Metadata capture in an electronic notebook: How to make it as simple as possible?

Metadatenerfassung in einem elektronischen Laborbuch: Wie macht man es so einfach wie möglich?

Julia Menzel[1,2]
Philipp Weil[1,2]
Sara Y. Nussbeck[1]

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

2 Department of Molecular Biology, University Medical Center Göttingen, Germany

Abstract

In the last few years electronic laboratory notebooks (ELNs) have become popular. ELNs offer the great possibility to capture metadata automatically. Due to the high documentation effort metadata documentation is neglected in science. To close the gap between good data documentation and high documentation effort for the scientists a first user-friendly solution to capture metadata in an easy way was developed. At first, different protocols for the Western Blot were collected within the Collaborative Research Center 1002 and analyzed. Together with existing metadata standards identified in a literature search a first version of the metadata scheme was developed. Secondly, the metadata scheme was customized for future users including the implementation of default values for automated metadata documentation.
Twelve protocols for the Western Blot were used to construct one standard protocol with ten different experimental steps. Three already existing metadata standards were used as models to construct the first version of the metadata scheme consisting of 133 data fields in ten experimental steps. Through a revision with future users the final metadata scheme was shortened to 90 items in three experimental steps. Using individualized default values 51.1% of the metadata can be captured with present values in the ELN.
This lowers the data documentation effort. At the same time, researcher could benefit by providing standardized metadata for data sharing and re-use.

Keywords: electronic laboratory notebook, metadata scheme, metadata capture, data documentation, default values

Zusammenfassung

In den letzten Jahren sind elektronische Laborbücher (ELNs) immer populärer geworden. ELNs bieten die Möglichkeit Metadaten automatisch zu erfassen. Durch den hohen Dokumentationsaufwand wird die Erfassung von Metadaten in den Naturwissenschaften oft vernachlässigt. Um die Lücke zwischen guter Datendokumentation und einem hohen Dokumentationsaufwand zu schließen, wurde ein nutzerfreundliches System in einem ELN entwickelt, um Metadaten vereinfacht zu erfassen. Zuerst wurden Western Blot Protokolle im Sonderforschungsbereich 1002 gesammelt und analysiert. Zusammen mit in einer Literaturrecherche identifizierten existierenden Metadatenstandards wurde eine erste Version des Metadatenschemas entwickelt. Im zweiten Schritt wurde das Metadatenschema durch die Umsetzung von individualisierten Default-Werten an die zukünftigen Nutzer angepasst.
Zwölf Protokolle des Western Blot wurden auf ein Standardprotokoll aus zehn unterschiedlichen experimentellen Schritten reduziert. Drei bereits existierende Metadatenstandards wurden als Vorlage verwendet, um die erste Version des Metadatenschemas zu erstellen. Dieses bestand aus 133 Datenfeldern in zehn experimentellen Schritten. Während

der gemeinsamen Überarbeitung mit den zukünftigen Nutzern wurde die finale Version des Metadatenschemas auf 90 Datenfelder in drei experimentellen Schritten gekürzt. Durch die Integration von Default-Werten können 51,1% der Metadaten vereinfacht erfasst werden. Dadurch wird der Dokumentationsaufwand verringert und gleichzeitig können Wissenschaftler durch die Bereitstellung von standardisierten Metadaten bei Datenaustausch oder der Datennachnutzung profitieren.

Schlüsselwörter: elektronische Laborbücher, Metadatenschema, Metadatenerfassung, Datendokumentation, Default-Werte

1 Introduction

For centuries laboratory notebooks have been hard-cover bound books with numbered pages into which scientists note their daily lab work [1]. A laboratory notebook is quite more than just a notebook for experiments; it is also a notebook for great, spontaneous ideas, for reminders or to organize samples and chemicals. Due to this documentation it becomes an important legal document for the reproducibility of experiments, patents and publications [2]. One of the main purposes of a laboratory notebook is the reproducibility and reusability of data. Due to the help of a laboratory notebook entry every experiment becomes reproducible and therefore usable and verifiable for every scientist [3]. This supports the general idea of science, that science should be available and useable for everyone. The detailed note-keeping in a laboratory notebook promotes the standards of good scientific practice [4].

Every Principal Investigator (PI) owns several hardcover paper-based laboratory notebooks of his or her students. There are some draw-backs to paper-based laboratory notebooks: It is difficult to reconstruct data with handwritten notes of others. Paper-based notebooks are not consistent due to the personal and specific style every researcher has. Moreover, they don't offer any possibility to search automatically for methods or results. Most importantly, one of the main problems in labs nowadays is that many data are generated in a digital format by lab equipment. Thus, many scientists have a mixed documentation of paper-based and electronic data [5].This is certainly one of the reasons for the increasing number of electronic laboratory notebooks (ELNs). Their annual growth per year is still above 20%, making ELNs one of the fastest growing informatics technologies [6].

