Taylor & Francis
Taylor & Francis Group

Synchrotron Radiation News

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/gsrn20

Specifications for Electronic Laboratory Notebooks (ELN) in the Photon and Neutron Community

P. Jordt, M. Osterhoff, Y. Tymoshenko, B. Hakim, P. Dolcet, F. Maurer, V. Biniyaminov, L. Amelung, F. Dall'Antonia, J.-D. Grunwaldt, F. Weber, W. Lohstroh & B. M. Murphy

To cite this article: P. Jordt, M. Osterhoff, Y. Tymoshenko, B. Hakim, P. Dolcet, F. Maurer, V. Biniyaminov, L. Amelung, F. Dall'Antonia, J.-D. Grunwaldt, F. Weber, W. Lohstroh & B. M. Murphy (07 Jan 2025): Specifications for Electronic Laboratory Notebooks (ELN) in the Photon and Neutron Community, Synchrotron Radiation News, DOI: 10.1080/08940886.2024.2432265

To link to this article: https://doi.org/10.1080/08940886.2024.2432265

Published online: 07 Jan 2025.

OPEN ACCESS TECHNICAL REPORT

Specifications for Electronic Laboratory Notebooks (ELN) in the Photon and Neutron Community

P. JORDT,[1,2] M. OSTERHOFF,[3] Y. TYMOSHENKO,[4] B. HAKIM,[6] P. DOLCET,[4,7] F. MAURER,[5] V. BINIYAMINOV,[4] L. AMELUNG,[8] F. DALL'ANTONIA,[9] J.-D. GRUNWALDT,[5] F. WEBER,[4] W. LOHSTROH,[10] AND B. M. MURPHY[1,2]

[1]*Institut for Experimental and Applied Physics, University of Kiel, Germany*
[2]*Ruprecht-Haensel-Laboratory, Germany*
[3]*Institute for X-ray Physics, University of Göttingen, Germany*
[4]*Institute for Quantum Materials and Technologies, Karlsruhe Institute of Technology, Karlsruhe, Germany*
[5]*Institute for Chemical Technology and Polymer Chemistry, Karlsruhe Institute of Technology, Germany*
[6]*Institut for Crystallography and Structural Physics, University of Erlangen, Germany*
[7]*University of Padua, Italy*
[8]*DESY, Hamburg, Germany*
[9]*European XFEL, Hamburg, Germany*
[10]*Forschungsneutronenquelle Heinz Maier-Leibnitz (FRM II), Technical University of Munich, Germany*
jordt@physik.uni-kiel.de

Introduction

Electronic Laboratory Notebooks (ELN)[1] are a building block of the recently establishing FAIR[2] data management procedures in research and the publication process of research data, by documenting the experimental details, data collection and analysis process. They enable the capture and management of experimental results in a collaborative approach. Furthermore, they provide an interlinking with the utilized techniques and holding information's on the investigated specimens. Thus, ELNs are the entrance point to an experiment for any scientist keen to understand and reproduce an experiment. The Data from Photon and Neutron Experiments[3,4] (DAPHNE4NFDI) data life cycle is depicted in Figure 1, ELNs are part of the documentation step during the data collection. DAPHNE4NFDI is one of the 26 consortia within the German National Research Data Infrastructure (Nationale Forschungsdaten Infrastuktur[5], NFDI). The aims of this initiative are the development and improvement of data management tools for research data from large scale Photon and Neutron (PaN) research infrastructures; similar activities are undertaken at other PaN facilities at EU[6,7] level and worldwide.[8] Research at PaN facilities is always a joint undertaking of the facilities operating the instruments and the external user groups which poses additional challenges for collaborative research data management and comprehensive experiment documentation.

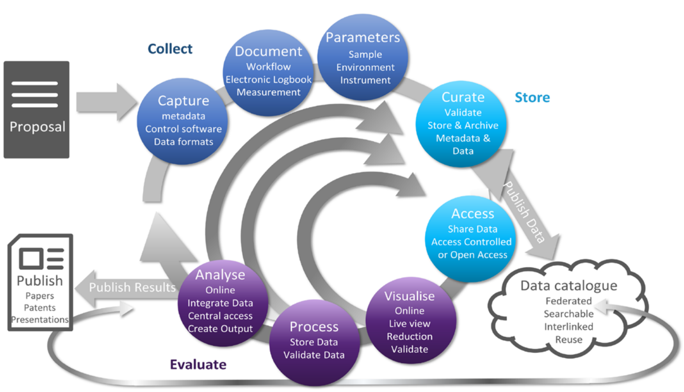

Figure 1: DAPHNE4NFDI FAIR data life cycle for data collected at photon and neutron sources.

