



# Electronic Lab Notebooks and Experimental Design Assistants

Björn Gerlach, Christopher Untucht, and Alfred Stefan

## Contents

1 Paper vs. Electronic Lab Notebooks .....	258
2 Finding an eLN .....	262
3 Levels of Quality for eLNs .....	265
4 Assistance with Experimental Design .....	266
5 Data-Related Quality Aspects of eLNs .....	267
6 The LN as the Central Element of Data Management .....	269
7 Organizing and Documenting Experiments .....	271
References .....	274

## Abstract

Documentation of experiments is essential for best research practice and ensures scientific transparency and data integrity. Traditionally, the paper lab notebook (pLN) has been employed for documentation of experimental procedures, but over the course of the last decades, the introduction of electronic tools has changed the research landscape and the way that work is performed. Nowadays, almost all data acquisition, analysis, presentation and archiving are done with electronic tools. The use of electronic tools provides many new possibilities, as well as challenges, particularly with respect to documentation and data quality. One of the biggest hurdles is the management of data on different devices with a

B. Gerlach (✉)

PAASP GmbH, Heidelberg, Germany

e-mail: [bjoern.gerlach@paasp.net](mailto:bjoern.gerlach@paasp.net)

C. Untucht

AbbVie Deutschland GmbH, Neuroscience Discovery, Ludwigshafen am Rhein, Germany

e-mail: [christopher.untucht@abbvie.com](mailto:christopher.untucht@abbvie.com)

A. Stefan

AbbVie Deutschland GmbH, Information Research, Ludwigshafen am Rhein, Germany

e-mail: [alfred.stefan@abbvie.com](mailto:alfred.stefan@abbvie.com)

© The Author(s) 2019

A. Bespalov et al. (eds.), *Good Research Practice in Non-Clinical Pharmacology and Biomedicine*, Handbook of Experimental Pharmacology 257, [https://doi.org/10.1007/164\\_2019\\_287](https://doi.org/10.1007/164_2019_287)

257

substantial amount of metadata. Transparency and integrity have to be ensured and must be reflected in documentation within LNs. With this in mind, electronic LNs (eLN) were introduced to make documentation of experiments more straightforward, with the development of enhanced functionality leading gradually to their more widespread use. This chapter gives a general overview of eLNs in the scientific environment with a focus on the advantages of supporting quality and transparency of the research. It provides guidance on adopting an eLN and gives an example on how to set up unique Study-IDs in labs in order to maintain and enhance best practices. Overall, the chapter highlights the central role of eLNs in supporting the documentation and reproducibility of experiments.

---

**Keywords**

Data management · Documenting experiments · Lab journal · Paper and electronic lab notebook · Study-ID

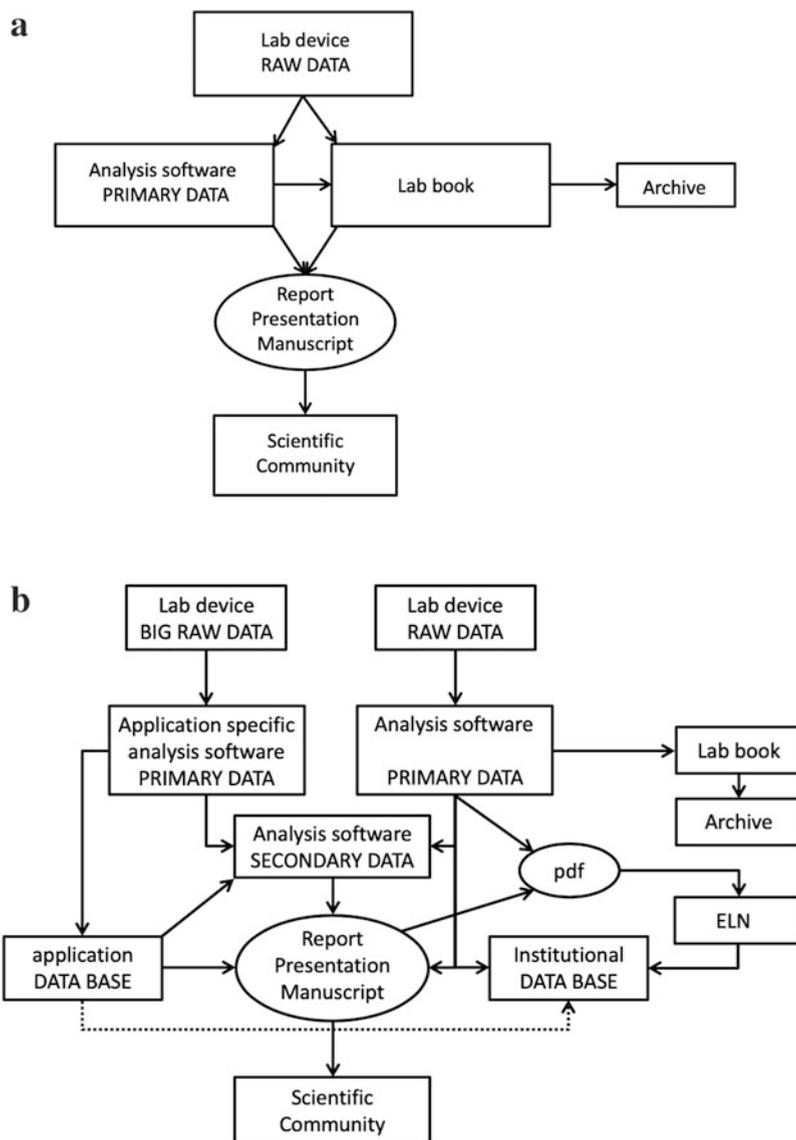
---

## 1 Paper vs. Electronic Lab Notebooks

Documentation of experiments is critical to ensure best practice in research and is essential to understand and judge data integrity. Detailed and structured documentation allows for researcher accountability as well as the traceability and reproducibility of data. Additionally, it can be used in the resolution of intellectual property issues. Historically, this had been performed by documentation in conventional paper laboratory notebooks (pLN). Lab notebooks (LN) are considered the primary recording space for research data and are used to document hypotheses, experiments, analyses and, finally, interpretation of the data. Originally, raw and primary data were recorded directly into the lab book and served as the basis for reporting and presenting the data to the scientific community (Fig. 1a), fulfilling the need for transparent and reproducible scientific work.