These days, many different types of ELNs are available on the market [6], [7]. An ELN is a software program designed to replace paper-based laboratory notebooks [8]. The software comes along with many advantages and new possibilities, which make the daily lab work much easier. ELNs help to keep all data produced during the daily lab work, like notes, images, protocols, tables in one place. Moreover, they store all data in a digital and searchable way. ELNs are well established as part of quality management systems in chemical and pharmaceutical industrial laboratories and were seen as responsible for a 20% growth in efficiency [5]. More than 50% of researchers in the pharmaceutical industry use an ELN, while only 4% of researchers in non-profit institutions use ELNs [9]. The specific challenge to introduce ELNs in academia are threefold: highly flexible working conditions, frequent employee changes and interdisciplinary collaborations [8].The highly flexible working structures in academia lead to dynamic protocols and working processes and therefore to a flood of data and metadata. Metadata are data about data [10], i.e. in science metadata provide information about experiments, e.g. experimental parameters, settings of used lab equipment or chemicals. Metadata make experimental data understandable and even more importantly reproducible [11]. Moreover, metadata help to structure, organize, and refund data making them indispensable for well-organized and structured data recording in science [12]. However, metadata documentation is rarely dealt with in most scientific labs. Many scientists associate good metadata capture with a high workload caused by increased documentation effort and are not aware of the consequences of bad metadata capture [13]. Scientific work or experimental results which are not well documented are worthless in the scientific community, even if they spark a great idea [7]. Metadata capture of experimental data needs to be done directly at the source where they are generated. This documentation is necessary to make a legal document out of a simple protocol.

Here, we analyzed how to achieve good data documentation without increasing the burden of documentation effort for the scientists. Is there an easy way to capture metadata with the help of default values?

2 Methods

To be able to answer the question of simplified metadata capture in ELNs in general, firstly the ELN was established in a Collaborative Research Center and a proper experiment for the start was chosen. Secondly, lab protocols for this experiment were collected from different labs and a first metadata scheme was developed. Afterwards, the metadata scheme was revised in close collaboration with further users.

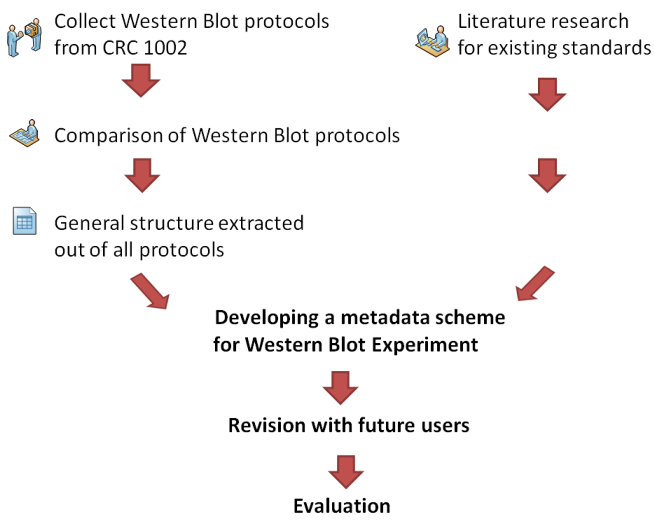

Figure 1: Generic workflow used to create a metadata scheme (here: for Western Blot)

2.1 Setting of the ELN in the framework of a Collaborative Research Center

The Collaborative Research Center 1002 (CRC 1002) at the University Medical Center Göttingen in Germany focuses on basic research in cardiovascular science and consists of research groups from very diverse research areas like pharmacology, immunology, molecular biology, biophysics, and cardiology. After performing a requirement analysis for a suitable software [3], the choice fell on eCat, an ELN developed by Axiope Limited. It is the first online ELN and the first ELN which was developed for academic research [8]. Before the introduction of eCat in four participating working groups of the CRC 1002 an acceptance survey regarding the expectations related to an ELN was performed. For this purpose eleven participants answered 18 questions in personal interviews.

