

The laboratory notebook in the 21st century

The electronic laboratory notebook would enhance good scientific practice and increase research productivity

Sara Y Nussbeck¹, Philipp Weil^{1,2}, Julia Menzel^{1,2}, Bartłomiej Marzec¹, Kai Lorberg¹ & Blanche Schwappach²

Detailed note-keeping is a prerequisite for, if not a key component of, scientific discovery. An unrecorded experiment is lost to the world even if it sparks a great idea in a scientist's mind: Additional work is needed to reproduce and confirm the original observation and to test the hypothesis by novel experimental strategies. Hence, a detailed record of the experimental setup, observations, and analysis is a crucial requirement for presenting a new discovery to the scientific community.

.....
“An unrecorded experiment is lost to the world even if it sparks a great idea in a scientist's mind...”
.....

For centuries, scientists have been using paper notebooks. However, the digital revolution has changed every aspect of data handling: Acquisition has become automated, primary data exist in a huge variety of formats and require vast memory space, and analysis increasingly uses sophisticated software. While electronic note-keeping has become state of the art in the pharmaceutical industry, it is by far not the standard in academic life-science laboratories. In fact, digital record-keeping has been controversial in academia, and hence, its implementation has been lagging.

We argue that many academic laboratories will soon abolish the current precarious mixture of digital data and paper-based annotation. We see three main reasons for this trend. First, the vast majority of the data generated by a scientist will be in digital form that has to be documented and archived; however, the paper-based laboratory notebook does not enable convenient documentation of folder structures and data paths. Second, ongoing efforts aim to standardize experimental protocols and data formats in order to improve their general comparability and re-analysis of data, which make standardization more attractive for scientists. In fact, standardization is a prerequisite for using an electronic laboratory notebook (ELN). Third, long-term data storage, an essential component of good scientific practice, easily turns into an absurd exercise when detailed and standardized annotations (metadata) are missing. Many institutions worldwide are therefore considering electronic note-keeping to support good scientific practice (<http://lib.guides.umd.edu/ELN>, <http://academictech.doit.wisc.edu/ideas/electronic-lab-notebooks>).

In 1985, Howard Kanare wrote a classic book *Writing the Laboratory Notebook* that has taught generations of scientists how to record their work [1]. The book features a whole chapter about the electronic notebook that carefully lists advantages and disadvantages. Many of the disadvantages are now anachronistic, but the main arguments

against the ELN remain technical issues and involve time requirements, costs, and security. Any laboratory interested in going paperless will have to make substantial investments into developing templates for notebook pages, protocols, and reagent stocks. The financial consequences of introducing an ELN are not limited to obtaining software licenses but include dedicated servers, backup systems, and IT staff. ELNs may introduce new error sources and security risks that did not exist for paper notebooks.

On the other hand, there are many substantial advantages of going paperless: In the long term, the use of templates for similar processes should considerably save time. Modern search functions can greatly aid the preparation of reports, manuscripts, and presentations. From the perspective of the PI, an ELN offers huge potential in standardized data collection and facilitated data management. Classically, a student who graduates after generating research data for years leaves several paper-based notebooks and a digital folder structure that contains raw and analyzed data on the group server. Without laboratory-wide nomenclature rules on how to name files and how to record data paths, it becomes extremely time-consuming to realign experiments and stored data. In fact, this material will often only be used in case of an emergency—failure to reproduce previous results or suspicions of scientific misconduct. Since entries into an ELN have to be standardized for each group, they

1 Department of Medical Informatics, University Medical Center, Göttingen, Germany

2 Department of Molecular Biology, University Medical Center, Göttingen, Germany. E-mail: sara.nussbeck@med.uni-goettingen.de, blanche.schwappach@med.uni-goettingen.de

DOI 10.15252/embr.201338358 | Published online 15 May 2014

would automatically provide a searchable description of all data stored on the group server. Ideally, the ELN would enable efficient use of data by successive generations of students and postdocs, would maintain technical knowledge in the laboratory, and allow detailed reconstruction of individual experiments.

