Lab Dragon: An Electronic Laboratory Notebook to Support Human Practices in Experimental Science

Marcos Frenkel
NCSA
University of Illinois
Urbana IL, USA
marcosf2@illinois.edu

Wolfgang Pfaff
Department of Physics and MRL
University of Illinois
Urbana IL, USA
wpfaff@illinois.edu

Santiago Núñez-Corrales
NCSA/IQUIST
University of Illinois
Urbana IL, USA
nunezco2@illinois.edu

Rob Kooper
NCSA
University of Illinois
Urbana IL, USA
kooper@illinois.edu

***Abstract*—Lab notebooks are an integral part of science, documenting and tracking research progress in laboratories. However, existing electronic solutions have not properly leveraged the full extent of capabilities provided by a digital environment, resulting in most physics laboratory notebooks merely mimicking their physical counterparts on a computer. To address this situation, we report here preliminary work toward a novel electronic laboratory notebook, *Lab Dragon*, designed to empower researchers to create customized notebooks that optimize the benefits of digital technology.**

***Index Terms*—collaborative research environments, electronic laboratory notebooks, experimental science support, laboratory notebooks**

I. Introduction

Reproducible experimental results constitute a fundamental pillar of scientific research [1]. An experiment lacks relevance unless its results can be replicated elsewhere. Real experiments happening in a laboratory are often messy, taking many winding paths until results are collected and analyzed. Because experiments can have branches where the research goes in different directions, it is important to keep track of the common path in the experiment and where the branches happen. It is the job of research notes inside a laboratory notebook to keep a faithful, actionable record of the events that transpired during the execution of an experiment.

Research notes are a special type of metadata that record the context of all other data and metadata, which is essential for collaboration and continuity in experimental research [2]. While most data and metadata associated with an experiment are usually values, parameters, images, etc., research notes store the human information associated with the experiment [3]. Research notes thus provide a most complete and detailed account of what happened during an experiment, with as much detail as possible, including every decision process and justification that a researcher made when collecting data [4]. Good research notes let other researchers know exactly what happened in the lab, why it happened, and what they would need to do to reproduce that experiment.

Furthermore, research notes are an invaluable tool for those supervising experimental work. Often, the researcher performing the experiments, such as a graduate student, is not yet an expert in the field, while the supervisors, such as postdocs or professors, may not be available at the time of the experiment. Therefore, it is extremely important for researchers to create a detailed log of the events and thoughts that led to the collected data. Since research is often an exploratory endeavor, the results that a researcher collects might not always be the ones that were expected.

Finally, research notebooks need to make it easier to collect the information right next to the experiment. When a researchers is working on a experiment they can be at different stations, and want to have their notes with them at all times, to record what happens, their thoughts, etc. A digital version of the notebook makes it easy to have access everywhere without ever forgetting where they placed the notebook.

A. What is a Lab Notebook

Lab notebooks are dedicated notebooks used to write and collect the previously described research notes [5]. In the past, a physical notebook was usually present in the laboratory where members would write the events that happened in the lab, their thoughts on the experiments, as well as draw plots or sketches to accompany the notes. Even today, when most work happens electronically on computers and data plotting is done through code, many research groups opt to print those plots and physically paste them in a paper notebook.

As a conceptual object, lab notebooks need to fulfill several objectives:

1) Tell the Story of an Experiment as a Whole: Another person should be able to read a notebook from start to finish and get a full picture of the research questions, the steps taken to answer those questions, the difficulties encountered, how they were solved, or if pivoting was necessary, why and how it was done.

2) Help Supervisors Verify the Research Happening in Their Laboratories: In most academic laboratories, the primary researchers performing experiments are graduate students or research assistants. Supervisors, such as professors or postdocs, often oversee multiple projects and may not be present during every experiment. Lab notebooks allow supervisors to review the detailed records of what was done, ensuring that proper procedures were followed, and providing a basis for feedback and further guidance.

3) Share Personal and Detailed Experiences with Other Team Members: Often, some procedures are repeated after a significant length of time has passed, where the original researcher performing that experiment is no longer a member of the team. Lab notebooks serve as a repository of these experiences, allowing new team members to learn from past experiments and avoid repeating mistakes, while at the same time offering insight into similar procedures.