The situation has become much more complex with the entry of electronic tools into the lab environment. Most experimental data is now acquired digitally, and the overall amount and complexity of data have expanded significantly (Fig. 1b). Often, the data is acquired with a software application and processed with secondary tools, such as dedicated data analysis packages. This creates different databases, varying both in volume and the type of produced data. Additionally, there can be different levels of processed data, making it difficult to identify the unprocessed, or raw, data. The experimental data must also be archived in a way that it is structured, traceable and independent of the project or the data source. These critical points require new data management approaches as the conventional pLN is no longer an alternative option in today's digital environment.

Current changes in the type, speed of accumulation and volume of data can be suitably addressed by use of an electronic laboratory notebook (eLN). An eLN that can be used to document some aspects of the processes is the first step towards improvement of data management (Fig. 1b). Ultimately, eLNs will become a central tool for data storage and connection and will lead to improvements in transparency and communication between scientists, as it will be discussed in the last two sections of this chapter.



**Fig. 1** (a) Traditional data flow: The lab book is an integral part of scientific data flow and serves as a collection of data and procedures. (b) An example of a more complex data flow might commonly occur today. Unfortunately, lab books and eLNs are often not used to capture all data flows in the modern laboratory

Although both paper and electronic LNs have advantages and disadvantages (Table 1), ultimately the overall switch of researchers from the use of pLNs to eLNs is inevitable. The widespread use of electronic tools for acquiring, analysing

**Table 1** Overview of advantages and disadvantages of paper and electronic lab notebooks

Characteristic	Paper lab notebooks	Electronic lab notebooks
Data accessibility	Restricted access, needs copying, often inaccessible once a researcher leaves a lab	Advanced data sharing, direct links to all types of files
Data organization	Chronological order of experiments with page numbers	Flexible organization of experimental data, parallel documentation of different projects
Data presentation	All electronic data has to be printed or referenced	Direct access to large data files via links
Templates	None	Documentation facilitated by templates
Peer review and sharing information	Difficult	Viewing/editing permissions can be granted to individuals and groups of people
Search for keywords	Not possible	Possible
Ease of use	Easy to use, intuitive, flexible	Many different solutions, DIY
Audit trail	Witnessing with a signature	Witnessing with an electronic signature, not always straightforward with DIY eLN
IT support	Not required	Needs IT infrastructure, accessibility across devices is not always possible
Costs	Low	Can be expensive

and managing data renders traditional pLNs impractical in modern science. Nevertheless, the switch to eLN usage is gradual and will take time. Several groups documented and have published their experiences and discussed the biggest barriers for establishing the standardized use of eLNs. Kanza et al. carried out detailed analyses, based on surveys, focusing on potential obstacles to the application of eLNs in an academic laboratory setting (Kanza et al. 2017). According to their results, the major barriers are cost, ease of use and accessibility across devices. Certainly, these aspects seem to be valid at first glance and need to be considered carefully. A more detailed examination provides means to address these issues as progress has been made in the development of tailored solutions.

As mentioned, cost is often the most significant hurdle to overcome when adopting the use of an eLN. In this regard, pLNs have a clear advantage because they require close to no infrastructure. In contrast, there are multiple different eLN device options which vary widely, from freeware to complex solutions, which can have costs of up to thousands of euros per researcher. This is, however, evolving with the use of cloud-based software-as-a-service (SaaS) products which are available for a monthly subscription. One of the major advantages of SaaS is that the software version will always be up-to-date. Today, many of these SaaS solutions are available as online tools which can also be used within a firewall to have maximum

**Table 2** Overview of different levels of complexity of laboratory notebooks

Type	Explanation
Paper lab notebook	Reporting of complete primary experimental data in paper format
Do-it-yourself electronic lab notebook	Most simple form of eLN. Many different possibilities exist that can be used as building blocks and combined to create a custom solution (e.g. OneNote, MS Office documents, Dropbox, OneDrive)
Dedicated electronic lab notebook	Sophisticated and dedicated software solutions for reporting and archiving complete primary experimental data
Systemic electronic lab notebook	As above including full lab inventory management system

protection of the data. Most digital solutions, however, require the appropriate hardware and infrastructure to be set up and updated.

In contrast to eLNs, pLNs do not require IT support and are, therefore, immediately usable, do not crash and are intuitive to use. However, in the majority of cases, the data cannot be used directly but must be printed, and a general search cannot be performed in the traditional lab book once the scientist who created it leaves the lab. Thus, the primary disadvantages of paper notebooks are the inability to share information, link electronic data, retrieve the information and search with keywords.