There are a variety of software products available. In order to select the optimal solution, institutions or research groups interested in going paperless would need to perform a careful analysis of their workflow to define their specific requirements. Decisions revolve around from where the software is run (local or company server), where the data are stored (locally or in a commercial cloud), who can access the data and documentation of all changes made. Critical issues are the design and administration of templates for standard operating procedures, data exchange with specified equipment and servers, search functions, and an inventory management system, for instance, for reagents or

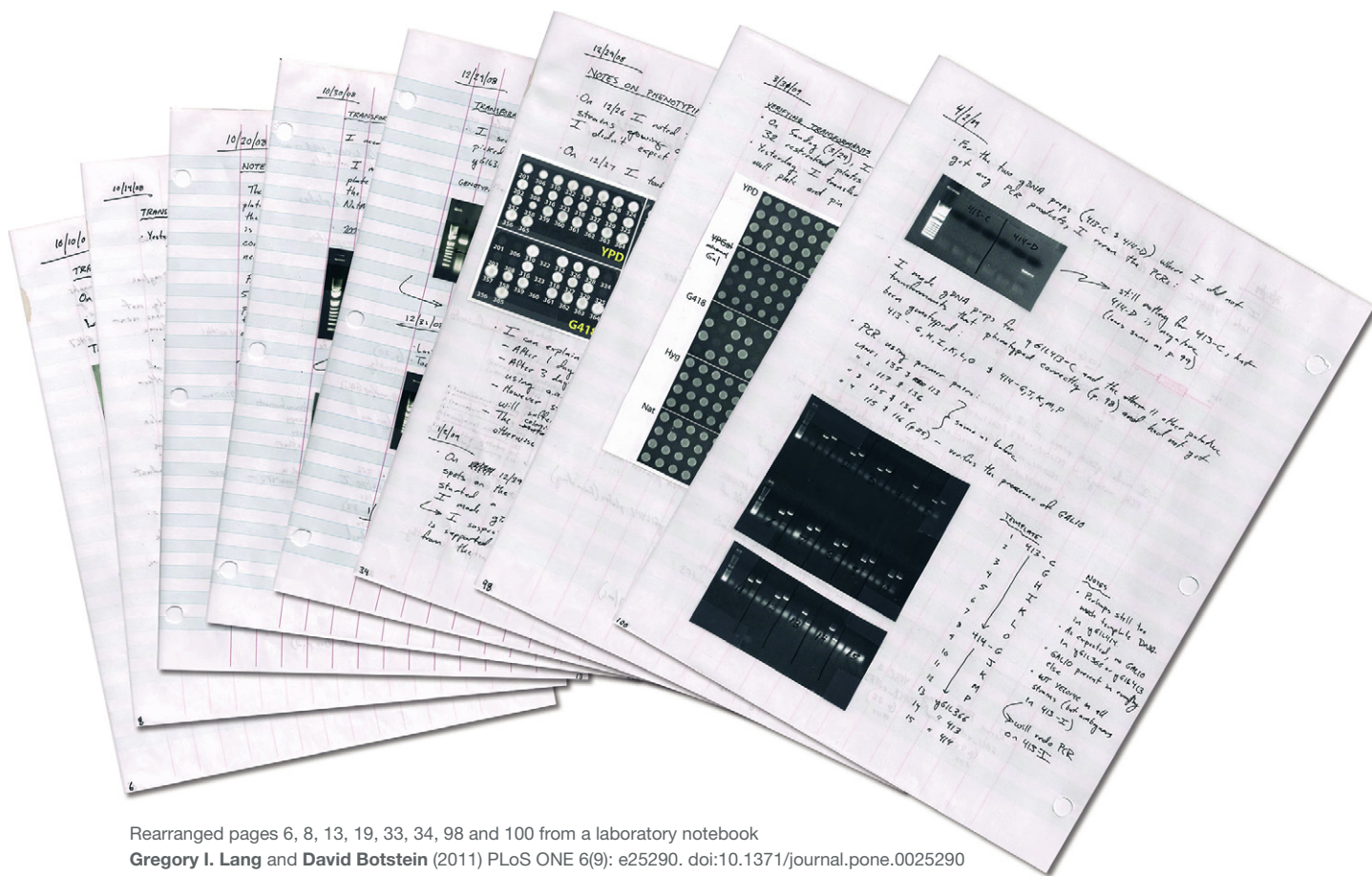
constructs stored in freezers. These aspects are critical because laboratory members will not use the ELN if their workflow becomes more tedious or time-consuming.

