



Section
**Common
Infrastructures**

Working Group Charter
Electronic Lab Notebooks

April 2022

Working Group Charter

Electronic Lab Notebooks (ELNs)

Version 1.0

Date: 28.4.2022

Contact Persons:

Karlsruhe Institute of Technology

Nicole Jung, nicole.jung@kit.edu

Christof Wöll, christof.woell@kit.edu

Authors:

Björn Brötz, DLR

Christof Wöll, Karlsruhe Institute of Technology (KIT)

Felix Bach, FIZ Karlsruhe

Florian Ott, GFZ Potsdam

Frank Tristram, Karlsruhe Institute of Technology (KIT)

Jan-Dierk Grunwaldt, Karlsruhe Institute of Technology (KIT)

Konrad Förstner, ZB MED - Information Centre for Life Sciences

Nicole Jung, Karlsruhe Institute of Technology (KIT)

Sonja Herres-Pawlis, RWTH Aachen University

1. Motivation

Electronic Lab Notebooks (ELNs) play an important role for many disciplines with respect to the digitalization of scientific workflows. ELNs provide the appropriate environment to collect information and data at the earliest stage of the research data lifecycle (e.g. experiment planning and data acquisition in a laboratory and at large scale research facilities) and can be used for the management of original data along with edited and annotated data. Additionally, ELNs can be used to collect and link data from different sources. They provide a significant contribution to protocolling experiment work flows and securing research results in line with FAIR data principles through supporting functions for data visualization, storage, discoverability and re-use. ELNs offer the option to organize information in a structured way, to standardize processes and data and to efficiently annotate data. This makes ELNs a highly valuable tool for NFDI since ELNs could be used to embed the results of the NFDI consortia for standardization, metadata schemes, ontologies/taxonomies and different other topics. Thus, the use of ELNs for scientific documentation can lead to a direct adoption of the standardization efforts of the NFDI at a very early stage of the scientific work. While many other tools for research data management are often seen as an additional burden, ELNs are in general widely accepted in the scientific community due to the manifold functionality that supports scientific work, facilitates the management of data and increases the scientists' efficiency. This offers a great chance for NFDI since ELNs can be used as a key instrument to foster the acceptance and use of the results of NFDI. In the end, ELNs are perfect tools to prepare data in a suitable way for further applications in the NFDI infrastructures. As an example, data captured and managed by ELNs could serve as a perfect source for a seamless, effortless transfer of the

data to repositories. A systematic use of ELNs could contribute to the success of the NFDI. Nevertheless, this requires the awareness of scientists to select suitable ELNs. Suitable solutions should allow a systematic standardization of data as well as clear and appropriate data structures and the adoption of ELNs to the achievements of NFDI working groups and their results. While a focus on discipline specific work and needs is important for the acceptance of ELNs, the standardization efforts need to be defined and discussed by NFDI as a whole. Potential to encompass future base services: Driven by the need for digitalization, ELNs are of high interest for many scientists as well as their institutions which are searching for suitable options for data management. Thus, ELNs are becoming a vital part of the scientists' world, being important for any digitalization strategy of the scientific institutions. This makes ELNs an important topic in almost all NFDI consortia. Unfortunately, scientists and institutions usually do not have the resources and skills to enable sophisticated investigations on ELNs, their potential and data structures. Finding the right ELN that fits the needs of the scientists in terms of functionality and meets the requirements to generate FAIR data is therefore hard to achieve. And while ELNs offer the foundation for an efficient digitalization of the research process, the current ELN ecosystem is dominated by commercial providers and interoperability is very limited. This harbors the serious risk of vendor-lock-in undermining the core idea of NFDI and an open research community. For this purpose, the promotion of open source ELN solutions but also common standards regarding meta data and exchange formats are crucial. Currently, the diverse efforts are not coordinated and joining forces for choosing the right ELN or adopting existing ones are needed. Also, a common initiative on the standardization of ELNs across different disciplines is a prerequisite to be able to efficiently reuse data from ELNs.

2. Objectives

The working group ELN will prepare and promote the installation of two main services. The first service will deal with the systematic assessment of different ELNs and their data structures, the second service will offer a platform to test different ELN software and similar (embedded) tools.