4) Help Researchers Organize Their Research and Extract Information: As time goes on, the amount of information that a lab notebook (or multiple notebooks) contains increases. A good lab notebook provides researchers with tools necessary to process all that information, helping them make informed decisions in future experiments.

B. An ideal electronic notebook?

While the digitalization of laboratory notebooks has been studied for some time [6], challenges in their implementation have been recognized early on as well [7]. Most notably, every different research group has different needs for their electronic lab notebook. As so, the following is simply a list of the common needs that a typical experimental physics laboratory needs out of their notebook.

One might be tempted to think that an ideal notebook is a tool that primarily focuses on documenting the experimental process in a structured and organized way is simple to make. However, even defining what structured and organized is, can be challenging when speaking with different research groups. This is why an ideal notebook should have extensive customization tools to let each individual group organize their own notes. At the same time, the structure and rules of a notebook should be defined by a few members, usually the principal investigator, given that if the lab is composed of ten people, having ten different organizational schemes often results in a mess.

Not less significant is to have strong linking options between notes and data. It is crucial to have hard links that remain over time between notes and the data they are referring to and vice-versa. Having a linking system that lets you identify, in both ways, what is the data that is referring to and what notes are referring to specific data, is fundamental in ensuring that collected data remains accessible after some time passes.

For a more in depth analysis of what an academic electronic lab notebook should look like, see [5].

II. Existing Options

Electronic lab notebooks have been around for a long time, with many options available on the market today. While many of them are excellent productivity tools and fulfill their roles well, most are not tailored to the specific needs of small research groups. Many of these notebooks are designed often focusing on fields like chemistry or biology, where the requirements for lab notebooks differ significantly from those in physics or other research laboratories.

A. OneNote

OneNote is currently one of the most popular tools for physicists to use as an electronic lab notebook [8]. Given that most universities already provide Office subscriptions, OneNote becomes a natural tool to use, especially since it makes entering new notes and reading them from any device a trivial affair. Additionally, OneNote is extremely versatile in its ability to record rich text, allowing users to write notes with images, tables, bullet points, numbered lists, and more. All of this makes OneNote the chosen tool for many physics laboratories.

While its flexibility in how notes are written is one of its best features, it is also one of the most detrimental aspects for research, especially for new lab members. Having free-flowing notes can be great, but they often become stream-of-consciousness documents, making it difficult for others to decipher the actual events that happened during the experiment. Similarly, it relies heavily on the discipline of users to keep it organized.

Given that OneNote only provides a space for notes to be written and stored, it is difficult to maintain direct relations to data. Currently, the way for researchers to show data in a note is through screenshots or by writing numbers in the note itself. This means that if researchers want to find the data again, they depend entirely on the discipline of the writer to record file paths or other identifiers used to generate those images. As time goes on and the lab notebook grows, the number of files and data measured also increases, exacerbating this issue.

B. Other Alternatives

Several other alternatives exist on top of OneNote, but most are focused on the needs of chemistry or biology labs. Electronic lab notebooks like Lab Folder [9] or Labii [10] offer numerous additional features on top of their note-taking capabilities. These features often include inventory management, task management, and strict experiment definitions and more. While these features may be necessary for some labs to function, they often get in the way in smaller groups like an academic physics laboratory. Jupyter notebooks, thanks to their re-executability and storage format, have gained traction. However, they lack enabling functionality for the research process overall [11] and only become another tool in the research kit.

In smaller academic research groups, experiments are often never repeated. Once an experiment is done, the researchers learn what they can from it and move on to the next iteration. The scale of their experiments and the tools also make inventory management relatively simple, eliminating the need for sophisticated solutions. As Harvard states [12], no electronic lab notebook is a one-size-fits-all solution, and most of these extra features end up making the notebook more difficult to use as well as to read.