The DAPHNE4NFDI consortium brings together partners from the photon and neutron user community, conducting experiments at large scale facilities, and the facilities itself, operating the sources and instruments necessary for these experiments.

In recent years, the dissemination of ELNs has significantly increased. They have evolved into a powerful tool for experimental record, due to the development and integration of new features that sets them apart from conventional paper notebooks. All popular ELN solutions have in common that they enable real time collaboration, data integration, fast search mechanism, APIs and mobile accessibility. One may differentiate between multiple available solutions by dividing them into closed and open-source projects and into self-hosted versus cloud solutions, but many of these solutions provide a similar basic foundation. There are many obstacles in the process of finding a sustainable ELN solution for a particular use case, these may be the costs, migration and integration of legacy notebooks, user training to exploit the full potential, data security and integrity concerns and long-term support.

This paper aims to give a concise overview of possible specifications, that can be of use in a modern ELN for use at large scale facilities. In the context of this paper, a *specification* denotes a single functionality/feature/specification of an ELN that is necessary for it to either be usable, enables a higher productivity, or an enhanced user experience. This document has a special emphasis towards solutions for large photon and neutron research infrastructures[9] and special requirements arising from deploying ELN solutions at these infrastructures, e.g. automation of information ingestion, huge workload, integration with additional technologies and user access management. Specifications range from *indispensable for the deployment at PaN facilities*, to, *enabling a productive user interface and user experience*. A ranking by the importance of the different specifications is included at the time of writing bearing in mind that the demands and capabilities on experiments are constantly changing due to the digital transformation and rapidly increasing data collection rates. Further, this paper presents a guideline for the evaluation of existing solutions and the development of new ELNs to come in the future. Additional details can be found in the supplementary information[10].

Specifications for ELN

In the context of ELN, one may discuss specifications of features, options, functionality and other terms, describing similar concepts. To describe technologies, ideas, or standards that are relevant for ELNs. Here, we will use the term *specification* in this sense. In this project, the process of finding and defining ELN specifications was split into two paths. First, existing solutions were tested to gather insights and ideas, what may be of interest and what is absolute key to a user friendly and productive ELN. This was, undertaken to understand what kind of solutions are already available and evaluate if these may satisfy the demands of PaN experiments. Second, experienced experimenters and scientists from the PaN community discussed about the findings and added further specifications. From these insights and discussions, a list of necessary, useful and helpful specifications was compiled. The full list is shown in Table 1.

Specifications are the building blocks of an ELN, they define its capabilities and extent. There may be general specifications that are of importance to many scientific domains, but a list of specifications must be carefully tailored for any new field and is the starting point for the search of an appropriate ELN for the respective field of interest. Figure 2 shows the exemplary information flow from the data generation at the instrument and the experiment control into an ELN. In a modern ELN the human user typically interacts with it through a web application, additionally, the applications often possess an application programming interface (API) for automatic information ingestion from the instrument or the experiment control. Later on in this document, we present a list of exemplary specifications for an ELN in the scientific domain of neutron and photon experiments with a focus on experiments at large scale facilities.

Table 1 includes all relevant ELN specifications that were found and discussed in work program TA1 of DAPHNE4NFDI while evaluating ELNs for large research facilities, these are defined by a title, a description of the specification, accomplished by examples where applicable, and ranked by their importance for the success of an ELN in future deployment at large research infrastructures.

How to: Identify suited ELN

Based up on the specifications from the section above, see Table 1, a work flow may be implemented to rank evaluated ELNs according to their concordance with the desired specifications. Here, we propose a formula deriving a single number P_n, figure of merit (FOM), for every tested ELN, describing the extent and maturity of its specifications. Enabling a fast and simple comparison between multiple solutions. However, this number may be a starting point for a preselection of suitable ELNs, but should always be accompanied by detailed discussions and extensive testing phase, leading to a final decision for an ELN. The specifications in Table 1 and the formula are meant as a guideline and may be altered or complemented, depending on the actual use case for the ELN.

