





RESEARCH ARTICLE OPEN ACCESS

Towards Defect Phase Diagrams: From Research Data Management to Automated Workflows

Khalil Rejiba¹  | Sang-Hyeok Lee¹  | Christina Gasper¹  | Martina Freund¹  | Sandra Korte-Kerzel¹  | Ulrich Kerzel² 

¹Institute of Physical Metallurgy and Materials Physics, RWTH Aachen University, Aachen, Germany | ²Faculty of Georesources and Materials Engineering, RWTH Aachen University, Aachen, Germany

Correspondence: Ulrich Kerzel (ulrich.kerzel@rwth-aachen.de)

Received: 31 October 2025 | **Revised:** 2 March 2026 | **Accepted:** 2 March 2026

Keywords: automated workflow | defect phase diagram | electronic laboratory notebook | FAIR data | research data management

ABSTRACT

Defect phase diagrams provide a unified description of crystal defect states for materials design and are central to the scientific objectives of the Collaborative Research Centre (CRC) 1394. Their construction requires the systematic integration of heterogeneous experimental and simulation data across research groups and locations. In this setting, research data management (RDM) is a key enabler of new scientific insight by linking distributed research activities and making complex data reproducible and reusable. To address the challenge of heterogeneous data sources and formats, a comprehensive RDM infrastructure has been established that links experiment, data and analysis in a seamless workflow. The system combines: (1) a joint electronic laboratory notebook and laboratory information management system, (2) easy-to-use large-object data storage, (3) automatic metadata extraction from heterogeneous and proprietary file formats, (4) interactive provenance graphs for data exploration and reuse and (5) automated reporting and analysis workflows. The two key technological elements are the openBIS electronic laboratory notebook and laboratory information management system, and a newly developed companion application that extends openBIS with large-scale data handling, automated metadata capture and federated access to distributed research data. This integrated approach reduces friction in data capture and curation, enabling traceable and reusable datasets that accelerate the construction of defect phase diagrams across institutions.

1 | Introduction

The field of materials science and engineering is undergoing a fundamental transformation driven by data-centric methodologies and digitalisation, which enable deeper insights by integrating complex, multidimensional datasets from experiments and simulations [1, 2]. Crystal defects, including point, line and planar defects, such as vacancies, solute atoms, dislocations, stacking faults and grain boundaries, play a decisive role in determining strength, ductility and corrosion resistance [3]. Defects can adopt different structural and chemical states depending on environmental factors, with transitions that resemble phase changes [4]. This motivates the concept of a defect phase, which

is a structurally and chemically distinct atomic-scale configuration that is spatially confined to the vicinity of a defect, such as a grain boundary or dislocation [5–7]. Within this context, the Collaborative Research Centre (CRC) 1394 is developing defect phase diagrams that are analogous to thermodynamic bulk phase diagrams and are designed to allow us to harness distinct defect states to engineer materials with tailored properties.

Realising such defect phase diagrams, which give the atomic-scale composition and structure of the most stable zero-dimensional (0D), 1D or 2D defect states of observed lattice defects as a function of chemical potential (Figure 1), requires the comprehensive integration of experimental and computational data across diverse instruments and methods. This introduces challenges in

 This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Advanced Engineering Materials* published by Wiley-VCH GmbH.

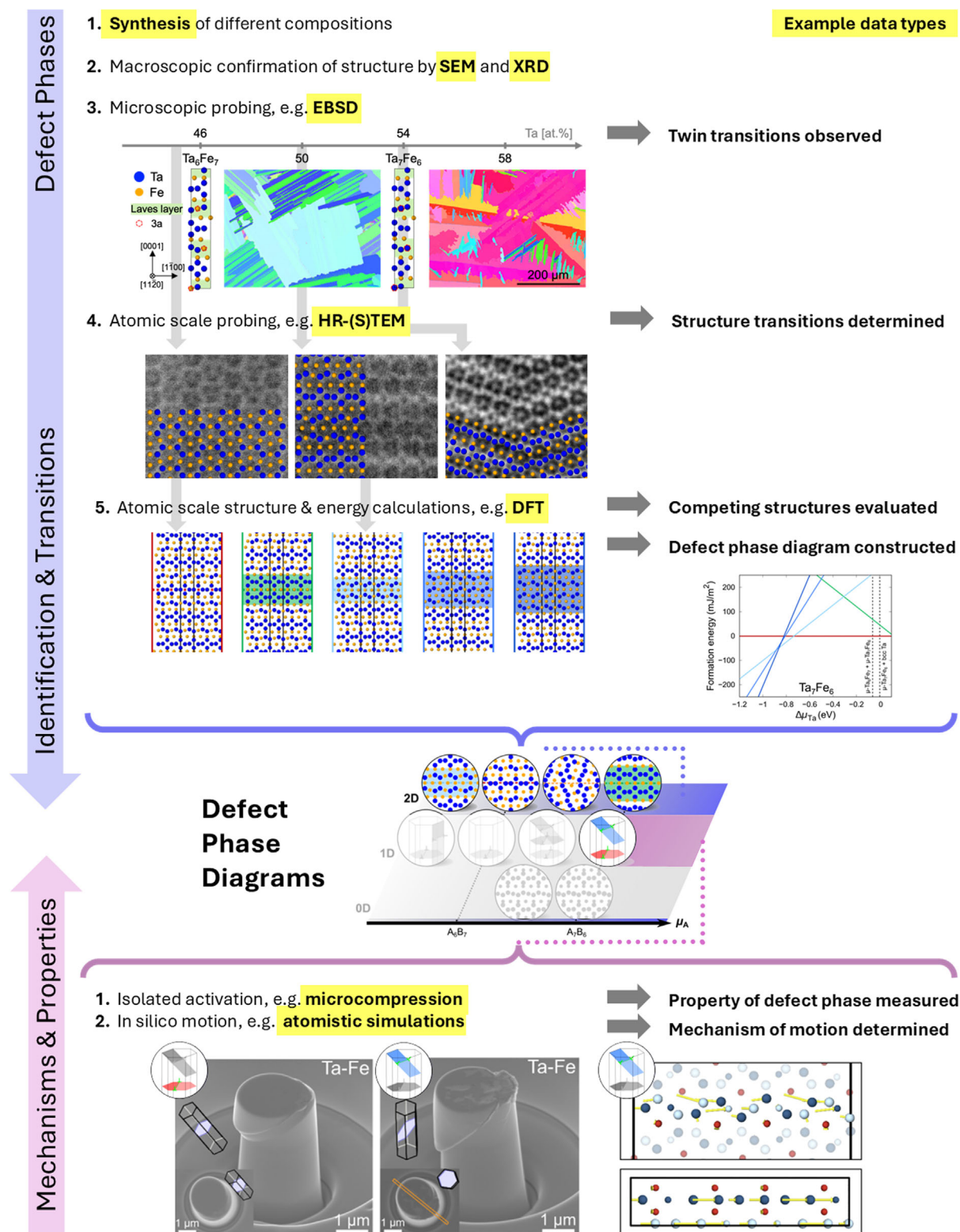


FIGURE 1 | Examples of the fundamental steps to determine defect phase diagrams and connect them to mechanisms and properties for materials design by joint experimental and computational studies. Here, defect phases of different dimensionalities are considered as a function of chemical potential in μ -phases with the two stoichiometric compositions A_6B_7 and A_7B_6 as reference points. For a given type of defect, the atomic-scale structure of the most stable configuration—the “defect phase”—changes, allowing the prediction as well as tailoring of the dominant lattice defects and the resulting material properties in alloys. Images reprinted from [8] and [9], originally published under a CC BY 4.0 licence.

data heterogeneity, provenance and reproducibility: Each step in experimental workflows—from sample synthesis and preparation to characterisation—uses diverse instruments that generate a wide

range of proprietary file formats. Computational methods increase the complexity further, as data from atomistic simulations to continuum modelling need to be included [4, 10].

In such large-scale, interdisciplinary collaborations, common practices such as local storage, transfers via physical media or email exchanges often impede systematic documentation of data provenance, experimental details and simulation parameters, thereby constraining collaboration. These limitations motivate the need for a robust research data management system that integrates heterogeneous data sources, supports automated workflows, preserves end-to-end provenance and enables reproducible analyses across distributed research groups. The National Research Data Infrastructure for Materials Science (NFDI-MatWerk) [11] was established to address the challenges of data-intensive and data-driven materials science as a concise, national initiative. Rather than a centralised ‘top-down’ approach, NFDI-MatWerk pursues a ‘bottom-up’ initiative where so-called ‘infrastructure use-cases’ develop research data management solutions that can then be made available to the materials science and engineering community aligned with the FAIR principles (Findable, Accessible, Interoperable and Reusable) [12].

In this work, we describe the solution developed and implemented by CRC 1394, as an infrastructure use-case IUC04 in NFDI-MatWerk, that allows the connection of heterogeneous data and sources crucial to our scientific goals, and results in FAIR data contributed to the scientific community. The process of creating defect phase diagrams has often been shaped by different workflows and involved varied methods of data exchange, making integration challenging [4, 8, 10]. The approach developed in this work allows researchers to collect, share and process data across the collaboration, and introduces automated workflows, thereby allowing researchers to handle the data more efficiently. Ultimately, this infrastructure will support the creation of defect phase diagrams from a wide range of experimental and simulated data.

