Using *OpenBIS* as Virtual Research Environment: An ELN-LIMS Open-Source Database Tool as a Framework within the *CRC 1411 Design of Particulate Products*

FABIAN PLASS

SILVAN ENGLISCH

BENJAMIN APELEO ZUBIRI (iD)

LUKAS PFLUG (iD)

ERDMANN SPIECKER

MICHAEL STINGL (iD)

*Author affiliations can be found in the back matter of this article

CORRESPONDING AUTHORS:
Michael Stingl

Competence Center Scientific Computing (CSC), Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany

michael.stingl@fau.de

Erdmann Spiecker

Institute of Micro- and Nanostructure Research (IMN) & Center for Nanoanalysis and Electron Microscopy (CENEM), IZNF, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany

erdmann.spiecker@fau.de

ABSTRACT

The digital transformation and consequent use of new digital technologies not only have a substantial impact on society and companies, but also on science. Analog documentation as we have known it for centuries will eventually be replaced by intelligent and FAIR (Findable, Accessible, Interoperable, and Reusable) systems. In addition to the actual research data and results, metadata now plays an important role not only for individual, independently existing projects, but for future scientific use and interdisciplinary research groups and disciplines as well. The solution presented here, consisting of an electronic laboratory notebook and laboratory information management system (ELN-LIMS) based on the *openBIS* (open Biology Information System) environment, offers interesting features and advantages, especially for interdisciplinary work. The *Collaborative Research Centre* (CRC) *1411 'Design of Particulate Products'* of the German Research Foundation is characterized by the cooperation of different working groups of synthesis, characterization, and simulation, and therefore serves as a model environment to present the implementation of *openBIS*. OpenBIS, as an open source ELN-LIMS solution following FAIR principles, provides a common set of general entries with the possibility of sharing and linking (meta-)data to improve the scientific exchange between all users.

1. INTRODUCTION

Digital transformation is a key challenge that impacts our entire society. The main players here are companies that use intelligent information technologies (IT) and networks for machines and processes. Starting with flexible production via modular, changeable production processes, to customer-specific solutions and products requires the help of sophisticated data acquisition, processing, and analysis (Bauernhansl et al. 2014; Lasi et al. 2014). Further examples of digital transformation are automation processes, machine-to-machine communication, internet-of-things (IoT) process implementation, or even augmented-reality-based workflows (Bauernhansl et al. 2014; Egger & Masood, 2020; Lasi et al. 2014; Li et al. 2015).

However, not only companies are subject to this change, but government institutions (eGovernment) (Gisler 2001), and science itself (Kimmig et al. 2021) are as well. Thus, the topic of Open Science has become a growing movement (National Academies of Sciences, Engineering, and Medicine (U.S.) et al. 2018). Open Science has been supported in the European context of the *Open Research Data and Data management plans* of the *European Research Council* (*ERC*), established by the *European Commission* for almost five years. However, the ERC has been promoting the causes of Open Science since 2007 (ERC Scientific Council 2021). This also manifests that open access publications from funded projects have already become mandatory to a certain extent. Exemplarily, the *DFG* (*German Research Foundation*) supports infrastructure projects within *Collaborative Research Centres* (*CRC*s), which are long-term university-based research institutions established for up to 12 years, and whose funding objective is to establish powerful information systems for research in a holistic perspective (German Research Foundation 2021). Accordingly, new infrastructure on a national (german *Nationale Forschungsdateninfrastruktur*, *NFDI*) and European level (*European Open Science Cloud*, *EOSC*) (European Commission 2016; Mons et al. 2017) have been established, fostering the subject of research data management, data publications, and Open Science. Nevertheless, the topic of Open Science includes more than just the public provision of data in the sense of open access publications: it also includes the approaches of open methodology, sources, and data. The necessary implementation and representation of good scientific data management, data quality, and stewardship (data governance) are tremendously important (Brous et al. 2016; Hildebrand et al. 2011; Ladley 2020; Wilkinson et al. 2016). The resulting benefits can be measured directly, such as in terms of improvements in process efficiency or cost and risk reductions, and indirectly, such as increased acceptance, perception, and trust (Brous et al. 2016; Hildebrand et al. 2011; Tallon 2013).

This paper presents *openBIS*, an Electronic Laboratory Notebook and Laboratory Information Management System to support open science broadly, including data management, handling, storage, and publishing within a scientific laboratory environment.

2. CURRENT STATE OF RESEARCH DATA MANAGEMENT

2.1. FAIR AS PART OF RESEARCH DATA MANAGEMENT

One of the cornerstones of this overall research data management (RDM) is the FAIR principles (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable). The emphasis placed on growing awareness of FAIRness is, however, more than just an essential duty that public funding agencies impose on research. Moreover, it is the key to conduct knowledge discovery, innovation, and transfer, as well as the subsequent integration and reuse by the scientific community (Wilkinson et al. 2016). Events such as the global Coronavirus pandemic demonstrate the need for, and the overall benefits of, making data available online (Besançon et al. 2021; Tse et al. 2020). This leads not only to efficient research and increased innovation, but also to fair and transparent use of public funds and tax capital, as well as increased visibility and scientific reputation and reliability, to name just a few benefits of Open Science (Janssen et al. 2012).

The FAIR Data Principles propose that all scholarly output should embody the characteristics of being Findable, Accessible, Interoperable, and Reusable. While these principles provide guidance on the expected behaviors of data resources, their practical implementation has been subject to varying interpretations. As the support for these principles has grown, so has the diversity of interpretations surrounding their application (Mons et al. 2017). FAIR principles recognize the need for data accessibility under defined conditions, but do not necessitate

complete openness. While transparency and clarity are required for accessing and reusing data, restrictions can be placed based on privacy, security, and competitive reasons. FAIR promotes a balanced approach that allows diverse participation and partnerships while ensuring the availability of data within specified guidelines (Mons et al. 2017).

