

# The importance of research data management: The value of electronic laboratory notebooks in the management of data integrity and data availability

Matt Dunie

*President and Co-Founder, LabArchives, 1915 Aston Ave, Carlsbad, CA 92008, USA*

*E-mail: [mdunie@labarchives.com](mailto:mdunie@labarchives.com)*

**Abstract.** Laboratory data – the data produced in the practice of the scientific method – is not consistently managed among academic labs within or external to many academic research institutions. As data on which research is based upon becomes more openly available, Data Management Plans will more often be enforced. Thus, Data integrity, Data lifecycle, Data Security, Perpetual Revision History, Permanence, and unchangeable time stamps, will be concerns that will evolve into in the management of laboratory research data. Proof of research and discovery is a major concern among researchers and there are more instances of research fraud, unintended or intentional, than most people realize. The use of a Digital Lab Notebook can help prove discoveries, protect intellectual property, and provide the tools necessary to defend or audit research activities and preserve research integrity. As the collaborative nature of scientific research continues to become more easily executed with the continual advancement of research technology, it becomes essential for researchers, funding agencies, publishers, and institutions to protect and defend the work produced in research laboratories.

Keywords: Data management, digital laboratory notebooks, data integrity

## 1. Introduction

There has been a growth in research-related data for several reasons. But the advancement and use of sophisticated research technology combined with the advent of immediate and low-cost communications and collaboration technology are enabling the creation of massive amounts of research data. These data must be managed properly in order to facilitate quality research. Research funding agencies make statements regarding data management and publishers provide for the submission of supplemental data. Yet there are research retractions for faulty and falsified data every month. Good data management tools and process can help to limit “bad” research.

According to a report published by the International Association of Scientific, Technical and Medical Publishers, there are more than twenty-eight thousand one hundred English language peer-reviewed journals in publication with an output of an estimated 2.5 million articles [3]. In addition, the report mentions that surveys of researchers suggest as many as 1%–2% of scientists have fabricated or falsified research data. The report states there were more than four million unique authors in 2014. Simple arithmetic then suggests that anywhere from forty thousand to eighty thousand of the named authors of the 2014 sample may have used incomplete or inaccurate data at some point in their career. Some of

them may have used it in the research for the papers published that year. It is very difficult to determine the degree of research fraud, but we do know how many papers are retracted (although incomplete or falsified data is only one potential reason for retraction). In 2011, according to the report, there were four hundred retractions of published papers. Generally, retractions occur two years after publication. Yet, in the time from publication to retraction, the paper can be cited many times, and those papers can be cited, and so on. Without studying every retraction, every citation of those retractions, and the reasons and basis for the citation in the first case, there is no way to calculate the potential fault of research caused by improper data management of the first paper in the chain.

## **2. Data management and laboratory notebooks**

Data Management Plans have been a requirement of various funding agencies for most of the past decade. Governmental agencies have been issuing policies on data and management for much longer. The basics are the same across all of the plans: preserve data & provide access. There are other components to be considered, however the basic premise is to make the data available for future research and evaluation for the purposes of reproducibility, research integrity, further research, or challenge.

The reasons for keeping a lab notebook are well known. The U.S. Health & Human Services Office of Research Integrity web page sums up the reasons very nicely [1]:

- To establish good work practices.
- To teach people in your lab.
- To meet contractual requirements.
- To avoid fraud.
- To defend patents.
- To allow work to be reproduced by others.
- To facilitate preparation of formal reports, presentations and papers.
- To validate your research.
- To serve as a source for assigning credit to lab members.

The sixth bullets above may be the most important. Among the most basic underpinnings of science is the reproducibility of research. Without strong data management policies, documentation, and data management, reproducibility is at risk. Research labs of all disciplines have varying types of equipment, but there is at least one standard among them: Research is to be documented in accordance with the scientific method. Good data is data that is documented, stored, and accessible.

