Data management for image-based characterisation of 2D materials

Pascal P. Ankli[1], Asmaa Ali[1], Senid Hodzic[1], Artem Logachov[1], Krzysztof Maciejczuk[1], Andrew Milovich[1], Spela Hardy[1], Damjana Drobne[2], Sara Novak[2], Eva Kranjc[2], Francesco Pellegrino[3], Dan Hodoroaba[4], Jörg Radnik[4], Mary Gulumian[5], Elliot Jones[6], Eugenia Valsami-Jones[7], Kerstin Jurkschat[8], Meike Van Der Zande[9], Daniel Fernandez-Poulussen[10] and Barry Hardy[1]

1. Introduction

ACCORDs is a Horizon Europe project with the goal to develop an image-based characterisation (ACCORDs) framework for the holistic and correlative assessment of Graphene Family Materials (GFMs) as a representative of 2D materials to assess and predict the 2D material's risk on health and environment.

During the development of the ACCORDs framework, a data management workflow and infrastructure guarantees the organisation of data which has been used and generated throughout the project, thereby adhering to Findable, Accessible, Interoperable and Reusable (FAIR) principles. This data includes protocols, SOPs, generated images, results and collected metadata from partners, which is correlated to parameters throughout the project to ensure proper characterisation of the materials in the end. In particular, the infrastructure includes:

1. A data management and Knowledge Infrastructure (KI).
2. An image repository which allows the uploading of image data and image data sets in an organised and FAIR manner.
3. Metadata collection forms associated with the submitted data, including images, within the knowledge infrastructure.
4. An analysis tool, which characterises materials based on images.

[1] EdelweissConnect GmbH, Switzerland, barry.hardy@edelweissconnect.com
[2] University of Ljubljana, Slovenia
[3] Università Degli Studi Di Torino, Italy
[4] Bundesanstalt für Materialforschung und -prüfung, Germany
[5] North West University, South Africa
[6] Haydale LTD, UK
[7] University of Birmingham, UK
[8] University of Oxford, UK
[9] Wageningen Food Safety Research, Netherlands
[10] Fundación Idonial, Spain

This infrastructure allows the easy as well as user-friendly organisation of data within the ACCORDs project.

2. . Objectives

Our objectives are separated into the following categories:

The Knowledge Infrastructure

The knowledge infrastructure allows users to upload, share, search and download data. It includes:

- A Content Management System (CMS).
- A Database Selection using a relational database such as PostgreSQL.
- A Search Functionality.
- An Access Control using Django's authentication system.
- A User Interface using HTML and CSS.
- Continuous Integration/Continuous Deployment (CI/CD) to automate testing and deployment of updates to our knowledge base.

Image repository

We use OMERO as the image repository. It allows users to upload and organise image data, annotate metadata, access the images for analysis and to prepare them for publication.

Data collection forms

To ensure data to be FAIR, it is crucial to collect and annotate datasets with comprehensive metadata. This includes detailed annotation of image datasets to capture essential parameters effectively. To facilitate this process and enhance user experience, we have developed an easy-to-use submission form, aligning with the Recommended Metadata for Biological Images (REMBI)[2] guidelines. This approach streamlines the metadata collection process, ensuring that the datasets are both rich in information and standardised across the project.

Image analysis

For feature extraction from images, we aim to analyse these images to identify and quantify various components such as graphene, functionalized graphene and the supporting film. The analysis is designed to detect features such as particle sizes and their distribution, edge roughness of particles, porosity of materials at a granular level, particle overlap and various artefacts including overcharging artefacts, preparation artefacts and issues related to morphology characterisation.

3. Ongoing activities

Knowledge Infrastructure

The KI was established as visualised in Figure 1 and above mentioned features are being implemented.