2.2. DATA REPOSITORIES AND DATA PUBLICATIONS

Data repositories are a key component of the digital transformation of science. Well-known examples are the commercial data repository service *figshare,* open access archives like arXiv. org, or platforms like *Dataverse* (Crosas 2011), *EUData* (Lecarpentier et al. 2013), and *Zenodo,* which is maintained by *CERN* and funded by the *EU Commission.* In fact, most of the known repositories already consider the high-level FAIR principles. So, in the case of *Zenodo,* the uploaded data is provided with a digital object identifier (DOI) and can optionally be published as open-accessible and viewable. This, in turn, leads to simplified findability, accessibility, and usability for the scientific community, as well as to the quotability of individual datasets, whose content no longer has to lead directly to a complete publication. This means that even data, methods, or code that initially received little attention can now be found by the general public and are not lost.

Despite all of this, justified doubts exist regarding the Open Science policy. Especially, there are questions and concerns about the security of data against external interference and possible compliance and regulatory requirements, especially in healthcare, such as the protection of relevant patient data. These issues circling around the subject of data sovereignty need to be clarified during the planning and before introducing digital technologies such as data management or cloud-based systems (Clayton et al. 2019; Hummel et al. 2021).

Good research data management does not start with the publication and archiving process of the work or the (meta-)data, but with their initial collection. This is because not only content-related data/information is of importance in the sense of RDM, metadata is as well. Metadata describes data or in general (additional) information about the described data(set). Metadata-specific information can be the authors, the creation time/date, or the type of the dataset, as well as the DOI of the dataset. In fact, many distinct types of metadata exist, including descriptive, structural, administrative, and process data.

2.3. ELECTRONIC LABORATORY NOTEBOOKS AND LABORATORY INFORMATION MANAGEMENT SYSTEMS

The scientific questions, choices of experimental procedures, materials and methods, data analyses, and interpretation of the results were traditionally recorded in detail in notebooks on paper, and not electronically (Barillari et al. 2016). Not only is this approach incomprehensible today, since most data are generated electronically or stored as code on a network anyway, but this concept vehemently contradicts the overarching principles of FAIR and Open Science in general. Neither are the data easy to find, nor are they accessible or usable by scientists outside the local system in which they are stored. Inevitably, the use of paper-based notebooks should be avoided, and a shift made to electronic systems such as a laboratory information management system (LIMS) and/or an electronic laboratory notebook (ELN).

An ELN and LIMS is a software solution designed to facilitate the documentation and management of laboratory processes and data. An ELN is used for capturing and organizing experimental data, while a LIMS supports the management of laboratory resources, sample tracking, quality assurance, and other laboratory functions. Together, ELN and LIMS provide a comprehensive platform for efficient and secure management of laboratory information, promoting compliance with best practices and regulatory requirements (Barillari et al. 2016; Bespalov et al. 2020; Machina & Wild 2013). ELN systems can play a major role in a successful RDM. In this way, a continuous workflow under FAIR conditions can be guaranteed from the very beginning, starting with the collection of data, the use of the data by oneself or the research group and other researchers, the publication of the data, and, finally, a superior archiving, for example of data repositories like *Zenodo* or *RADAR* (Kraft et al. 2016). Further advantages of an ELN system are: (i) the easy and metadata-based collection of information, as well as their shareability, (ii) an ensured long-lasting data storage on a secured server, (iii) simplified accessibility via a global or local network, (iv) the archiving possibilities on open data

repositories, and (v) self-implementable applications with the system (Barillari et al. 2016). Currently available project management tools connecting classical, collaborative project, and data management, include platforms like OSF (https://osf.io/). Classical ELN systems range from commercial applications like *CERF* (https://cerf-notebook.com/), *Benchling* (https://www.benchling.com/), *labfolder* (https://www.labfolder.com/), to open-source-based ones like *chemotion* (https://www.chemotion.net/chemotionsaurus/), *eLabFTW* (https://www.elabftw.net/), and *openBIS* (https://openbis.ch/).

3. CURRENT SITUATION, METHODOLOGY, AND STRATEGY

3.1. CURRENT SITUATION FROM AN INTERDISCIPLINARY POINT OF VIEW

The scope of Open Science and research data management, as well as a clear and well-defined data stewardship and governance, has been clearly recognized from the perspective of the DFG-funded *Collaborative Research Centre (CRC) 1411 Design of Particulate Products*. However, an ELN-LIMS is currently lacking for the use within an interdisciplinary field between engineering, material sciences, natural sciences, mathematics, theoretical modeling, and simulation. Furthermore, commercial systems are generally not recommended by the possible financial conditions, in addition to a thereby complementary Open Science policy. Open-source products like *chemotion* are, again, too subject-specific (in this particular case, chemistry), which would lead to a less beneficial impact on interdisciplinary work and research, such as lower usability, and, thus, in the end, also acceptance in other subject groups. On the other hand, the *openBIS* system of the *Department of Biosystems Science and Engineering and Biology* of the *ETH Zurich* (Barillari et al. 2016; Bauch et al. 2011) shows promising features that are worth a closer look, even though this system was designed primarily for biological and medical disciplines and is not only used by academic institutions, but also by companies in the industry sector (Bauch et al. 2011).

Accordingly, we describe our system in several subsections, starting with the basic principles and working features of *openBIS*, continuing with the goals and requirements of researchers within the *CRC* and the scientific community for a well-functioning ELN tool, and ending with its implementation and usage. *OpenBIS* is, besides other ELN-LIMS examples, a good starting point and framework to foster cooperativity via a digital environment using an electronic lab notebook version and by fulfilling the requirements by the *DFG* and *NFDI* consortia, such as *FAIRmat* (for materials science, physics, chemistry, and mathematics) with respect to data management, handling, storage, and publishing. In addition, we also want to share our experiences in further developing the system and its implementation and daily use.