2.2 Development of a metadata scheme

In order to analyze whether simplified metadata capture is possible in an ELN, one general experiment had to be chosen for a start as an example, which is being performed at most working groups of the CRC 1002. For that purpose different options were discussed at one of the regular project meetings and a decision for the Western Blot experiment [14] was made. A Western Blot is an experiment in which a protein mixture is separated based on its molecular weight through gel electrophoresis and afterwards transferred to a membrane. On this membrane single proteins can be identified using specific antibodies [14]. Figure 1 illustrates the generic workflow, which was used to create a metadata scheme (here: for Western Blotting). After making the decision, protocols for this experiment were collected from all working groups within the CRC 1002. The protocols were compared with regard to their content. The most detailed protocol was subdivided into different experimental steps and all other protocols were assigned to these experimental steps. Thereby, one standard protocol was developed out of all different protocols. Subsequently, a literature search was performed using the standard search function of the websites “Google” (http://www.google.de) and “Pubmed” (http://www.ncbi.nlm.nih.gov/pubmed) with the search terms listed below. The search was performed between August 15th, 2013 and August 29th, 2013.The following search terms were used in the literature search for a metadata scheme for a Western Blot experiment:

- Metadata scheme for a western blot experiment
- Metadata scheme for western blot
- Metadata scheme for a western blot experiment
- Metadata scheme for a western blot
- Metadata western blot experiment
- Metadata western blot
- Minimal information about western blot experiment
- Minimal information about western blot
- Metadata scheme for science
- Metadata scheme in science

In several iterations the development of a metadata scheme was discussed with the PhD students of the CRC 1002, who will be the future users of the ELN [15]. On basis of the results of the literature search and the standard protocol for a Western Blot experiment the first metadata scheme for a Western Blot was developed as a simple Excel list. For this purpose, all experiment steps out of the standard protocol were taken, structured and transferred into the Excel list. In parallel, metadata which should be captured during the experiment were added in a separate column. To define the to-be-included metadata, participating PhD students were asked for relevant metadata from their side. Afterwards, the first metadata scheme was presented to future users and modifications were discussed and integrated, including individual adjustments for future users.
Next the metadata scheme had to be implemented in the ELN. The ELN has a modular system for the design of templates. The following fields can be added to a template: text, text box, choice, checkbox, date, number, ra-

dio, reference, resource and time. For each line in the excel sheet a suitable field for the implementation in the ELN was selected, the field was named and a short description about the entry as a help function was added. Afterwards, the templates were reviewed by the future users and the last step was an evaluation.

2.3 Development of default values

For capturing metadata (in the Western Blot) as simple as possible default values were used as a first step. These default values were generated in close collaboration with future users. Firstly, future users were interviewed about their most used values for every data field in the Western Blot protocol. Secondly, a range of individual values, e.g. for drop down lists or radio buttons were offered for each data item in the metadata scheme and needed confirmation by the future users. Due to specific work processes and special user requirements it was necessary to customize default values for every working group using controlled vocabulary. Thus, it was decided to offer a range of all possible values for each working group. This meant that different default values had to be implemented for every working group.

2.4 Literature review for controlled vocabulary

For achieving comparable data sets for easy data search, sharing and exchange the same vocabulary and terms to describe the work processes, equipment, experiments, etc. should be used [16]. Another literature search was performed using the search engines "Google" and "Pubmed" as described above using the search terms listed below. The search was performed between August 23rd, 2013 and September 23rd, 2014 and refined between December 1st, 2014 and December 5th, 2014. The following search terms were used in the literature search for a metadata scheme for a Western Blot experiment:

- Controlled vocabulary for a Western Blot experiment
- Controlled vocabulary for a Western Blot
- Controlled vocabulary in molecular biology
- Controlled vocabulary
- Controlled vocabulary in laboratory work
- Controlled vocabulary for science
- Ontology for a Western Blot experiment
- Ontology for a Western Blot
- Terminology vocabulary for a Western Blot experiment
- Terminology vocabulary for a Western Blot
- Ontology in laboratory work
- Terminology in laboratory work

The results of the literature search will be used in the future to adapt already existing metadata schemes from one working group for another ensuring the use of controlled vocabulary for data sharing and exchangeability.

2.5 Creation of more metadata schemes for standard experiments

An ELN can only be a suitable tool for daily lab work, if metadata schemes exist within the ELN for all experiments routinely performed in this lab. Thus, after the implementation of a metadata scheme for a Western Blot experiment more metadata schemes for other protocols and experiments needed to be designed in the same fashion. For this purpose, protocols from the four working groups of the CRC 1002, who volunteered to work with the ELN first, were collected and further metadata schemes were generated using the above described and established workflow.

