Data Management According to the Good Scientific Practice

Jan Potthoff, Marius Walk, Sebastian Rieger
Steinbuch Centre for Computing
Karlsruhe Institute of Technology
Karlsruhe, Germany
[name.surname]@kit.edu

***Abstract*—Data Management in scientific and research processes needs to comply with rules defined in the Good Scientific Practice polices, e.g., issued by scientific institutes and research organizations. Typically these regulations include definitions to protect the data privacy and security throughout the entire lifecycle of the scientific data. Beginning from the generation of the data and ensuring its provenance up to the final archival and corresponding long-term data storage management. This paper introduces a simple way to manage the scientific data that ensures long-term protection and provability of the data quality. Furthermore an easy to use implementation which offers a possibility to structure the scientific data and to archive it according to Good Scientific Practice is presented.**

Keywords-ELN; scientific data; data formats; long-term preservation; evidence

I. INTRODUCTION

During the research process scientists have to deal with electronic data and its management. Complex applications like electronic laboratory notebooks (ELN) and laboratory information and management systems (LIMS) are developed to support the scientists during day-to-day work. In addition several rules, e.g., the good scientific practice (GSP) defined by scientific institutes and research organizations, have to be respected in the research process and the entire lifecycle of the scientific data. In some research areas, e.g., social science or computer science, the documentation with a laboratory notebook is not common and sometimes inefficient regarding the amount of data that is processed in the research process. In this case the documentation of the research process is typically not based on laboratory notebooks and applications such as ELN or LIMS are less frequently used. Nevertheless a data management according to GSP is needed.

The paper is outlined as follows. Section II describes the requirements for good scientific practice (GSP) and its implications for data management. In Section III different forms and formats of scientific data and related work in this research area are shown. Using the requirements given in Section II and the various forms of scientific data described in Section III, Section IV focusses on appropriate research data management tools and practices. Section V contains a software implementation that facilitates data management according to GSP requirements introduced in this paper. Finally in Section VI a conclusion, major strengths and weaknesses of the solution along with future work are presented.

II. THE GOOD SCIENTIFIC PRACTICE

A sustainable documentation of the research results requires keeping several rules, e.g., the rules of GSP. The intention is to guarantee a high quality in the research area and the work of scientists to prevent scientific deception or fraud. The spectrum of scientific misconduct ranges from several violations of scientific ethics to criminal intentions [1]. Hence, the transparency in dealing with primary data is a basic claim of GSP. With the term “primary data” all data from an experiment or a scientific survey is covered [1].

By the rules of GSP the following requirements are addressed: traceability, long-term interpretability and sustainable archiving. The traceability is deemed necessary when the scientific process is documented in a way such that the results need to be completely reproducible. Hence, a long-term interpretability to understand and reproduce the results is important as well. For the sustainable archiving, corresponding techniques are required. For example, the German research foundation (DFG) requires archiving for 10 years [2]. The DFG proposes the documentation based on a laboratory notebook, but nowadays especially for large amounts of data digital archiving is needed as well. In addition the rules of the DFG are used by several organizations as well [1][3]. Especially for the work in a laboratory there are further regulations, e.g., the good laboratory practice [4]. There are also legal regulations, for example in the healthcare sector, e.g., rules with regard to the archiving period for documents like the digital patient file.

So, the data management in the research process has not only to deal with data structures, but also with several regulations relating to the respective research areas.

III. SCIENTIFIC DATA

Establishing policies for GSP has an influence on the data being generated and processed in laboratories and scientific processes. The following sections give an overview of typical forms and formats of scientific data and related work that focusses on the proper management of scientific data and metadata.

A. Data Formats and the Long-term Preservation

In the research process electronic data is generated and used in varying volumes depending on the research field and approach. Based on this data existing research results are confirmed or new approaches are elaborated. For example, in

 ISBN: 978-1-61208-247-9

social science, data, which was generated in the research process, is used, e.g., in surveys, experiments or observations, or in non-scientific works, e.g., official statistics [5]. The scientific data and its data formats vary over different kinds of approaches and the software used in the research area.