3.2. GENERAL OVERVIEW AND STRUCTURE OF OPENBIS

Before we shed light on how *openBIS* can help us with a successful implementation of a beneficial RDM, we first clarify i) what *openBIS* is, ii) how it works, and iii) what technical requirements need to be met. *OpenBIS* is an open-source platform that functions both as an Electronic Laboratory Notebook (ELN) and as a Laboratory Information Management System (LIMS). Developed by ETH Zurich, *openBIS* provides an open-source database for research laboratories, especially designed and implemented in and for life sciences (Barillari et al. 2016). The goal was to build a simple and efficient, yet comprehensive ELN-LIMS system that meets the daily needs of a research institution. Everyday things like storage of materials, instrumental setups and devices, the acquisition, description, and processing of large amounts of data, and the forwarding to possible users and scientists within the *openBIS* system should be possible (Bauch et al. 2011).

However, questions arise regarding what kind of technical conditions are needed to build *openBIS* within a research group, whether *openBIS* is scalable, and whether such system in the area of a scientific network (EU or DFG fund), or even a comprehensive database within an entire university, is conceivable at all. All of these questions are important to know before starting with an ELN-LIMS and will be discussed in the following subsections.

The history of *openBIS* and its platform, on which it is based, started in 2007 and is still actively maintained, nowadays by the *Scientific IT Service Team* of the *ETH Zurich*. As operating system (OS), *openBIS* requires a modern Unix-like OS, for instance Linux systems. However, *openBIS* can

be run on virtual machines and docker containers and therefore is mainly platform-independent. A very interesting and detailed look on the general technical background of *openBIS* is provided by the developers, accordingly (Bauch et al. 2011). Therefore, shortly summarized, in *openBIS*, one or several *Data Store Server*(s) (*DSS*) and an *Application Server* (*AS*) are introduced. On the *AS*, data provenance actions like metadata handling are conducted, while on the *DSS*, the raw data is managed. While the user access is facilitated via a web browser, the *AS* uses a *Relational Database Management System* (*RDBMS*) to generate persistent information like index information about all data sets, the data itself is covered and stored within the (several) *DSS* system(s). The latter is responsible for creating, querying, and visualizing data while they are mediated by the *AS* (Bauch et al. 2011). At the application side, the ELN-LIMS system is accessible via browser-based tools (the recommended ones are Chrome, Firefox, and Safari) and is reachable from any electronic device and operating system.

Moreover, as part of our INF project of the *CRC 1411* (see also Funding information), we want to further extend the concept of ELN-LIMS to a more virtual research environment in the following years (see section 4.2 for further information). Our goal is for VREs to be collaborative and present requirements-tailored tools that support web-based research environments (Allan 2009; Candela et al. 2013; Lave & Wenger 1991). The DFG defines "Virtuelle Forschungsumgebung" (a literal translation of 'Virtual Research Environment') as a platform for internet-based collaborative working that enables new ways of collaboration and a new way of dealing with research data and information (Reimer & Carusi 2010).

4. RESULTS

4.1. FEATURES OF OPENBIS AS ELN-LIMS SYSTEM WITHIN CRC 1411

As *openBIS* is accessible for users via a web app, and even from outside the university network, access to the user's own data and data provided by others is guaranteed always and everywhere. It is not only crucial to be able to create and implement (meta-)data, but doing so further enables the function that every user can permit access to its projects to selected users. This covers different *roles* (observer, user, admin) with different rights like reading, writing, or even deleting. Here, the authorization process lies completely in the hands of each user and can be adjusted independently–with respect to other users within the network or other projects by the same user. In general, *openBIS* has a predefined hierarchical structure, which is divided into several levels. The first and main level is the *Work* or *Data Space* level, which, in addition to general rights and access, primarily contains all *Projects*. The second level pertains to the projects themselves. These, in turn, have *Collections* (third level), which consist of different *Object Types* (fourth level), with/without *Datasets* (see Figure 1), such as *Experimental Step*, *Entry*, or *Instrument* (Bauch et al. 2011). This is also valid for *Collections*. In fact, the first and second level contain a persistent identifier with *one* specific path using the nomenclature "*/user(workspace)/projectx*". *Collections* and *Objects* extend the identifier by a unique code. However, these are moveable between *Projects* and *Spaces*. In the example of Figure 1, we can move *Collection 1* with all its entries to *Project 2*, which can be shared, but the *Projects* and *Spaces* are not moveable.
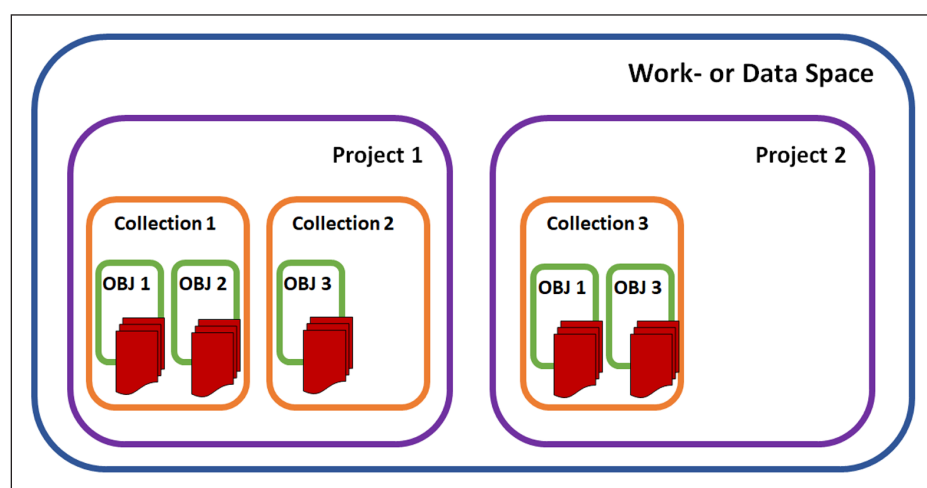

Figure 1 Logical structure layers of *openBIS* containing the *Work-* or *Data space*, *Projects*, *Collections*, *Object Types* (*OBJs*), and attached *Datasets* (red). This figure is adapted from Barillari et al. (2016).