2.6 Evaluation of the metadata scheme

After introducing the ELN in four participating groups an evaluation was performed with three researches from one group and one researcher from another participating group addressing the handling with the ELN. These four scientists worked with eCat for one week and during this period documented all experiments and results with the ELN. Templates, which were required for the documentation in the ELN were designed ahead in close collaboration with the scientists. Subsequently, the participants were asked to fill in an evaluation sheet with 35 questions followed by a short personal dialogue.

3 Results

3.1 Introduction of the ELN in the framework of the CRC 1002

Before the introduction of eCat in the four participating work groups an acceptance survey was performed. The three most frequently mentioned requirements related to an ELN were "function of automatic data entry", "simplified data entry" and "faster data entry".

3.2 Development of metadata scheme for a Western Blot experiment

A discussion with the scientists of the CRC 1002 at a project meeting identified the Western Blot [17], [18] as a suitable candidate experiment for starting with automatic metadata capture. The Western Blot is a widespread yet complex experiment in molecular biology, which takes at least two days and generates metadata throughout all its different steps. Requesting Western Blot protocols from all working groups resulted in twelve protocols from 16 groups. Only four research groups of the CRC 1002 did not perform any Western Blot experiments at all. The comparison of the twelve protocols showed significant differences with regard to length, degree of details and included steps. The most detailed protocol was chosen

Table 1: Ten experimental steps of a Western Blot protocol used to create a metadata scheme for a Western Blot experiment

Step	Description
1	Dispense SDS-gels (recipes)
2	Preparation of SDS PAGE for electrophoresis (amount of loaded samples, sample order)
3	Electrophoresis (used running buffer, run time, run conditions)
4	Coomassie staining of the SDS gel
5	Destaining of the SDS gel
6	Blotting
7	Staining of the membrane
8	Blocking
9	Antibody staining
10	Film development

Table 2: First version of a metadata scheme for a Western Blot experiment (excerpt). This metadata scheme includes the columns: "Work steps" in which every working step of the experiment which generates data is listed; the second column "Value" describes the work step and clearly describes the value(s) which should be documented. The third column "Selection values" offers a range of values which could be selected by the user, and the last column "Acronym" includes an acronym where suitable.

Work Steps	Value	Selection values	Acronym
Experiments			
Gel electrophoresis SDS-PAGE			
GEL_ELECTROPHORESIS_PROTOCOL_RUNNING_BUFFER_COMPOSITION:	The composition of the running buffer used for gel electrophoresis	Enter the prescription of the used buffer, for example: 5x Running buffer: 15 g Trizma Base 72 g Glycine 5 g SDS add 1 L H_2O 1x Running buffer: Dilute 5x Running buffer 1:5	
GEL_ELECTROPHORESIS_PROTOCOL_RUNNING_BUFFER_PRODUCER:	The name of the company, laboratory or person that provided the running buffer used for gel electrophoresis	Enter the name of the provider, who produced the running buffer, for example: Technical Assistant Julia Menzel	TA JM

and subdivided into ten experimental steps (see Table 1); all other protocols were mapped to these steps.

Thus, the first standardized protocol was constructed following the workflow depicted in Figure 1 and consisted of 133 data fields in the ten experimental steps listed in Table 1. The literature search for existing metadata standards for a Western Blot experiment resulted in five papers/standards. Out of these, the "IHEC metadata standard" [19], the "Sample Pre-analytical Code (SPREC)" [20], and the "Dublin Core Standard" [21] were considered useful for the construction of a Western Blot metadata scheme. These standards together with the aforementioned standardized protocol led to a Western Blot metadata scheme as partially shown in Table 2. From the originally five identified papers/standards only three standards were used, the other two standards/papers were not considered to be helpful because of their complexity. The standards were used as model and especially the design of the IHEC metadata standard was used as a template for the implementation of the first version of the Western Blot metadata scheme.

The future users assessed the first metadata scheme as not usable in daily practice because of its length, complexity, and the high number of 133 data fields to document. The revised second version of the metadata scheme consisted of 90 data fields in only three experimental steps and included a detailed experimental protocol as an attachment (see Figure 2). Finally, the metadata scheme was approved in a discussion by the four users out of one participating work group.

