, Hinxton, UK

² Royal Microscopy Society, Oxford, UK

³ University of Dundee, Dundee, UK

⁴ King's College, London, UK

Introduction

Following the survey (10.5281/zenodo.16873139) conducted during April - May 2025, we devised a series of follow up questions and activities to be rolled out during the BioImagingUK meeting and the Microscience Microscopy Conference (MMC) 2025 that took place from 30th of June till the 03rd of July 2025. The project was introduced in four short presentations during the BioImagingUK meeting followed by an activity/discussion session. Matthew Hartley and Aybuke Yoldas gave an overview of the DRI-Skills Data Stewardship project, Melina Velasquez gave a summary of the results of the initial survey (Imaging Data Management survey) and proposed training modules and next steps. Josh Moore was invited to give an overview of the work done by NFDI4Bioimage, and to chair the discussion session.

Participants were asked 2 follow-up questions via Slido:

- Please rank the four proposed training modules in order of need (1 = most important, 4 = least important)
- Would you attend a 1-day in-person course if it requires less than a 2 hour journey? (Either type No or the name of your City if yes)

The session activities consisted of:

- Adding green, yellow and red dots to two Data Life Cycle posters located in the room, where green represented bad, yellow worse and red worst pain points/problematic areas in the data life cycle. Each participant could only add three dots (one for each pain category).
- Adding post-its to the data life cycle posters with comments about the proposed training modules, or any other suggestion that we failed to cover in our initial survey questions or proposed next steps. Participants could add as many post-its as they wanted.

During the rest of the conference, we ran a booth together with BiolmagingUK and EuroBiolmaging, where we displayed another Data Life Cycle poster and asked visitors to the booth to do the same activity; adding stickers for their pain points and using post-its for other comments or suggestions. The participants at the booth were also restricted to a maximum of three dots (one for each colour/pain category) but weren't required to add all three colours/pain categories

We had 80 participants who answered the follow-up questions and took part in the activities at the Biolmaging UK meeting and we estimate around 30 participants put up pain point stickers and/or post-it notes at our booth at the MMC, bringing the total to 110 participants.

Analysis and Results

Data Life Cycle Pain Points

Participants were asked to identify their top three pain points within the data life cycle by assigning green (bad), yellow (worse), and red (worst) dots to three large posters representing the full data cycle. Analysis of the combined responses revealed that the *Analyse* phase was the most frequently cited area of difficulty overall (Fig. 1). However, when broken down by severity, the *Preserve* stage received the highest number of "worst" (red) votes (Fig. 2). This underscores the need for more targeted support in long-term data preservation strategies which is covered in the proposed training modules 2 and 3. While image analysis training is not within the scope of this project as it is already covered by multiple existing courses and resources, the data made it clear that there are ongoing challenges and needs. Our modules will address the analysis difficulties that are related to (un)standardised metadata and data formats, or choosing models based on similar training data, in training modules 1, 2 and 3. We will also ensure to link to the appropriate training materials and courses on bioimage analysis.

The *Collect* stage attracted the most "bad" (green) votes, however overall Collect is one of the least painful stages of the data life cycle (Fig 1 and 2). *Process* and *Share* on the other hand have a higher number of total votes, both of which are covered in the training modules 2 and 3. Similarly, *Plan* got the third highest "worst" pain point votes and it is covered in training module 1.