There are no limitations on the maximum number of users working on one *openBIS* ELN system. The only limitations are connected to the available and used server resources. It can also link data on samples, materials, instruments, and experiments. Here, *openBIS* features the *parent-children relationship* model (see Figure 2 for further detail): this means that every created *Object type* is logically interconnected between other *Object types*. As an example of our *CRC*, synthesis and characterization groups are trying to develop specific and well-defined particles considering different synthetic samples, materials, and processes. To ensure that the creation, modification, or deletion of any electronic records is traceable, computer-generated and time-stamped audit trails are used to record the date and time of any user interventions. As an example, to create an *Experimental Step X* and *Y* (see Figure 2), the researcher/user requires a different amount of *Samples A* and *B*, and one specific *Instrument I*. Moreover, for a specific *Simulation Z*, one uses *Sample B* and *Software S*. Here, *Sample*, *Instrument*, *Software*, and the *Experimental Step* and *Software* or later on *Publication* are all different created *Object types* that are accessible and existing for every user of our *openBIS* system. Furthermore, all *Object Types* within our system are classified into three different categories and differ if they are generalized entries (*Entry* and *General Type*), experimental procedures or analysis (like *Experimental Step* or *Simulation*), or properties and (real existing) objects (like *Instrument* or *Software*). In our basic examples, it is valid that *Sample A* and *B*, *Instrument I*, and *Software S* are all *parents* of the *Experimental step X*, *Y*, and *Simulation Z*, respectively, which are automatically the *child(ren)* of the prior *Object types*. If we now create an *Object type Publication P*, which is linked and fed via our *Experimental step X*, *Y*, and *Simulation Z*, then, accordingly, *P* is the *child* of *X*, *Y*, and *Z*, and vice versa. This results in a clear and well-understandable line of ancestry, which fulfills the FAIR concept. In our example, even after publishing (*Publication P*), one can clearly find and reuse(meta-)data of the corresponding project. Moreover, one does not have to create object types (and their underlying (meta-)data multiple times, as one can reuse them if needed. This reduces redundancy of (meta-)data that is used across *Spaces*, *Projects*, and *Collections*, and saves storage and data maintenance time. Additionally, in an interdisciplinary environment, multiple groups utilize a common set of *Object types*, which would otherwise appear in a redundant manner in different fields. This can save a lot of time and lead to the possibility to create and implement a more comprehensible laboratory and research system.
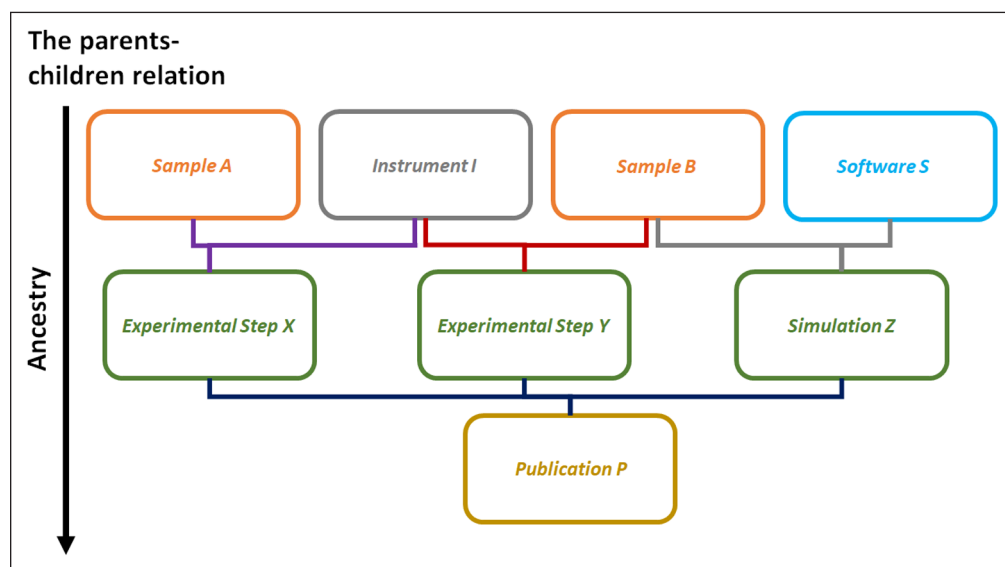

Figure 2 The hierarchical structure of the *parents-children relation* in action. Every object type (for example *Sample*, *Experimental Step*, and *Publication*) is *child* and *parent* depending on its ancestry relation. In our example, *Publication P* is child of every other object type like *Instrument I* or *Experimental Step X*.

OpenBIS can also be used for administrative and lab-specific workflow processes. To understand how to do this, we have to take one step back and consider how *openBIS* works from each user's point of view. First, every user possesses its own *Workspace*, which can be compared to its own desktop or computer storage. Within this *Workspace*, projects, metadata, and others, can be created and adjusted without necessarily being connected to other users. This is possible using the *Manage Access* option, a function that allows users to simply grant access to those users within the system who should be allowed to read, write, and/or delete information in your project. All permitted users will appear as folder icons below the own *Workspace* folder icon. However, one researcher or user is connected to a specific working group or institute containing

several other researchers/users. Finding a way to bring the entire working group together without repeatedly pressing the *Manage Access* button and effectively defining user roles is a key consideration. Especially for general administrative or instruments within the institute, which are accessible for every researcher of the corresponding group, creating for example several *Instrument* object types is not only time-consuming, but it also holds the danger of creating different metadata for the same, here exemplarily, instrument. This would result in problems with the FAIR principle.

Moreover, *openBIS* includes an additional working area called *Inventory*, which serves as a 'common work environment' for user-groups to access information and data stored within this section. This area houses data and information that is universally relevant or of interest to all users or larger groups, and it contains spaces for the *CRC*, such as an *openBIS* on-boarding or a common folder. Furthermore, we have made further adjustments to the *Inventory* section by implementing sub-*Inventories* that are specific to the working groups/institutes within our *CRC*.

These different spaces, including the *Inventory* and our institute-based *Inventories*, provide fields for collaborative projects and administrative workflows, offering a general and overall usage area, along with a sub-*Inventory* containing commonly used, institute-dependent instruments and materials, among others. With a collaboratively-used *Inventory*, each user has access to collections of items like Instrumentation, where necessary object types (e.g., *Instrument I1*, *I2*, …) are created once and can be utilized by any user within the workgroup (those with access permission to the specific folder of the *Instrumentation* section). The same concept applies to collections of samples or documents/protocols that are employed in large collaborative projects.