In experimental research electronic data is generated in varying file formats in each process phase as well. For example, at the Max Planck Institute for Dynamics and Self-Organization (Göttingen, Germany) experiments are carried out in a wind tunnel using three high speed cameras. Each of these cameras generates 10,000 images per second [6]. While digital cameras usually take pictures in widely-used file formats other measurement instruments generate data whose proprietary format is specified by the manufacturer. It is even possible that the format is dependant on the device. This could be common formats in ASCII code or XML, but also complex formats specified individually by applications.

The scientific data is managed using its metadata in the research process. With the metadata the corresponding data will be described (descriptive metadata) and technical information will be documented (administrative metadata).

Although standards such as Dublin Core, Metadata Object Description Schema (MODS) or Metadata Encoding and Transmission Standard (METS) aim to uniform the format of metadata, the volume and form of metadata strongly vary throughout the research areas [7].

Metadata is very important for the long-term preservation as well. Using the metadata, the corresponding data can be found even if a unique identifier does not exist, is unknown or has been lost. Therefore the long-term preservation data formats such as XML formatted Archival Information Package (XAIP) or the universal object format (UOF) usually contains a combination of data and metadata. For example, XAIP was designed for an archive system, whose construction is based on the technical directive of the Federal Office for Information Security (BSI, Germany) [8]. The archival information package is an XML file that contains the data in the Base64 format and the corresponding metadata. In the UOF the data is stored as a file in a tar archive instead of the Base64 format. To save the metadata, a separate file, based on METS, is also stored in the tar archive [9].

B. Related Work

The data handling in the research process [7] and the requirements of the scientists [10] are current research topics. Because of the amount of information and data [28], respectively, the goal is to find a way to support the scientists in their day-to-day work [30].

Many software applications for the documentation of the research process, e.g., ELN or LIMS, with different emphasis are available today [11]. The requirements of scientific data management are addressed in several research projects [12][13]. Even if the documentation varies with the research area, all research fields need sustainable archiving of the scientific data [14]. The use of designed applications and archiving solutions has to be kept simple [10].

We address these requirements with an intuitive and simple application that will be described in Section V.B. Using this application a simple data collecting and management is achieved as well as a sustainable archiving mechanism. Other projects in the same area can be found, e.g., in [27][29]. Compared to our solution these approaches do not offer explicit protection of the probative force of the scientific data. Somehow [27] focusses on the provenance of scientific data without implementing long-term preservation of the probative force. Our solution offers a small light-weight client that can be simply used like a file explorer by the scientists. In addition, by using the BeLab system as basis for our implementation, as explained in Section V.A., the data will be long-term archivable according to GSP.

IV. RESEARCH DATA MANAGEMENT

As described in Section III.A, data and corresponding metadata is produced and used in a variety of formats throughout scientific processes. As the processing and interpretation of the data is essential for generating results from the scientific process (e.g., for publications), efficient storage and management of the generated research data is necessary in scientific environments. Several management tools and frameworks for scientific data have been developed in the last decades. Especially web-based information and document management tools have evolved and extended to fulfill the requirements of scientific processes. To allow a profound integration with specific scientific processes a new category of software products has been formed, e.g., ELN, that is described in the next section. These tools enhance the management of research data by facilitating its retrieval and processing beginning from the generation (e.g., by directly importing data from sensors during experiments), modification (e.g., data processing or manual interaction) up to publication, archival and deletion of data forming the scientific data lifecycle.

A. Electronic Documentation

The increased use of computers and the corresponding amount of electronic data led to the need of electronic documentation in the research process. By using a paper-based laboratory notebook and storing the electronic data separately, the danger of losing data is increased. Nevertheless this kind of documentation is used in several institutions [15]. A central electronic data management reduces these risks and also offers further advantages. For example, a collaboration with other colleagues is facilitated, the search for data becomes easier and faster and the research process gets more efficient [16].