Templates within the ELN are electronic forms that guide the scientist through the various steps of an experimental protocol and make it convenient to enter all the specific information while providing standard values by default. A major advantage of the ELN is the inclusion of all digital data via links—hence the software must be capable of linking to any storage location that a research group uses. Another major asset of the ELN is full digital and fast access to stored information. The respective search algorithms must be powerful and reliable to fulfill this expectation. A product-requirements document is essential to assess these and other desired features. In most cases, the final decision regarding ELN software will ultimately rely on a test by some members of the laboratory.

The laboratory notebook is the scientist's permanent companion. It lies open on the bench next to reagents and experiments and

is being carried around from room to room wherever its owner conducts experiments. This mobility allows for continuous recording. To achieve the same degree of mobility for the ELN, a tablet PC is possibly the best choice. In principle, any device capable of displaying a website could run most ELN software. However, empirical testing of the specific combination of software and tablet in a standardized case scenario is necessary as certain combinations of tablet PCs and browsers may greatly reduce functionality. In addition to choosing the optimal software and hardware, the ELN has to be integrated into the local network infrastructure with associated storage space that can be securely assessed to enter data using different hardware devices.

Switching to an ELN requires motivation and substantial time and financial investments. But once adopted, the ELN can benefit the scientific community as a whole by catalyzing the development and use of metadata schemes. In fact, digitalization at the source, that is, right at the experiment,



Rearranged pages 6, 8, 13, 19, 33, 34, 98 and 100 from a laboratory notebook

Gregory I. Lang and David Botstein (2011) PLoS ONE 6(9): e25290. doi:10.1371/journal.pone.0025290

could greatly aid reproducibility, data quality, long-term storage, and the lifetime and impact of research data.

Even more importantly, metadata—data about data—are becoming increasingly essential. In the experimental sciences, metadata would describe the parameters of an experiment, such as the temperature at which a reaction took place, the name of the scientist performing the experiment, or the details of the equipment and reagents just to name a few examples. Metadata aim at fully describing the context of how the data were acquired to make them reproducible, interpretable, and amenable to re-analysis. Metadata are also key to assembling useful databases and to comparing large-scale experiments.

“... many academic laboratories will soon abolish the current precarious mixture of digital data and paper-based annotation”

Traditionally, metadata have been recorded in paper-based laboratory notebooks implicitly rather than explicitly. Use of an ELN is therefore a great opportunity to make metadata explicit and to record them in the only way they can be useful: as detailed and as standardized as possible. Practically, this means that the templates and forms which research groups design as they customize the ELN should enable, encourage, and—where appropriate—enforce metadata recording. Once useful metadata schemes for standard procedures exist, we anticipate that some of them will spread in the scientific community because using them will benefit individual scientists (for instance when publishing their data), research laboratories (when training new laboratory members), collaborating research groups (when sharing or comparing data), and whole research fields (for instance, when referring to the literature or databases).

The simple example of Western blotting, a standard procedure in molecular cell biology and many related disciplines, illustrates how metadata could help to solve a number of problems in current scientific practice. Proteins are separated by gel electrophoresis, transferred to a carrier membrane, and

visualized by an antibody-based detection technique. A Western blot yields information on the size and abundance of a protein in complex protein mixtures, including variations owing to differential gene expression or posttranslational modifications. While good standard protocols for Western blotting exist in books and on the Internet, it is surprisingly difficult to find even two laboratories that perform this method in exactly the same way. Moreover, as it is a standard technique, publications rarely describe the experimental details of how Western blotting was performed. Thus, many relevant details of the experiment, such as the gel system or protein transfer system, will never be accessible to other scientists. The poor documentation of the immuno-detection part of the procedure is an Achilles' heel of modern biomedical science. The indiscriminate and unprofessional use of antibodies has confused whole fields by myriad misinterpreted or irreproducible experiments involving antibodies [2] and wasted time and money. Poor documentation accounts for a large part of this confusion.

A standard metadata scheme for Western blotting represents an ideal way for recording all this information in a concise yet versatile manner. The metadata can immediately be used as Supplementary Material for publishing and enable future inclusion into databases as appropriate (<http://www.antibodypedia.com>). If provided as a routine template in the framework of an ELN, such a metadata scheme would be integral to the experimental workflow. Hence, it would not have to be painstakingly created at the time of publication to avoid the deposition of another poorly documented blot into the scientific literature.