1.1 | Requirements for Research Data Management Systems in Collaborative Environments

As the very first step, we need to consider the requirements for a research data management system, particularly in a collaborative environment. In this setting, these systems are pivotal to the digital transformation in materials science and engineering. They help researchers manage, organise and share data across a project’s life cycle. Most available systems either provide an electronic laboratory notebook (ELN) or a laboratory information management system (LIMS). However, researchers need both components to track the complete analysis chain end-to-end, from sample creation to final publication, including physical assets, instruments, consumables, standardised preparation steps and experimental methods. An ELN replaces paper-based records and enables systematic documentation of experiments and results. A LIMS monitors the usage of scientific instruments, tracks physical samples, records consumables such as chemicals used in experiments and preparation, and stores standardised methods.

Building on these basic components, we identify key requirements as follows:

- Tracking of physical samples and performed analyses

- Physical or computational samples from synthesis (or sample acquisition) to final analysis, including derived and intermediate samples,
- Processing steps applied to samples or used for their creation,
- Tools employed, such as scientific instruments and software,
- Data acquired from devices and generated in further analysis steps,
- Consistent extraction of metadata for efficient and repeatable analysis workflows,
- Scalable and collaborative data storage allowing collaborators to share data,
- Intuitive graphical user interface (GUI) with extensive automation,
- Role-based access control and audit trails that satisfy institutional data governance,
- Efficient publication of relevant and curated research data according to FAIR data principles.

Several research data management systems are available as commercial and open-source solutions. In this work, we focus on open-source systems because they can be operated and maintained by the community and avoid recurring licencing costs that are often difficult to include in funding proposals.

Within the open-source landscape, we evaluated Coscine [13], PASTA-ELN [14], eLabFTW [15], NOMAD Oasis [16], and openBIS [17]. At the time of our decision (end of 2022), only openBIS aligned reasonably well with the combined requirements of an ELN and a LIMS as outlined above. OpenBIS is developed and maintained by ETH Zurich as open-source software with a large user community across the European research landscape [17–23]. Originally employed in biological research, it now has a substantial user base in materials science and engineering, including national research institutions such as the Federal Institute for Materials Research and Testing in Germany [19] and the Swiss Federal Laboratories for Materials Science and Technology in Switzerland [23]. This broad community helps address the specific needs of materials science and engineering and benefits the software’s development through cross-disciplinary exchange.

1.2 | Contribution of This Work

This work describes how we built a comprehensive research data management system for the CRC 1394, with a focus on experimental workflows.

In Section 2, we outline the boundary conditions, challenges and opportunities in a heterogeneous, distributed research environment, including diverse instruments, proprietary file formats and multi-institutional data governance policies. In Section 3, we show how openBIS was customised for materials science by extending schemas for samples, instruments, protocols and datasets; introducing controlled vocabularies and validation scripts; and implementing QR code-based sample tracking to preserve end-to-end provenance. We illustrate these capabilities with case studies in scanning electron microscopy (SEM) and

In this way, information is gathered across a range of specimens of decreasing size from an original casting, and different methods are used jointly to extract data on structure and properties with correlated conditions under which these observations took place. Together, this ensemble of physical specimens, acquired data and their connections allows us to extract relevant information for the creation of a defect phase diagram and mechanism/property maps—if we can record, store and access all other experimental and computational data sources and methods (density functional theory, molecular dynamics, mathematical image analysis, machine learning,...) relating to the same set of thermodynamic equilibria.

The diagram in Figure 2 distinguishes equipment and consumables with light blue borders (e.g., microscopes, preparation equipment, abrasives and lubricants), emphasising the importance of recording device settings, attachments and materials used to ensure reproducibility and comparability across institutions. Orange-bordered elements denote examples of data products generated along the workflow, such as micrographs, orientation maps, spectra and mechanical response curves.

The devices and experimental setups used by each team are distributed across institutions. Each group employs specialised and often highly individualised instruments for preparation and analysis. Each device has unique attachments, capabilities, limitations, calibration states and setup requirements, often coupled with specific software versions and environmental conditions. Central documentation of these details ensures that variations between experiments driven by device settings and configurations can be tracked and accounted for, thereby enhancing reproducibility and comparability across institutions.

2.1.2 | Data and Information Technology

One of the main challenges in multi-institutional collaboration is that research groups often employ similar experimental and computational methods but use different instruments and software environments. This diversity in hardware, analysis tools and computational infrastructure results in non-uniform or non-interoperable data formats, metadata structures and access protocols, which complicate seamless data exchange and joint analysis. Institution-specific data governance policies further constrain data management and sharing across sites.

To address these challenges, the research data management system must provide a unified framework that accommodates all researchers regardless of location or local regulations. It should integrate all data formats used by the collaboration and handle them consistently, shifting the burden of data availability and exchange from individual researchers to the system. Ideally, researchers interact with the system through common interfaces and workflows to contribute and access data, independent of data origin or local computing environment.

2.2 | Large-Scale Data Storage

One of the main challenges of establishing an integrated LIMS is to incorporate and handle all data recorded from the scientific instruments, make them accessible for the subsequent data analysis and track the provenance of the data and results throughout the life cycle of the entire process.

This challenge has two major aspects: the technical setup and user access. On the technical side, the storage system must accommodate multiple terabytes of data. While data from nano-indentation may be relatively small, techniques such as SEM, EBSD, TEM and APT produce substantial datasets; for example, EBSD measurements can reach hundreds of gigabytes per file.

The storage system should be resilient against failure, provide backups and ideally be geo-redundant to safeguard against catastrophic data loss. On the researcher side, the data storage system needs to be available across a wide range of devices, from the scientific instruments from which the data are uploaded, to the individual analysis workstations. The data storage system needs to be accessible for all researchers, regardless of their location and their IT setup as discussed in Section 2.1.2.

While openBIS allows the integration of a large storage system as part of the instance setup, the requirements are best met by Coscine [13]. The Coscine project was established in a collaboration of IT centres from multiple universities in the state of North Rhine-Westphalia (NRW), Germany. DataStorage.nrw (previously RDS-S3) is one of the provided solutions and is built on S3 storage technology [26]. This service offers storage capacities of up to 125 terabytes per project and per application, backups and geo-redundant storage.

While openBIS does not natively support the S3 storage technology, it does support the concept of 'Linked Data'. In this model, files remain in an external system, and openBIS stores links and metadata describing the content and where to find it. Because this mechanism was developed in a different context, further integration is required to use it effectively in our setup; we address this in the companion application described in Section 4.2.

3 | Customising OpenBIS for Materials Science

3.1 | Initial Setup and Custom Objects

The default installation of openBIS provides a wide range of tools to describe and represent all stages of a scientific experiment, from sample synthesis to the experiment itself, and the subsequent analysis. However, it does not include domain-specific metadata tailored to the materials science community. Therefore, the generic openBIS setup needs to be configured and extended to reflect the concepts, workflows and standards specific to the concrete environment it is used in. As a first step, all relevant object and dataset types are defined together with the fields to be completed by the researchers or populated through automated processes [27]. This process is time-intensive in a collaborative setting, where a joint set of definitions and terminology must be agreed upon as an important initial step, but it also provides an opportunity to engage not only internally but also with the larger materials science and engineering community to harmonise terminology and to prepare links to emerging knowledge graphs and ontologies.

Figure 3 shows a tailored openBIS instance: the user-facing ELN and LIMS interface (a) and the administrator view of object types (b). Researchers can reuse the inventory of instruments, methods and consumables by linking the relevant parts to their ELN entries. This creates a comprehensive record of the work and

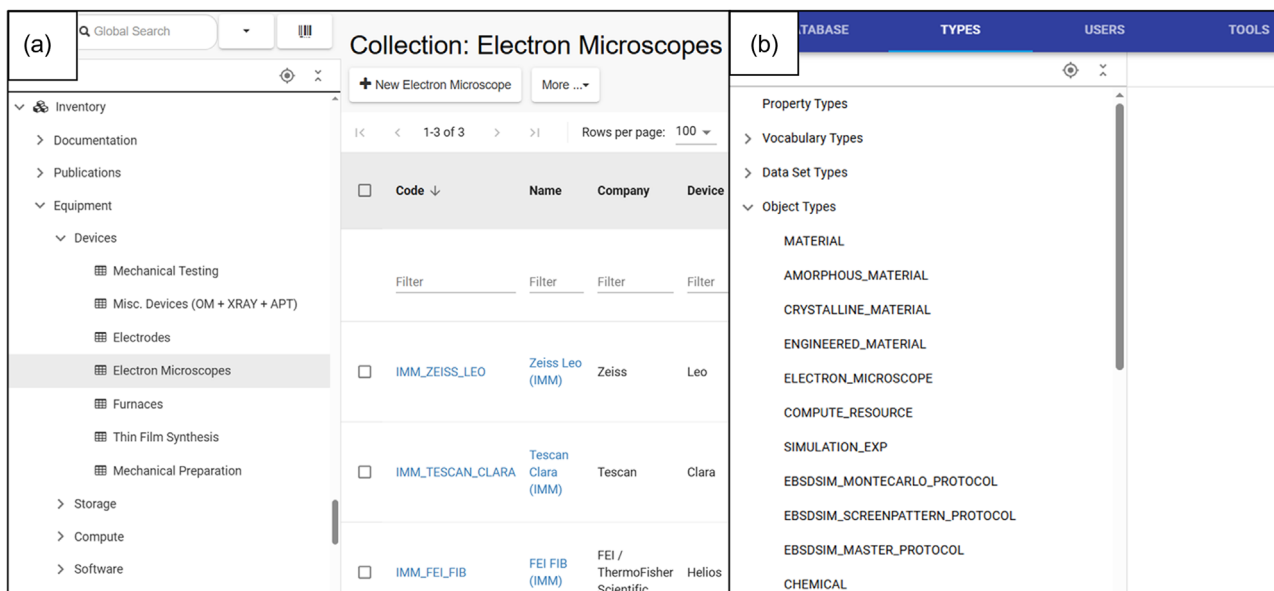


FIGURE 3 | Tailored openBIS setup for materials science. (a): ELN-LIMS graphical interface highlighting the (reusable) inventory of instruments, methods and consumables visible to the researcher. (b): Examples of relevant object types displayed in the Admin GUI.

reduces manual data entry, while preserving provenance and enabling automation based on structured metadata.