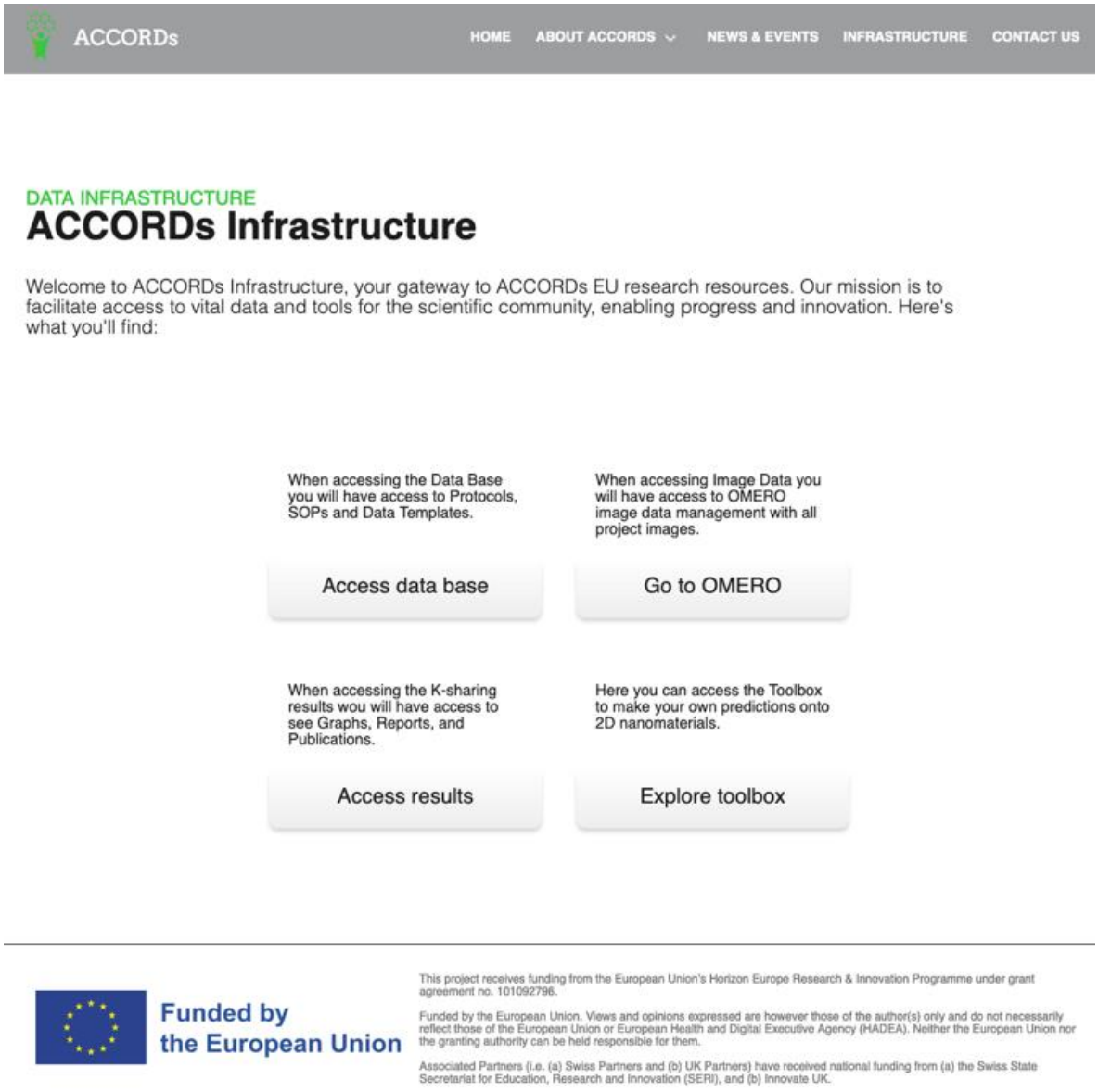


Fig. 1: The ACCORDs knowledge infrastructure portal.

OMERO image repository

The OMERO image repository is set up as visualised in Figure 2 and can be accessed through the ACCORDs KI. First datasets have been uploaded. We are currently working on standardising data to make it more «FAIR» and to implement automated uploading of images and annotated metadata.