C. Why create a new alternative?

Despite the existence of many viable options, particularly in fields like biology or chemistry, the main reason for creating a new electronic lab notebook is the lack of focus—or rather, the focus on non-ideal features that are not required or prioritized by small physics research groups or similar teams. On one hand, tools like OneNote do fulfill this role, allowing researchers to quickly and effortlessly write and read information from the notebook. However, they are usually too flexible in what they allow. OneNote's flexibility can lead to notes that are sometimes not ideal. It is easy to place images and use rich text, but it is also easy for researchers to dump every thought they have in an unstructured manner, leaving other researchers confused about what is happening. On the other hand, other existing tools also come with a large set of features that are not only not needed for smaller teams, but get in the way from the notes themselves.

Additionally, most existing notebooks store information in proprietary formats, which poses a risk of losing data if those products disappear in the future. At the same time, this storage method makes it extremely difficult to tightly integrate the notes from an experiment with the data itself, hindering seamless transitions between the two.

III. Lab Dragon

Lab Dragon [13] is a new kind of electronic lab notebook designed from the ground up to meet the needs of experimental research laboratories, offering extensive customization to accommodate various fields. Researchers write annotations on different *entities*, and a collection of these entities forms a notebook. Entities come in different types and categories, each with specific meanings and behaviors. By providing a flexible and adaptive customization system, Lab Dragon allows every group to tailor their logging processes in the way that best suits them.

Additionally, each field within a entity contains specific types of data, ensuring that the information is structured and accurate. For example, a field designated for content blocks will only accept text, a field for users will require valid usernames, and a field for time will need to be in the correct date format. This feature allows for a high degree of customization while maintaining the integrity and consistency of the data in the notebook.

At its core, an entity is simply a key-value store that allows you to store values for specific keys. Every entity has a name/title, a creation timestamp, and an author (or multiple authors for collaborative work), but each type of entity also has its own specific fields and requirements. All types of entities can be broadly categorized into two main groups: story entities and data entities.

A. Story Entities

Story Entities are created by researchers as they conduct their research. They provide the context necessary to understand the events that occur in the laboratory. Lab Dragon natively provides three different types of story entities: Projects, Tasks, and Steps.

1) Projects: Project entities represent broad objectives or research directions. They are organizational entities used to group other entities. They can be as broad as a place to store all the entities of a notebook or more focused, like a specific experiment. They fulfill the role of a folder in a file system, a section in OneNote, or a notebook in Evernote. Projects can have any story entity as their children.

2) Tasks: Task entities represent complex actions with specific intents and results. This does not mean that the researcher needs to know exactly what to expect from experiments, but rather, there should always be a goal or objective guiding their actions. Examples include "characterizing an object" or "figuring out how to operate an instrument". Tasks can have other Tasks and Steps as their children, but not Projects.

3) Steps: Step entities represent atomic actions. The goal of Steps is to encourage researchers to think about every individual action they take to make an experiment possible. Steps can be detailed or brief, but they should represent an event that cannot be broken down any further.

The main difference between story and data entities, is that story entities contain *content blocks*. Writing content blocks in story entities is how the story of the experiment is recorded. Every story entity allows the user to write as many content blocks as they want. Content blocks support various rich text features such as bold text, italics, bullet points, numbered lists, tables, and images. They also allow for mentions of other entities in the form of links or by directly embedding plots or images associated with Instances.

We believe that these three native types are general enough to describe any range of actions and thoughts that occur in the lab, while also gently guiding researchers toward better note-taking practices.

B. Data Entities

Data entities, as their name implies, are entities that are used to annotate data. By default, Lab Dragon comes with two separate types of data entities, *Instances* and *Buckets*.

Instances are used to represent specific measurements or "instances" of an experiment. We have found that when taking data, experiments tend to have multiple small measurements that happen, which when combined form the big picture. The goal of Instance entities is to be able to reference whatever the specific research group considers a unit of data. Instances have an identifier field that links them to the real data, this can be either a file path pointing to the path of where the data lives, an ID that points to a specific line or table in a database or any other identifier needed to directly link it to that specific unit of data. Instances also hold specific information about the data, like the axis that the data is sweeping through, the number of points that are in that data, or any other descriptors for the shape of the data. Instances also have fields for arbitrary parameters, these working as general key-value pairs that are used to store extra metadata associated with that instance.