Different types of LNs are summarized in Table 2. A pLN can be seen as the primary place for reporting research. Simple electronic tools such as word processors, or do-it-yourself (DIY) solutions (Dirnagl and Przesdzing 2017), provide this in electronic format. A DIY eLN is a structured organization of electronic data without dedicated software, using instead simple software and data organization tools, such as Windows Explorer, OneNote, text editors or word processors, to document and organize research data (see the last two sections of this chapter for tips on how to create a DIY eLN). In contrast, dedicated and systemic versions of eLNs provide functionality for workflows, standard data entry as well as search and visualization tools. The systemic eLNs provide additional connections to laboratory information management systems (LIMS), which have the capability to organize information in different aspects of daily lab work, for example, in the organization of chemicals, antibodies, plasmids, clones, cell lines and animals and in calendars to schedule usage of devices. Following the correct acquisition of very large datasets (e.g. imaging, next-generation sequencing, high-throughput screens), eLN should support their storage and accessibility, becoming a central platform to store, connect, edit and share the information.

The traditional pLN is still widely used in academic settings and can serve as a primary resource for lab work documentation. Simple versions of eLNs improve the data documentation and management, for example, in terms of searching, accessibility, creation of templates and sharing of information. However, with an increasing amount of work taking place in a complex digital laboratory environment and the associated challenges with regard to data management, eLNs that can connect, control, edit and share different data sources will be a vital research tool.

## 2 Finding an eLN

Many different eLNs exist, and finding the appropriate solution can be tedious work. This is especially difficult when different research areas need to be covered. Several published articles and web pages provide guidance on this issue (are summarized in the following paragraphs with additional reading provided in Table 3). However, a dedicated one-size-fits-all product has, to date, not been developed. Several stand-

**Table 3** Summary of published data describing selection and usage of eLNs

Author and year	Title	Description
Rubacha et al. (2011)	A review of electronic laboratory notebooks available in the market today	Very comprehensive analysis of eLNs on the market in 2011; distinguishes between five different categories
Nussbeck et al. (2014)	The laboratory notebook in the twenty-first century	A general advocacy for eLNs; good background information about the advantages
Oleksik and Milic-Frayling (2014)	Study of an electronic lab notebook design and practices that emerged in a collaborative scientific environment	Comprehensive user experience of establishing a DIY eLN (OneNote) in a physics research lab
Guerrero et al. (2016)	Analysis and implementation of an electronic laboratory notebook in a biomedical research institute	Quite comprehensive comparison of six eLNs with a more in-depth consideration of two devices (PerkinElmer and OneNote)
Kanza et al. (2017)	Electronic lab notebooks: can they replace paper?	Broad and informative background information on the use of eLNs
Riley et al. (2017)	Implementation and use of cloud-based electronic lab notebook in a bioprocess engineering teaching laboratory	Practical guidance on setting up an eLN (LabArchives) as an example in a teaching lab
Dirnagl and Przesdzing (2017)	A pocket guide to electronic laboratory notebooks in the academic life sciences	Experience report about setting up an eLN in an academic lab
May <sup>a</sup>	Companies in the cloud: Digitizing lab operations	Short description of several commercial tools connecting lab equipment with eLNs
Kwok <sup>b</sup>	How to pick an electronic laboratory notebook	Collection of several aspects to be considered when choosing an eLN
The Gurdon Institute <sup>c</sup>	Electronic lab notebooks – for prospective users	Comprehensive and up-to-date guidance on different eLNs
Atrium Research <sup>d</sup>	Electronic laboratory notebook products	Most comprehensive list of eLNs that are currently available on the market

<sup>a</sup><https://www.sciencemag.org/features/2017/02/companies-cloud-digitizing-lab-operations>

<sup>b</sup><https://www.nature.com/articles/d41586-018-05895-3>

<sup>c</sup><https://www.gurdon.cam.ac.uk/institute-life/computing/elnguidance>

<sup>d</sup><http://atriumresearch.com/eln.html>

alone solutions have been developed for dedicated purposes, and other eLNs provide the possibility to install “in programme applications”. These additional applications can provide specific functionality for different research areas, e.g. for cloning.

Thus, the first course of action must be to define the needs of the research unit and search for an appropriate solution. This solution will most likely still not cover all specific requirements but should provide the functionality to serve as the central resource for multiple different types of data. In this way, the eLN is a centralized storage location for various forms of information and is embedded in the laboratory environment to allow transparency and easy data exchange. Ideally, an eLN is the place to store information, support the research process and streamline and optimize workflows.

The following points must be considered in order to identify the best solution:

- *Does the organization have the capability to introduce an eLN?* A research unit must judge the feasibility of eLN implementation based on different parameters:
  - *Financial resources:* Setting up an eLN can be cost intensive over the years, especially with the newly emerging model of “SaaS”, in which the software is not bought but rented and paid for monthly.
  - *IT capability:* It is important to have the knowledge in the research unit to set up and provide training for users to adopt an eLN or at least to have a dedicated person who is trained and can answer daily questions and provide some basic training to new staff.
  - *Technical infrastructure:* The technical equipment to support the eLN has to be provided and maintained for the research unit, especially when hosted within a research unit and not on external servers.
- *At which level should the eLN be deployed and what is the research interest?* Clarification on these two considerations can dramatically reduce the number of potential eLNs which are suitable for application in the research lab. For example, a larger research unit covering more areas of research will need a general eLN, in contrast to a smaller group which can get an eLN ready tailored to their specific needs. An example of a field with very specific solutions is medicinal chemistry: An eLN for application in a lab of this type should have the functionality to draw molecules and chemical pathways, which will not be needed for an animal research or image analysis lab. Having a clear idea about the requirements of a research unit and necessary functionality will help in the choice of the correct eLN resource.
- *Which devices will be used to operate the eLN software?* Many researchers require the facility to create or update records “live” on the bench or other experimental areas (e.g. undertaking live microscopy imaging), as well as on other devices inside and outside of the lab (e.g. running a high-throughput robotic screening in a facility). Alternatively, researchers may want to use voice recognition tools or to prepare handwritten notes and then transcribe into a tidier, more organized record on their own computer. It should also be noted that some vendors may charge additional fees for applications to run their software on different device types.