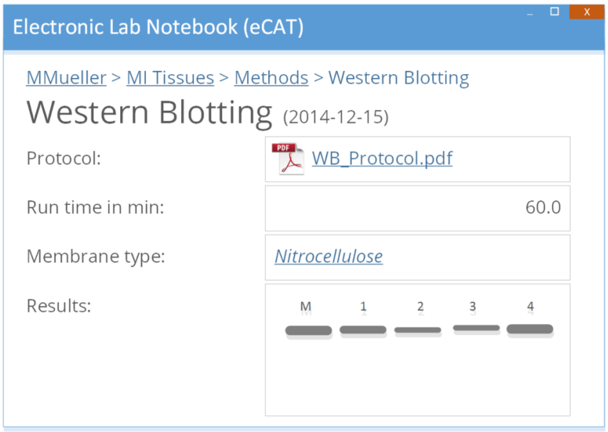

Figure 2: Structure of the metadata scheme for a Western Blot experiment implemented in the ELN. This figure shows an excerpt from the metadata scheme for a Western Blot experiment which was implemented in the ELN. Firstly the detailed protocol of the experiment is attached as a pdf file; secondly all variable data which is produced during the experiment could be documented. Results in form of images could be easily attached to the protocol.

3.3 Simplify metadata capture with default values

The metadata scheme for a Western Blot covers the experiment in a detailed way including the sample itself, all working steps, all chemicals and substances used and all data produced during this experiment. To simplify the metadata capture default values were implemented according to the needs and habits of the working groups. All in all, the metadata scheme for a Western Blot experiment in its recent version consists of 90 data fields out of which 46 can be prefilled with default values. Out of these 46 default values, 66.1% were prefilled free-text data fields; in the other cases the user can choose from up to five different values. To implement the options in the ELN for documentation, the field types "date", "text", "choice", "string", "checkbox", "radio button" and "number" were used. For data management reasons, the implementation of the Western Blot experiment had a high number of "radio buttons", "checkboxes" and "choice fields" rather than free text fields. Altogether, 51.1% of the metadata fields of the Western Blot experiment could be prefilled by using default values. The remaining 44 data items include only four free text fields, the other data items can be filled-in with just one click.

3.4 Literature search for controlled vocabulary for the Western Blot metadata scheme

The analysis of the twelve different Western Blot protocols from within the CRC 1002 already indicated that every single working group uses its own laboratory slang, e.g. the Western Blot sometimes was also referred to as Immunoblot. To avoid misunderstanding in data sharing and data chaos controlled vocabulary should be used in the ELN. The literature analysis for controlled vocabulary revealed a lack of controlled vocabulary for biomedical experiments. No specific controlled vocabulary, terminology or ontology for a Western Blot experiment could be found using the terms described in the methods section. Only a detailed instruction on how to develop a controlled vocabulary called "Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies" [22] could be identified. For the already developed Western Blot metadata scheme a terminology was defined by the authors of this manuscript and afterwards all users agreed on this terminology. With the help of this guideline a controlled vocabulary for the metadata schemes within the CRC 1002 including the Western Blot metadata scheme should be developed in the future.

3.5 Creation of more metadata schemes for standard experiments

After the development of the metadata scheme for a Western Blot experiment 13 more schemes were developed. An overview of implemented schemes can be found below. For example, for the polymerase chain reaction (PCR) [23] a metadata scheme with 27 data fields was developed. Out of these 27 data fields 18 could be prefilled with default values; thus 66.6% of the metadata can be captured using default values. So far, the following 13 metadata schemes were developed and implemented in the ELN for only one working group:

- Cell line sample
- Tissue sample
- Yeast sample
- Mini Prep
- NaOH lysis of yeast cells for protein purification
- Qualitative PCR
- Recombinant protein expression in *E. coli*
- Transformation of target DNA into *E. coli*
- Western Blot – Immunodetection
- Western Blot – Protein gel electrophoresis
- Western Blot – Blotting
- Yeast immunoprecipitation
- Yeast transformation

3.6 Evaluation of the metadata scheme

Until now the metadata scheme and the use of the ELN was evaluated by four scientists with an average age of 33.25 years. None of the researchers had previous experiences with ELNs. Two researchers found the ELN easy to use, while the other two researchers found it difficult. The ELN met the expectations of two researchers, while one researcher was unsure and one researcher was disappointed of the ELN. The data management using an ELN was significantly improved for one scientist, was improved a little for two scientists and one scientist was unsure about this point. Two scientists would like to continue with the ELN used so far, while the other two scientists would like to try a different ELN (see Table 3).