By following the principle of the parent-child relation depicted in Figure 2, we ensure that instruments (e.g., Instrument I from our example) within an Instrumentation folder can now be used for tracking as part of the *Inventory* folder, while avoiding redundant information and directly providing the required documentation. This makes it possible to combine a classical working group environment with interdisciplinarity and FAIR data management within an overall framework.

4.2. OPENBIS AS A MULTI-INSTITUTIONAL FRAMEWORK WITHIN A MATERIAL SCIENCE-BASED ENVIRONMENT: ADJUSTMENTS AND EXPERIENCES

This multi-institutional framework can now be implemented and customized to specific user needs. Via an additional overlay (*Core UI*), the implementation of different object types (from simple general types to specifically customized types) is possible. The implementation process will involve questions regarding the number and types to be implemented. Whether more specific and fine-tuned types are preferred or not needs to be clarified in advance. In our case, we shifted from specific object types (as the pre-defined openBIS system for a bio-medical environment provides with, e.g., bacteria, yeast etc.) to more generalized object types such as sample or simulation, which are usable in an interdisciplinary environment by multiple groups at once. We decided to use *openBIS* as foundation or 'hollow' version, while winnowing object types that exist by default but may be not required in our digital illustration of our *CRC*-based workflows, and to develop new ones using our own *CRC*-based ontology, accordingly. To achieve this, we decided to form a pilot group for all users within the *CRC 1411* about six months before handing the *openBIS* system over to the entire group. Since our interdisciplinary working environment consists of different disciplines, from pure particle synthesis to their characterization, analysis, simulation, and theoretical optimization, we invited about fifteen people from the different working groups/disciplines and implemented, discussed, and modified our preliminary system up to this point, so that everyone was satisfied with the result. It turned out that implementing dozens of different and fine-tuned object types for each user, while appearing functional at first sight, also brings some disadvantages, because while each object type is visible in the drop-down menu, not every object type is used by every user. For example, theoretical scientists, who use mathematical models to simulate and predict physical properties of potential nanoparticles, rarely use synthesis-based object types like *General Protocol* or *Experimental Step*. This means that any user from a particular discipline will only stick to the types that are useful for their own research. Furthermore, since each type is always present in the system's drop-down menu, it quickly becomes cluttered and confusing, leading

to poorer usability and, consequently, a less accepted package. Therefore, we finally decided to use only eleven object types, two of which are already preset by the developers (*Entry* and *Experimental Step*), eight of which contain specific (meta-)data, and one which is defined as an overall or general object type (*General Type*).

Meeting requirements (such as scientific or administrative needs) that may exist within the same discipline can be a challenge. Again, our implemented multi-institutional framework could help. Now, since each workgroup/environment has its own 'space' within the overall *openBIS* system, it is possible to implement specific 'workgroup-defined object types'. For example, in our *CRC 1411*, synthesis and processes can vary between workgroups, so implementing finely tuned objects, which can only be used by users within that workgroup or users with access privileges, is preferable. Furthermore, there are no negative effects, such as usability on interdisciplinary work or traceability of (meta-)data, since, within the entire *openBIS* system (and users with access rights), the new (and now workgroup-specific) types can still be viewed. As a result, this leads to a balanced combination of a universal and well-usable *openBIS* data management system and a well-tailored work environment.
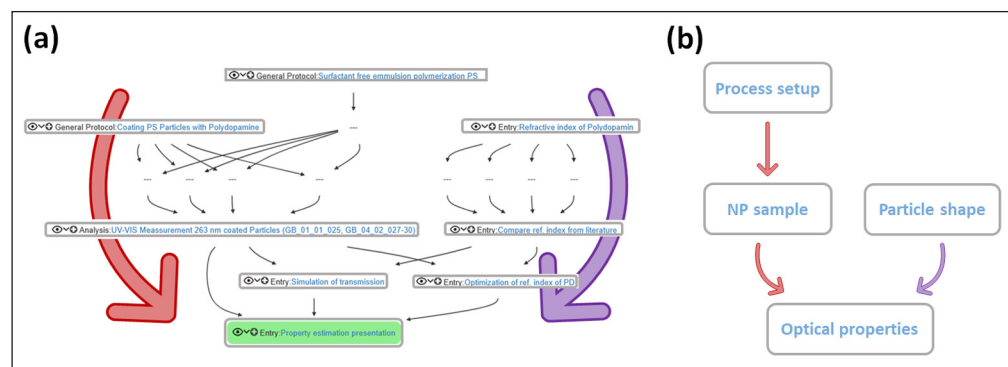

Figure 3 The current joint venture project within *CRC 1411* involves collaboration between synthesis (project process related to the red arrow) and theoretical simulation (project process related to the red arrow). The 'real' scheme (hierarchy graph) in (a) provides a comprehensive representation of complete process workflows during this cooperative project. For better clarity and visibility, we have filtered and shortened (a), while (b) illustrates the proposed structure of our *CRC* in relation to the joint venture project between our synthetic (red arrow) and theory (violet arrow) research groups. Both the proposed (b) process structure, and the implemented (a) structure, exhibit similarities even though (a) is more complex overall.

