1st Conference on Research Data Infrastructure

Engineering Sciences

https://doi.org/10.52825/CoRDI.v1i.229

Published: 07 Sept. 2023

The BAM Data Store

Piloting an openBIS-Based Research Data Infrastructure in Materials Science and Engineering

Rukeia El-Athman[1] [https://orcid.org/0000-0003-0749-160X], Jörg Rädler[1] [https://orcid.org/0000-0002-8935-6059], Oliver Löhmann[1] [https://orcid.org/0000-0002-1527-1319], Angela Ariza[1] [https://orcid.org/0000-0002-1005-5780] and Thilo Muth[1] [https://orcid.org/0000-0001-8304-2684]

[1] Bundesanstalt für Materialforschung und -prüfung, Germany

Abstract. As a partner in several NFDI consortia, the Bundesanstalt für Materialforschung und -prüfung (BAM, German federal institute for materials science and testing) contributes to research data standardization efforts in various domains of materials science and engineering (MSE). To implement a central research data management (RDM) infrastructure that meets the requirements of MSE groups at BAM, we initiated the Data Store pilot project in 2021. The resulting infrastructure should enable researchers to digitally document research processes and store related data in a standardized and interoperable manner. As a software solution, we chose openBIS, an open-source framework that is increasingly being used for RDM in MSE communities.

The pilot project was conducted for one year with five research groups across different organizational units and MSE disciplines. The main results are presented for the use case “nanoPlattform”. The group registered experimental steps and linked associated instruments and chemicals in the Data Store to ensure full traceability of data related to the synthesis of ~400 nanomaterials. The system also supported researchers in implementing RDM practices in their workflows, e.g., by automating data import and documentation and by integrating infrastructure for data analysis.

Based on the promising results of the pilot phase, we will roll out the Data Store as the central RDM infrastructure of BAM starting in 2023. We further aim to develop openBIS plugins, metadata standards, and RDM workflows to contribute to the openBIS community and to foster RDM in MSE.

Keywords: Research Data Infrastructure, ELN, openBIS, Materials Science and Engineering.

1 Premise and Motivation

In the Bundesanstalt für Materialforschung und -prüfung (BAM, German federal institute for materials science and testing), research and development, as well as scientific and technical services are conducted in 10 departments divided into ~60 scientific units, comprising a scientific staff of more than 1,000 employees. As a partner in several NFDI consortia (currently in NFDI-MatWerk, FAIRmat, and DAPHNE4NFDI; prospective partner in NFDI4Chem), BAM contributes to research data standardization in various domains of materials science and engineering (MSE). Before introducing the Data Store, no software solutions for central research data management (RDM) or electronic lab notebooks (ELN) were established, resulting in institutional “data silos” that hampered the generation of FAIR datasets [1]. To fill this gap, we

launched the BAM Data Store project in 2021 to build an infrastructure that enables BAM researchers to document research processes and store associated data in a standardized and interoperable manner.

To collect RDM requirements and to test the implementation in the Data Store, we conducted a one-year pilot project with five research groups at BAM, consisting of 50 members from 15 scientific units. The pilot groups represent multidisciplinarity within the MSE domain and routinely apply a broad range of methodologies.

2 openBIS: An Open-Source RDM and ELN Software

The Data Store is based on the open-source software openBIS [2], [3]. Developed by the Scientific IT Services of ETH Zurich, openBIS was originally intended as a framework to support RDM in life science laboratories but is increasingly being used in the MSE community. openBIS offers a web-based graphical user interface (GUI) for the digital representation of laboratory inventory, and an ELN for the documentation of experimental procedures. Data files of any format can be imported either via the GUI or via programming interfaces and linked to inventory elements and experimental steps. In addition, openBIS provides various interfaces for data transfer, e.g., for exporting meta(data) to the data repository Zenodo [4] as well as for analyzing data in the web-based environment Jupyter Notebook [5].

