TUM School of Nature Science &
TUM School of Engineering and Design
Technical University of Munich

Data Steward Training – Final Report
Olaf Ackermann and Annika Schneider

FAIR Research Data Management with Electronic Lab Notebooks (ELNs)

Abstract

This report discusses the importance of Electronic Lab Notebooks (ELNs) in facilitating FAIR (Findable, Accessible, Interoperable, and Reusable) Research Data Management. It highlights the limitations of traditional paper-based laboratory notebooks and the advantages of transitioning to an electronic format. It proposes ideas of how to enable FAIR research data management with ELNs.

1. Introduction

Laboratory notebooks serve as a crucial tool for documenting and providing evidence of experiments, thereby eliminating the need for repeat attempts. However, paper-based laboratory notebooks present several disadvantages, including the lack of a search function, limited metadata, susceptibility to damage, and lack of durability. Moreover, they do not support copy-pasting or integration of analytical data.

This is why nowadays, ELNs are much preferred over paper-based laboratory books. There are several ELN providers, which all have different strengths and weaknesses and there is no one-size-fits-all solution when it comes to choosing an ELN. The choice of an ELN should be based on its compatibility with the workflow. Moreover, different technical approaches exist, including file-based, database, and cloud solutions.

Furthermore, the different ELNs are not compatible with each other. This means that data and recordings from one ELN cannot be transferred to another. If the research group moves, this can lead to problems.

For synthetic chemistry, one free, open-source option is "Chemotion". Later, we will validate, to what extend Chemotion enables FAIR research data management.

2. Ideas for FAIR ELNs

FAIR research data management is crucial for transparent and re-usable research. We will now state some ideas, of how ELNs can support their users in a FAIR workflow.

(1) ELNs should aid in all steps of the research data lifecycle:

a. Integrate a planning tool which allows you to predefine experiments, alongside with some FAIR documentation standards.

b. Collection, Processing, Analysis, and Saving is typically covered in ELNs (that's what they are for).

c. Integrate seamless publishing, e.g. via enabling to mark collections publicly available and citable, assign DOIs.

d. To enable reusability, the ELN must have an understandable, intuitive, and well documented structure of how data and results are presented.

(2) ELNs should facilitate writing RDM plans:

a. The ELN planning tool should contain every step relevant for an RDM plan.

b. There should be several RDM plan templates available.

c. One might even consider auto-generation of RDM plans based on the ELN planning tool, and integrated LLM assistance when writing the plan.

(3) ELNs should work with controlled metadata / ontologies.

a. Each element should be documented with rich metadata, following industry standards.

b. Ideally, metadata is embedded in an ontology.

(4) ELNs should ensure data quality.

a. ELNs could implement automated data quality checks, warning users when data has many outliers or missing values.
b. ELNs could provide automatic data visualizations, helping the user to quickly detect data quality issues.

In the next section, we will evaluate to what extend Chemotion already supports FAIR research data management.

3. Case Study - Chemotion

Chemotion already successfully provides support in the key steps of the research data lifecycle. The user can design a research plan, can collect, process, analyze and save data, and Chemotion robustly supports the publication workflow. Furthermore, Chemotion's data annotation is embedded in an industry-standard ontology. This means that the data stored in Chemotion is well-structured and follows a standardized format, making it easier for researchers to understand and use the data.

Despite its strengths, there are several areas where Chemotion could be improved. Chemotion is not very intuitive to use, which can make it difficult for new users to navigate and utilize its features. Additionally, it lacks interfaces, making it a silo software. This means that it does not integrate well with other software, limiting its versatility. Specifically, it does not have an interface to "ChemDraw", a popular tool for drawing chemical structures.

Another area where Chemotion falls short is in its support for writing Research Data Management (RDM) plans. RDM plans are crucial for outlining how data will be managed during a research project, and the lack of support for this in Chemotion is a significant drawback. Lastly, Chemotion does not help to ensure data quality. It lacks features that could automatically check the quality of data, such as warning users when data has many outliers or missing values. This means that users must manually check their data for quality, which can be time-consuming and prone to errors.

4. Tips for Introducing ELNs

When implementing a new electronic lab notebook (ELN) such as Chemotion, it is crucial to ensure that the users are taken along in the process. This means that users should be involved in the decision-making process, given adequate training on how to use the software, and provided with ongoing support to address any issues or concerns they may have. Before settling on a particular ELN, it is advisable to test different ones to see which best fits the needs and workflow of the users. Each ELN has its strengths and weaknesses, and what works well for one group may not work as well for another. Therefore, it is important to evaluate several options, taking into consideration factors such as ease of use, compatibility with existing systems, and the specific needs of the users.

References

https://websites.fraunhofer.de/lost-in-life-sciences/?p=52

https://www.g2.com/categories/electronic-lab-notebook-eln

https://www.publisso.de/fileadmin/user_upload/PUBLISSO/PUBLISSO_ELN-Wegweiser_2020-10-05_view.pdf

https://chemotion.net/