Now each user can edit and share their own data, both scientific and non-scientific in nature/origin, with themselves or with other users within the *openBIS* system by creating and storing data and metadata that can be exported in turn. An exemplary workflow, in which data are shared and exchanged among CRC1411 users, is illustrated in Figure 3. The reasons for this can be quite different: more and more journals are demanding that special attention must be paid to Open Science and research data management, for instance, by using a data repository to which publication-specific (meta)-data must be uploaded and made freely available to all. In some journals, for example, it is common to prepare a data availability statement when submitting the publication. This statement includes information about the data itself, where to find it, an identifier for the data (DOI or persistent identifier), and how to access the data. The function used for re-exporting stored and listed (meta)data in *openBIS* is the *Export Metadata & Data* function. The *openBIS* server now plays the role of a kind of e-mail distributor. That is, *openBIS* exports the project folder selected by the user as a *.zip* file, which now contains all (meta)-data, as *.txt*, *.doc*, *.html*, and *.json* files, as well as the introduced structure by the researcher and the system. In fact, the system uses a terminology adapted for our workflows within the *CRC*. This means that the default syntax and terminology of our *openBIS* system, along with the modifications and additions made to display workflows for various research areas within our *CRC*, will be included in the *.zip* file. The management of the researcher's data will therefore be displayed in the folder as well. Since each user has its own account in *openBIS*, the user can send the exported file to others directly to their e-mail inbox via a download link, with no data size limitations, by entering their e-mail address. This *.zip* file can now be attached to a publication as Supplementary or Supporting Information, or uploaded to a data repository, and the DOI obtained can be noted in the actual publication. This closes the entire research data management data lifecycle, from project inception through data production and analysis to long-term archiving. Moreover, as stated above, the pre-structured organization of your project(s) on *openBIS* (and via the *Object types*) stays the same after export and may be re-imported by writing a parser to another ELN system, such as *openBIS*. That means that no additional structuring process is needed. Furthermore, after implementation of the gathered publication or research papers that have been published under the *CRC 1411* roof, we track all of them via *openBIS* now as well.

As mentioned at the beginning, the ELN-LIMS system *openBIS@CRC* is designed to adhere to FAIR principles. This is evident in various aspects of the system. Firstly, the data stored in the *openBIS* database can be easily found through persistent identifiers and a search function. Additionally, the data is easily accessible via the internet without the need for a virtual private network (VPN), allowing users to access the webapp from anywhere; the webapp GUI is illustrated in Figure 4. The system is also interoperable, allowing external and internal scripting to interact with the data. Moreover, the data and metadata can be exported, making it reusable. Alongside the FAIR principles, the system also promotes good scientific practice and follows the data management lifecycle, covering aspects such as documentation, tracking of projects and data, and archiving. Collaboration is facilitated through role management and the availability of separate workspaces for individuals and working groups within *openBIS*.
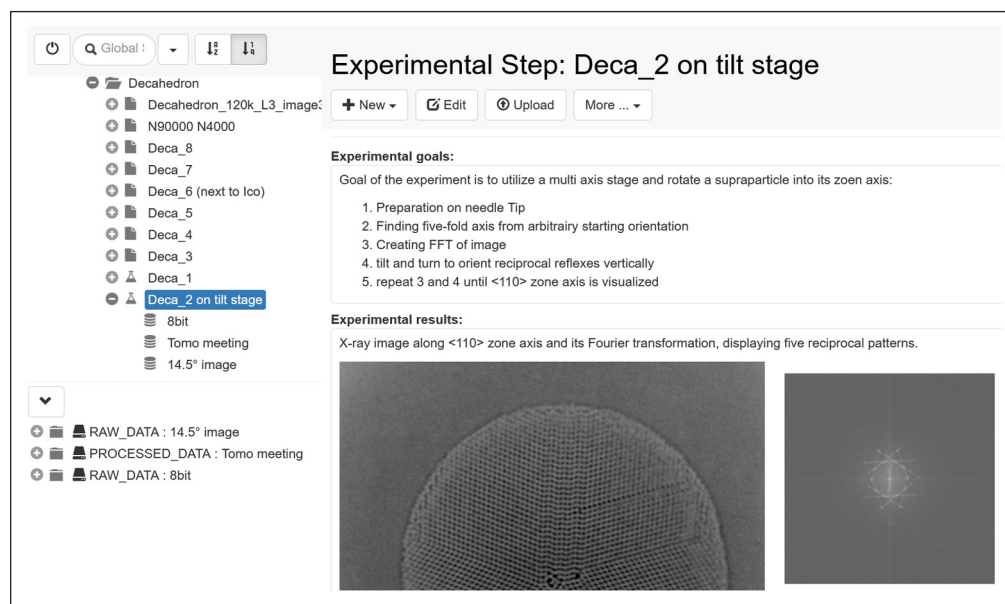

Figure 4 The depicted view represents the graphical user interface of *openBIS* from the researcher's perspective. The top-left section clearly displays the hierarchical folder structure shown in Figure 1, while the right side allows the researcher to fill in relevant metadata and to link it to the corresponding raw or processed data, as shown in the bottom-left part.

Our current *openBIS* system has been up and running since end of December 2021, but we are still testing new features and making minor and major changes. Following the rollout of the 'final' version of the *openBIS* system to our users within the *CRC 1411*, we have organized several introductory events within our *iRTG* (integrated Research Training Group). In addition, we host monthly support meetings where questions or comments can be brought forward, and support is offered. Furthermore, we did not want to develop an exclusive data management system, so young research students, for instance in the context of their Bachelor's or Master's thesis, will have access as new users to get in touch with research data management. In addition, the purchase of lab notebooks and tablets has increased the adoption of *openBIS*, as data creation, processing, and sharing within the labs is now very fast and easy. The use of paper-based laboratory notebooks has been decreasing significantly over time.

Our *CRC* aims to expand the capabilities of the ELN-LIMS concept in the future. As described in section '*3.1 Current situation, methodology, and strategy*', our vision of a virtual research environment includes integration of data repositories like Zenodo, enhanced visualization options including AR/VR technologies, and post-processing tools for stored raw data. This can be achieved through Python scripting using openBIS's internal API, enabling the implementation of scripts and other post-processing functionalities. The flexibility and range of possibilities offered by *openBIS* played a significant role in our decision to select this ELN-LIMS for our *CRC*.