- *Does the funding agreement require specific data security/compliance measures?* Some funding agencies require all data to be stored in a specific geographic location to ensure compliance with local data protection regulations (e.g. General Data Protection Regulation (GDPR)). Some eLN systems, however, are designed to store content and data only on their own servers (i.e. in “the cloud”), in which case a solution could be negotiated, or the provider cannot be used.
- *Consider exporting capabilities of the eLN.* This is a very critical point as the eLN market is constantly changing and companies might be volatile. The eLN market has existed for almost 20 years, during which time several solutions emerged and disappeared or were bought by other companies. With these changes, there are certain associated risks for the user, such as, for example, the fact that pricing plans and/or data formats might change or the eLN could be discontinued. Therefore, we consider it as an absolute requirement that all data can be easily exported to a widely readable format that can be used with other applications.

It is advisable to identify a dedicated person within the research unit to prepare an outline of the lab requirements and who will be educated about the available technical solutions and prepare a shortlist. Subsequently, it is necessary to test the systems. All eLN providers we have been in contact so far have been willing to provide free trial versions for testing. Additionally, they usually provide training sessions and online support for researchers to facilitate the adaptation of the system more quickly.

When choosing an eLN, especially in academia, an important question will always be the cost and how to avoid them. There is a debate about the usage of open-source software, that is, free versions with some restrictions, or building a DIY solution. A good resource in this context is the web page of Atrium Research & Consulting, a scientific market research and consulting practice which provides a comprehensive list of open-source solutions (<http://atriumresearch.com/eLN.html>). It is important to note that open source is not always a truly open source – often it is only free to use for non-commercial organizations. They can also come with some limitations, for example, that all modifications must be published on the “source website”. However, the industry is ready to accept open source with many examples emerging, such as Tomcat or Apache, which are becoming in our experience de facto industry standards. Important factors to be considered for choosing open-source software are:

- The activity of the community maintaining the system (e.g. CyNote (<https://sourceforge.net/projects/cynote/>) and LabJ-ng (<https://sourceforge.net/projects/labj/>) was not updated for years, whereas eLabFTW (<https://www.elabftw.net>) or OpenInventory (<https://sourceforge.net/projects/enventory/>) seems to be actively maintained.
- Maturity and structure of the documentation system.
- If the intended use is within the regulated environment, the local IT has to take ownership.

Open-source software, particularly in the light of best documentation practice, will generate risks, which need to be handled and mitigated. This includes ensuring the integrity of the quality system itself and guaranteeing data integrity and prevention of data loss. To our knowledge, most open-source solutions cannot guarantee this. This is one reason why a global open-source eLN doesn't exist: there is no system owner who could be held accountable and would take responsibility to ensure data integrity.

The SaaS approach often uses cloud services as an infrastructure, e.g. Amazon Web Services or Microsoft Azure. These services provide the foundation for the web-based eLNs and are very easy to set up, test and use. These systems have a significant benefit for smaller organizations, and it is worth investigating them. They provide truly professional data centre services at a more affordable price, attracting not only smaller research units but also big pharmaceutical companies to move into using SaaS solutions to save costs.

---

### 3 Levels of Quality for eLNs

In principle, the “quality features” of eLNs can be related to three categories which have to be considered when choosing or establishing an eLN: (1) System, (2) Data and (3) Support.

System-related features refer to functionality and capability of the software:

- *Different user rights*: The ability to create different user levels, such as “not share”, read-only, comment, modify or create content within the eLN, is one of the big advantages. The sharing of information via direct access and finely tuned user levels allows for individual sharing of information and full transparency when needed.
- *Audit trail*: Here, a very prominent regulation for electronic records has to be mentioned, the CFR21 Part 11. CFR stands for Code of Federal Regulations, Title 21 refers to the US [Food and Drug Administration](#) (FDA) regulations, and Part 11 of Title 21 is related to electronic records and electronic signatures. It defines the criteria under which electronic records and electronic signatures are considered trustworthy, reliable and equivalent to paper records (<https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=11>). Hence, when choosing an eLN provider, it is important that the product adheres to these regulations since this will ensure proper record keeping. Proper record keeping in this case means that a full audit trail is implemented in the software application.
- *Archiving*: Scientists need to constantly access and review research experiments. Therefore, accessing and retrieving information relating to an experiment is an essential requirement. The challenge with archiving is the readability and accessibility of the data in the future. The following has to be considered when choosing an eLN to archive data:

- There is always uncertainty about the sustainability of data formats in the future, especially important for raw data when saved in proprietary formats (the raw data format dilemma resulted in several initiatives to standardize and harmonize data formats).
- The eLN can lead to dependency on the vendor and the system itself.
- Most current IT systems do not offer the capability to archive data and export them from the application. This results in eLNs becoming de facto data collection systems, slowing down in performance over time with the increased data volume.

Data-related features concern the ability to handle different scientific notations and nomenclatures, but most importantly, to support researchers in designing, executing and analysing experiments. For scientific notations and nomenclatures, the support of so-called namespaces and the ability to export metadata, associated with an experiment in an open and common data format, are a requirement. Several tools are being developed aiming to assist researchers in designing, executing and analysing experiments. These tools will boost and enhance quality of experimental data, creating a meaningful incentive to change from pLNs towards eLNs. The subsequent sections focus in detail on that point.