Objectives of a potential service ELN assessment (ELN-assessment):

We would like to establish a service that provides information on the strengths of different ELN software and similar documentation tools depending on discipline specific and interdisciplinary requirements. The service will offer information to evaluate ELNs based on the implemented data models, supported data structures, available interfaces (to other tools, ELNs, and repositories), the description of APIs, export and import options and exit strategies which have a direct influence on the re-usability of the data that is collected and generated via ELNs. The information is relevant for all scientists that are currently using a dedicated tool or will make the decision on one as it describes the value that data has for its use within NFDI and also other services.

(1) This service will support scientists in making their choice for a suitable ELN that allows the generation of FAIR data. In addition to the already ongoing activities for ELN description focusing on functionality, this service supports scientists with information they cannot gain easily on their own (e.g. which kind of data model is implemented and how well does it meet current standards for data) due to a necessary deep insight into the structure of an ELN. In particular with regard to the systematic description of data structures, the service will allow a direct comparison of ELNs and describe the results of choosing a distinct ELN solution.

(2) The service will offer information on how to structure data in ELNs to meet the requirements of FAIR data and to benefit from the services of NFDI. Guidelines on how to structure data will be presented with examples from different consortia and communities to enable the development of documentation

standards based on the available software tools. In the long run, the work of NFDI may help to improve the available options in ELNs, thus improving the available data models for different software tools.

(3) The service will support the efforts for interoperable ELNs by collecting the supported standards and by listing the mapping options for data stored in different ELNs. In combination with policies on how to use standard data models in different ELNs (2) the service will offer a process for gaining interoperability of ELNs. This will be the basis for an ELN converter that allows scientists to use ELNs without being limited to one solution and without the risk of losing data.

Objectives of a potential service ELN test environment hosting (ELN-host):

The choice for ELN defines the way of how to capture data and the options to re-use data for many years. Therefore this choice is an important task that requires a detailed knowledge of the ELN's functionality and dependencies and needs to be done by or at least with disciplinary scientists as they need to provide the discipline-specific requirements and check the suitability of the software.

(1) The service will provide an open platform to test different ELN software, to compare their functionality and to assess the results of different implemented data models. The effort to install, configure and host the tools will be done centrally, guaranteeing the availability of latest developments and the error-free operation of the software. The gained reference instances can be used as a direct, effortless entry into ELN testing.

(2) The test instances of different ELNs can serve as demo instances for activities of the section education and training, allowing access to the software in parallel to guidance and teaching that is provided in workshops and roadshows. In combination with a diverse set of use cases/examples coming from different communities and embedded in the different ELN solutions, the service demonstrates the benefits for different software and can advise scientists "hands on" how to use ELNs and how to organize data. The collection of examples will result in an exchange of knowledge and best practice information for scientists by scientists.

(3) The centrally hosted reference instances can be used to implement results of the NFDI developments, in particular the work on standards and ontologies. This allows a direct assessment of the tools generated within NFDI by the development teams and the user communities. The service will also provide a platform to promote the embedding or linkage of further software tools that facilitate the analyses of data. This will foster the use of available data management and analysis tools as the service generates awareness of the supported functions.

3. Work Plan

Our work plan includes different steps (work package = WP) to define and establish the two services ELN-assessment and ELN-host.

ELN-assessment WP 1: For an assessment of ELNs in detail, scientists being most familiar with the structure and details of ELNs and ELN developers that contribute directly to the project, need to be identified. Together, we will then describe the ELNs' data structure, supported data models and we will provide a comparison of the same/similar components that can be mapped (e.g. according to the ISA model).

ELN-assessment WP 2: Further work will include the establishment of a web-based platform that offers the description of (meta)data schemes and data structures required for different disciplines and a description of the existing standards used in different communities. We will initiate the description of mapping schemes for data exchange among ELNs. For a detailed understanding of the suitability of

certain ELNs for distinct domains, examples from different disciplines and a mapping to different ELNs will highlight options and differences of ELNs.

ELN-assessment WP 3: The operation of the service will result in continuously maintained standards to map and transfer data and data models from an ELN to others reflecting the special needs of different disciplines and their data structures. A description of interfaces to other services such as repositories and analysis tools should be available and a recommendation of entity-specific standards for ELNs should be given. The service will support potential ELN users with a detailed assessment of exit strategies of ELNs and will summarize the review for data re-use options.