One specific example of collaboration and post-processing within our *CRC* involves the implementation of a script for color calculations of nanoparticles. This script will assist our synthesis groups in developing nanoparticle synthesis routes even before they begin their experiments. By utilizing the *openBIS* server, the synthesis groups can directly access the theoretical calculations typically performed by our theory groups, eliminating any time delays. This iterative process of development, deployment, implementation, and utilization accelerates research within our *CRC* and highlights the value of the *openBIS* ELN system. Furthermore, we are planning to integrate a Jupyter Hub system with our *openBIS@CRC* system in the

5. CONCLUSIONS

We demonstrate an ELN-LIMS solution that addresses the challenges and needs of multiple disciplines within a technical and natural science faculty on the *openBIS* system. In fact, the versatility and modifiability of the system formed the basis for a system for multiple working groups of the *CRC 1411 Design of Particulate Products* that meets the documentation needs of groups focusing on synthesis, characterization, and simulation, and, in particular, supports collaboration between different disciplines within the *CRC*. The structure was developed and adapted via several steps and provides both a common set of *Object types* and the ability to include more specific ones for individual working groups without disrupting common workflows. As a result, we discovered that the balance between providing users with flexibility and customization in an ELN-LIMS system and maintaining a commonly used and understandable interface, particularly concerning the number of object types, is delicate and requires careful consideration. Indeed, both inadequate and overwhelming design and functionality lead to a reduction in potential use by researchers and, therefore, require accurate design and communication between users and the administrators of the ELN-LIMS, especially in a highly interdisciplinary project like the presented *CRC*. Ultimately, this ELN-LIMS could serve as a template solution for other similarly structured collaborative research centers or research groups.

SUPPLEMENTARY MATERIALS

The aim of this research article is to show not only what *openBIS* can do in the interdisciplinary research environment, but also how metadata and data are ultimately presented. Therefore, as an example project, the development of the research article *'Using openBIS as virtual research environment: An ELN-LIMS open-source database tool as a framework within the CRC 1411', 'Design of Particulate Products'* was connected via *openBIS* to the necessary/received metadata and data and a clear structure was developed. The data received through *openBIS* as exported *.zip* files contain four different data types of each existing/used object type and folder, organized in a clear folder-subfolder structure. Typical standard data types used here include *.doc*, *.txt*, *.json*, and *.html*. Attached data (as *Datasets*) to its corresponding object type is additionally linked as subfolder. DOI: https://doi.org/10.5334/dsj-2023-044.s1

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

FP and SE contributed equally to this work. FP and SE prepared the initial draft of the manuscript and all authors reviewed and revised the manuscript prior to submission, as well as contributed to the conceptualizing of the paper. All authors contributed to the research and investigation process. FP and SE revised the manuscript to address recommendations offered by the anonymous Data Science Journal reviewers and all authors reviewed the revised manuscript.

AUTHOR AFFILIATIONS

Fabian Plass
Competence Unit for Scientific Computing (CSC), Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany

Silvan Englisch
Institute of Micro- and Nanostructure Research (IMN) & Center for Nanoanalysis and Electron Microscopy
(CENEM), IZNF, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany

Benjamin Apeleo Zubiri orcid.org/0000-0003-0498-7271
Institute of Micro- and Nanostructure Research (IMN) & Center for Nanoanalysis and Electron Microscopy
(CENEM), IZNF, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany

Lukas Pflug orcid.org/0000-0001-8001-5832
Competence Center Scientific Computing (CSC), Friedrich-Alexander-Universität Erlangen-Nürnberg,
91058 Erlangen, Germany; Department of Mathematics, Chair of Applied Mathematics (Continuous
Optimization), Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany

Erdmann Spiecker
Institute of Micro- and Nanostructure Research (IMN) & Center for Nanoanalysis and Electron Microscopy
(CENEM), IZNF, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany

Michael Stingl orcid.org/0000-0002-3626-0723
Competence Center Scientific Computing (CSC), Friedrich-Alexander-Universität Erlangen-Nürnberg,
91058 Erlangen, Germany; Department of Mathematics, Chair of Applied Mathematics (Continuous
Optimization), Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany

REFERENCES

Allan, RJ. 2009. *Virtual research environments: from portals to science gateways*. 1. publ. Oxford: Chandos
Publ.

Barillari, C, Ottoz, DS, Fuentes-Serna, JM, Ramakrishnan, C, Rinn, B and **Rudolf, F.** 2016. openBIS ELN-
LIMS: an open-source database for academic laboratories. *Bioinformatics*, 32(4): 638–640. DOI:
https://doi.org/10.1093/bioinformatics/btv606

**Bauch, A, Adamczyk, I, Buczek, P, Elmer, F-J, Enimanev, K, Glyzewski, P, Kohler, M, Pylak, T,
Quandt, A, Ramakrishnan, C** and **others.** 2011. OpenBIS: a flexible framework for managing and
analyzing complex data in biology research. *BMC bioinformatics*, 12(1): 1–19. DOI: https://doi.
org/10.1186/1471-2105-12-468

Bauernhansl, T, Ten Hompel, M and **Vogel-Heuser, B.** 2014. *Industrie 4.0 in Produktion, Automatisierung
und Logistik – Anwendung · Technologien · Migration*. Wiesbaden: Springer Fachmedien Wiesbaden.
DOI: https://doi.org/10.1007/978-3-658-04682-8

Besançon, L, Peiffer-Smadja, N, Segalas, C, Jiang, H, Masuzzo, P, Smout, C, Billy, E, Deforet, M and
Leyrat, C. 2021. Open science saves lives: lessons from the COVID-19 pandemic. *BMC Medical
Research Methodology*, 21(1): 117. DOI: https://doi.org/10.1186/s12874-021-01304-y

Bespalov, A, Michel, MC and **Steckler, T.** (eds.) 2020. *Good research practice in non-clinical pharmacology
and biomedicine*. Cham: Springer International Publishing. DOI: https://doi.org/10.1007/978-3-030-
33656-1

Brous, P, Janssen, M and **Vilminko-Heikkinen, R.** 2016. Coordinating decision-making in data
management activities: A Systematic review of data governance principles. In: Scholl, HJ, Glassey, O,
Janssen, M, Klievink, B, Lindgren, I, Parycek, P, Tambouris, E, Wimmer, MA, Janowski, T and Sá Soares,
D (eds.), *Electronic Government*. Cham: Springer International Publishing. pp. 115–125. DOI: https://
doi.org/10.1007/978-3-319-44421-5_9