Support-related quality features affect the stability and reliability of the infrastructure to support the system and data. The most basic concern is the reliability and maintenance of the hardware. In line with that, fast ethernet and internet connections must be provided to ensure short waiting times when loading data. The security of the system must be ensured to prevent third-party intrusions, which can be addressed by including appropriate firewall protection and virus scans. Correspondingly, it is also very important to include procedures for systematic updates of the system. In contrast, this can be difficult to ensure with open-source software developed by a community. As long as the community is actively engaged, the software security will be normally maintained. However, without a liable vendor, it can change quickly and without notice. Only a systematic approach will ensure evaluation of risks and the impact of enhancements, upgrades and patches.

---

## 4 Assistance with Experimental Design

The design of experiments (DOE) was first introduced in 1920 by Ronald Aylmer Fisher and his team, who aimed to describe the different variabilities of experiments and their influence on the outcome. DOE also refers to a statistical planning procedure including comparison, statistical replication, randomization (chapter “Blinding and Randomization”) and blocking. Designing an experiment is probably the most crucial part of the experimental phase since errors that are introduced here determine the development of the experiment and often cannot be corrected during the later phases. Therefore, consultation with experienced researchers, statisticians or IT support during the early phase can improve the experimental outcome. Many tools supporting researchers during the DOE planning phase already exist. Although

most of these tools have to be purchased, there are some freely available, such as the scripting language R, which is widely used among statisticians and offers a wide range of additional scripts (Groemping 2018).

An example of an interactive tool for the design and analysis of preclinical *in vivo* studies is an online tool based on the R application (R-vivo) MANILA (MAtched ANImaL Analysis). R-vivo is a browser-based interface for an R-package which is mainly intended for refining and improving experimental design and statistical analysis of preclinical intervention studies. The analysis functions are divided into two main subcategories: pre-intervention and post-intervention analysis. Both sections require a specific data format, and the application automatically detects the most suitable file format which will be used for uploading into the system. A matching-based modelling approach for allocating an optimal intervention group is implemented, and randomization and power calculations take full account of the complex animal characteristics at the baseline prior to interventions. The modelling approach provided in this tool and its open-source and web-based software implementations enable researchers to conduct adequately powered and fully blind preclinical intervention studies (Laajala et al. 2016).

Another notable solution for *in vivo* research is the freely available Experimental Design Assistant (EDA) from the National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3R). NC3R is a UK-based scientific organization aiming to find solutions to replace, refine and reduce the use of animals in research (<https://www.nc3rs.org.uk/experimental-design-assistant-eda>). As one of the goals, NC3R has developed the EDA, an online web application allowing the planning, saving and sharing of individual information about experiments. This tool allows an experimental plan to be set up, and the engine provides a critique and makes recommendations on how to improve the experiment. The associated website contains a wealth of information and advice on experimental designs to avoid many of the pitfalls that have been previously identified, including such aspects as the failure to avoid subjective bias, using the wrong number of animals and issues with randomization. This support tool and the associated resources allow for better experimental design and increase the quality in research, presenting an excellent example for electronic study design tools. A combined usage of these online tools and eLN would improve the quality and reproducibility of scientific research.

---

## 5 Data-Related Quality Aspects of eLNs

An additional data-related quality consideration, besides support with experimental design, concerns the metadata. Metadata is the information important for understanding experiments by other researchers. Often, the metadata is generated by the programme itself and needs a manual curation by the researcher to reach completeness. In principle, the more metadata is entered into the system, the higher the value of the data since it makes it possible to be analysed and understood in greater detail, thus being reproducible in subsequent experiments.

Guidance on how to handle metadata is provided in detail in chapter “Data Storage” in the context of the FAIR principles and ALCOA. In brief, ALCOA, and lately its expanded version ALCOAplus, is the industry standard used by the FDA, WHO, PIC/S and GAMP to give guidance in ensuring data integrity. The guidelines point to several important aspects, which also explain the acronym:

- *Attributable*: Who acquired the data or performed an action and when?
- *Legible, Traceable and Permanent*: Can you read the data and any entries?
- *Contemporaneous*: Was it recorded as it happened? Is it time-stamped?
- *Original*: Is it the first place that the data is recorded? Is the raw data also saved?
- *Accurate*: Are all the details correct?

The additional “plus” includes the following aspects:

- *Complete*: Is all data included (was there any repeat or reanalysis performed on the sample)?
- *Consistent*: Are all elements documented in chronological order? Are all elements of the analysis dated or time-stamped in the expected sequence?
- *Enduring*: Are all recordings and notes preserved over an extended period?
- *Available*: Can the data be accessed for review over the lifetime of the record?

These guiding principles can be used both for paper and electronic LNs, and the questions can be asked by a researcher when documenting experiments. However, eLNs can provide support to follow these regulations by using templates adhering to these guiding principles, increasing transparency and trust in the data. The electronic system can help to manage all of these requirements as part of the “metadata”, which is harder to ensure with written lab notebooks or at least requires adaptation of certain habits.