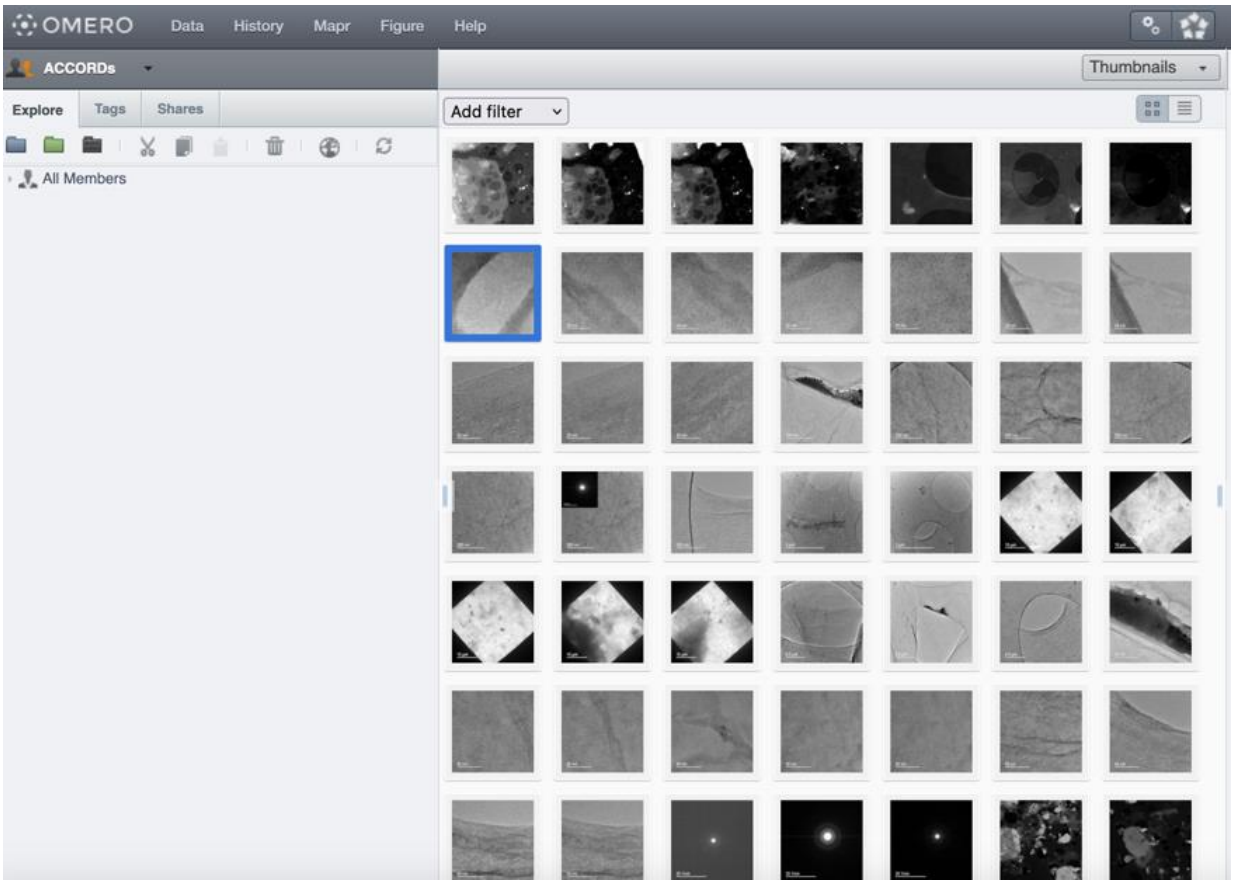


Fig. 2: Initial image dataset generated on ACCORDs and uploaded to the OMERO image repository.

Data Collection Forms

We streamlined the data collection process by implementing REMBI guidelines and incorporating partner feedback to ensure all essential details are captured. Our method involves analysing REMBI guidelines, categorising them into distinct, non-redundant entities for a database schema, and adding unique identifiers. We then structured the data collection into phased, related entries, creating Google Forms for each phase to gather partner feedback , illustrated in Figure 3. Next, we plan to integrate these forms into the ACCORDs knowledge Infrastructure for data submission by partners.

