OPEN RESEARCH NOTES: AN ELECTRONIC LAB NOTEBOOK

Key words: data management, web development, open source

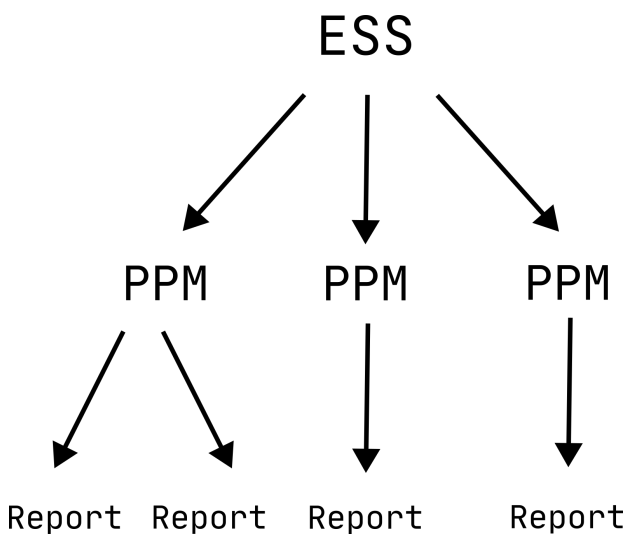
Authors:
Pedro Sader Azevedo (student), IC – UNICAMP
Prof. Dr. Christoph Friedrich Deneke (supervisor), IFGW - UNICAMP
Prof. Dr. André Santanchè (reviewer), IC - UNICAMP

INTRODUCTION:

Collecting and analyzing data is one of the core tasks in exact sciences like Physics, Chemistry and Biology as well as various engineering disciplines. As most research data is either directly acquired with a computer (connected to some equipment) or later entered into one for analysis and visualization, internal management of digital data has become an important requirement for the research that takes place inside universities. Nevertheless, the handwritten lab book is still the most common tool for keeping track of research data, despite its shortcomings in handling and interacting with digital research data.

This form-factor makes it harder to organize data semantically, since notes are necessarily written in chronological order. In addition, sharing information written on a physical notebook is unreliable and cumbersome, burdening the communication between research students and mentors. More importantly, the difficulty in sharing raw data makes it harder for the supervisor to check the integrity and the ethics of a student's research.

An electronic lab notebook (ELN) can solve all these problems by storing research data (i.e. measurements collected from a sample) and metadata (i.e. what sample was investigated, how the measurement was performed etc.) at the same place, in an orderly manner. Although some commercial solutions exist, we are not aware of any that is open source, manageable for smaller research groups found in universities, and includes a complete suite of data management tools for sample, measurement and report data, mirroring a typical workflow in applied Physics. And thus began our project, Open ResearchNotes (ORN).



In ORN, experiments, samples, and simulations (ESS) are understood as the base entities of research. Each ESS has all raw data associated with its procedures, processings and measurements (PPM) stored within it. Finally, the PPMs contain one or more Reports that encompass processed data or analyses of the data (i.e. what was learned). Most importantly, this hierarchical structure allows raw data to be easily traced back to the base entity that originated it, thus avoiding the separation of metadata and raw data.

GOALS:

As we discussed the disadvantages of the handwritten lab notebook, we arrived at a few requirements that we wanted our project to fulfill. We grouped requirements that satisfied the same value, and summed up our discussion in the table below. This table represents all the goals of the project for the time being, and can be thought of as a general high-level roadmap for the project.

Value	Requirements
Organization Data stored in ORN is guaranteed to be tidy	• Store experiment data (raw measurements) and metadata (how, where, when, etc, the measurements were taken) inseparably • Enforce a strict hierarchy of entities (E/S/S → P/P/M → Reports) to standardize data management • Real world mapping: store the real life locations (which lab, which drawer) and identifiers of samples and instruments
Cooperability The program enables working with others	• Users can be organized into groups, assuming different roles with different levels of privilege (e.g. student, supervisor, admin, etc) • Samples and instruments can be shared between groups and users
Permanence Data won't be mistakenly removed or lost	• Destructive actions are restricted to certain roles of a group (e.g. supervisors) • Research materials are kept around even when a student leaves the group

Independence The platform is designed to avoid vendor lock-in	• The project is open source and can be hosted by anyone who has the proper infrastructure • Data can be exported in a "machine-readable" format and imported back into any instance of ORN • Data can be exported in a "human-readable" format, and be used regardless of ORN
Retrievability Data stored in ORN is always easy to find	• Users can search for terms in a global search bar, that indexes all data accessible to that user • Measurements and instrumentation journal entries can be tagged with custom labels to facilitate grouping • All data and metadata can be sorted and filtered using a graphical user interface
Convenience Using ORN saves time and avoid stress	• Allow users to define templates, so that they can avoid repeatedly typing out boilerplate • The user interface is simple and focused, compelling the user to concentrate on the task at hand

METHODS:

As for our techstack, we employed Python's Flask framework for the server-side backend, SQLAlchemy for the database abstraction layer, Meilisearch/Elasticsearch for the full-text search engine, and Bootstrap for the frontend.

RESULTS:

Our project has seen significant embracement in the Department of Applied Physics (DFA) within Unicamp's Institute of Physics (IFGW). ORN is currently being used to safely and conveniently store data by more than twenty researchers. They spread across multiple different labs, such as the Laboratory of Preparation and Characterization of Materials (LCPM), the Laboratory of Photovoltaic Research (LPF), and the Laboratory of Device Research (LPD). In two of those laboratories, usage of ORN has become mandatory.

In addition to that, ORN was featured in a few conferences. Last year, it made its debut in the WTD 2022 and, this year, it will be presented in Python Brazil 2023.

CONCLUSION:

Developing an Electronic Lab Notebook turned out to be a great opportunity to learn the inner workings of science as well as a multitude of software engineering skills. The project was well received by the Unicamp community and has proved to be a very successful initiative.

REFERENCES

IOP Publishing data availability policy. Available on: <https://publishingsupport.iopscience.iop.org/iop-publishing-data-availability-policy/>. Accessed on: 29 nov. 2022.

FAPESP. **Gestão de dados**. Available on: <https://fapesp.br/gestaodedados>.

The Electronic Lab Notebook in 2022: A comprehensive guide. Available on: <https://labfolder.com/electronic-lab-notebook-eln-research-guide/>. Accessed on: 29 nov. 2022.

How to use fixtures — pytest documentation. Available on: <https://docs.pytest.org/en/latest/how-to/fixtures.html#how-to-fixtures>. Accessed on: 26 Nov. 2022.

Utilities for with-statement contexts — Python 3.11.0 documentation. Available on: <https://docs.python.org/3/library/contextlib.html>. Accessed on: 29 Nov. 2022.

How to use skip and xfail to deal with tests that cannot succeed — pytest documentation. Available on: <https://docs.pytest.org/en/7.1.x/how-to/skipping.html>. Accessed on: 29 Nov. 2022.

ATLASSIAN. **Gitflow Workflow | Atlassian Git Tutorial**. Available on: <https://www.atlassian.com/git/tutorials/comparing-workflows/gitflow-workflow>. Accessed on: 27 Nov. 2022.

REDHAT. **What is a REST API?** Available on: <https://www.redhat.com/en/topics/api/what-is-a-rest-api>.