The formula calculates a number P_n taking all desired specifications F into account. Every single specification F_i is identified by an index i and accompanied by its importance parameter I_i, see Table 1. Additionally, a satisfaction parameter S_i is introduced, ranking the actual usability and maturity of the respective specification on a range from 1 to S_{max}, e.g. on a range from 1 to 10. This parameter heavily relies on the personal flavor and perception of the tester. The product term is introduced to the formula to ensure that all necessary specifications are met. Hence, leads to $P_n = 0$ if a single necessary functionality is missing or not satisfying (e.g. if $S_i < 8$). The division by the sum ensures a normalization and the sum at the end weights the respective specification by its importance I_i and satisfaction S_i parameter.

$$p_n = \frac{\prod_{necessaryF_i} \begin{cases} F_i satisfied : 1 \\ else : 0 \end{cases}}{\sum I_i} \frac{\sum_{i=1}^{n} I_i * S_i}{S_{\max}}$$

The number P_n may be incorporated in a work flow to decide for an ELN. An exemplary work flow could look as follows:

1. Identify eligible ELNs;
2. Identify desired specifications and define importance;
3. Decide for a cut-off criterion on P_n (e.g., 0.7);
4. Test and evaluate ELN functionalities;
5. Calculate P_n's;
6. Check cutoff criterion;
7. Further test and discuss on ELN's passing the criterion;
8. Decide for an ELN or refine work flow and repeat (if necessary).

It may happen that none of the tested ELNs is passing the cut-off criterion in step 6. If this is the case it may be necessary to identify further ELNs, make some compromises on the necessary specifications and the satisfaction level, or, if these options don't work for a specific use case, it could be appropriate to think about the development of a new ELN. An exemplary evaluation and application of the work flow is given in the SI[11].

Survey on existing ELNs

Based on the defined ELN specifications a survey on existing ELNs or solutions under development was conducted in 2022[12]. This survey

Table 1: Contains a descriptive name of the respective ELN specifications, a description of the feature and their application in an ELN.

Specification	Description	Imp
Interface	Application Programming Interface (API) to access and modify content of an ELN. e.g. Auto ingestion of (meta) data	5
Intuitive use	Intuitive User interface enabling a productive user experience with a professional programmed and designed GUI	5
Reliability	Enabling legally binding documentation of research by introducing a record of changes, e.g. audit trail.	5
Open source	The ELN is developed and shared in an open manner, so that all willing can read, test and further develop the code base. e.g. source code on gitlab.	5
Documentation	Documentation of the status, capabilities and functionality of an ELN, e.g. training for new users.	5
Content export	Content of the ELN may be exported in a free and well established data format, e.g. as pdf or html.	5
Parallel editing	Multiple actors can input information into the ELN, simultaneously. e.g. user input and automated ingestion from control systems.	5
Authentication	User authentication in a reliable and secure manner.	5
Authorization	Define access rights for users and user groups by access management.	5
Searchable	Information is searchable for humans and machines, e.g. full text search, structured data, keywords or tags.	5
Scalable	Software is designed to support numerous users across an entire facility, compatible with a facility-wide deployment.	5
Image import	Import of wide spread image file formats into the ELN, from cloud or file system.	5
Server deployment	Deployment of the software at large scale research infrastructures as a central service.	5
Free-form input	Free-form input capable of containing any kind of information represented by characters and numbers, further, elements to organize and style these e.g. tables, lists, headings, fonts and colors.	5
Recovery	Reliable mechanism to store the content of an ELN and restore it in case of data loss.	5
Auto save	Mechanism to automatically store and synchronize the input.	5
Identity management	Federated identity management for authentication and authorization e.g. ORCiD, Helmholtz ID.	4
Links/tags	Possibility to use links and tags in the ELN.	4
Software image	Software image enabling an easy and reliable deployment of the ELN.	4
Local installation	Installation of the software on personal computers or lab instruments.	4
Image input	Interface to ingest photos into the ELN e.g. of equipment, setup or samples.	4
Comment	Convenient and fast option to comment on information in the ELN.	4
Schematic input	Schematic input to enter structured data into the ELN, e.g. templates.	3
Detail level	Control the detail depth that is shown, e.g. show more technical details for beamline scientist.	3
Platform independent	Accessible on multiple devices and operating systems, e.g. by an application based on web technologies.	3
Voice input	Recorded voice files or voice to text transcripts in the ELN e.g. for preparation in wet labs.	3
Summary table	Spreadsheet like overview of key inputs, e.g. important parameters from each scan.	2
Experimental management	Track the current status of an experiment and define the following steps, e.g. by to-do lists or project boards.	2
Sketches	Drawing of illustrations, e.g. sample structures or experimental setup.	2
Data import	Import of raw and derived data into the ELN, e.g. as a table.	1
Integration of software	Possibility to include and execute software code inside the ELN, e.g. by inclusion of ipython notebook features.	1

The importance (Imp) is represented by an integer from 5 (essential feature) down to 1 (useful) for an ELN.