Having these principles in place is the first important step. Then, it needs to be considered that the data will be transferred between different systems. Transferring data from one specialized system to another is often prone to errors, e.g. copy-and-paste error, loss of metadata, different formats and so on. Thus it must always be considered when the information is copied, especially when data are processed within one programme and then stored in another. It has also to be ensured that data processing can be reproduced by other researchers. This step should be documented in LNs, and eLNs can provide a valuable support by linking to different versions of a file. There are myriad of applications to process raw data which come with their own file formats. Ideally, these files are saved in a database (eLN or application database), and the eLN can make a connection with a dedicated interface (Application programming interfaces, API) between the analysis software and the database. This step ensures that the metadata is maintained. By selecting the data at the place of storage yet still using a specialized software, seamless connection and usage of the data can be ensured between different applications. For some applications, this issue is solved with dedicated data transfer protocols to provide a format for exporting and importing data. The challenge of having a special API is

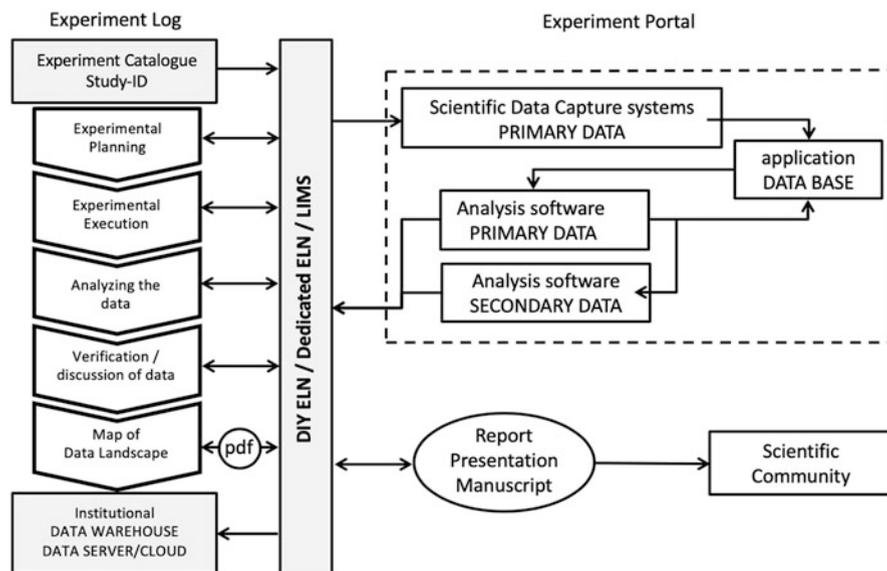
recognized by several organizations and led to the formation of several initiatives, including the Pistoia Alliance (<https://www.pistoiaalliance.org>) and the Allotrope Foundation (<https://www.allotrope.org>). These initiatives aim to define open data formats in cooperation with vendors, software manufactures and pharmaceutical companies. In particular, Allotrope has released data formats for some technologies which will be a requirement for effective and sustainable integration of eLNs into the lab environment. Therefore, it is worth investing time to find solutions for supporting storage, maintenance and transfer of research data.

---

## 6 The LN as the Central Element of Data Management

Today's researcher faces a plethora of raw data files that in many cases tend to stay within separated data generation and analysis systems. In addition, the amount of data a scientist is able to generate with one experiment has increased exponentially. Large data sets coming from omics and imaging approaches generate new data flows in scientific work that are not captured by eLNs at all, primarily due to a lack of connectivity. Thus, the eLN needs a structured approach to connect all experimental data with the raw data (Fig. 2). This can most likely still be achieved with a pLN, but support from digital tools seems to be obviously advantageous, which can be in the form of a dedicated eLN or even a DIY approach. In any case, the approach will have to evolve from a scientist's criteria of an integrated data management system meeting several requirements: Documentation of intellectual property generation, integrated raw data storage and linking solutions and enhanced connectivity with lab equipment, colleagues and the scientific community.

Several eLNs provide a user interface for document generation: Independent of the kind of document created, the eLN should be the starting point by creating a unique Study-ID and will add directly, or at least link, the credentials of the scientist to the document. Some eLNs require the input of some basic information such as the project, methods used or co-workers involved and will compile a master project file with some essential metadata. Existing files like protocols or data packages can then be linked in the master file. The master project file within the eLN acts as a map for the landscape of scientific experiments belonging to one project. In a DIY approach, these steps can also be created manually within a folder system (e.g. OneNote, Windows Explorer, etc.) and a master file created in the form of a word processor file (e.g. MS Office Word) or a page within OneNote. Of course, this can also be achieved within a pLN, but, again, it is more effort to structure the information and keep an overview. In all cases, each step of the scientific experiment process can then be automatically tracked either by the eLN or manually to allow for fast data location and high reproducibility. Even if copied elsewhere, recovery and identification of ownership is easy with the help of the unique Study-ID. At the end of a series of experiments or projects, the researcher or an artificial intelligence within the eLN decides which files to pool in a PDF file, and after proofreading, somebody can mark the project as completed. Thereby, the final document, best in the form of a PDF file, will not only contain some major files but also the master file with the Study-IDs of



**Fig. 2** Workflow for the lab environment with the eLN/LIMS being the central element. Each experimental setup will start with a unique Study-ID entry into the eLN in the experimental catalogue which will be used throughout the whole experiment and allow for tagging during all steps. The eLN will be the hub between the experimental procedure (left) and the data collection and reporting (right). The eLN should collect all different types of data or at least provide the links to the respective storage locations. One of the last steps is the summary of the experiment in the “Map of the Data Landscape” in the form of a PDF file. Next-generation dedicated eLNs could themselves be used to create such a document, thereby providing a document for reporting with the scientific community and storage in the data warehouse

all data sets that are searchable and identifiable across the server of an organization. This will also allow smart search terms making the organization of experimental data very straightforward.