3.2 | Sample Tracking

In the physical world, experiments are centred around the physical specimen that is used and the instruments or methods it is subjected to. Similarly, from a research data management perspective, samples must also be addressed in two ways. First, the physical sample needs to be tracked and linked to the steps, data and results produced during the experiment (this section). Second, the sample must be represented in the ELN and LIMS and linked to its physical counterpart (Section 3.3).

In openBIS, each object is associated with an immutable identifier (*permId*) that can be rendered as a QR code label. By creating an object for each sample and attaching the QR code label to the physical specimen, researchers link the physical item to its electronic record and can retrieve provenance via mobile scanning in the openBIS interface. This enables tracking as samples move between groups and supports later archiving. Each entry captures key metadata such as dates and comments, strengthening the overall traceability and interpretability of the stored information.

A key part of the scientific work in the CRC 1394 is to compare and connect experimental results with simulation studies. Computational and experimental data are produced under different conditions and assumptions. The infrastructure represents each category independently, along with its corresponding contextual metadata, leaving the assessment and analysis to the researcher. To enable consistent handling, a shared digital representation is used for both experimental and computational samples, combining compulsory fields common to both with optional fields that preserve type-specific information.

We define this Object Type as follows: the schema includes compulsory fields, such as dimensions and location, so that the server rejects entries when critical information is missing. Composition is stored as separate fields, one per chemical element and

validation scripts perform automated data quality checks; for example, the total composition must sum to 100%.

Establishing a common terminology across research groups enables a shared language, which, in turn, makes sharing data and establishing workflows easier across groups. Such a common terminology can be formalised as a ‘controlled vocabulary’ and represented as drop-down menus to help the researchers to choose the agreed terminology for specific items or actions. An example is shown for the case of experimental samples in Table 1. Because defects are central to the research questions, researchers can annotate samples with defect tags. Examples of metadata stored in the ‘Sample’ *Object Type* are included in Figure 4.

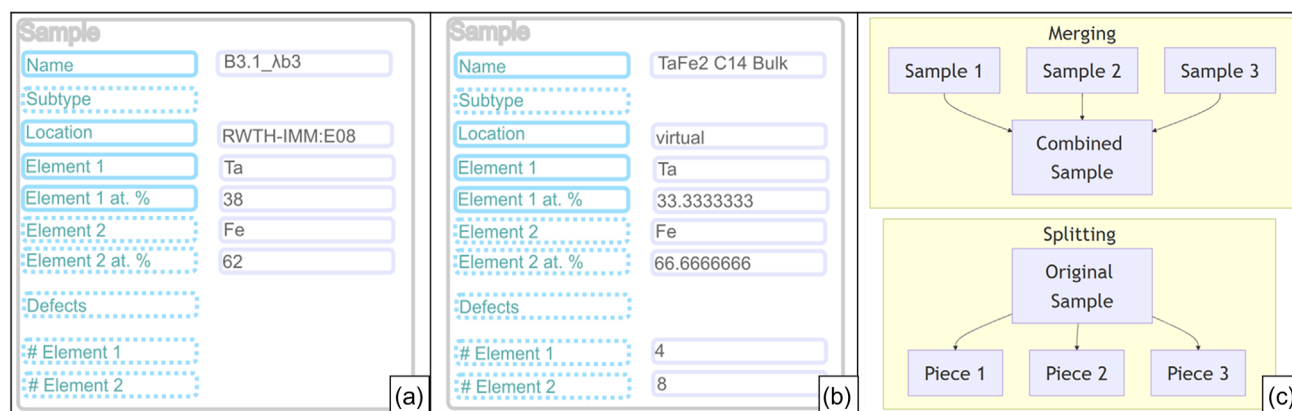
Documentation within the openBIS instance guides researchers in assigning parents and children consistently across groups, see Figure A4 for further details. By preserving these relationships, researchers can systematically reconstruct a sample’s entire experimental history and correlate experimental outputs with corresponding simulations.

Figure 4 illustrates the concept of the sample Object Type developed in this work, using data from [8, 28] as representative examples. The left panel shows the digital representation of an experimental specimen, whose electronic record links to a QR code that can be attached to the corresponding physical sample. The middle panel depicts the equivalent digital object used to represent a simulation input derived from the same material system. The right part of the figure illustrates how parent–child relationships can be used to describe how samples are made. Multiple samples can be merged together, for example during welding. More commonly, a large sample is cut into several pieces to allow different characterisations.

As data generation within the CRC 1394 continues to accelerate, we have focused on building and validating the underlying research data management framework in parallel. Using representative, well-characterised datasets from [8, 28], we developed and tested the workflows and digital representations of samples in openBIS. This proactive approach ensures that the

TABLE 1 | Non-exhaustive list of terms used to categorise experimental samples.

Sample Type	Description
Rod	Long and narrow sample produced by extrusion
Sheet	Flat and thin solid sample
Bulk	Generic macroscopic specimen
APT Tip	Sharpened needle for Atom Probe Tomography
TEM Lamella	Ultra-thin slice for Transmission Electron Microscopy
Thin Film	Thin coating on a substrate
Micro Pillar	Small vertical structure fabricated to investigate mechanical properties

**FIGURE 4** | The sample Object Type. Comparison of an experimental sample (a) and a computational sample (b), the dashed line highlights optional fields. Examples of how multiple samples can be linked through parent–child relationships (c).

infrastructure is fully prepared to handle and interlink forthcoming experimental and simulation data as the collaborative research advances.

3.3 | Electronic Laboratory Notebook

ELNs enable researchers to document their work in a detailed and organised manner, replacing physical paper notebooks to track experiments across synthesis, preparation, characterisation and analysis. In the following, we use micro-mechanical testing combined with SEM to illustrate how openBIS, tailored to materials science, supports end-to-end documentation and provenance.

Researchers start by creating a new ‘Experiment’ in openBIS, either in their personal space, or, ideally, in a shared space. The shared space allows multiple researchers to work in the same environment and remains accessible even as collaborators join or leave the project.

Samples used in the experiment can be sourced externally or synthesised in-house and added either to the same space where the experiment entry is created or to a general inventory available to all researchers, which additionally stores devices and consumables. In the case of the CRC 1394, we have created a common area for all the samples from the collaboration. This enables use of a single openBIS instance across multiple research groups while maintaining access to common sample repositories. Researchers can then add all relevant details regarding sample preparation, for example consumables and instruments used

in grinding and polishing, as well as a session at the electron microscope to record the data in this specific example.

Because many details are defined during the initial setup of the openBIS installation, researchers need to provide relatively few new entries. They reuse pre-defined entities for consumables, devices and standardised preparation methods, link them to their experiments and record only parameters that they vary, along with environmental conditions such as temperature and humidity when suitable sensors are not available. This also includes records that are not contained in the files produced by the scientific instruments but that are essential for the subsequent analysis.

Figure A3 shows a controlled vocabulary for experimental techniques used throughout the collaboration. Using controlled vocabularies as drop-down menus, as well as spreadsheets in the ELN, the researchers can be supported in adding the required information manually, providing a structured approach to guide them through the process.

Even when researchers need to develop a custom approach, they benefit from the pre-defined entities. They can copy an existing object and modify only the fields that require change, for example starting from a metallographic preparation protocol and extending the etching time by editing the timing field. The ELN entry also contains records of the data captured from scientific instruments described in more detail in Section 3.4.