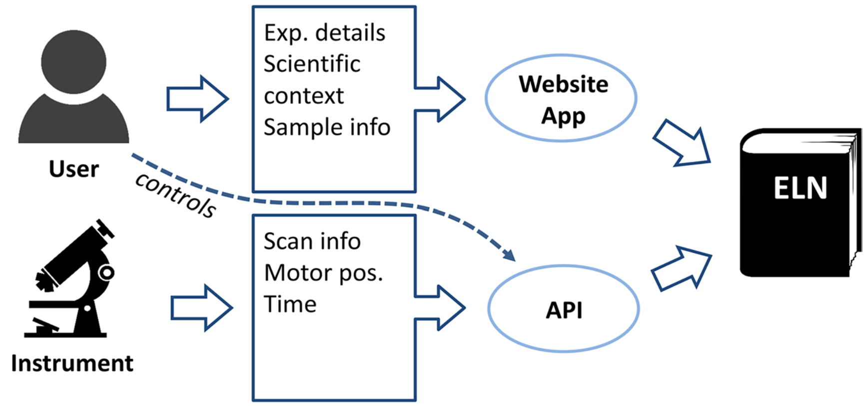

Figure 2: Information flow from an experiment into an ELN. [parts from pixabay.com].

was held internally in the DAPHNE4NFDI project, 17 replies were collected containing 12 different ELNs. It should be noted that the results are highly dependent on the personal flavor of the evaluating persons and are representing a rather small number of samples. In the survey a range of the satisfaction parameter from 1 (fully) to 7 (not) was employed to rate the functionalities, especially the suitability and the maturity, as two key parameters.

Suitability describes the overall readiness of an ELN.

Maturity rates the technical readiness of the solutions.

The survey indicates that 65% of all tested ELNs are rated as suitable for usage and 82% are seen as mature enough by the evaluating persons, see Figure 3. This results in a combined 53% of all ELNs that are suitable, as well as mature enough for usage as ELN at an large scale facility. Here, the satisfaction parameter had to be 1 (fully) or 2 (very) to be deemed as suitable or mature enough for usage as ELN, see Figure 4.

Figure 4 depicts the results for the suitability and maturity with an increased level of detail, showing the distribution on the scale from fully to not suitable/mature. Both graphs indicate that the tested solutions are slightly more technical mature than they are suited for the usage as ELNs. Furthermore, both charts highlight the fact that the majority of the evaluated ELNs were deemed to be very or fully suitable/mature for the usage at PaN facilities, as evaluated by the survey participants.

It is quite illustrating and easy to look at the suitability and maturity and rank these to get a first insight into the available ELN solutions. However, as laid out before an ELN consists of many different specifications that play a role and have a different importance for the individual success for different use cases, see Table 1. To get a more complete and quantitative evaluation of the available solutions we propose to evaluate as many different specifications as needed independently and calculate a FOM for comparison, see former section. Applying this fine granulated approach to the survey data gives a more punctuated result. From our survey we see that out of the 12 surveyed ELNs only two pass the FOM cut-off of 0.6, which is a 17% share. This is a much lower share of usable ELNs compared to 53%, when solely taking the evaluated maturity and suitability parameter into account. This huge discrepancy may be interpreted as a sign that the evaluating persons are losing the overview over all the criteria for a successful ELN at large research facilities or even don't have all important and necessary specifications in mind when rating the overall suitability and maturity. Thus, the compilation of a list of specifications, beforehand, can be seen as a way to introduce a guideline for the evaluation and as an approach to reduce bias.