Unique Study-IDs for projects across an organization are key prerequisites to cross-link various sets of documents and data files into a project (see the next section on how to set up). Using an eLN as the managing software connecting loose ends during the data generation period of a project and even right from the beginning will free the researcher from the hurry to do so at the end of the project. This type of software can easily be combined with an existing laboratory inventory management system (LIMS) not only fulfilling the documentation duties but also adding the value of project mapping.

In any case, enhanced connectivity is key for successful transparency. Enhanced connectivity can be understood as connectivity across different locations and between different scientists. In terms of different locations, a researcher can access data, such as the experimental protocol, within the eLNs from the office and lab areas. A tablet can be placed easily next to a bench showing protocols that could be amended by written or spoken notes. Timer and calculation functions are nothing

new but form an essential part of daily lab work and need to be easily accessible. In very advanced applications, enhanced connectivity is implemented using camera assistance in lab goggles capturing essential steps of the experiment. If coupled to optical character recognition (OCR) and artificial intelligence (AI) for identifying the devices and materials used and notes taken, a final photograph protocol might form part of the documentation. Information on reagent vessels would be searchable and could easily be reordered. Connectivity to compound databases would provide information about molecular weight with just one click and can be used to help a researcher check compound concentrations during an experiment. A tablet with the facility to connect to lab devices can be used to show that the centrifuge next door is ready or that there is available space at a sterile working bench or a microscope. These benefits allow seamless, transparent and detailed documentation without any extra effort.

Connectivity between researchers within an organization can be optimized in a different way by using electronic tools. At some point, results may need to be presented on slides to others in a meeting. Assuming best meeting practice is applied, these slides will be uploaded to a folder or a SharePoint. After a certain period, it may become necessary to access again the data presented in the files uploaded to the SharePoint. SharePoint knowledge usually lacks connectivity to raw data as well as to researchers that have not been part of a particular meeting. However, a unique identifier to files of a project will allow for searchability and even could render SharePoint data collections redundant. Feedback from a presentation and results of data discussion might be added directly to a set of data rather than to a presentation of interpreted raw data. The whole process of data generation, its interpretation and discussion under various aspects followed by idea generation, re-analyses and new experiments can become more transparent with the use of electronic tools. An eLN may also potentially be used to enhance clarity with respect to the ownership of ideas and experimental work.

---

## 7 Organizing and Documenting Experiments

Unique Study-IDs are important components for appropriate and transparent documentation and data management. Creating a structured approach for the accessibility of experimental data is reasonably straightforward, although it requires discipline. Unique Study-IDs should be used in all lab environments independent of the LN used. The DIY eLN is discussed here as an example by which to structure the data, which can be achieved with a generic software such as OneNote, as described by Oleksik and Milic-Frayling (2014), or can be implemented within Windows Explorer and a text editing programme, such as Word. Dedicated eLNs will provide the Study-ID automatically, but in the case of a DIY eLN (or LN), this must be set up and organized by the researcher. Several possibilities exist, and one can be creative, yet consistent, to avoid any confusion. The Study-ID will not only help to identify experiments but will also assist a scientist in structuring their data in general. Therefore, it should be kept simple and readable.

**Table 4** Implementing a unique Study-ID system, an example

Number of experimental categories	Description of the experimental category	Number of the experiment within the category	Example file name	Unique identifier (further differentiation possible to label material)
01	IHC – brain	01	01.01 effect of treatment X	01.01
		02	01.02 confirmatory experiment for 01.01	01.02
		...	...	...
02	SH-SY5Y	01	02.01 titration of rapamycin to test fibre length	02.01
		02	02.02 analysis of gene expression after treatment Y	02.02
		...	...	...
03	Primary cortical cells	01	03.01 Western blot analysis to determine expression of proteins after treatment G	03.01
		02	03.02 pull-down assay of X to test interaction with Y	03.02
		...	...	...
04	...	...	...	...

A possible labelling strategy could be to include the date, with several formats possible, depending on preference (e.g. DDMMYY, YY-MM-DD, YYMMDD). It is, however, advisable to agree on one format which is then used throughout the research unit. The latter one (YYMMDD), for example, has the advantage that files are directly ordered according to the day if used as the prefix of a file name or, even better, if used as a suffix. When used as a suffix, files with the same starting name are always ordered according to the date. If several experiments are planned on the same day, a short additional suffix in the form of a letter can distinguish the experiments (YYMMDD.A and YYMMDD.B). Such a unique identifier could be created on the day when the experiment was planned and should always be used throughout the experiment to identify experimental materials (such as in labelling cell culture plates, tubes, immunohistochemistry slides, Western blots and so on). It is also advisable to add some more information to the name of the folder to get a better overview and use only the Study-ID and short numbers for temporary items or items with little labelling space (e.g. for cell culture plates and tubes).

Another possibility to categorize experiments is by using “structured” numbers and categories created by the researcher and then using running numbers to create unique IDs. Table 4 provides examples of potential categories. The researcher can be creative here and set up a system to fit their requirements. The main purpose is to create a transparent system to be able to exactly trace back each sample in the lab.

This was found to be especially useful for samples that might need to be analysed again, e.g. cell lysates which were immunoblotted a second time half a year later for staining with a different antibody. To achieve this level of organization, the tubes only need to be labelled with one number consisting of a few digits. To take an example from Table 4, six tubes were labelled with the Study-ID (03.01) plus an additional number differentiating the tube (e.g. 03.01.1 to 03.01.6). This approach clearly identified the lysates from experiment 03.01 which is the lysates of primary cortical cells treated with substance “G” under six different sets of conditions. If such a labelling system is used for all experimental materials in the lab, it will ensure that each item can be unambiguously identified.