Figure 5 illustrates the final entry in the ELN for a micro-pillar compression test. The left-hand side of the figure shows the navigation bar, which contains multiple micro-mechanical tests

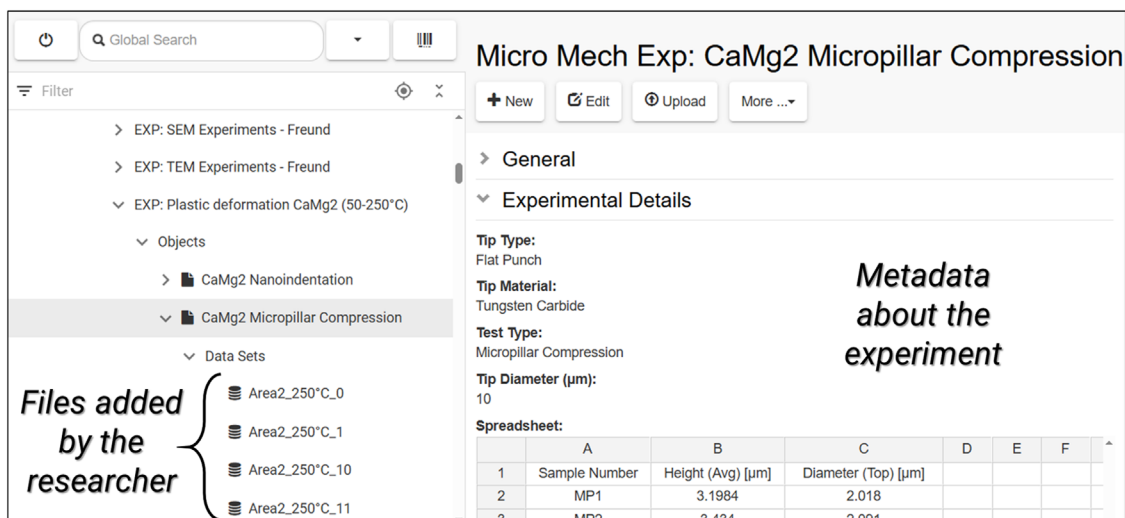


FIGURE 5 | An example of how a micro-mechanical testing step can be represented in openBIS. Details about the micro-pillar compression experiment are recorded in a standardised manner.

below the 'Objects' folder. The right-hand side shows the experiment entry with the object type and device details used in this test. A spreadsheet captures key information such as micro-pillar dimensions, enabling automated workflows as highlighted in Section 5.2.

Since all parts of the analysis life cycle can be captured in a similar way, the relations between these entries as well as the instruments, methods and consumables used across the chain can be tracked to establish full provenance. Figure 6 shows a hierarchical graph of linked samples, processes and devices and represents a concrete example from the workflow shown earlier in Figure 2. Each node of the graph can be explored by clicking on the respective element in the graphical interface. The figure is built around the sample highlighted in green in the second-to-last row of the graph. Moving upwards, researchers can follow the sample's provenance from synthesis to the experiments where it was used. 'Devices' represent scientific instruments, together with their description, such as the nano-indentation device. 'Exp' or 'Experimental Step' represent entries in the ELN. The experimental step is a generic template, whereas 'Exp' has been specifically tailored for micro-mechanical testing or SEM to capture relevant information. Parts of the graph have been omitted to keep the figure readable. In Figure 7, we highlight a different

sample from an earlier study [29]. Moving downwards from the indented sample highlighted in green, researchers can follow subsequent analysis and processing steps. On the left-hand side, micro-pillars are created in a bulk sample using ion milling and are imaged in a SEM before compression. On the right-hand side, we highlight the fabrication of an electron-transparent lamella, from which data is collected in a transmission electron microscope on two separate instances.

3.4 | Handling Data and Metadata

Data captured from scientific instruments or simulation outputs typically contain rich metadata that describe instrument settings during acquisition, software configurations, sample identifiers and contextual information such as environmental conditions. Rather than asking the researchers to add these metadata manually, the data-handling process should include automated parsing to extract relevant metadata and register them in the ELN for consistent, reproducible analysis.

There are two main challenges in metadata extraction: first, materials science and engineering data are heterogeneous, spanning many methods and instruments, and vendors seldom provide standardised, open file formats. As a result, each vendor

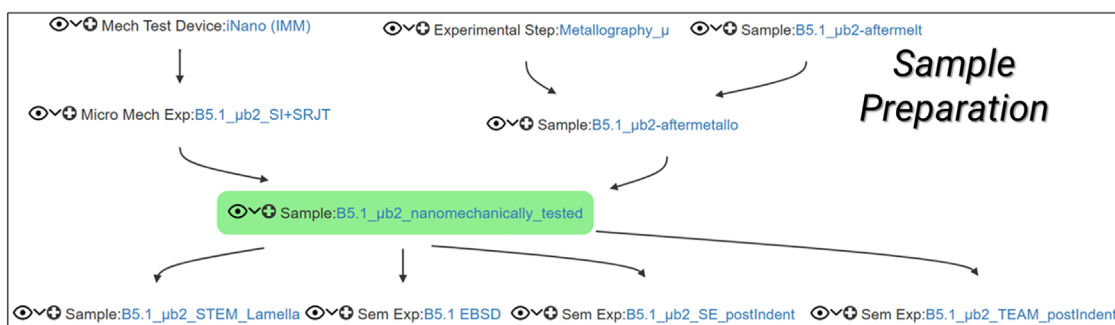


FIGURE 6 | Graphical representation of the provenance and history of a sample that underwent nano-indentation and was characterised by SEM and EBSD. The sample is then transferred across laboratories and undergoes further characterisation. The sample from which the query starts is highlighted in green.

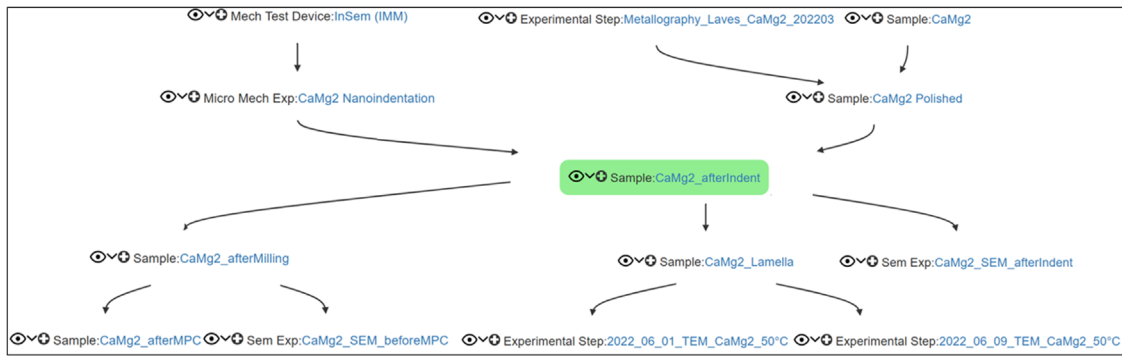


FIGURE 7 | Graphical representation of the provenance and history of a sample that underwent both nano-indentation and micro-pillar compression and characterised by SEM and conventional TEM. The sample from which the query starts is highlighted in green.

typically uses proprietary, often binary formats that must be decoded before metadata can be extracted; some vendors publish specifications, while others do not, requiring reverse engineering. For each instrument and each software version, the community must determine whether the format is known and supported, whether existing tools can read the files and extract metadata or whether new software needs to be developed. This challenge affects the academic community as a whole and is best addressed through coordinated efforts such as the NFDI-MatWerk. Starting from the initial metadata schema developed earlier in the CRC [25], definitions are improved and expanded as more data formats become available, which are published online [27].

The second challenge concerns technical integration when files are registered in openBIS. OpenBIS provides ‘dropboxes’ that assume that data are stored locally within the server and can be copied to dedicated folders on the local file system. This mechanism relies on Jython [30] which bridges Java (in which openBIS is developed) and Python. However, Jython has not been updated since Python 2.7 and does not allow integration of external libraries. Consequently, the stack provided for metadata extraction is outdated, and community tools for file parsing and analysis cannot be integrated through this route [31–33]. Although the openBIS development team has recognised this limitation, a workable replacement is not available in the near term. As a result, the existing mechanisms do not support metadata handling for cloud-based object storage, which is required by the distributed nature of CRC 1394. We, therefore, implement metadata extraction outside the legacy stack, via a Python workflow in a companion application described in Section 4 below.

4 | Companion Application for OpenBIS

In order to overcome the challenges that remain after tailoring openBIS for materials science and engineering, we have developed a companion application that integrates the following functionality:

- Data upload and retrieval from cloud-based object store (S3).
- Registration of externally stored files as linked data in openBIS.
- Extraction of metadata from proprietary formats and registration in openBIS.

The companion application is built with the Streamlit web development framework [34] and can be hosted on a dedicated server or local machines. It operates separately from openBIS and communicates with the server through the Python application programming interface, pyBIS. While this requires the researcher to interact with two graphical interfaces, it enables integration of interactive visualisations, community-developed tools and project-specific features, as well as tailoring the functionality to the specific needs of the researchers of the collaboration. In the following, we describe the key aspects of the companion application in more detail.

4.1 | Adding Data and Metadata to Electronic Laboratory Notebook Entries with the Companion App

The companion application allows researchers to add relevant data and metadata in the ELN. Figure 8 summarises the overall process, and Section 4.2 below describes data handling and metadata extraction in more detail.

Researchers start from an existing experimental sample in the openBIS web interface either by navigating the sidebar or by scanning the physical specimen’s QR code label (Step 1). In Step 2, they create an entry in the ELN, to store, for example, images acquired by a SEM or measurements using EBSD, EDS/EDX, or electron channelling contrast imaging. Using parent–child relationships in openBIS, this notebook entry is linked to the sample, establishing provenance. In Step 3, researchers complete the notebook entry. For SEM, a customised template with a moderate number of fields reduces manual input because most fields can be extracted automatically from the uploaded files (see Section 4.2). In Step 4, the scientific instrument is linked (e.g., a specific SEM model), which automatically adds instrument details to the experiment without requiring manual intervention. Finally, in Step 5, the researchers use the companion application to upload the data to the corresponding entry in the ELN, specifying the *Dataset Type*, which defines the fields that are filled automatically during metadata extraction.