ELN-host WP 1: A suitable institution for the centrally hosted service needs to be identified. A selection of ten open source ELNs from international developments will be installed and made available for testing purposes based on the feedback of all NFDI consortia. The WG will establish contacts to vendors and developers of commercially available ELNs and comparable tools to evaluate their openness for cooperation and participation.

ELN-host WP 2: Additionally available open source tools that may complement the functionality of ELNs with respect to data conversion, data analysis, data visualization and others will be made available. Also, the installation of commercial ELNs that are of relevance in the NFDI consortia and that agree on the free hosting of a testing instance, is offered. To guarantee close interaction with the community, we will offer online feedback forms that collect further needs and help to continuously evaluate the offered software. The online forms will be complemented by regular surveys from the NFDI-communities.

ELN-host WP 3: The service will start its operation directly with the availability of the first ELNs, being gradually extended thereafter. For being able to offer a trustable service in terms of robustness, security and scalability, additional measures have to be prepared. A strategy for a regular update and upgrade of the provided systems has to be developed. Further, a network of developers and users will be built up and maintained, providing the community to integrate best practice examples. A concept to continuously support users and to offer information will be developed. A combination of existing services (such as helpdesks installed with the NFDI consortia) as well as lead-by-example scientists will be considered.

4. Collaboration Plan

The *WG ELN* has common interests and challenges with the sections *metadata*, *edutrain* and *ELSA* and will closely work with them. There are also several links to other working groups in the section common infrastructures. *Section-metadata*: The assessment of supported metadata schemes and standards in the form of e.g. ontologies can directly contribute to the work of section metadata. Also, with a hosting service for ELNs, the *WG ELN* can offer a platform to integrate international ontologies and metadata standards. *Section-Edutrain*: We will closely work together with *edutrain*, in particular for the establishment of the *service training infrastructure* which can re-use the assessment information gained in this *WG*. *Edutrain* can use the hosted ELN service for its teaching and training efforts. *Section-ELSA*: ELNs as a work instrument with optional capture of users' data and the storage of critical research data are an important topic to be discussed in *ELSA*. Results and recommendations of the ELN and *ELSA* interface will be integrated into the activities of *edutrain*. The *WG ELN* will benefit directly from various connections to other working groups in CI: Synergies are in particular expected for collaboration with RSE (research software engineering) to facilitate good software development practice, with PID (Persistent Identifier Services): the *WG PIDs* will offer services that need to be usable by the software considered in the *WG ELNs*. *ELN host* will further benefit directly from connections to the working groups RKG (Knowledge

Graphs), LTA (long term archival), IAM (identity management), DMP (data management planning) in CI.

5. Initial Membership List

	Name	Institution	NFDI consortia
1	Nicole Jung	Karlsruhe Institute of Technology (KIT)	NFDI4Chem
2	Christof Wöll	Karlsruhe Institute of Technology (KIT)	FAIRmat
3	Felix Bach	FIZ Karlsruhe	NFDI4Chem, NFDI4Culture
4	Sonja Herres-Pawlis	RWTH Aachen University	NFDI4Chem
5	Florian Strauß	TU Clausthal	NFDI4Ing
6	Jan-Dierk Grunwaldt	Karlsruhe Institute of Technology (KIT)	DAPHNE4NFDI, (NFDI4CAT)
7	Frank Tristram	3DMM2O (KIT/University HD)	NFDI4Phys
8	Joachim Mayer	RWTH Aachen, FZ Jülich	NFDI MatWerk NFDI4Phys
9	Bridget Murphy	CAU - Kiel	DAPHNE4NFDI
10	Matthias Lange	IPK	FAIRAgro
11	Thomas Rose	Goethe University Frankfurt	NFDI4Earth
12	Justine Vandendorpe	ZB MED - Information Centre for Life Sciences	NFDI4Microbiota
13	Björn Brötz	DLR	NFDI4Earth
14	Florian Ott	GFZ Potsdam	NFDI4Earth
15	Olaf Deutschmann	KIT	NFDI4Cat
16	Konrad Förstner	ZB MED - Information Centre for Life Sciences	NFDI4Microbiota