Our direct experience in academic research labs transitioning to LabArchives indicates more researchers are using paper notebooks to document their finding than anything else. But, they are also using some digital substitutes for paper and a plethora of home grown solutions. Some of what we found: blogging and wiki software, custom applications, flash drives, old PCs running unsupported software, various word processors and note taking applications, and... paper file folders. Almost all research note taking in paper notebooks is contemporaneous or transferred to an official “research notebook” from other media. Paper lab notebooks can cost as much as seventy-five U.S. dollars. The security measures currently taken are just as variable as the tools used in the lab. Other than in highly secure facilities, we found solutions ranging from keeping research notebooks in heavy-duty safes and lockable file cabinets to research data kept on media, which is then carried in a briefcase or backpack. Consider the potential implications that are caused when a junior researcher, unaware of the need to manage research data security, uses a smart phone to take a photo of research output or notes and then uses an unsecure chat

application to send the data to someone he is working with across the continent in a café! Or someone taking home a paper notebook, in an honest effort to continue their work while in transit, and mistakenly leaving it in the seatback pocket of a plane or train!

There is a great dichotomy: The research taking place in academia, which may include the use of amazingly complex and sophisticated equipment, is documented using pen and paper. And, the tools used to collaborate: email attachments, online document editors sophisticated image viewers, data analysis software, and various spreadsheet applications all output digital files or images – are then printed and glued into paper notebooks.

LabArchives was established to solve the workflow issues produced by using a combination of old and new technologies. It was a simple issue to describe: “Let’s provide a workflow tool to make researchers more efficient.” We learned over the years that not only would we make them more efficient, but we can also solve other research concerns as well. Research fraud and integrity issues become more apparent. Companies and organizations like ours endeavor to assist the research community in the management of rapidly growing research data; in the education of the next generation of researchers in research data management; and in the enablement of more researchers using technology to manage the output of an increasingly technical workbench.

The scientific method at the highest level, appears quite simple. It is a circular process by which a researcher goes through at least four phases: ideation, research, analysis, and conclusion. The output is typically either a peer reviewed, or “grey” paper or report. However, when evaluated in more detail, the process is much more complex than the four steps just mentioned.

In granularity, good research requires skill, tools, and rigor. One of the more granular components is data management. . . and possibly, the definition of data. To the lay person, the data is limited to the results of an experiment or survey. But in reality, and depending upon the discipline, the data may include documentation of process and procedure. A research notebook may include information on failed experiments and documentation of missteps or unresolved questions. Research notes, environmental observations, lab notes and evaluations are all part of the research data. All of this can be “templated” into an electronic Laboratory Notebook (ELN), ensuring that the complete research record is preserved and available for future use.

This is especially true when considering that a major element of validation in the research process is the reproducibility of previous work. Many retractions are the result of irreproducible research. According to some in the research community, the community itself has conflicting objectives. The pressure placed on researchers to increase their research output, to win more grant proposals, and to regularly publish their research can be at odds with solid scientific research management. Indeed, a recent article details two peer reviewed articles which “. . . urge scientists to make research more reproducible” [2].

It appears Research Data Management practices are highly-variable. If you want proof, walk the halls on different floors of a research building in academia and visit a dozen research labs. Depending on institutional policies and the level of independence provided principal investigators, you may see as many as a dozen different mechanisms to manage research data in a dozen different labs. But, research labs must be different. . . that is the nature of the business of research, especially in Academia. And those differences are the basic nature and strength of academia.

### **3. Electronic laboratory notebooks**

But there are tools in the marketplace that can help researchers maintain their independence, provide researchers and their administrators with the ability to protect their work product *and* enable scientific

reproducibility. They are Electronic Laboratory Notebooks (ELN – a terrible 1980’s-era name describing a product that essentially replaces the analog paper notebook with a much-improved digital version). And, while many of these products can be used in multiple disciplines, some have dedicated discipline-specific capabilities.