To implement a central approach of documentation and data handling a corresponding software solution is needed. Hence, various systems are available on the market today. ELN can be defined as “a secure system assembling scientific content from multiple sources related to each other, allowing for contextual annotation, and packaging it in a legally acceptable document to be searched, mined and collaborated” [17]. In general, ELN can be understood as “an electronic embodiment of what is currently being done in a paper laboratory notebook” [18] whereas LIMS is more integrated in the research process. It offers the possibility to collect data from connected measurement instruments, such that the systems naturally have to deal with more scientific data. Additionally, LIMS is used for administrative processes as well as for documentation. It is also possible to integrate a separate ELN in LIMS. A clear separation between these systems is therefore difficult because they deal with similar use cases and the precise definition depends on the range of use [19].

In addition to the applications named here, other programs that are not designed for this purpose are used in the research process. For example, web-based information and document management systems (e.g., Wiki systems) or even simple office applications are used [20]. In other cases, an individual ELN solution has been developed by the research group. For example, the solution "open enventory" has been developed as an administrative system for chemical substances in the first phase. Later on the system has grown to an ELN that is tailored to address the requirements of research groups in chemistry [21].

B. *Practical Requirements*

As the workflows that generate and process data in institutions and groups differ, it is hard to establish a unified data management, e.g., using ELN and LIMS as described in the previous section. The trade-off between individual optimization of the data processing workflows and using standardized software solutions holds for both scientific and corporate scenarios. Often the users simply want to archive the data during or at the end of the scientific process. Moreover not every scientific institute has the necessary resources to implement ELN or LIMS systems for their research environments.

Establishing GSP for scientific data does not require specific frameworks like ELN or LIMS. The benefit of using ELN and LIMS regarding the GSP depends on the integration of workflows and, i.e., measuring the probative value of the data in an early phase during the data generation in the scientific process (e.g., combing and verifying information from different sensors and user interaction). However this is typically not required by GSP regulations. Additionally it is nearly impossible to get the entire raw data during its generation and verify it in real-time, i.e., because of its increasing volume, e.g., due to high resolution sensors. For example, the regulations by the DFG require the scientific institutions to protect the integrity of the data on the long-term, e.g., while being stored in a digital archive, and not to include the data of every single sensor and verify the entire workflow of data generation which would be rather complex both during archival and verification. This could also be achieved by checking the integrity and sign the data before archiving it using a simple client that is able to store generic files and thus data and metadata containers. To allow a long-term preservation of the data and proving its probative value, standards for the signature and archival format should be used. Also the client application should offer a simple way for verifiers and scientists to search retrieve and verify the data. It should also handle the evaluation of the probative value and the import and export of data and metadata extraction in a highly automated manner. The effort necessary to evaluate the probative value and archive the data should be as minimal as possible to allow for a higher acceptance. The majority of the data is stored in files, so the application should be file-based.

V. UNIFORM DATA PREPARATION

As described in Section III, specialized applications can be used for the documentation of the research process and to manage the scientific data. For many research areas these applications are over-sized, because the scientist simply has to deal with a few files during a workday. In this case individual programs, e.g., office applications, are used [20]. However, this case requires a solution to address the regulations of the GSP. In the following section a system that can be used for the long-term preservation of the scientific data and its probative value is described. An application that is based on this system is presented in Section V.B. The application enables the scientist to use the system in an easy way to manage the data and to ensure the GSP requirements.

A. *Evidential Long-term Preservation*

In many research areas a long-term preservation of scientific data is required, e.g., by German law or internal regulations of research organizations like the GSP, as described in Section II. The goal of the BeLab project founded by DFG is to develop a concept for the long-term preservation of scientific data and its probative value to secure the quality of complex data. Therefore the requirements of the long-term preservation of scientific data, the probative value of electronic data and the possibilities of its conservation are analyzed.

Even if paper-based laboratory notebooks are still used in the scientific process [15], today's scientific documentations include an increasing amount of electronic data. Also, in research areas in which the documentation is typically not based on the use of an ELN or a paper-based laboratory notebook, the volume of electronic data is constantly growing, as described in Section III.