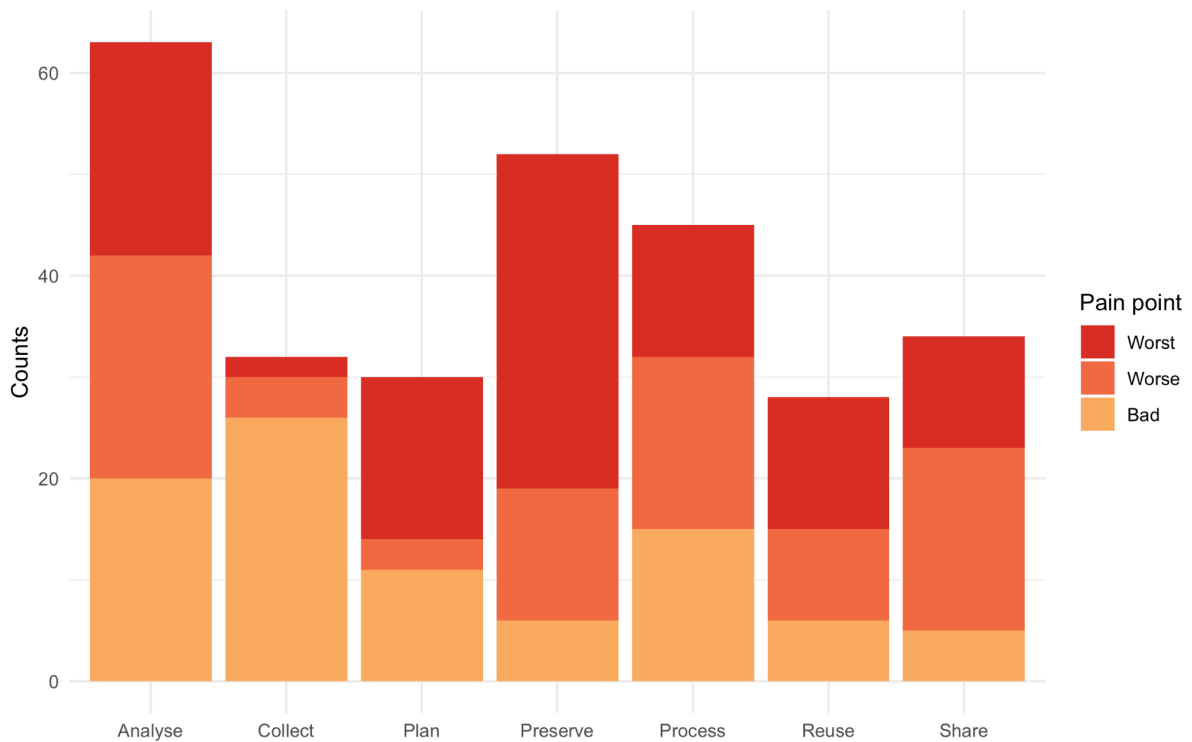


Figure 1. Bar chart showing number of dots (pain points) per data life cycle section. Participants were asked to put dots to a data life cycle poster: Bad (green dot), Worse (yellow dot) and Worst (red dot).

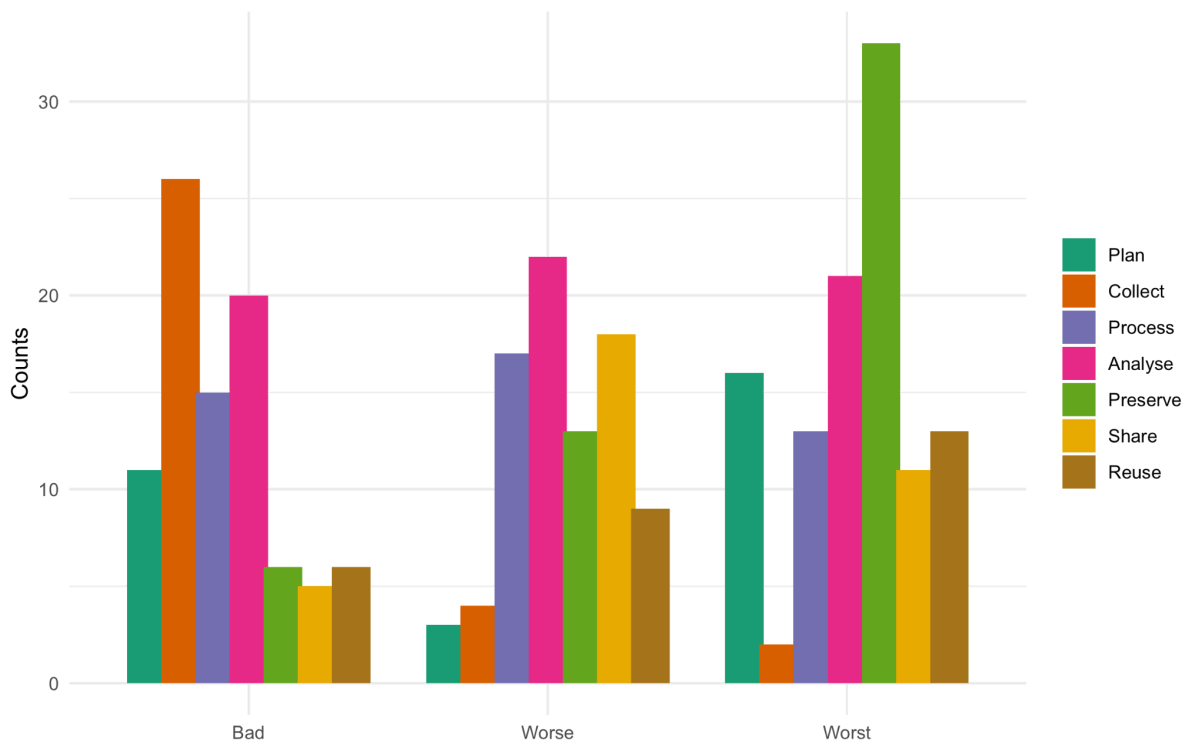


Figure 2. Bar chart showing number of dots (pain points) per pain category (Bad, Worse, Worst) subdivided by data life cycle section. Participants were asked to put dots to a data life cycle poster: Bad (green dot), Worse (yellow dot) and Worst (red dot).