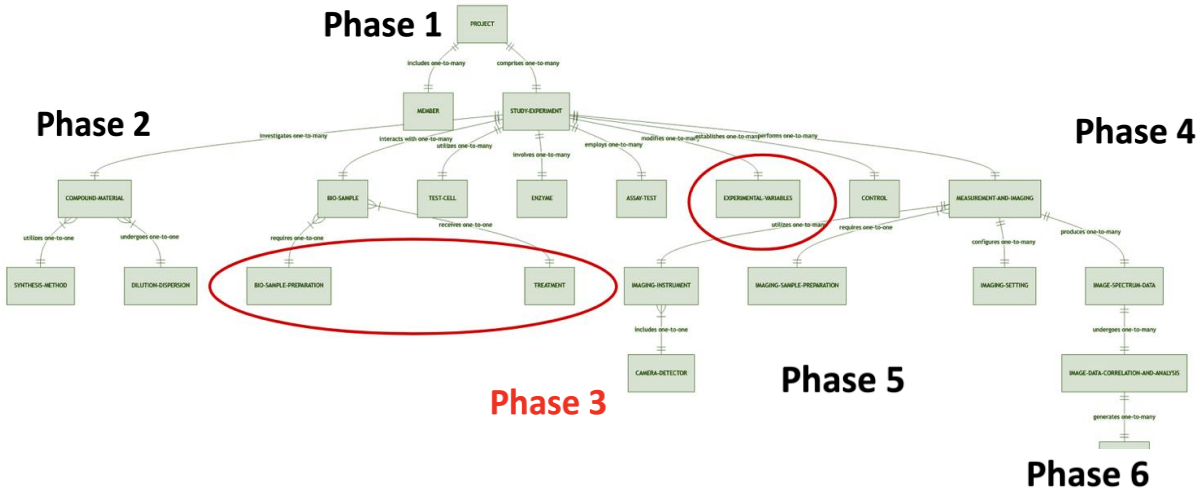


Fig. 3: Identification of the different entities throughout the project. Related entities are grouped into phases, which are then implemented in online forms.

Image Analysis

We started with thresholding methods for segmentation to create a dataset with labelled pixels, facilitating the training of machine learning segmentation models. This technique is valuable for developing a small, labelled dataset for trainable segmentation, utilising pixel labelling techniques and region labelling techniques. In scenarios where the dataset is unlabeled or when facing increased complexity, unsupervised deep learning becomes a viable option. This approach requires a substantial number of example images but eliminates the need for pre-labeled data. On the other hand, for more structured and detailed analyses, we plan to adopt supervised deep learning, which depends on a large repository of labelled data for effective training. Examples of results from the ongoing image analysis work are illustrated in Figures 4 and 5.

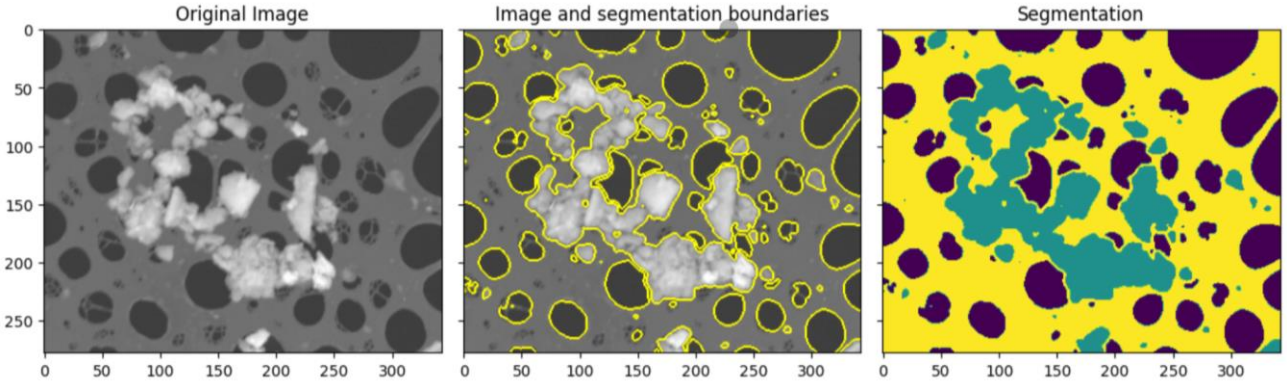
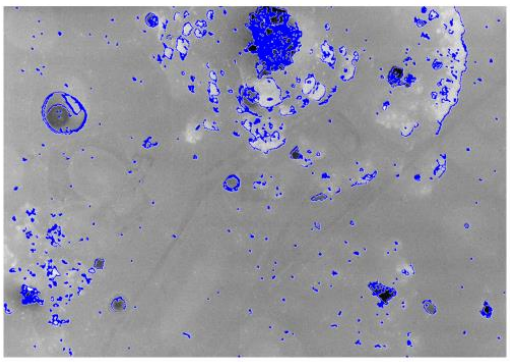
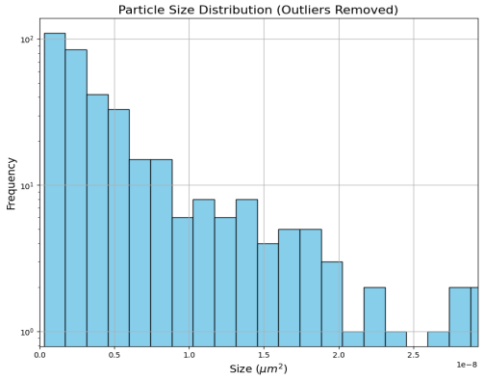


