

Description of Research Data in Laboratory Notebooks: Challenges and Opportunities

Nishida, Emi
Kyushu University

Ishita, Emi
Kyushu University

Watanabe, Yukiko
Kyushu University

Tomiura, Yoichi
Kyushu University

<https://hdl.handle.net/2324/4479591>

出版情報 : Proceedings of the Association for Information Science and Technology. 57 (1), pp.e388-, 2020-10-22. Association for Information Science and Technology

バージョン :

権利関係 :



Description of Research Data in Laboratory Notebooks: Challenges and Opportunities

Emi Nishida

Kyushu University, Japan, nishida.emi.497@s.kyushu-u.ac.jp

Emi Ishita

Kyushu University, Japan, ishita.emi.982@m.kyushu-u.ac.jp

Yukiko Watanabe

Kyushu University, Japan, watanabe.yukiko.935@m.kyushu-u.ac.jp

Yoichi Tomiura

Kyushu University, Japan, tomiura.yoichi.802@m.kyushu-u.ac.jp

Abstract

In scientific research, laboratory (lab) notebooks have traditionally been used to record experiments and their results. Lab notebooks also act as an important record of the generation, processing and analysis of data over the research data lifecycle. As research data becomes increasingly digitized and voluminous, it becomes harder to record and describe research data, especially in paper lab notebooks (PLNs). This paper addresses the challenges of recording contemporary research data in lab notebooks and discusses the requirements in lab notebooks for describing research data. Two basic requirements of lab notebooks are the completeness of the data description and the ability to link the experiment records with their corresponding research datasets. Descriptions of research data should also document the provenance of the research data so that the original data can be retrieved. Guidelines for consistent file naming and systematic directory structures for saved research data can support efficient research data retrieval.

Keywords

Paper lab notebooks (PLNs), research data

Introduction

Research data are the “recorded factual material commonly accepted in the scientific community as needed to document and support research findings” (NIH, 2003). In recent years, as part of movements to promote open science or open data, scientists have been encouraged to share and reuse research data. The research data lifecycle consists of six stages: planning, data collection, data processing and analysis, publishing and sharing, long-term management, and using data (Corti, Eynden, Bishop, & Woollard, 2014). To share and reuse research data effectively, complete records and thorough descriptions of research data should be captured when the data are produced. This level of detail also provides proof of the fairness and authenticity of that research.

In scientific fields, laboratory (lab) notebooks are commonly used to record experiments. Lab notebooks record the title, date, and the purpose of each experiment, the materials used and their quantities, the procedure and manipulation, and the results (Kanare, 1985). Because result data is one kind of research data, a lab notebook is also an essential tool for describing research data. Research data should be recorded as accurately and completely as possible (Day, n.d.).

Increasing digitization of the research environment allows the generation of massive amounts of research data. For example, the raw data produced from an experiment can be processed using different tools, then converted into different formats, such as images, recordings, waveforms, and numerical values. Lab notebooks are ill-suited to directly capture the full set of associated data in their various formats. Nevertheless, lab notebooks can still play a role in the documentation of research data. In particular, it is important that lab notebooks accurately record and fully describe the various sets of research data produced during an experiment. This paper addresses the challenges of recording and effectively describing research data in lab notebooks.

Lab notebooks are used in a variety of situations and have traditionally been paper notebooks. In recent years, electronic lab notebooks (ELNs) such as SciNote and RSPACE have been developed (SPLICE, 2019). A 2008 survey conducted by Atrium Research showed that about 33% of biopharmaceutical organizations reported the installation of at least one ELN system, and that the non-profit segment of the market was the slowest to adopt ELNs, with an estimated 5% penetration at the time (R & D Editors, 2008). One reason for this difference could be budget-related, because the research budgets of universities are typically smaller than those of companies. Another possible reason is the relatively long learning curve to use an ELN. A researcher could spend up to 3 months training to use an ELN while continuing to record experiments on paper (Kwok, 2018). Because undergraduate and master’s students are often the main writers of lab notebooks and generally leave the university within a

few years, it is inefficient for them spend their limited research time learning how to use an ELN. Although the transition from paper lab notebooks (PLNs) to ELNs is underway, these considerations suggest that it may take some time to complete the transition. The remainder of this paper discusses the basic requirements of lab notebooks in the role of recording and describing research data, assuming the continued use of PLNs in university laboratories.