Training Module Prioritisation

In the Slido activity, attendees were asked to rank four proposed training modules. The highest priority was given to *Imaging Data Management & Metadata in Imaging*, followed closely by *Image Data Transfer, Storage and Sharing* (Fig. 3). These selections reflect strong interest in improving data handling practices and ensuring metadata is captured and preserved appropriately. The *Train-the-Trainer Programme* was ranked third, indicating recognition of the need to build local capacity. Although *Public Repository Submission and Data Re-use* was ranked lowest, its relevance remains significant given that the *Preserve* phase was simultaneously identified as the most problematic. This mismatch highlights a possible underestimation of the skills required for effective data preservation and suggests that the benefits of data sharing need to be communicated more persuasively to end users.

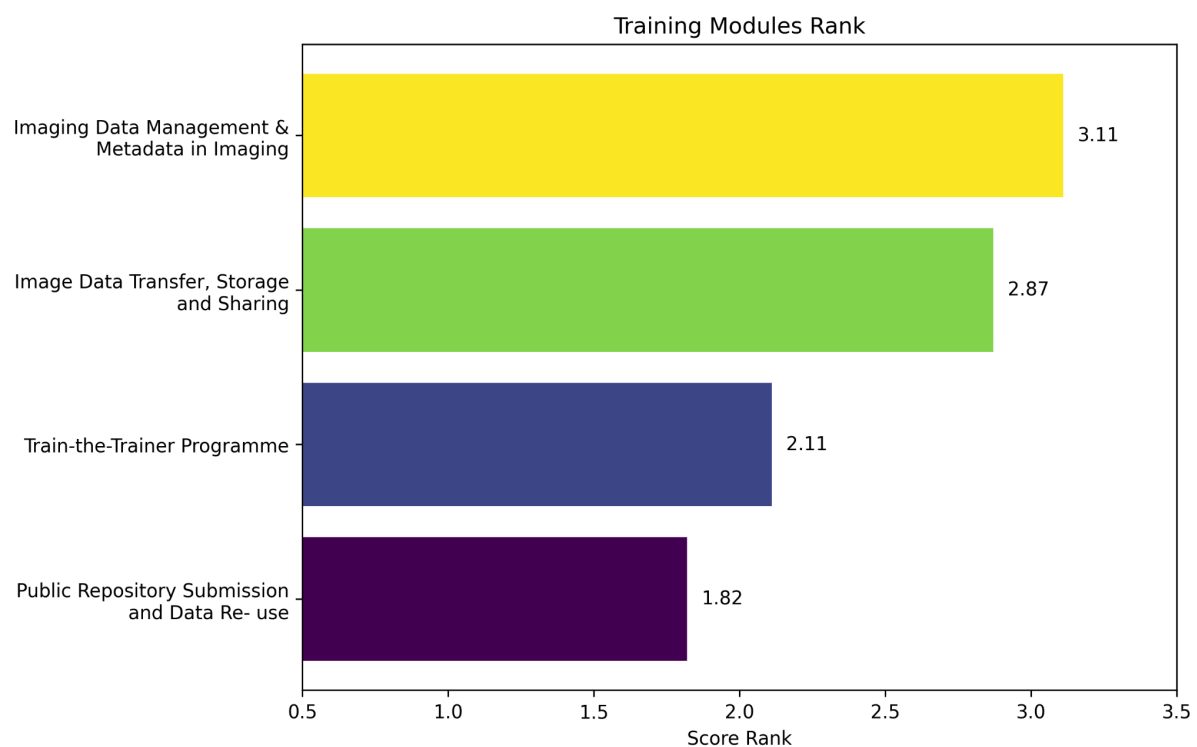


Figure 3. Bar chart showing the ranking of the training modules. Participants were asked to vote on all 4 modules: 1 being important and 4 being least important.

Analysis of the comments and suggestions

Participants provided written comments via post-its placed on the data life cycle posters. We collected 79 post-it notes throughout the conference. A word cloud analysis of these revealed analysis, management, metadata, staff, storage, training, and funding and software as repeated terms (Fig.4).