Good scientific practice mandates that data should be kept for at least 10 years [3,4]. Many laboratories retain all the documentation to reconstruct experiments and to revisit the associated primary data much longer—but tracing and using it can be anything from time-consuming to impossible. In the worst case, the group server turns into a ‘black-hole’ of key information without truly making it accessible. The ELN can be used to address this problem by virtue of simultaneously arranging and structuring the recorded information into person- and project-based, searchable records without requiring additional work while entering the data. Each scientist in the

research group can use the ELN for daily record-keeping as if using a paper-based notebook. In addition, project identifiers connect the described experiments to a particular project, which allows to store, organize, and retrieve all records in a chronological or project-based fashion. All primary data will be directly linked to these records—either as files uploaded into the ELN database or as links connecting to network-attached storage space, that is, the group server that will now no longer be a ‘black-hole’. Data structure, such as names of directories or levels of organization, and access rights for individual files have to be carefully adapted to the ELN as the main record-keeping tool. At least the outline of a pertinent data structure and access rights needs to be designed by the respective PI and the more long-term laboratory members such as laboratory managers or senior technicians. An intuitive and time-saving concept to data management and backup is the only way to ensure compliance by the whole team.

“... the evolution of scientific methodology [...] makes it increasingly difficult to keep all relevant information in the respective notebook”

Many scientists collaborate with colleagues from other laboratories, often funded by joint research grants. Thus, research data from these projects are usually relevant to others beyond the own laboratory; on the other hand, it also raises the question of protecting unpublished results. Still, sharing of data seems to be a problem in some research areas [5]. As soon as several scientists from different laboratories begin to store their data electronically on one server, sooner or later the question of ‘whose is it?’ will come up. Normally, issues of intellectual property will be covered by local legislation, and the data policy can refer to this framework but it would also need to include a separate data policy signed by all PIs in a joint project. By default, scientists should only have access to their own data in an ELN. For sharing data with other colleagues, they can have assigned additional rights, such as only reading or reading and writing.

PIs of individual laboratories will have to decide on who can share the data generated by their team members.

Although there seem to be many advantages inherent to an ELN, every new technology entails new challenges. These can arise from the infrastructure, storage media, software, or human error. The arguably biggest challenge is the change of data formats over time; a problem that of course does not affect paper since it has been around for 4000 years. All data have to be migrated to new media or new formats once certain standard formats for texts, images, or other forms of data become outdated. The software provider's concept or solution for data migration during software updates should be an important aspect when choosing ELN software.

Many academic laboratories still prefer the paper-based laboratory notebooks [6]. However, the evolution of scientific methodology, which almost exclusively relies on computer-controlled equipment that generates digital data, makes it increasingly difficult to keep all relevant information in the respective notebook. Furthermore,

modern concepts of sharing and comparing publicly funded research necessitate digitalization. Digital recording at the source is the most effective and scientifically most appropriate way of further developing and implementing these concepts. While the transition to completely digital note-keeping represents a formidable challenge that requires substantial investments into infrastructure, software, and staff, it also represents an opportunity to revive and where necessary reinforce a digital version of the standards and skills that Howard Kanare so successfully disseminated.

Acknowledgments

This work was funded by the German Research Foundation (DFG) within the grant for the Collaborative Research Centre (SFB) 1002 on Modulatory Units in Heart Failure, subproject INF. We thank Johann Bernhardt, Anne Clancy, Detlef Doenecke, Stephan Lehnart and David Schwappach for helpful discussions.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Kanare HM (1985) *Writing the Laboratory Notebook*. Washington, DC: American Chemical Society
2. Rhodes KJ, Trimmer JS (2006) Antibodies as valuable neuroscience research tools versus reagents of mass distraction. *J Neurosci* 26: 8017–8020
3. European Science Foundation (2000) Good scientific practice in research and scholarship. *ESF Sci Policy Brief* 10: 1–16
4. Bosch X (2010) Safeguarding good scientific practice in Europe. *EMBO Rep* 11: 252–257
5. Nelson B (2009) Data sharing: empty archives. *Nat News* 461: 160–163
6. Wright JM (2009) Make it better but don't change anything. *Autom Exp* 1: 5



License: This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.