Candela, L, Castelli, D and **Pagano, P.** 2013. Virtual Research Environments: An overview and a research
agenda. *Data Science Journal*, 12(0): GRDI75–GRDI81. DOI: https://doi.org/10.2481/dsj.GRDI-013

Clayton, EW, Evans, BJ, Hazel, JW and **Rothstein, MA.** 2019. The law of genetic privacy: applications,
implications, and limitations. *Journal of Law and the Biosciences*, 6(1): 1–36. DOI: https://doi.
org/10.1093/jlb/lsz007

Crosas, M. 2011. The dataverse network®: an open-source application for sharing, discovering and
preserving data. *D-lib Magazine*, 17(1/2). DOI: https://doi.org/10.1045/january2011-crosas

Egger, J and **Masood, T.** 2020. Augmented reality in support of intelligent manufacturing – A systematic
literature review. *Computers & Industrial Engineering*, 140: 106195. DOI: https://doi.org/10.1016/j.
cie.2019.106195

ERC Scientific Council. 2021. Open research data and data management plans. Available at https://erc.
europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_
Management_Plans.pdf [Last accessed 15 February 2022].

European Commission. Directorate General for Research and Innovation. 2016. *Realising the European
open science cloud: first report and recommendations of the Commission high level expert group
on the European open science cloud*. LU: Publications Office. Available at https://data.europa.eu/
doi/10.2777/940154 [Last accessed 13 July 2023].

German Research Foundation. 2021. Guidelines collaborative research centres. Available at https://www.
dfg.de/formulare/50_06/50_06_en.pdf [Last accessed 17 October 2022].

Gisler, M. (ed.) 2001. *E-Government. 1: Eine Standortbestimmung / [1. Schweizer e-Government-Symposium am 22. August 2000 in Zürich]. Michael Gisler … (Hrsg.)*. 2., aktualisierte Aufl. Presented at the Schweizer eGovernment-Symposium, Bern Stuttgart Wien: Haupt.

Hildebrand, K, Gebauer, M, Hinrichs, H and **Mielke, M.** (eds.) 2011. *Daten- und informationsqualität*. Wiesbaden: Vieweg+Teubner. DOI: https://doi.org/10.1007/978-3-8348-9953-8

Hummel, P, Braun, M, Tretter, M and **Dabrock, P.** 2021. Data sovereignty: A review. *Big Data & Society*, 8(1): 205395172098201. DOI: https://doi.org/10.1177/2053951720982012

Janssen, M, Charalabidis, Y and **Zuiderwijk, A.** 2012. Benefits, Adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4): 258–268. DOI: https://doi.org/10.1080/10580530.2012.716740

Kimmig, J, Zechel, S and **Schubert, U S.** 2021. Digital Transformation in Materials Science: A paradigm change in material's development. *Advanced Materials*, 33(8): 2004940. DOI: https://doi.org/10.1002/adma.202004940

Kraft, A, Razum, M, Potthoff, J, Porzel, A, Engel, T, Lange, F, Van den Broek, K and **Furtado, F.** 2016. The RADAR Project—A Service for Research Data Archival and Publication. *ISPRS International Journal of Geo-Information*, 5(3). DOI: https://doi.org/10.3390/ijgi5030028

Ladley, J. 2020. *Data governance: how to design, deploy and sustain an effective data governance program*. Second edition. London [England]; San Diego, CA: Academic Press.

Lasi, H, Fettke, P, Kemper, H-G, Feld, T and **Hoffmann, M.** 2014. Industry 4.0. *Business & Information Systems Engineering*, 6(4): 239–242. DOI: https://doi.org/10.1007/s12599-014-0334-4

Lave, J and **Wenger, E.** 1991. *Situated learning: legitimate peripheral participation*. Cambridge [England]; New York: Cambridge University Press.

Lecarpentier, D, Wittenburg, P, Elbers, W, Michelini, A, Kanso, R, Coveney, P and **Baxter, R.** 2013. EUDAT: a new cross-disciplinary data infrastructure for science. *International Journal of Digital Curation*, 8(1): 279–287. DOI: https://doi.org/10.2218/ijdc.v8i1.260

Li, S, Xu, LD and **Zhao, S.** 2015. The internet of things: a survey. *Information Systems Frontiers*, 17(2): 243–259. DOI: https://doi.org/10.1007/s10796-014-9492-7

Machina, HK and **Wild, DJ.** 2013. Laboratory Informatics Tools Integration Strategies for Drug Discovery: Integration of LIMS, ELN, CDS, and SDMS. *SLAS Technology*, 18(2): 126–136. DOI: https://doi.org/10.1177/2211068212454852

Mons, B, Neylon, C, Velterop, J, Dumontier, M, da Silva Santos, LOB and **Wilkinson, MD.** 2017. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use*, 37(1): 49–56. DOI: https://doi.org/10.3233/ISU-170824

National Academies of Sciences, Engineering, and Medicine (U.S.), National Academies of Sciences, Engineering, and Medicine (U.S.), National Academies of Sciences, Engineering, and Medicine (U.S.), and **National Academies of Sciences, Engineering, and Medicine (U.S.).** (eds.) 2018. *Open science by design: realizing a vision for 21st century research*. Washington, DC: The National Academies Press.

Reimer, T and **Carusi, A.** 2010. Virtual research environment collaborative landscape study. Imperial College London. DOI: https://doi.org/10.25561/18568

Tallon, PP. 2013. Corporate governance of big data: perspectives on value, risk, and cost. *Computer*, 46(6): 32–38. DOI: https://doi.org/10.1109/MC.2013.155

Tse, EG, Klug, DM and **Todd, MH.** 2020. Open science approaches to COVID-19. *F1000Research*, 9: 1043. DOI: https://doi.org/10.12688/f1000research.26084.1

Wilkinson, MD, Dumontier, M, Aalbersberg, Ij J, Appleton, G, Axton, M, Baak, A, Blomberg, N, Boiten, J-W, da Silva Santos, LB, Bourne, PE and **others.** 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific data*, 3(1): 1–9. DOI: https://doi.org/10.1038/sdata.2016.18