On top of that, data analysis results can be stored in the instance itself. Often, the raw data collected is not what researcher cares about, but the result of data analysis performed on the raw data. These results can be fit parameters, plot images, interactive plots (stored in an HTML file), or other files like, code scripts or Jupyter notebooks that were used to perform the analysis. By including these in the Instances themselves, we can create a very clear connection between the source of the analysis and their results. Having plots directly in the notebook allows the researcher to seamlessly integrate them in other entities, which allows them to display them directly between other notes and entities while keeping track of where this data came from and the analysis process it took to get there.

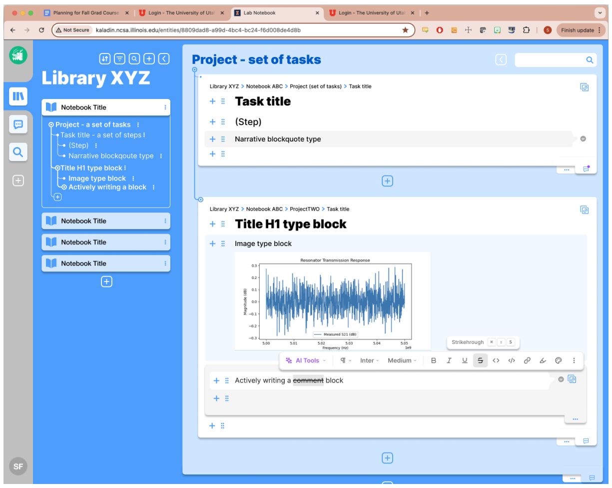

Fig. 1. Early design prototype screenshot. The screenshot shows a tree representation of a group of different entities. In the main page we can see a Project showing two different Tasks with a sequence of Steps. The content in them is a placeholder to illustrate how notes will be displayed in the a filled Lab Dragon Notebook.

As an experiment advances, Instances can become too numerous to easily handle. For this, Lab Dragon provides a second data entity: a Data Bucket. Data Buckets are used to group Instances together into a single manageable unit that is easier to handle. It is often the case that a single experimental setup can produce different experimental data. This way, the researchers can group and specify more general information about their setup, for example: certain instruments were used for all of the following instances or certain events might have compromised the following instances. Data Buckets also provide an easy way of linking data to story entities. Instead of individually marking all related instances to any number of entities describing a whole experiment, simply link the Data Bucket that holds all the generated Instances to a Project or Task, and any page that is a child of that Project or Task is going to be able to see any Instance contained in the bucket.

C. Mentions, Links, Tree Structure

The structure of a notebook is as important as the individual entities themselves. Lab Dragon notebooks have a tree structure. Every notebook has a root entity with several children, and depending on the type of those children, they themselves contain more children and so on. A single group can have multiple notebooks for different research

In contrast to a usual tree structure, a child entity can have multiple parent entities. There might be times when some action in the lab is being performed by more than a single experiment or is directly related to more than a single entity. Even though there are direct parental relationships between entities, events in the lab are usually very interconnected. For situations like this, entities can *mention* other entities to establish a link to a different part of the notebook. This can simply be a mention of another story entity to reference a previous event or to note that this action is related to or influenced by it. Mentions also allow us to directly embed data results into our story entities. If an Instance contains images, plots, parameters, or other information, and that Instance is in a Data Bucket attached to any parent entity, we can embed the images or plots themselves into the story entities. This way, we can have explicit links to where the data we are seeing comes from and all the steps that were taken to generate it.

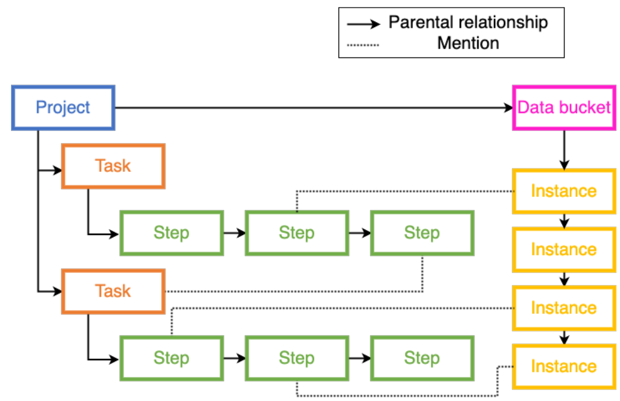

Fig. 2. Figure showing the tree structure of a small Lab Dragon notebook. In it we can see all five native types. The three story types: Projects in blue, Tasks in orange, Steps in green. We can see the two data types: Data Buckets in pink and Instances in yellow. We can also see the two types of relation that entities can have, a parental relationship representing ownership and indicating the order of how things happened, as well as mentions, used for indicating a relation between entities without changing the order of events or the harder structure of the notebook.