3 Pilot Project: Testing openBIS with Five MSE Research Groups

3.1 IT Infrastructure

openBIS is designed to run on Linux servers and requires a Java environment and a PostgreSQL database system. To gain maximum flexibility of the configuration, every pilot group was equipped with a dedicated virtual Linux server containing an openBIS installation managed by the central IT department. While all datasets are stored as regular files on an openBIS server (which may require a large amount of local storage), the data models, metadata, relations, and permissions are stored in the database. A JupyterHub installation was added to each server for testing purposes and configured to work seamlessly with openBIS. Central file shares and individual import scripts (a.k.a. Dropboxes) were used to automate the import and partial analysis of data from lab equipment.

3.2 Onboarding Process

The onboarding to the Data Store was conducted individually for each pilot group. In a first workshop, the group members defined and prioritized their RDM requirements as testable user stories along the phases of the data lifecycle. Next, the group's admins learned about the underlying data hierarchy of openBIS and were tasked with the creation of a data model to digitally represent their group's inventory and experimental steps. After the inventory was completed and the building blocks for the lab notebook were defined, the whole group was trained to use the Data Store for their daily work. The pilot phase was concluded by testing the user stories. The implementation of the Data Store by the pilot group "nanoPlattform" is described below.

3.3 Use Case "nanoPlattform"

The "nanoPlattform" is an internally funded BAM project with the aim to provide reference materials, reference data, and reference protocols in the field of nanoscience. Within the project, more than 30 scientists and technical staff of different disciplines from 10 different scientific units currently use the Data Store to ensure data traceability and data transfer. Within one year, about 400 nanomaterials were synthesized and more than 1000 experimental steps were

registered using more than 90 instruments and more than 150 different chemicals. All these objects are stored and connected in the Data Store, ensuring full traceability of the scientific process and full accessibility for all group members (Figure 1).

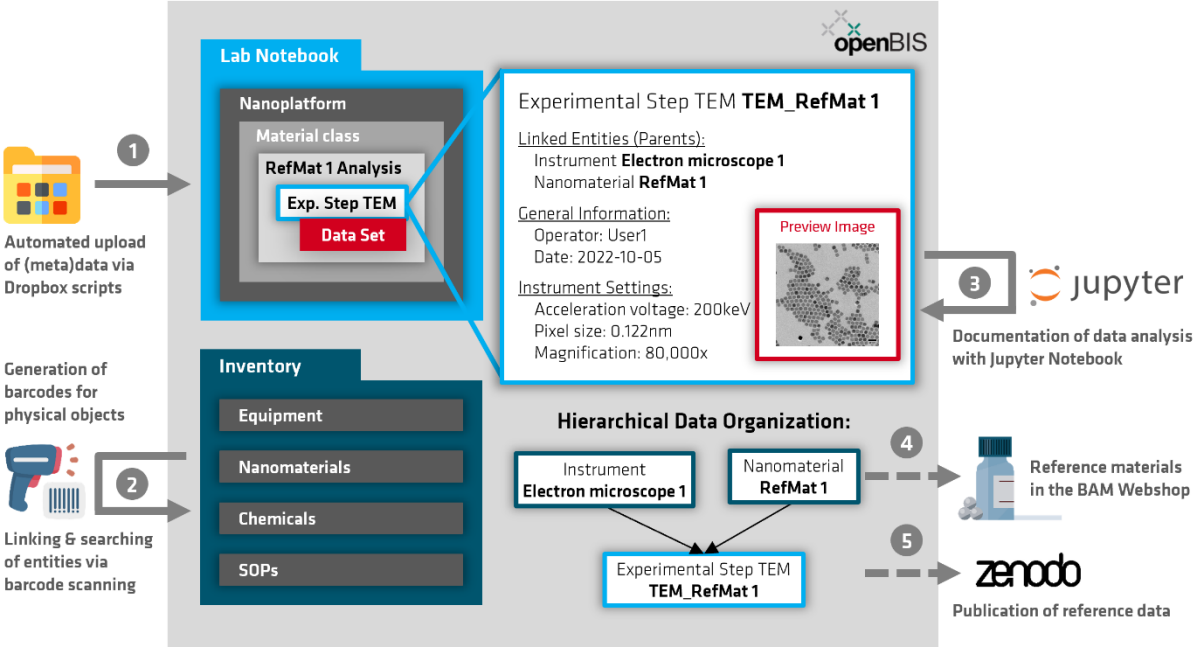

Figure 1. Data Store implementation of the pilot group “nanoPlattform”.