Table 3: Overview about the evaluation of the metadata scheme and the use of the ELN. N=4

Topic out of the evaluation	Answer possibilities	Choice
Previous experiences with an ELN	Yes	0%
	No	100%
Easy or difficult to use	Easy	0%
	Somewhat easy	50%
	Somewhat difficult	50%
	Difficult	0%
Expectations fulfilled	Yes	50%
	No	25%
	Unsure	25%
Improved data management	Yes	25%
	Little improvement	50%
	Unsure	25%
	No	0%
Continue working with ELNs	Continue with eCat	25%
	Use eCat when missing requirements are fulfilled	25%
	Try a different ELN	0%
	Try a different ELN sometime in future	50%
	No ELN anymore	0%

4 Discussion

A metadata scheme for a Western Blot experiment was developed capturing 51.1% of the metadata in a simplified way by using default values. The metadata scheme was developed on the basis of twelve protocols for a Western Blot experiment which were used in different biomedical working groups departments of the CRC 1002. Notably, every group used just one protocol for this experiment. However, the comparison of the protocols showed significant differences between these protocols related to the length and the comprehensiveness. This is not astonishing as every working group of course has specific requirements, habits and peculiarities.

From the experience made, using the example of the Western Blot, we think it is impossible to construct one general metadata scheme for an experiment which should be used by different working groups out of different research fields; even if it is a standard experiment. The customizing of the metadata scheme itself as well as the default values is of uttermost importance for the user acceptance. This individual customizing comes along with increased work load for the administrator, if the customizing is not done by the users themselves. However, only by customizing it can be ensured that the metadata scheme and the default values fit the requirements and the working style of the users. This ensures the practical use of the metadata templates within an ELN. It is known that technological improvements in the lab especially about documentation are seen with some suspicion [24]. Many researchers fear a higher work load or a data chaos due to the conversion to new data capture systems. Personal support by the IT for ELN users and paying attention to the researchers' requirements are the most important points for the introduction of new technologies. The acceptance of a new technology is a point that should not be underestimated [24], [25]. Therefore, much attention was spent by the IT to the acceptance of the ELN by future users.

However, for accurate data documentation metadata are indispensable. Many researchers don't know or don't care about the potential importance of metadata for scientific progress [4], [12]. One of the first steps to develop the metadata scheme for a Western Blot experiment was to identify existing experimental metadata standards by literature search. All identified standards were general standards for the documentation of different data types or the documentation of experimental data in general but no standard was specific for a Western Blot experiment. Remarkably, even the highly recommended "Minimum Information about proteomics experiment (MIAPE) [22] does not provide the "how to document" the data items. All in all, this supports the assumption that either metadata are neglected and underestimated in biomedical research or that no one thinks they deserve publications.

Nevertheless, the metadata scheme for Western Blot developed and described here captures all accrued important metadata by using default values and ensures the correct, detailed and accurate documentation of scientific data. With default values it captures previously selected metadata in an easy way. The greatest advantage of default values is that they can be easily adapted individually, even by users. Science is a fast growing and frequently changing field, resulting in the need to make adjustments constantly. There are disadvantages as well: default values pre-fill an experimental protocol, this invites for fast and careless documentation or it causes documentation mistakes because of overseen selected checkboxes, choice fields, etc. The risk of wrong documentation or documentation mistakes exits all the time but every researcher has taken over the responsibility for correct and detailed documentation by following the rules of good scientific practice, which are enforced at most academic institutions, at least in Germany and Europe. It is never possible to avoid wrong or manipulated scientific work by higher control mechanisms. Thus, it is every scientist's responsibility to act scientifically correct. The fear of consciously mediated documentation mistakes or manipulations should not avoid the technological progress.

The evaluation has been performed with only four researchers using the ELN for one week so far. Moreover, the evaluation addressed the handling with the ELN and not specifically working with the metadata scheme. Only the questions on the easiness to work with the ELN and the improved data management indirectly address the usefulness of the metadata scheme. Therefore, it was difficult to give a general statement about how good the ELN and the metadata scheme work for the academic researchers. In addition, during the evaluation phase the researchers used only templates with default values available. To address the usefulness and checking of preset default values a different setting would be necessary. This could be a comparison of the use of metadata schemes with default values in comparison to the free-text entry function of the ELN and/or the paper laboratory. Altogether, two scientists were satisfied with the ELN, while two participants had problems with the handling and wish to try another ELN. However, an evaluation with all working groups of the CRC 1002 could lead to different results but it would be interesting to see how useful default values set for one working group would be for another.

Beside the simplified metadata capture another goal of developing the metadata scheme was to improve the quality of records in the ELN. Therefore, it is important to always use the same terminology and descriptions. Every lab has its own specific terms or names for different experiments, equipment or processes, which makes data incomprehensible or prevents their reusability. Controlled vocabularies define terms in an unambiguous way and by this prevent misunderstandings or unusable data. A literature search on controlled vocabulary for the Western Blot indicated a lack of controlled vocabularies using the indicated search terms, which might be true for life sciences in general. The results out of the literature search did not meet the expectations of the PhD students as well as their PI's. All standards were really universal; none of the identified metadata standards was specific for just one experiment. In contrast every lab produces specific metadata depending on specifically adapted methods or different used machines. Therefore, one of the next goals of this infrastructure project within the CRC 1002 is to develop controlled vocabulary with the help of the "NISO – Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies" [22]. As this needs to be done centrally for all different working groups of the CRC for the ELN, a tight collaboration between IT and the wet-lab scientists is indispensable, even if the latter should become more independent with regard to using the ELN with SOPs.