D. Inheritance and types

The needs of different research groups, or even different members and activities within the same research group, vary greatly. Therefore, groups need to tailor their research needs accordingly. For its customization, Lab Dragon entities follows the conceptual ideas of object-oriented programming, where new types of entities can be created based on a parent entity, and the child inherits every property from the parent.

All entities are children of the base ”Entity”. While the *Entity* type is hidden from users (they cannot create entities of Entity type), all the other entity types that users interact with are based on that one. As previously mentioned, the notebook comes with three different story entities and two different data entities. From these, every group can create their own types of entities to better organize and represent their specific research.

Custom entities can have arbitrary fields, which can hold arbitrary data; the researcher only needs to assign a name to a field and specify what kind of data it accepts. There is no limit on how many fields or what type of data those fields can accept (a single custom field can accept more than one data type). Custom fields can also have different structures for their values. They can accept a single value, a list of values, a dictionary of values, or file paths to other files in which the notebook knows how to interpret (a path to an image is understood to represent the image itself). Additionally, fields can be marked as mandatory or optional, allowing researchers to ensure that specific entities have the necessary data associated with them. This can be useful for stricter recording scenarios.

E. Current Implementation

At the time of writing this paper, we have a working proof-of-concept prototype for Lab Dragon [13]. While conceptually we believe the design is extremely robust, many aspects will evolve during the development of Lab Dragon.

Much of the data collected as part of the experiments is done on different systems, all around the lab. The data is written to a shared storage that all machines in the lab have access to. The notebook is built with the assumption that the team works and stores its data on a shared storage resource notebook has access to this data as well. The specific form this resource takes is not relevant; it can be a small NAS in a closet or hosted on a platform offsite, as long as the computer running the server can mount the resource, the notebook will function.

As mentioned before, we wanted an open format that can easily be converted to other formats and can be parsed by many program languages, we decided to use TOML files for easy reading and writing, but any common plain text file format can be used seamlessly. When creating a new notebook, a script is executed to generate a new root entity. This entity serves as the starting point for a notebook, and all other entities and paths are defined in relation to this root entity.

Once the root entity is created, the server hosting the back-end service can be run. The server is a Connexion REST API using a Flask web server, which loads the initial entity, retrieves all its children, and any mentioned files, creating an index of every item in the notebook. Once the back-end server is running, both the front-end server and other code can make regular HTTP requests. An example of other uses for the API is the automatic creation of Instances. Whenever a new measurement is taken in the lab, the computer making that measurement sends a POST request to create a new Instance. New analyses and plots generated from this data also get automatically added to Instances through HTTP requests, allowing researchers to mention and display plots as soon as they are created.

The front-end is built on top of a Next.js app. At the moment, it is a simple prototype interface that allows the user to create new entities within their selected parent and write content blocks inside any entity they want.

Whenever a new entity is created by the front-end, the back-end server creates a new TOML file in a selected file path and updates it live as the user continues to modify the notebook. Every modification, including changes to the entity name or any content block, is stored within the TOML file, along with whether the entity has been deleted or not. This is done to preserve the history of the file, with the idea of adding a history view later on.

IV. Conclusion

Lab Dragon positions itself in a unique niche among the existing options for electronic lab notebooks by addressing the specific needs of small, experimental research groups, fields like physics where the requirements for lab documentation differ from those in other sciences. Unlike many existing solutions that either offer too much flexibility or an excess of unnecessary features, Lab Dragon strikes a balance by providing a customizable, structured platform that integrates seamlessly with the research workflow.