But there are commonalities among them – data integrity; revision history (perpetual); time stamps; compliance with regulations related to data accessibility and the security of data and research; access rights management; and institutional controls – are all key capabilities of an ELN. Some ELNs integrate with other products used in research labs and in data analysis. Some allow for publishing data and large group research collaboration.

When data are entered into an ELN – and data means research data, notes, observations, formulas, equations, sketches, data sets, images. . . any type of data – the product must have the ability provide the ability to support several interested parties: 1) a funding agency’s requirement for a Data Management Plan; 2) the researcher’s need to document their research; 3) the administration’s need to be able to protect and prove discovery; and 4) the publisher’s need to review and publish data supporting the research.

ELNs, however, unlike other work-flow software, sometimes do not show benefits until they have been in use for some time. But the benefits that they provide eliminate many of the research community’s pain points. ELNs protect data from “walking” – research work can be shared properly and securely, and as researchers change positions and institutions, the data can both stay and travel in accordance with the institutional policies. The best example of a benefit which grows over time is when somebody is trying to search through hundreds of pages of notes in a lab notebook. Or better yet, trying to search for a chemical structure or a term used to tag a drawing. With the paper medium, the only way is to thumb through all the pages until you find the page you are looking for, which, if you remember correctly, includes the notebook entry that you need. With the digital medium (an ELN), you can search your own notes in a secure system, the same way you would search research literature. . . but in a full functioning ELN, you will be able to search all the data – even inside data files. (LabArchives enables researchers to search inside attached data files as well as the text of notes, comments, pdf files, and chemical structures.)

The data upon which research is published needs to be kept in a secure environment, but that environment must provide for research collaboration. Some cloud-based ELNs provide such a capability, some even to the point of enabling thousands of researchers to collaborate on a single platform for the purposes of data gathering. But, security must be contemplated. The ELN should enable different levels of access, comply with industry standard security protocols and, when used in conjunctions with institutional policies, protect personally-identifiable information. Finally, an ELN must have the ability to scale to support thousands of users combined with the ability to be completely customizable for any user and multiple disciplines.

LabArchives is a cloud-based scientific research platform. It is the leading ELN (as measured by web traffic) and used by more than one hundred and ninety thousand researchers and more than 200,000 students in lab course instruction around the world at the time of publication of this article. In the past year, LabArchives users logged seventy million research activities in research labs (for more information on LabArchives go to: <http://www.labarchives.com/>).

ELNs can support the Scientific Method in ways traditional paper notebooks cannot. They also support institutional research policies and objectives and provide a platform for institutional data management and research support. A robust ELN supports Data Integrity, Data Lifecycle Management, Data Management, Data Accessibility, Collaboration, and Research Reproducibility. In today’s world of global

collaborative research, digital information, and robust, advanced information technology, ELNs are becoming the “must have” tool for researchers.

### About the author

Matt Dunie has founded three information services companies (Insight Publications, RefWorks, and LabArchives), and has held several executive-level positions. His professional experience includes the senior management roles CSA, ProQuest, Data-Planet, and currently LabArchives. Dunie has managed more than twenty acquisitions. He is the inventor of two patents, a Director on the Board of ThirdIron ([www.thirdiron.com](http://www.thirdiron.com)), and a founding partner of the angel group, Riverbend Capital.

### References

- [1] [https://ori.hhs.gov/education/products/wsu/data\\_lab.html](https://ori.hhs.gov/education/products/wsu/data_lab.html).
- [2] Two manifestos for better science, *Discover Magazine* **January 11** (2017), <http://blogs.discovermagazine.com/neuroskeptic/2017/01/11/manifestos-better-science/#.WT6mvmjyuUk>.
- [3] M. Ware and M. Mabe, *The STM Report: An Overview of Scientific and Scholarly Journal Publishing*, 4th edn, International Association of Scientific, Technical and Medical Publishers, Netherlands, 2015, [http://www.stm-assoc.org/2015\\_02\\_20\\_STM\\_Report\\_2015.pdf](http://www.stm-assoc.org/2015_02_20_STM_Report_2015.pdf).