4.2 | Data Handling and Automatic Metadata Extraction

Figure 9 illustrates the process from the researcher’s perspective: first, the researchers need to authenticate both with openBIS and the cloud object storage to obtain write access. This also gives the

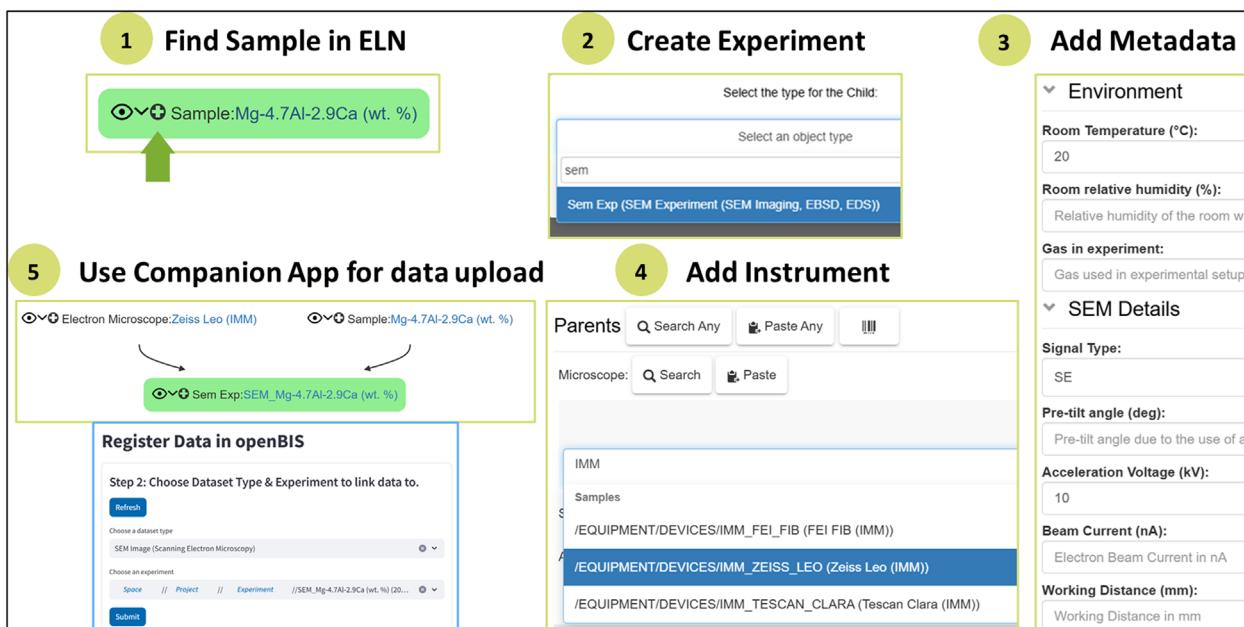


FIGURE 8 | Example workflow of how researchers use the ELN with the companion application to add data and metadata to ELN entries.

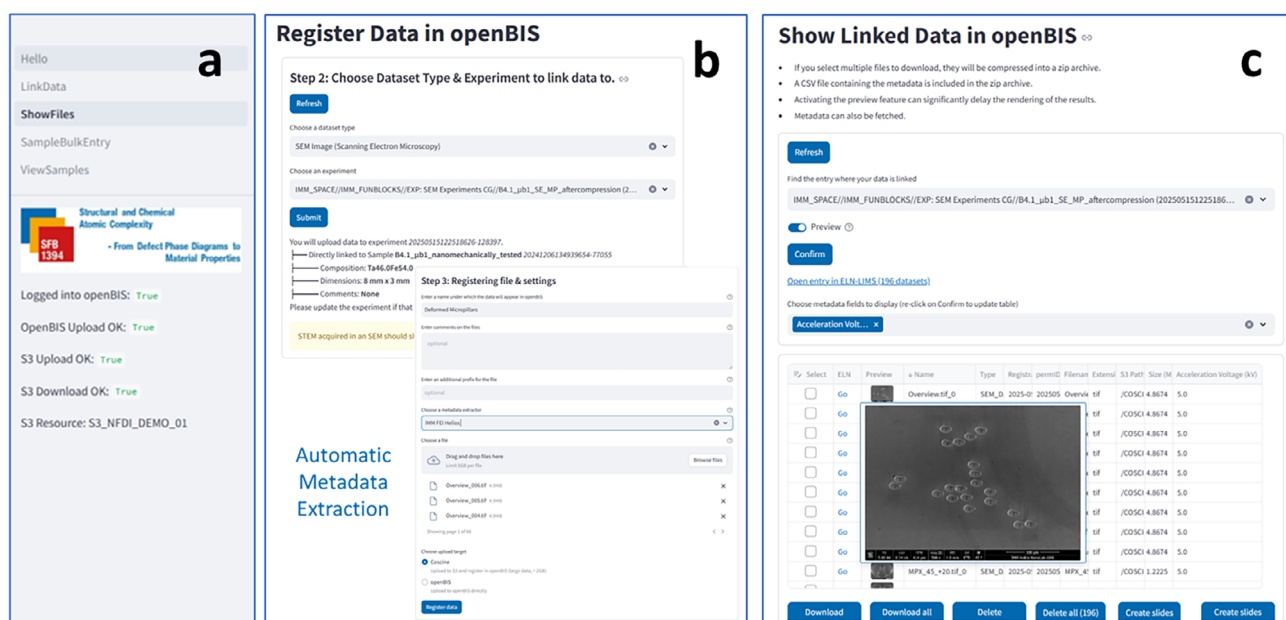


FIGURE 9 | An overview of the companion application. A sidebar (a) enables the researcher to navigate between pages and actions. One page (b) is dedicated to registering data and another (c) is used for data retrieval.

researchers access to the relevant parts in openBIS inside the companion application and allows them to upload files to the external storage. The companion application then retrieves the ELN entries that the researcher is authorised to access, and the researcher selects the experiment to which the data should be linked.

Because a file type alone does not uniquely identify the scientific instrument or the software used to generate it, researchers must select the appropriate metadata parser from a list of available options before uploading data. Each file format requires a dedicated method to extract information, and the internal data structure often differs between vendors, even for the same instrument

type. Vendors also use different terminology for equivalent physical quantities. In SEM, for example, Table A1 shows that Vendor A labels the acceleration voltage as 'EHT', Vendor B as 'HV' and Vendor C as 'Beam.HV'. To ensure consistent metadata representation within the research data management system, we first define a unified superset of terms and then map vendor-specific names to this vocabulary in the metadata parsers. This vocabulary can subsequently link to a formal ontology [35, 36], which supports semantic interoperability and enables integration of the data into knowledge graphs. Figure 10 illustrates how openBIS displays the extracted metadata together with the corresponding micrograph, allowing researchers to access both data and

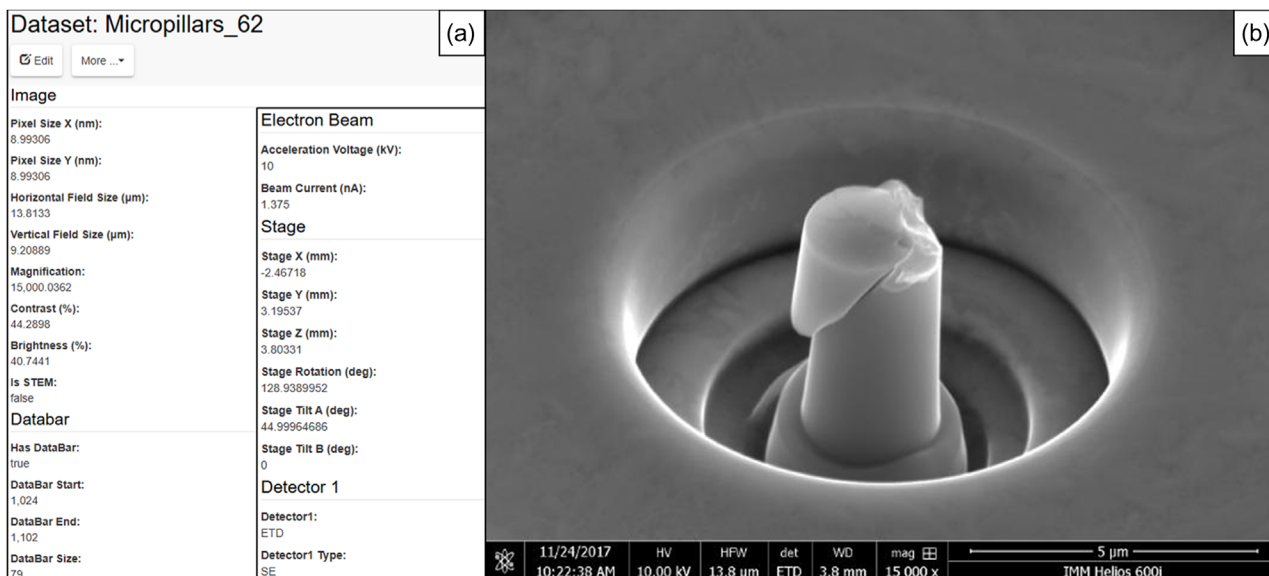


FIGURE 10 | Extracted SEM metadata view in the ELN-LIMS GUI (a) alongside corresponding image of a deformed micro-pillar (b).

metadata directly. When no parser exists, researchers can still register and link the file; they can later enrich it with metadata once a parser becomes available.