Basic Requirements of Lab Notebooks as a Description of Research Data

A basic requirement of lab notebooks is to record, accurately and completely, the details of experiments and all research data. Experiments produce three main types of data: raw data, processed versions of the data, and result data. When a researcher processes raw data, the process should be fully described to enable reproducibility of the experiment. The more complicated the processing, the more difficult it is to fully document the procedure in a PLN. Despite the existence and availability of many writing manuals for PLNs, these manuals rarely address the level of detail required in describing how data are processed. Because the description of the processing of research data is left to the writer's discretion, it is difficult for inexperienced students to have a consistent description of what detail is needed. A potential solution involves creating a set of metadata items to describe research data in each laboratory. A template with these metadata items can be pasted and filled when research data are documented in a PLN.

Another recording challenge concern missing links between experiment records and their corresponding research datasets. All research data cannot be directly attached in the PLNs. Instead, usually the final version of processed data is attached in the PLN, and other kinds of data are saved on the researcher's own computer. Even when all the research data are saved, however, after a few years, researchers struggle to retrieve datasets associated with a particular experiment, especially when these are stored on other computers. To solve this problem, experiments and their associated research datasets could be assigned a unique ID by a university organization. These IDs could be recorded in the PLNs, to enable researchers to track all datasets associated with an experiment. In addition, the final published data could also be assigned a digital object identifier (DOI).

Other Requirements

All research data are the common property of the laboratory, so the provenance of research data should be clearly documented even after individual researchers leave. Occasionally, an experiment cannot be reproduced because of insufficient information in the PLNs about how to conduct the experiment, or because the raw data cannot be retrieved (Bungers, n.d.). Retrieval problems occur because of incorrect file names, insufficient directory structure descriptions, and/or inconsistent naming of the sets of raw data, processed data, and final results. Recommendations to avoid these problems include establishing consistent file-naming conventions and developing guidelines or rules for contents of readme files and directory structures.

Future Work

We plan to develop guidelines for describing research data in PLNs, especially for research data intended to be offered as open data.

REFERENCES

- Kanare, H. (1985). *Writing the Laboratory Notebook*. American Chemical Society.
- National Institutes of Health. (2003). NIH Data Sharing Policy and Implementation Guidance [web log]. Retrieved June 2, 2020, from https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm
- R&D Editors. (2008). Electronic Laboratory Notebooks Enter Mainstream Informatics [web log]. Retrieved June 2, 2020, from <https://www.rdworldonline.com/electronic-laboratory-notebooks-enter-mainstream-informatics/>
- Kwok, R. (2018). Lab Notebooks Go Digital. *Nature*, 560, 269-270.
- SPLICE. (2019). The 9 Best Electronic Lab Notebooks (ELN) Review for Your Research [web log]. Retrieved June 2, 2020, from <https://splice-bio.com/the-7-best-electronic-lab-notebooks-eln-for-your-research/>
- Beavon, R. (n.d.). Writing the Laboratory Notebook [PDF]. Retrieved June 2, 2020, from http://www.rod.beavon.org.uk/lab_book.htm
- Bungers, S. (n.d.). The Electronic Lab Notebook in 2020: A Comprehensive Guide [web log]. Retrieved June 2, 2020, from <https://www.labfolder.com/electronic-lab-notebook-eln-research-guide/>
- Corti, L., Van den Eynden, V., Bishop, L., & Woollard, M. (2014). *Managing and Sharing Research Data: A Guide to Good Practice*. SAGE.