The synthesis of nanoparticles is documented in the lab notebook of the Data Store and linked to the chemicals registered in the inventory. (1) Measurement data are imported and attached to the respective experiment by Dropbox scripts which also parse and register the associated metadata. (2) Synthesized samples are labeled with barcodes generated by openBIS, allowing all project members to access sample-related metadata and data simply by scanning a barcode. Sample characterizations are also documented in the Data Store and linked to the respective samples, instruments, sample preparation. (3) Data are analyzed using Jupyter Notebooks that access data via the Python extension pyBIS. The resulting Notebooks are saved in the Data Store and linked to the respective experiments. (4) The synthesized reference material is available for purchase in the BAM Webshop. (5) The publication of reference data and protocols are currently in preparation. (Left-side icons made by Freepik from www.flaticon.com.)

4 Conclusion and Outlook

As illustrated by the example of the “nanoPlattform”, many of the RDM requirements of MSE research groups can be met by openBIS functionalities. The pilot project showed that the transition to a digital system for the representation of laboratory notebooks requires researchers to invest time and effort during the implementation and customization phase. However, the Data Store can ultimately save time by automatizing daily tasks, e.g., the barcode-supported documentation of consumables used for measurements, and by the automatic import of standardized (meta)data from instruments. In addition, the Data Store enables researchers to implement RDM practices according to the FAIR principles by integrating software tools such as Jupyter Notebook to document data analysis, associated code, and underlying experimental data.

Altogether, the Data Store facilitates cooperation between scientists and across research communities within BAM by providing a common infrastructure and shared metadata schemas.

Based on the successful implementation of openBIS in research workflows during the pilot project, the Data Store will be rolled out as the institute's central RDM infrastructure from 2023. We expect the Data Store to become the backbone of digital RDM at BAM and we aim to develop additional openBIS plugins, metadata standards, and RDM workflows that will benefit various MSE communities both within and beyond BAM.

Data availability statement

No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Author contributions

Rukeia El-Athman: Methodology (lead); conceptualization (equal); writing – original draft (lead); writing – review and editing (equal); visualization (lead). Jörg Rädler: Software (lead); conceptualization (equal); writing – original draft (supporting); writing – review and editing (equal). Oliver Löhmann: Methodology (supporting); writing – original draft (supporting); writing – review and editing (equal). Angela Ariza: Writing – original draft (supporting); writing – review and editing (equal). Thilo Muth: Supervision (lead); project administration (lead); conceptualization (equal); writing – review and editing (equal).

Competing interests

The authors declare that they have no competing interests.

Funding

The BAM Data Store project is funded by internal BAM resources.

Acknowledgement

We would like to thank Caterina Barillari from the Scientific IT Services of ETH Zurich for her support and advice, and for conducting the openBIS webinars during the pilot phase.

References

1. M. D. Wilkinson et al., "The FAIR Guiding Principles for scientific data management and stewardship," Sci Data, vol. 3, no. 1, Art. no. 1, Mar. 2016, doi: 10.1038/sdata.2016.18.
2. A. Bauch et al., "openBIS: a flexible framework for managing and analyzing complex data in biology research," BMC Bioinformatics, vol. 12, no. 1, p. 468, Dec. 2011, doi: 10.1186/1471-2105-12-468.
3. C. Barillari, D. S. M. Ottoz, J. M. Fuentes-Serna, C. Ramakrishnan, B. Rinn, and F. Rudolf, "openBIS ELN-LIMS: an open-source database for academic laboratories," Bioinformatics, vol. 32, no. 4, pp. 638–640, Feb. 2016, doi: 10.1093/bioinformatics/btv606.
4. European Organization For Nuclear Research and OpenAIRE, "Zenodo." CERN, 2013. doi: 10.25495/7GXK-RD71.
5. M. Beg et al., "Using Jupyter for reproducible scientific workflows," Comput. Sci. Eng., vol. 23, no. 2, pp. 36–46, Mar. 2021, doi: 10.1109/MCSE.2021.3052101.

6. “BAM-N012 - Nanomaterialien - Referenzmaterialien.” https://webshop.bam.de/webshop_de/referenzmaterialien/nanomaterialien/bam-n012.html (accessed Apr. 18, 2023).