These challenges also imply that researchers typically cannot open files easily and verify that they have accessed the right data they are looking for. The companion application addresses this limitation with an integrated preview function similar to modern file browsers. It automatically generates thumbnails for image data and tailored previews for non-image data where possible. As shown in Figure 11, the system can extract an orientation map from EBSD data and store it in the cloud as a preview image. This functionality allows the researcher to confirm file content before downloading and processing. Researchers can select and download multiple files simultaneously. The companion application also includes a ‘Create slides’ feature that lets researchers compile selected images into a slide deck for sharing. Each slide contains one image and a corresponding table of metadata, as shown in Figure A6.

4.3 | Finding Samples through an Extended Hierarchical Graph

The openBIS system itself already provides hierarchical graphs as discussed earlier and illustrated in Figures 6 and 7. However, the native graph’s appearance depends on where the researcher opens it in the interface, and its interaction model is relatively static. Although researchers can expand linked entities and view metadata, expanded nodes quickly consume screen space and make navigation cumbersome.

To improve sample discovery and provenance exploration, the companion application includes an extended, interactive hierarchical graph as shown in Figure 12. The graph uses colours to distinguish entity types such as samples, protocols, experiments and instruments, and it presents a consistent layout regardless of the entry point. Starting from material elements, researchers can filter to find samples that contain a specific chemical element and then traverse linked protocols, instruments, experiment entries and associated datasets. Researchers can hover over any node

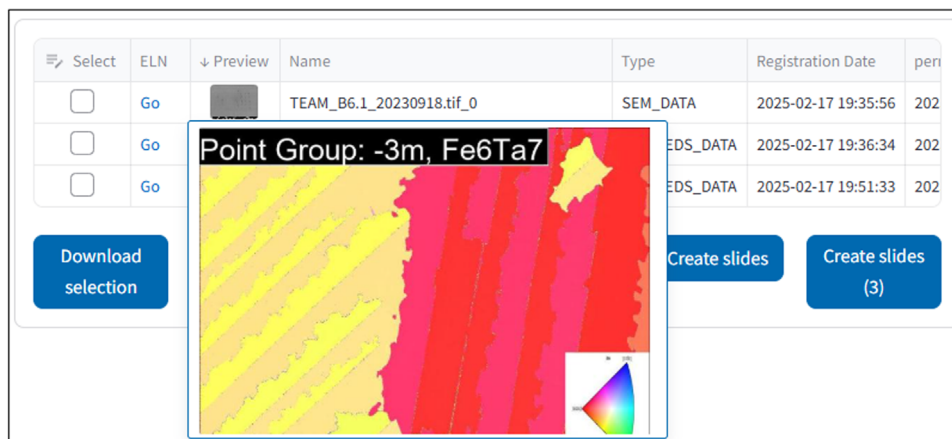


FIGURE 11 | Example of EBSD data visualised in the companion application. The orientation map is extracted from the EBSD data and stored as a preview image. Then, clicking on this thumbnail, a larger version, shown here in the screenshot, is presented in the companion application. Researchers can then quickly verify that the files hold the relevant information before downloading and analysing them in more detail. Data taken from [37].

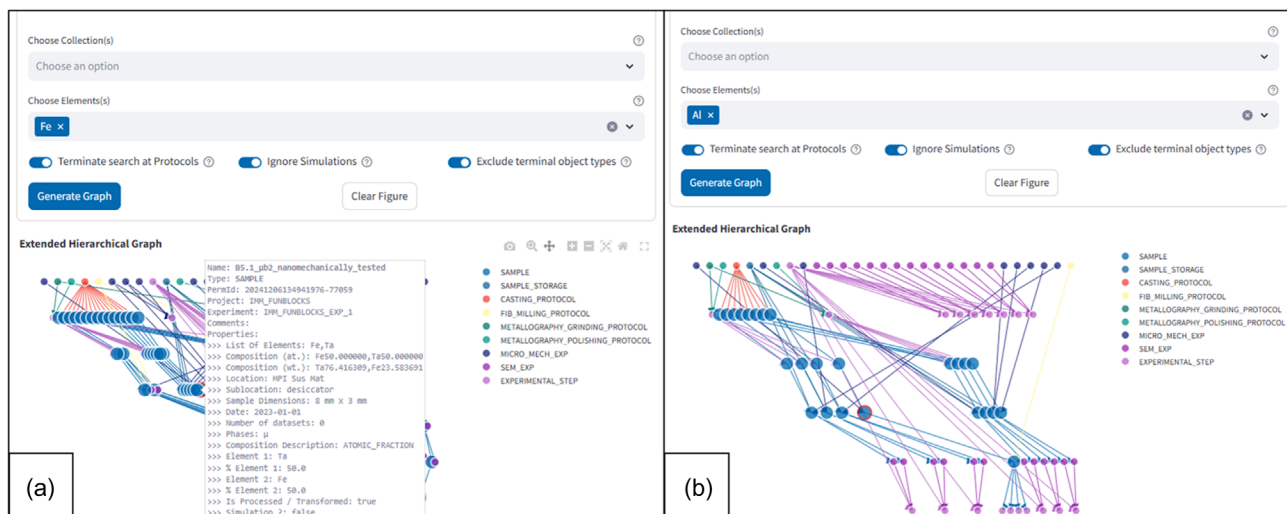


FIGURE 12 | Extended hierarchical graph viewed in the Companion App. The graph shows all samples, methods and ELN entries for all samples containing iron (a) and aluminium (b). The additional metadata that is accessible by hovering the mouse cursor over the respective element is additionally illustrated in (a). This figure uses the same experimental sample as Figure 6.

to reveal key metadata in a tooltip, such as composition, dimensions, instrument model, acquisition date and dataset counts and click on nodes to open the corresponding record.

5 | Workflows and Automation

Once the research data management system is established in the everyday practice of the researchers, its functionality can be extended beyond a documentation and archival tool: because data and their provenance are accessible within one environment, automated workflows can be built on top of the established structures. For example, routine analysis steps can be integrated into the data registration or executed as scheduled tasks. This benefits researchers in several ways. First, standardised data quality checks and baseline analyses are executed automatically, ensuring consistency across researchers and reducing manual effort. Second, a common software environment delivers identical results for identical inputs, improving reproducibility. Third, automation alleviates the burden of running routine analyses manually, allowing researchers to focus on specialised tasks aligned with their scientific objectives.

In the following, we provide two examples that illustrate the benefits of this approach: automated reports and automated standard analysis.

5.1 | Automated Reports

A key part in making research reproducible is to establish the provenance of not only the scientific results and the samples, instruments and software used but also of the preparation steps required to obtain those results. For example, before a metallic sample can be used in micro-mechanical testing, electron microscopy or other analysis, it must be prepared metallographically as appropriate for the analysis. The detailed steps can be stored as a ‘standard operating procedure’ (SOP) or a ‘protocol’ in openBIS and linked to the relevant entry in the ELN. Several protocols have been defined to capture details related to metallographic

preparation. To make this more accessible to the researchers, we have implemented an automated report that summarises the procedure in a table (Figure 13). Preparation details are consolidated from different objects linked through parent–child relationships and displayed in a tabular format. Researchers can quickly verify that all steps are appropriate for the planned work, and the report provides a convenient template for documenting methods in subsequent publications.

5.2 | Automated Analysis in Micro-Mechanical Testing

Micro-pillar compression testing [38] is a micro-mechanical testing technique in which focused ion beam fabricated pillars, typically a few micrometres in diameter, are compressed using a nano-indenter to probe the plastic deformation behaviour of small volumes of material. It enables quantitative investigation of size effects, slip mechanisms and local microstructural phenomena under well-controlled, near uniaxial stress conditions, bridging the gap between nano-indentation [39] and standardised bulk mechanical testing.

A standard baseline analysis is the construction of the stress–strain curve, which shows the sample’s response under mechanical loading. Because this curve should be derived as a first step in further analysis, we implemented an automated workflow (Figure 14). The researcher creates an entry in the ELN using a customised ‘Micro Mech Exp’ template and fills in the relevant details as defined by the template. As shown in Figure 14a, multiple micro-pillars milled from the same specimen are documented, and the researcher uploads the raw load–displacement data as described in Section 4.2.