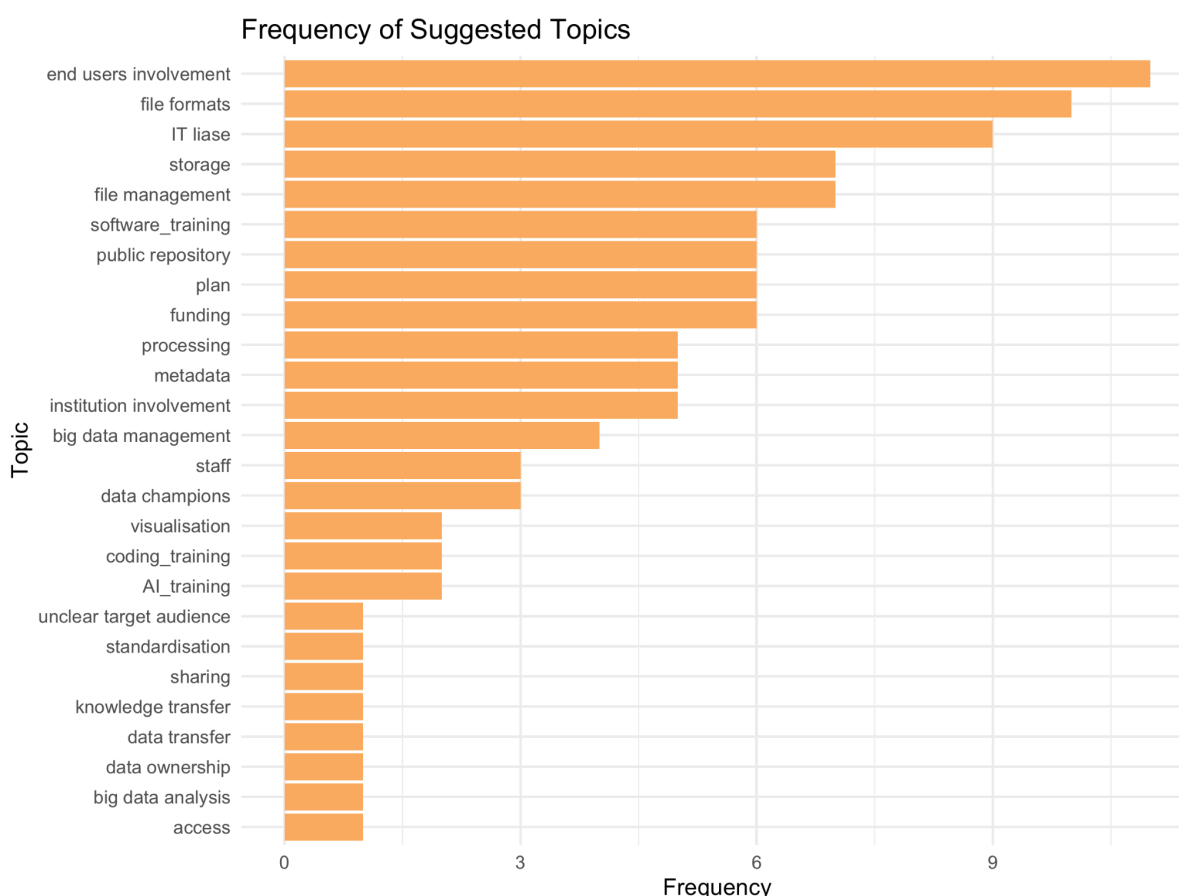


Figure 5. Bar chart showing the frequency of topics per response. One or more topics were assigned to each response.

Other highly cited themes included file management, software training (especially data management software), and storage, indicating a broad need for practical, hands-on guidance in managing and manipulating imaging data. Notably, topics such as institutional involvement, funding, and staff were also well represented, suggesting that participants view organisational capacity and resources as integral to successful data stewardship. Less frequently mentioned, though still present, were emerging areas such as AI training, data visualisation, and big data analysis, hinting at growing interest in future-proof competencies. These also underline the importance of public repositories serving standardised FAIR open data, even though Public Repository Submission and Data Re-use was ranked lowest as a training module. Overall, the distribution of topics reflects both foundational needs in data handling and a recognition of the broader institutional and technical ecosystems that support research data management.

Crucially, all thematic concerns could be mapped onto one or more of the proposed training modules (Figures 6,7,8, 9). Topics relating to institutional involvement and end-user engagement were deemed most appropriate for inclusion in the first module (Imaging data management and metadata) and third module (Public repository submission and data re-use).. These findings suggest that while the training design is broadly robust, more explicit framing around stakeholder engagement and institutional buy-in may be needed to ensure the training resonates with participants' lived challenges.



Figure 6. Word cloud showing the frequency of topics in training module 1.



Figure 7. Word cloud showing the frequency of topics in training module 2.



Figure 8. Word cloud showing the frequency of topics in training module 3.



Figure 9. Word cloud showing the frequency of topics in training module 4.

Training Delivery Preferences

When asked whether they would be willing to travel up to two hours for an in-person course, 97% of participants responded positively, indicating a very strong preference for face-to-face

Annexes

All the raw data used to generate the above charts and word clouds are in the linked spreadsheets.

[LifeCycle_dots](#) (Figs. 1 - 2)

[Ranking_modules](#) (Fig. 3)

[Other_comments](#) (Figs. 4 - 8)

[Cities](#) (Fig. 9)