5 Conclusion

Here, we could show that capturing metadata by using default values, can simplify the data documentation effort. At the same time, the researcher could benefit by providing standardized metadata for data sharing and re-use.

Altogether, using metadata schemes saves time, records the data in a standardized and organized way and captures the right data. This simplifies the everyday working of the scientists and improves the data quality. Experience has demonstrated that it is indispensable for the usage of the ELN to respect the wishes and requirements of the users. Metadata schemes must be adapted individually to fit the individual work processes of a user and to ensure that the schemes will be constantly used in the frame of an ELN. A close collaboration with the future users is a key to achieve acceptance.

The next step is to extend the evaluation to all participating working groups of the CRC 1002. Moreover, more metadata schemes need to be developed. The logical next step after using default values to capture metadata will be to design and implement interfaces between laboratory equipment and the ELN for an even more accurate and automated data capture.

Notes

Competing interests

The authors declare that they have no competing interests.

Acknowledgments

We like to thank all participating scientists of the CRC 1002. Especially we would like to thank Blanche Schwappach, Anne Clancy and Bartlomiej Marzec. This project was funded by the Deutsche Forschungsgemeinschaft (DFG) SFB 1002, TP-INF.

References

1. Kanare HM. Writing the Laboratory Notebook. Washington, DC: American Chemical Society; 1985.
2. Potthoff J. Beweiswerterhaltendes Datenmanagment im elektronischen Forschungsumfeld. In: Müller P, Neumair B, Reiser H, Rodosek GD, eds. 5. DFN-Forum Kommunikationstechnologien: Fachtagung; Regensburg; 21.-22.05.2012. Bonn: Gesellschaft für Informatik; 2012. (Lecture Notes in Informatics; P-203). p. 109-18. Available from: http://cs.emis.de/LNI/Proceedings/Proceedings203/article6710.html
3. Menzel J, Weil P, Bittihn P, Hornung D, Mathieu N, Demiroglu SY. Requirement analysis for an electronic laboratory notebook for sustainable data management in biomedical research. Stud Health Technol Inform. 2013;192:1108.
4. European Science Foundation. Good scientific practice in research and scholarship. ESF Sci Policy Brief. 2000;10:1-16. Available from: http://www.esf.org/fileadmin/Public_documents/Publications/ESPB10.pdf
5. Kloeckner F, Farkas R, Franken T, Schmitz-Rode T. Development of a prediction model on the acceptance of electronic laboratory notebooks in academic environments. Biomed Tech (Berl). 2014 Apr;59(2):95-102. DOI: 10.1515/bmt-2013-0023
6. Rubacha M, Rattan AK, Hosselet SC. A review of electronic laboratory notebooks available in the market today. J Lab Autom. 2011 Feb;16(1):90-8. DOI: 10.1016/j.jala.2009.01.002
7. Fischer S, Stewart TE, Mehta S, Wax R, Lapinsky SE. Handheld computing in medicine. J Am Med Inform Assoc. 2003 Mar-Apr;10(2):139-49. DOI: 10.1197/jamia.M1180
8. Goddard NH, Macneil R, Ritchie J. eCAT: Online electronic lab notebook for scientific research. Autom Exp. 2009;1(1):4. DOI: 10.1186/1759-4499-1-4
9. Nussbeck SY, Weil P, Menzel J, Marzec B, Lorberg K, Schwappach B. The laboratory notebook in the 21st century: The electronic laboratory notebook would enhance good scientific practice and increase research productivity. EMBO Rep. 2014 Jun;15(6):631-4. DOI: 10.15252/embr.201338358
10. National Information Standards Organization. Understanding metadata. Bethesda, MD: NISO Press; 2004. ISBN: 1-880124-62-9. Available from: http://www.niso.org/publications/press/UnderstandingMetadata.pdf
11. Willoughby C, Bird CL, Coles SJ, Frey JG. Creating context for the experiment record. User-defined metadata: investigations into metadata usage in the LabTrove ELN. J Chem Inf Model. 2014 Dec;54(12):3268-83. DOI: 10.1021/ci500469f