A set of Python scripts, executed as a scheduled job on a dedicated server, retrieves the relevant data and metadata from the object storage and openBIS for each ‘Micro Mech Exp’ instance. The workflow converts load–displacement to stress–strain using the recorded geometry and generates the stress–strain curves. The figures are then uploaded as datasets to the

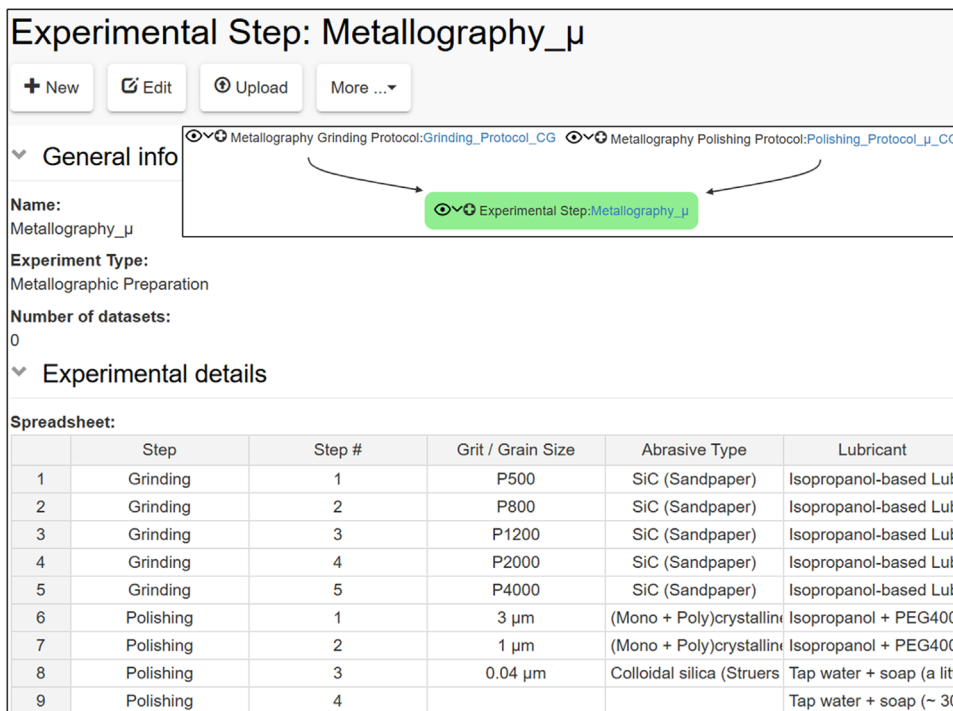


FIGURE 13 | Automatically generated summary report of metallographic preparation steps in the ELN-LIMS graphical interface. All details are extracted from data entered by the researcher, showing the abrasive materials and lubricants used in the preparation of the sample. Data taken from [37].

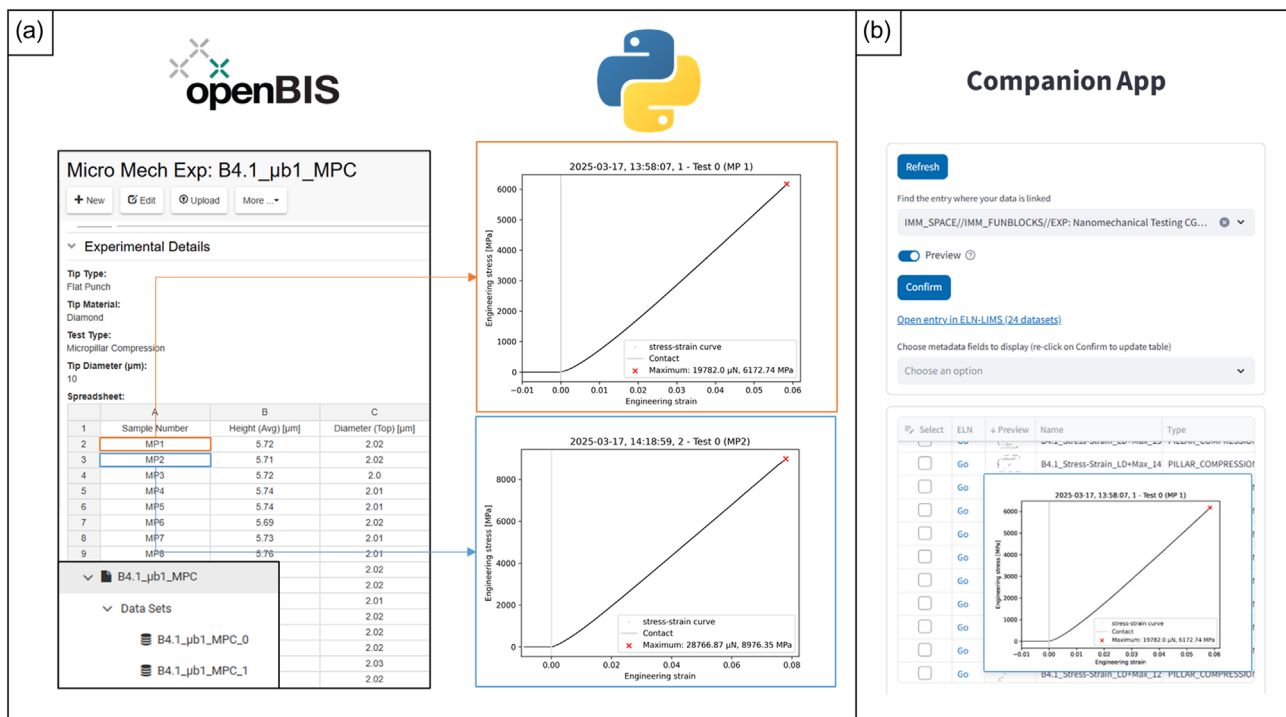


FIGURE 14 | Automated analysis in openBIS for the example of micro-pillar compression tests. An ELN entry for micro-mechanical testing ('Micro Mech Exp'), containing details for multiple micro-pillars milled in the sample (a). Using two examples (MP1, orange, and MP2, blue), the stress–strain curves can be automatically derived from the data uploaded by the researcher and shown in the bottom left corner. The axes for each stress–strain curve have been rescaled using the dimensions of the micro-pillar. The figures are uploaded back to the ELN entry and can be viewed in the companion application (b).

same entry in the ELN and become available in the companion application (Figure 14b). These scripts provide a starting point for researchers to perform more specific analyses in Jupyter notebooks, ensuring consistent, reproducible baseline analyses.

6 | Conclusion and Outlook

Establishing a comprehensive research data management system for heterogeneous collaborations requires integrating data from a wide range of scientific instruments with proprietary formats, aligning metadata schemas across disciplines, accommodating diverse IT setups and linking physical entities such as samples and instruments to their digital representations. Within the CRC 1394, combining openBIS as an ELN and a LIMS with a companion application addresses much of this need. Tailored electronic records, controlled vocabularies, validation scripts, QR code labels and parent–child relationships enable end-to-end provenance and customisation to match research group requirements. The companion application complements openBIS by providing cloud storage integration, automated extraction of metadata from uploaded files and automated reports and workflows that support structured documentation and standardised analyses. By providing the described framework, we aim to establish an infrastructure tailored to the needs of researchers in materials science and engineering, enabling them to accelerate scientific discoveries related to defects. Specifically, our infrastructure aids in curating and organising datasets, building the foundation researchers need to construct defect phase diagrams.

In future work, we will expand the set of vendor-specific file formats from which metadata are extracted automatically and integrate Jupyter Hub into the openBIS setup to allow researchers to access standardised analyses and workflows directly from a web browser, with convenient access to all relevant data and metadata for a given analysis. We will also extend the rollout of tablets across research groups to make data entry and access more convenient when moving between different parts of the laboratories without consistent access to desktop computers. Finally, although openBIS already provides export functionality to repositories such as Zenodo [40], we will extend this to include data stored in the cloud and additional details currently available through the companion application, ensuring that published datasets are findable, accessible, interoperable and reusable.

Author Contributions

Khalil Rejiba: methodology, software, data curation, writing – original draft. **Sang-Hyeok Lee:** software, writing – review & editing. **Christina Gasper:** data curation. **Martina Freund:** data curation. **Sandra Korte-Kerzel:** supervision, project administration, funding acquisition, writing – review & editing. **Ulrich Kerzel:** conceptualisation, methodology, software, supervision, project administration, funding acquisition, writing – original draft, writing – review & editing.

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as part of Collaborative Research Centre CRC 1394—Structural and Chemical Atomic Complexity—From Defect Phase

Diagrams to Material Properties (project number 409476157). Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the National Research Data Infrastructure – NFDI 38/1 – project number 460247524. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreements No. 852096 FunBlocks and No. 101168203 TailorPlast). The data used in this publication was managed using the research data management platform Coscine (<http://doi.org/10.17616/R31NJJZ>) with storage space of the Research Data Storage (RDS) (DFG: INST222/1261-1) and DataStorage.nrw (DFG: INST222/1530-1) granted by the DFG and Ministry of Culture and Science of the State of North Rhine-Westphalia.

Open Access funding enabled and organized by Projekt DEAL.

Funding

This study was supported by Deutsche Forschungsgemeinschaft (409476157, 460247524), European Research Council (852096, 101168203).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analysed in this study.