A few exemplary ELNs used or developed within the DAPHNE4NFDI community are the following. Snip[13] is a user-friendly ELN that integrates digital data with handwritten notes, fosters collaboration, and supports secure access control. The ELN currently developed at the Heinz Maier-Leibnitz Zentrum (MLZ) for research with neutrons is based on the software package TUM Workbench[14]. It combines powerful capabilities in data management, user access management and integration with the facility infrastructure. MyLog is an ELN based on the commercial Zulip[15] chat application, with its content structured in streams and topics (which make it suited for multi-proposal use), accompanied by a powerful text integration and programmable bots for automatic notifications, and interoperable with other services through an API. elabFTW[16] is a wide spread, open-source ELN, that offers a wide range of features for managing scientific data, including experimental recording, instrument scheduling and team collaboration. Chemotion[17] is an open-source ELN tailored for chemists, offering features like experiment tracking, data management, collaboration, and integration with chemical tools with a further developed version for catalysis (LabIMotion[18]). OpenBIS[19] is an open-source ELN, combining laboratory information management (LIMS) and electronic laboratory notebooks (ELN) in one application. The number of available ELNs is huge and in constant change. There are many different solutions available for almost any use case, completed by generic ELNs. These may be used for or adapted to multiple scientific use cases. Here is a non exhaustive list of available solution: Benchling, Chemotion, Confluence, elabFTW, ESRF logbook, Google Docs, Ipython notebooks, Labfolder, Labstep, MediaWiki, MyLog,

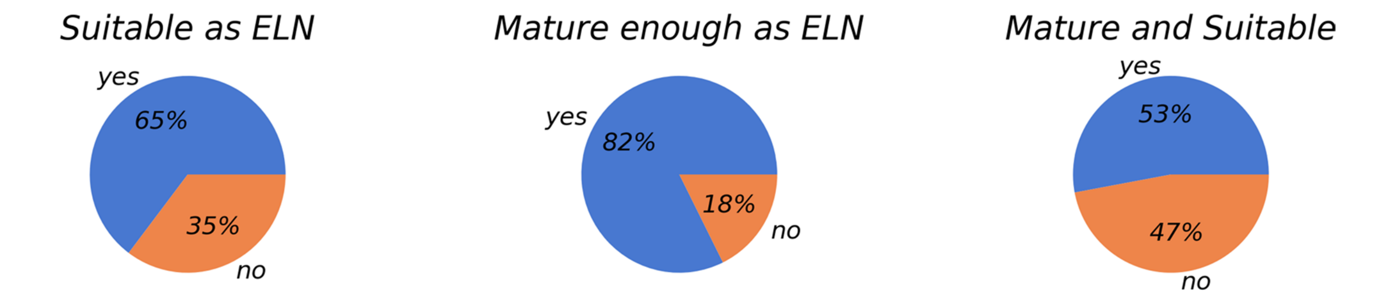

Figure 3: The overall rated suitability and maturity for ELNs. Suitability rates the capabilities and usefulness of the solutions as ELNs. Maturity describes the technical readiness.

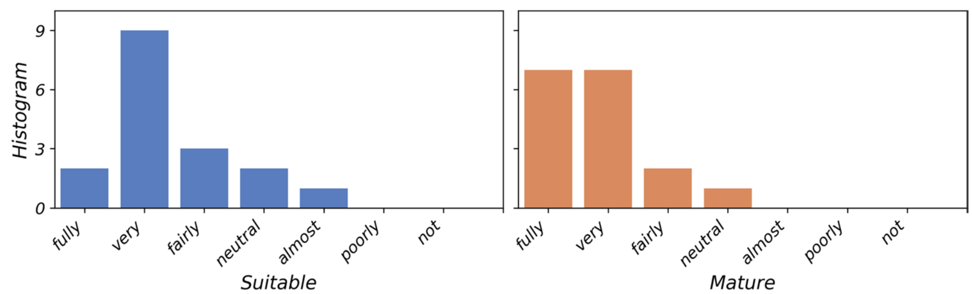

Figure 4: Detailed distribution of the evaluation results for the suitability and maturity.

Nomad, OpenBIS, RSpace, SciNote, SciLog, SampleDB, Snip, Workbench. An in depth description of ELNs in use within DAPHNE4NFDI is given in the SI[20]. For an extensive collection of available ELNs consult the ELN finder[21] of the University Darmstadt and the ELN product list[22] by the Cambridge University.

Summary

ELNs are a key resource for documentation and collaboration on scientific experiments, especially for generation of FAIR data. The discussion here is focused on data generation and metadata collection at the internationally connected large scale facilities. Specifications have been outlined and compiled as a ranked list of necessary and useful for ELNs. Based on the proposed list of specifications a work flow to evaluate and rank ELNs is presented.