12. Kroon-Batenburg LM, Helliwell JR. Experiences with making diffraction image data available: what metadata do we need to archive? Acta Crystallogr D Biol Crystallogr. 2014 Oct;70(Pt 10):2502-9. DOI: 10.1107/S1399004713029817
13. McNutt M. Journals unite for reproducibility. Science. 2014 Nov;346(6210):679. DOI: 10.1126/science.aaa1724
14. Mahmood T, Yang PC. Western blot: technique, theory, and trouble shooting. N Am J Med Sci. 2012 Sep;4(9):429-34. DOI: 10.4103/1947-2714.100998
15. Marzec B, Weil P, Nussbeck SY. Vorbereitungsprozess für die elektronische Dokumentation und Nachnutzbarkeit von FD in ELN. In: GMDS 2013. 58. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V. (GMDS). Lübeck, 01.-05.09.2013. Düsseldorf: German Medical Science GMS Publishing House; 2013. DocAbstr.85. DOI: 10.3205/13gmds051
16. Sousa AM, Pereira MO, Azevedo NF, Lourenço A. An harmonised vocabulary for communicating and interchanging biofilms experimental results. J Integr Bioinform. 2014;11(3):249. DOI: 10.2390/biecoll-jib-2014-249
17. Towbin H, Staehelin T, Gordon J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc Natl Acad Sci U S A. 1979 Sep;76(9):4350-4. DOI: 10.1073/pnas.76.9.4350
18. Burnette WN. "Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate--polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. Anal Biochem. 1981 Apr;112(2):195-203. DOI: 10.1016/0003-2697(81)90281-5
19. Data and Metadata Models Developed by the IHEC Metadata Standards Workgroup. 2013 July 3. http://ihec-epigenomes.org/research/operating-procedures/
20. Betsou F, Lehmann S, Ashton G, Barnes M, Benson EE, Coppola D, DeSouza Y, Eliason J, Glazer B, Guadagni F, Harding K, Horsfall DJ, Kleeberger C, Nanni U, Prasad A, Shea K, Skubitz A, Somiari S, Gunter E; International Society for Biological and Environmental Repositories (ISBER) Working Group on Biospecimen Science. Standard preanalytical coding for biospecimens: defining the sample PREanalytical code. Cancer Epidemiol Biomarkers Prev. 2010 Apr;19(4):1004-11. DOI: 10.1158/1055-9965.EPI-09-1268
21. Dublin Core Metadata Initiative (DCMI) Usage Board. DCMI Metadata Terms. 2012. Available form: http://dublincore.org/documents/dcmi-terms/
22. National Information Standards Organization. Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. ANSI/NISO Z39.19-2005. Baltimore, MD: NISO; 2005. ISBN: 1-880124-65-3. Available from: http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf
23. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science. 1985 Dec;230(4732):1350-4. DOI: 10.1126/science.2999980
24. Mückschel C, Nieschulze J, Weist C, Sloboda B, Köhler W. Herausforderungen, Probleme und Lösungsansätze im Datenmanagement von Sonderforschungsbereichen. Gesellschaft für Informatik in der Land-, Forst-und Ernährungswirtschaft e.V. (GIL), ed. eZAI. 2007;2. Available from: http://www.preagro.de/ezai/index.php/eZAI/article/view/21/22
25. Mückschel C, Schachtel GA, Wehum A, Nieschulze J, Koehler W. Grundlegende Anforderungen an das Datenmanagement in interdiziplinären Forschungsprojekten. In: Wenkel KO, Wagner P, Morgenstern M, Luzi K, Eisermann P, eds. Land- und Ernährungswirtschaft im Wandel – Aufgaben und Herausforderungen für die Agrar und Umweltinformatik. Referate der 26. GIL Jahrestagung, 06.-08. März 2006, Potsdam. Bonn: Gesellschaft für Informatik; 2016. (Lecture Notes in Informatics; P-78). p.185-8. Available from: http://subs.emis.de/LNI/Proceedings/Proceedings78/article3750.html

Corresponding author:

Julia Menzel

Institut für Medizinische Informatik, Universitätsmedizin Göttingen, Robert-Koch-Straße 40, 37075 Göttingen, Germany

julia.menzel@stud.uni-goettingen.de

Please cite as
Menzel J, Weil P, Nussbeck SY. Metadata capture in an electronic notebook: How to make it as simple as possible? GMS Med Inform Biom Epidemiol. 2015;11(1):Doc06.
DOI: 10.3205/mibe000162, URN: urn:nbn:de:0183-mibe0001626

This article is freely available from
http://www.egms.de/en/journals/mibe/2015-11/mibe000162.shtml

Published: *2015-09-15*