This unique identifier should be used for all files created on every device during experiments. In case this type of system is adopted by several researchers in a research unit, another prefix, e.g. initials, or dedicated folders for each researcher have to be created to avoid confusion. The system can be adopted for the organization of the folders housing all the different experiments. To more easily understand the entire content of an experiment, there should be a master file in each folder providing the essential information on the experiment. This master file can be a text file which is always built from the same template and contains the most important information about an experiment. Templates of this type are easy to create, and they can simplify lab work, providing an advantage for eLNs over the pLN. The template for the master file should contain at least the following sections or links to locations for retrieving the information:

- (a) Header with unique ID
- (b) Name of the researcher
- (c) Date
- (d) Project
- (e) Aim of the experiment
- (f) Reagents and materials
- (g) Experimental procedure
- (h) Name or pathway to the storage folder of raw data (if different to the parent folder)
- (i) Results
- (j) Analysis
- (k) Conclusions
- (l) References

The master file should be saved in PDF format and should be time-stamped to properly document its time of creation. The usage of a master file requires agreement by the organization, perseverance from the researcher and control mechanisms. Depending on the organization and the methods applied regularly, written standard operating procedures (SOPs) can ensure a common level of experimental quality. These files can be easily copied into the master file, and only potential deviations need to be documented. It is recommended to back up or archive the complete experimental folder. How this can be achieved depends on the infrastructure of the facility. One possibility would be that the files are archived for every researcher and

year or only if researchers leave the lab. This creates a lot of flexibility and makes searching for certain experiments much more convenient than in a pLN. This type of simple electronic organization combines the advantage of a pLN and an eLN: All electronic files can be stored in one place or directly linked, and it is electronically searchable, thus increasing transparency without the association of any additional costs.

Well-structured scientific data can be more easily transferred to an eLN, and even changes in the eLN system will not have a devastating effect on original file organization. Based on the requirements for data documentation, we suggest the following steps to a better data documentation policy that ultimately will improve data reproducibility:

1. Identify the best system for your needs:

There are different means by which data can be documented within a lab environment. Identifying the best approach for the specific requirements in a lab saves time and resources by optimizing workflows.

2. Structure your data:

Well-structured data increases the possibility to find and retrieve the information. Using a unique Study-ID is a good step towards achieving this.

3. Structure the data sources and storage locations:

Organizing data-storage locations, connecting them to experimental documentation and considering backup solutions are important for transparency.

4. Agree to and follow your rules:

Agree on minimal operational standards in the research unit that fulfil the long-term needs, e.g. adherence to ALCOA, design and documentation of experiments or IP ownership regulations.

5. Revise your data strategy and search for improvements:

Search for tools that allow for better connectivity and simplify documentation.

In summary, documentation is the centrepiece for best research practices and has to be properly performed to create transparency and ensure data integrity. The adoption of eLNs along with the establishment of routinely applied habits will facilitate this best practice. The researchers themselves have to invest time and resources to identify the appropriate tool for their research unit by testing different vendors. Once the right tool is identified, only regular training and permanent encouragement will ensure a sustainable documentation practice.

---

## References

- Dirnagl U, Przesdzing I (2017) A pocket guide to electronic laboratory notebooks in the academic life sciences [Version 1; Referees: 4 Approved]. Referee Status 1–11. <https://doi.org/10.12688/f1000research.7628.1>
- Groemping U (2018) CRAN task view: design of experiments (DoE) & analysis of experimental data. <https://CRAN.R-project.org/view=ExperimentalDesign>

- Guerrero S, Dujardin G, Cabrera-Andrade A, Paz-y-Miño C, Indacochea A, Inglés-Ferrándiz M, Nadimpalli HP et al (2016) Analysis and Implementation of an electronic laboratory notebook in a Biomedical Research Institute. PLoS One 11(8):e0160428. <https://doi.org/10.1371/journal.pone.0160428>
- Kanza S, Willoughby C, Gibbins N, Whitby R, Frey JG, Erjavec J, Zupančič K, Hren M, Kovač K (2017) Electronic lab notebooks: can they replace paper? J Cheminf 9(1):1–15. <https://doi.org/10.1186/s13321-017-0221-3>.
- Laajala TD, Jumppanen M, Huhtaniemi R, Fey V, Kaur A, Knuutila M, Aho E et al (2016) Optimized design and analysis of preclinical intervention studies in vivo. Sci Rep. <https://doi.org/10.1038/srep30723>
- Nussbeck SY, Weil P, Menzel J, Marzec B, Lorberg K, Schwappach B (2014) The laboratory notebook in the twenty-first century. EMBO Rep 6(9):1–4. <https://doi.org/10.15252/embr.201338358>.
- Oleksik G, Milic-Frayling N (2014) Study of an electronic lab notebook design and practices that 592 emerged in a collaborative scientific environment. Proc ACM Conf Comput Support Coop Work CSCW:120–133. <https://doi.org/10.1145/2531602.2531709>
- Riley EM, Hattaway HZ, Felse PA (2017) Implementation and use of cloud-based electronic lab notebook in a bioprocess engineering teaching laboratory. J Biol Eng:1–9. <https://doi.org/10.1186/s13036-017-0083-2>
- Rubacha M, Rattan AK, Hosselet SC (2011) A review of electronic laboratory notebooks available in the market today. J Lab Autom 16(1):90–98. <https://doi.org/10.1016/j.jala.2009.01.002>

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