The structure of Lab Dragon, with its distinct story and data entities, ensures that every aspect of the experimental process is meticulously documented and easily retrievable. By enabling researchers to create custom entity types, Lab Dragon offers the flexibility needed to adapt to various research environments while maintaining data integrity and consistency. The strong linking and mention systems ensure that the connection between data and the corresponding research notes remains clear and accessible, facilitating reproducibility and collaboration.

Looking forward, the development of Lab Dragon will focus on refining its features, improving user interface design, and expanding its capabilities. The goal is to create a robust, adaptable tool that not only meets the immediate needs of researchers but also evolves to support the dynamic nature of experimental science. To accomplish the best notebook experience we will keep working with researchers to add new features and test them.

In summary, Lab Dragon aims to empower researchers by providing a tailored, efficient, and reliable electronic lab notebook solution that enhances the documentation, organization, and reproducibility of scientific experiments. Through continuous development and community engagement, Lab Dragon aspires to become an indispensable tool in the modern researcher's toolkit.

Acknowledgment

The authors wish to thank support from the National Center for Supercomputing Applications and the Illinois Quantum Information Science and Technology Center at UIUC. They also thank Sebastian Frith for his design work on the UI.

References

[1] M. Schapira, R. J. Harding, O. L. N. Consortium, *et al.*, "Open laboratory notebooks: good for science, good for society, good for scientists," *F1000Research*, vol. 8, 2019.

[2] F. Wang, P. Liu, J. Pearson, F. Azar, and G. Madlmayr, "Experiment management with metadata-based integration for collaborative scientific research," in *22nd International Conference on Data Engineering (ICDE'06)*, pp. 96–96, IEEE, 2006.

[3] E. Garoufallou and C. Papatheodorou, "A critical introduction to metadata for e–science and e–research," *International Journal of Metadata, Semantics and Ontologies*, vol. 9, no. 1, pp. 1–4, 2014.

[4] A. Even, G. Shankaranarayanan, and S. Watts, "Enhancing decision making with process metadata: Theoretical framework, research tool, and exploratory examination," in *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, vol. 8, pp. 209a–209a, IEEE, 2006.
[5] S. G. Higgins, A. A. Nogiwa-Valdez, and M. M. Stevens, "Considerations for implementing electronic laboratory notebooks in an academic research environment," *Nature Protocols*, vol. 17, no. 2, pp. 179–189, 2022.
[6] C. N. Klokmose and P.-O. Zander, "Rethinking laboratory notebooks," in *Proceedings of COOP 2010: Proceedings of the 9th International Conference on Designing Cooperative Systems, May, 18-21, 2010, Aix-en-Provence*, pp. 119–139, Springer, 2010.
[7] H. K. Machina and D. J. Wild, "Electronic laboratory notebooks progress and challenges in implementation," *Journal of laboratory automation*, vol. 18, no. 4, pp. 264–268, 2013.
[8] S. Guerrero, A. López-Cortés, J. M. García-Cárdenas, P. Saa, A. Indacochea, I. Armendáriz-Castillo, A. K. Zambrano, V. Yumiceba, A. Pérez-Villa, P. Guevara-Ramírez, *et al.*, "A quick guide for using microsoft onenote as an electronic laboratory notebook," *PLoS computational biology*, vol. 15, no. 5, p. e1006918, 2019.
[9] "Electronic Lab Notebook - Labfolder — labfolder.com." https://labfolder.com/. [Accessed 03-06-2024].
[10] "Labii ELN & LIMS — labii.com." https://www.labii.com/. [Accessed 03-06-2024].
[11] I. Urcelay-Olabarria, R. Lazkoz, J. Urrestilla, A. Leonardo, and J. M. Igartua, "Jupyter notebook as the physics experimental laboratory's logbook-first approach," in *International Conference on Computer Supported Education*, vol. 2, pp. 458–463, SCITEPRESS, 2017.
[12] "Electronic Lab Notebooks — datamanagement.hms.harvard.edu." https://datamanagement.hms.harvard.edu/collect-analyze/electronic-lab-notebooks. [Accessed 03-06-2024].
[13] "Lab dragon repository github page." https://github.com/ncsa/lab-dragon. [Accessed 03-06-2024].