References

1. J. Kimmig, S. Zechel, and U. S. Schubert, “Digital Transformation in Materials Science: A Paradigm Change in Material’s Development,” *Advanced Materials* 33, no. 8 (2021): 2004940.
2. B. Bayerlein, M. Schilling, H. Birkholz, et al., “PMD Core Ontology: Achieving Semantic Interoperability in Materials Science,” *Materials & Design* 237 (2024): 112603.
3. G. Gottstein and S. Korte-Kerzel, “Kristallbaufehler,” *Materialwissenschaft Und Werkstofftechnik: Physikalische Grundlagen*, (Springer Berlin Heidelberg, 2025), pp. 85–128, ISBN: 978-3-662-68326-2, Available: https://doi.org/10.1007/978-3-662-68326-2_3.
4. X. Zhou, P. Mathews, B. Berkels, et al., “Materials Design by Constructing Phase Diagrams for Defects,” *Advanced Materials* 37, no. 3 (2025): 2402191.
5. S. Korte-Kerzel, T. Hickel, L. Huber, et al., “Defect Phases—thermodynamics and Impact on Material Properties,” *International Materials Reviews* 67, no. 1 (2022): 89–117.
6. T. Frolov, J. Neugebauer, and Y. Mishin, “Thermodynamics of Grain-Boundary Phases: T. Frolov, Et al,” *MRS Bulletin* 51, (2026): 152–167, <https://doi.org/10.1557/s43577-025-01022-6>.
7. H. Howard, W. Cunningham, A. Genc, et al., “Chemically Ordered Dislocation Defect Phases as a New Strengthening Pathway in Ni–Al Alloys,” *Acta Materialia* 289 (2025): 120887.
8. C. Gasper, N. Ulumuddin, S. Zhang, et al., “Chemically Tailored Planar Defect Phases in the Ta-Fe μ -Phase,” 2025, arXiv: 2510.17336 [cond-mat.mtrl-sci]. [Online]. Available: <https://arxiv.org/abs/2510.17336>,
9. W. Luo, C. Gasper, S. Zhang, et al., “Non-Basal Plasticity in the μ Phase at Room Temperature,” *Acta Materialia* 277 (2024): 120202.
10. A. Tehrani, S. Zhang, A. Zendegani, C. Scheu, T. Hickel, and J. Neugebauer, “Metastable Defect Phase Diagrams as Roadmap to Tailor Chemically Driven Defect Formation,” *Acta Materialia* 277 (2024): 120145.
11. C. Eberl, M. Niebel, E. Bitzek, et al., Consortium Proposal Nfdi-Matwerk, 2021, Available at: <https://doi.org/10.5281/zenodo.5082837>.

12. M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, et al., “The Fair Guiding Principles for Scientific Data Management and Stewardship,” *Scientific Data* 3, no. 1 (2016): 1–9.
13. I. Lang, M. Nellesen, and M. Politze, “RDM Platform Coscine-Fair Play Integrated Right from the Start,” *Ing. Grid* 2, no. 1 (2024).
14. H. Tsybenko, S. Brinckmann, and R. Schwaiger, Managing Locally Stored Scientific Data with PASTA-ELN, 2023, Available at: <https://doi.org/10.18154/RWTH-2023-10438>.
15. S. Herres-Pawlis, N. A. Parks, K. W. Kröckert, et al., “Data-Producing Methods in CRC 985: Recommendations for Research Data Management in Large Interdisciplinary Projects,” *Ing. Grid* 3, no. 1 (2025).
16. D. Speckhard, T. Bechtel, L. M. Ghiringhelli, M. Kuban, S. Rigamonti, and C. Draxl, “How Big Is Big Data?,” *Faraday Discussions* 256 (2025): 483–502.
17. C. Barillari, D. S. Ottoz, J. M. Fuentes-Serna, C. Ramakrishnan, B. Rinn, and F. Rudolf, “openBIS ELN-LIMS: An Open-Source Database for Academic Laboratories,” *Bioinformatics* 32, no. 4 (2016): 638–640.
18. L. K. Cuellar, A. Friedrich, G. Gabernet, et al., “A Data Management Infrastructure for the Integration of Imaging and Omics Data in Life Sciences,” *BMC Bioinformatics* 23, no. 1 (2022): 61.
19. R. El-Athman, J. Rädler, O. Löhmann, A. Ariza, and T. Muth, “The BAM Data Store,” in *Proceedings of the Conference on Research Data Infrastructure*, (2023), 1–5, 1.
20. F. Plass, S. Englisch, B. A. Zubiri, L. Pflug, E. Spiecker, and M. Stingl, “Using OpenBIS as Virtual Research Environment: An ELN-LIMS Open-Source Database Tool as a Framework Within the CRC 1411 Design of Particulate Products,” *Data Science Journal* 22, no. 1 (2023): 44.
21. M. Zaki, C. Prinz, and B. Ruehle, “A Self-Driving Lab for Nano- and Advanced Materials Synthesis,” *ACS Nano* 19, no. 9 (2025): 9029–9041.
22. G. Lileikyte, A. Bakochi, M. Isaksson, et al., “Proteomic Analysis of Serum Samples after Cardiac Arrest: Rationale and Design of a TTM-Trial Substudy,” *Resuscitation Plus* 25 (2025): 101014.
23. Y. Hemani, K. Koch, O. Mendo-Diaz, A. Bachhofner, S. Baffelli, and D. Bleiner, “The openBIS Digital Platform for Instrumentation and Data Workflow in the Analytical Laboratory,” *Chimia* 79, no. 1-2 (2025): 36–45.
24. M. Hans, P. Keuter, A. Saksena, et al., “Opportunities of Combinatorial Thin Film Materials Design for the Sustainable Development of Magnesium-Based Alloys,” *Scientific Reports* 11, no. 1 (2021): 17454.
25. N. Siemer, L. Huber, M. O. SFB, S. Korte-Kerzel, and J. Neugebauer, Metadata Schemes for Materials Science Data [Online], 2022, <https://doi.org/10.5281/ZENODO.6513745>, Available: <https://publications.rwth-aachen.de/record/861207>.
26. E. Bocchi, M. Mellia, and S. Sarni, “Cloud Storage Service Benchmarking: Methodologies and Experimentations,” in *2014 IEEE 3rd International Conference on Cloud Networking (CloudNet)*, (IEEE, 2014), 395–400.
27. K. Rejiba, S. Korte-Kerzel, and U. Kerzel, openBIS Schema Extension for Research in Materials Science and Engineering, 2025, Available at: <https://doi.org/10.5281/zenodo.15731469>.
28. C. Gasper, E. M. Soysal, N. Ulumuddin, et al., “Mechanical Properties and Deformation Mechanisms of the C14 Laves and μ -Phase in the Ternary Ta-Fe (-al) System,” *Materials & Design* 251 (2025): 113625.
29. M. Freund, D. Andre, C. Zehnder, et al., “Plastic Deformation of the CaMg₂ C14- Laves Phase from 50-250°C,” *Materialia* 20 (2021): 101237.
30. J. Juneau, J. Baker, F. Wierzbicki, et al., *The Definitive Guide to Jython: Python for the Java Platform*, (Apress, 2010).
31. D. N. Johnstone, B. H. Martineau, P. Crout, P. A. Midgley, and A. S. Eggeman, “Density-Based Clustering of Crystal (mis) Orientations and the Orix Python Library,” *Applied Crystallography* 53, no. 5 (2020): 1293–1298.
32. H. W. Ånes, L. A. H. Lervik, O. Natlandsmyr, et al., Pyxem/Kikuchipy: Kikuchipy 0.11.2, Version v0.11.2, 2025, <https://doi.org/10.5281/zenodo.14995525>.
33. F. de la Peña, E. Prestat, J. Lähnemann, et al., Hyperspy/Hyperspy: V2.3.0, Version v2.3.0, 2025, <https://doi.org/10.5281/zenodo.14956374>.
34. Streamlit Inc., Streamlit, Version 1.x, Python Library for Building Interactive Data Applications [Online], Available at: <https://streamlit.io>, 2025.
35. B. Bayerlein, M. Schilling, M. Curran, et al., “Natural Language Processing-Driven Microscopy Ontology Development,” *Integrating Materials and Manufacturing Innovation* 13, no. 4 (2024): 915–926.
36. V. Hofmann, O. Brendike-Mannix, A. Azócar, and A. A. Guzmán, “The EM glossary - a community resource providing harmonized terminology in electron microscopy,” in *Microscopy Conference 2025 (MC2025)* (Zenodo, 2025), <https://doi.org/10.5281/zenodo.17035738>.
37. C. Gasper, I. Gao, F. Busch, et al., “Preparation of Binary and Ternary Laves and μ -Phases in the Ta-Fe (-Al) System for Property Analysis at the Microscale,” *Metallurgical and Materials Transactions A* 55, no. 7 (2024): 2244–2263.
38. M. D. Uchic, D. M. Dimiduk, J. N. Florando, and W. D. Nix, “Sample Dimensions Influence Strength and Crystal Plasticity,” *Science* 305, no. 5686 (2004): 986–989.
39. W. C. Oliver and G. M. Pharr, “An Improved Technique for Determining Hardness and Elastic Modulus Using Load and Displacement Sensing Indentation Experiments,” *Journal of Materials Research* 7 (1992): 1564–1583, <https://doi.org/10.1557/JMR>.
40. European Organization For Nuclear Research and OpenAIRE, Zenodo, en, 2013, <https://doi.org/10.25495/7GXK-RD71>, [Online]. Available: <https://www.zenodo.org/>.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Supporting Fig. A.1:** Mechanical polishing details described in [37]. **Supporting Fig. A.2:** A protocol to represent the process of creating an electron-transparent membrane using focused ion beam milling. The protocol is used for different samples. **Supporting Fig. A.3:** A view of the Vocabulary Browser in the ELN-LIMS GUI show casing categories of experimental techniques. **Supporting Fig. A.4:** Screenshot showing documentation of micro-pillar compression tests stored in openBIS. **Supporting Fig. A.5:** Screenshot showing documentation centred on atom probe tomography stored in openBIS. **Supporting Fig. A.6:** Example showing slide deck creation from data registered in openBIS. **Supporting Table A.1:** A non-exhaustive list of metadata that can be extracted from SEM images.