To define appropriate specifications for ELNs an intense discussion among field experts and facility personal took place. The discussion addressed the purpose and content of an ELN as well as requirements for a great user experience and necessary properties for a successful deployment at large scale facilities. One important take away from these discussions, valid far beyond our use case for large facilities, is that there is not, and probably will not be, a single solution for all necessities and all the different experiments. In the end a list of ELN specifications was compiled from these discussions, listing many different aspects that may be important for an ELN. Many of the listed specifications may be of general interest for an ELN in the scientific context and a few more are probably of special interest for the deployment at large scale facilities.

During discussion on the ELN specifications a few of them were found to be of mayor interest and importance for deployment at large scale facilities, both on a national and international level. An application programming interface (API) is of great importance for the inter connectivity between the ELN application and the experimental setup, i.e. instrumentation, facility, servers, and, thus, of great importance for the DAPHNE4NFDI use case. Enabling an automatic flow of experimental information's into the ELN. Related is the specification of parallel editing, e.g. input from an experimenter in a laboratory about experimental details and, simultaneously, information about measurements from the instrument. Especially for large research facilities, user management, authentication and authorization are of great importance to keep data secure and ensure user access to individual experiments.

Following the discussion on the specifications an ELN may or must fulfill, a work flow is proposed to evaluate and rank multiple solutions in a quantitative and bias reduced procedure. It is important to find and use an appropriate work flow for individual use cases, to obtain the optimal solution for individual experimental methods. In the case, of photon and neutron experiments at large research facilities, the evaluation of the ELN solution should take place at the facilities in close cooperation with the users, so that the log book can be sustainably integrated in the facility infrastructure while meeting the community needs.

Acknowledgements

This publication was written in the context of the work of the consortium DAPHNE4NFDI in association with the German National Research Data Infrastructure (NFDI) e.V. NFDI is financed by the Federal

Republic of Germany and the 16 federal states and the consortium is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number 460248799. The authors would like to thank the DFG for the funding and support. Furthermore, thanks go to all institutions and actors who are committed to the association and its goals.

Disclosure statement

No potential conflict of interest was reported by the author(s). ■

ORCID

P. Jordt http://orcid.org/0000-0002-5402-7407
M. Osterhoff http://orcid.org/0000-0002-7865-515X
B. Hakim http://orcid.org/0000-0002-5841-739X
P. Dolcet http://orcid.org/0000-0001-9583-9375
F. Maurer http://orcid.org/0000-0002-3307-4132
L. Amelung http://orcid.org/0000-0002-8137-1553
F. Dall'Antonia http://orcid.org/0000-0003-0799-2244
J.-D. Grunwaldt http://orcid.org/0000-0003-3606-0956
F. Weber http://orcid.org/0000-0002-4256-1354
W. Lohstroh http://orcid.org/0000-0001-8404-2109
B. Murphy http://orcid.org/0000-0002-1354-2381

Notes

1. https://www.nature.com/articles/d41586-018-05895-3.
2. https://www.nature.com/articles/sdata201618.
3. https://doi.org/10.5281/zenodo.8040606.
4. https://www.daphne4nfdi.de/.
5. https://www.nfdi.de.
6. https://www.panosc.eu/.
7. https://expands.eu/.
8. https://www.energy.gov/science/ascr/articles/bringing-fair-principles-ai-models.
9. https://www.helmholtz.de/en/research/research-fields/matter/research-with-photons-neutrons-and-ions-pni/facilities/.
10. https://doi.org/10.5281/zenodo.13891465.
11. https://doi.org/10.5281/zenodo.13891465.
12. https://doi.org/10.5281/zenodo.10817031.
13. https://snip.roentgen.physik.uni-goettingen.de/frontpage.
14. https://www.ub.tum.de/en/workbench.
15. https://zulip.com/.
16. https://www.elabftw.net/.
17. https://chemotion.net/.
18. https://doi.org/10.52825/cordi.v1i.334.
19. https://openbis.ch/.
20. https://doi.org/10.5281/zenodo.13891465.
21. https://eln-finder.ulb.tu-darmstadt.de/home.
22. https://www.data.cam.ac.uk/data-management-guide/electronic-research-notebooks/electronic-research-notebook-products.