Fig. 4: Results from combining pixel labelling and region labelling techniques using thresholding approaches. Subsequently, trainable segmentation was used to label all pixels in the image to identify the different components within the image.






```
Image: AC_UT_GO_02-01.tif | Magnification: 1KX
Mean size: 1.3698655913978494e-08
Median size: 2.9000000000000003e-09
Standard Deviation: 9.278423533203485e-08
Range: 1.5988000000000003e-06
Mode: 1.4000000000000001e-09
Coefficient of Variation: 6.773236433901169
25th, 50th, 75th, 90th Percentiles: [1.600e-09 2.900e-09 6.725e-09 1.638e-08]
Skewness: 14.980263689027604
Kurtosis: 239.10658261317715
Geometric Mean: 3.768411030672972e-09
Geometric Standard Deviation: 3.026247069018761
```

Fig. 5: Analysis of Scanning Electron Microscope (SEM) images using an automatic thresholding technique. This method finds a threshold that minimises the intra-class intensity variance or, equivalently, maximises the inter-class variance, to detect flakes and obtain the particle size distribution within the image.

4. Expected Outcomes

The outcome of the work will be a knowledge infrastructure where documents such as protocols, SOPs and other data can be uploaded, searched for, shared and downloaded in a FAIR way. Data will be annotated to images which are stored in the OMERO image repository where they can be processed for publication and accessed for analysis. An overview over the workflow and infrastructure is given in Figure 6.

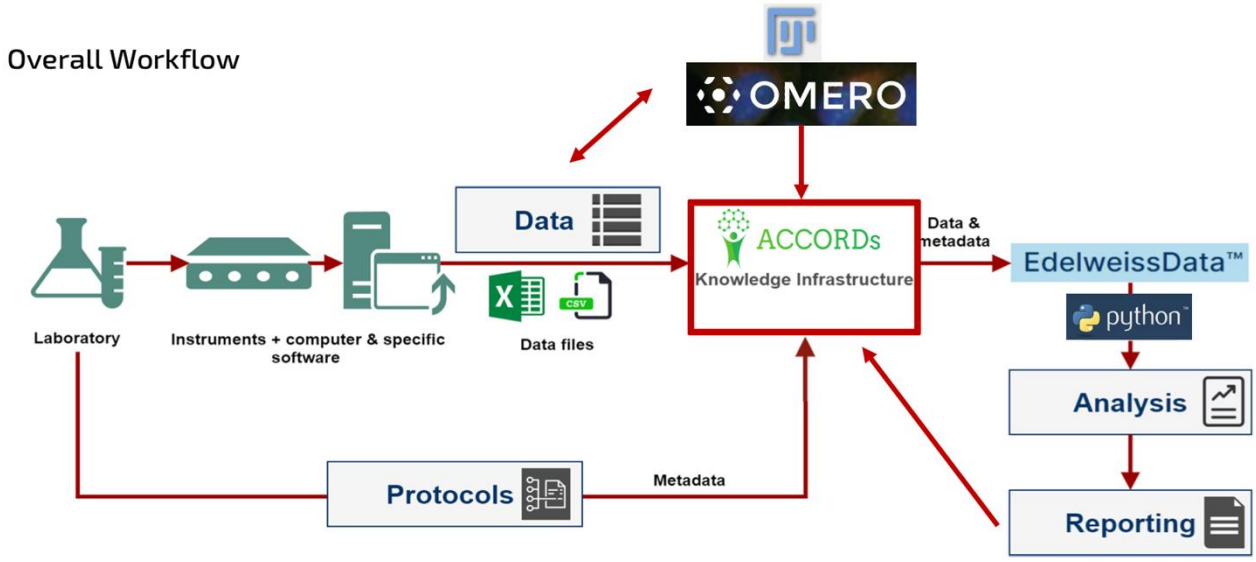


Fig. 6: ACCORDs workflow and infrastructure: Data is collected and organised in a knowledge infrastructure based on FAIR principles. It is related to images stored in OMERO as image repository. Data will also be accessible with computational tools for analysis and reporting.

5. Acknowledgements

This project receives funding from the European Union's Horizon Europe Research & Innovation Programme under grant agreement no. 101092796.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency (HADEA). Neither the European Union nor the granting authority can be held responsible for them.

Associated Partners (i.e. (a) Swiss Partners and (b) UK Partners) have received national funding from (a) the Swiss State Secretariat for Education, Research and Innovation (SERI), and (b) Innovate UK.

6. References

(1) Sarkans et al.. REMBI: Recommended Metadata for Biological Images—Enabling Reuse of Microscopy Data in Biology. *Nat Methods* **2021**, *18* (12), 1418–1422. https://doi.org/10.1038/s41592-021-01166-8.