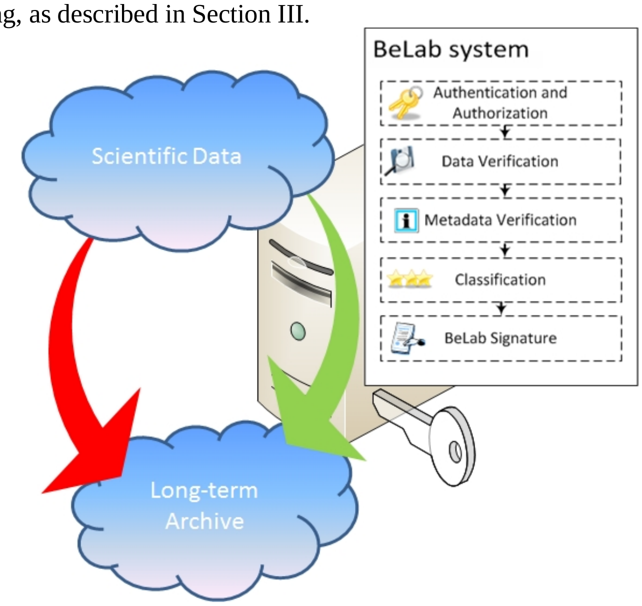

Figure 1. The BeLab system.

Electronic documents are considered to be insecure and to have less probative value in comparison to digital certificates relating to the German law [22]. By using electronic security technologies, the probative value can be increased to the same level as in paper-based documents. Especially to secure the integrity and authenticity, digital signatures can be used, based on the German Digital Signature Act. In the BeLab project a concept for a middleware (BeLab system) was designed and implemented. The BeLab system, as shown in Figure 1, can be used from various specialized applications, e.g., ELN, to prepare the scientific data for the evidential long-term preservation. After that the scientific data will be submitted to a connected archive system [23]. For example, the archive can be used as mentioned in Section III.A.

The technical directive (BSI TR-03125) defines an archives system, which is able to update digital signatures before they are invalid [8]. To enable a uniform processing of the data and evaluation of its probative value within the BeLab system, a data model based on UOF is used. Therefore the files to be submitted will be packed in a TAR or ZIP archive. Additionally, a metadata file (mets.xml) will be separately added to the archive [23]. The structure of the metadata file is based on METS that was designed to ensure a uniform management and exchange of electronic data [24].

The BeLab system executes verification modules, which are based on the file format [23]. These modules analyze the probative value, the suitability for long-term preservation and the characteristics of the data generation. The verification of the probative value is mainly based on the use of digital signatures but also on hash values of files being submitted, that are defined in the metadata file. For the verification of the suitability for the preservation, the current file format is analyzed and a verification of the data generation is performed.

In a subsequent classification process, as depicted in Figure 1, the result of the data and metadata verification is mapped to a category, which describes the degree of the probative value, the suitability for the long-term preservation and the security measures during the data generation [23].

The results of the data and metadata verification are recorded together with the classification in a log file as well as in the updated metadata. To ensure the integrity of the metadata file the results are added in a separate copy (belab.xml) of the file. The copy of the metadata will be signed by the BeLab system and is added to the data model (BeLab object), which contains the scientific data in the UOF. Using this digital signature the integrity and authenticity can be verified while the data is stored in the archive. Finally, the BeLab object will be submitted to the connected archive system. The user gets a unique identifier, which represents the archived data model. Additionally, the BeLab object is managed with an individual project, ELN and container identifier being defined by the user [23].

By using the BeLab system to archive the scientific data, as shown in Figure 1 with the right arrow, the scientist gets useful information about the probative value and the suitability for long-term preservation of the data being submitted, that would not be available upon archiving the data without any data verifications, as shown in Figure 1 with the left arrow.

B. Data and Metadata Detection

Based on the results of the BeLab project a client application was implemented, which offers a possibility for the management of scientific data and detection of corresponding metadata in the research process. Instead of a complex ELN or LIMS the client was designed for a simple use case which allows the scientist to archive the data with respect to the GSP, as described in Section IV.B. The graphical user interface is divided in three sections, as shown in Figure 2. In these sections the user has the possibility to collect files to be archived or to check out archived files and edit them. The section of metadata refers to the structure of data, i.e., project, ELN and general container id. Further metadata can be indicated for each added file in the data section, e.g., document title, author and creation date. This metadata is used by the archive system to manage the data and in the BeLab verification process, as described in Section V.A.. For example, the author will be compared with the given author in the document, e.g., a Word file.

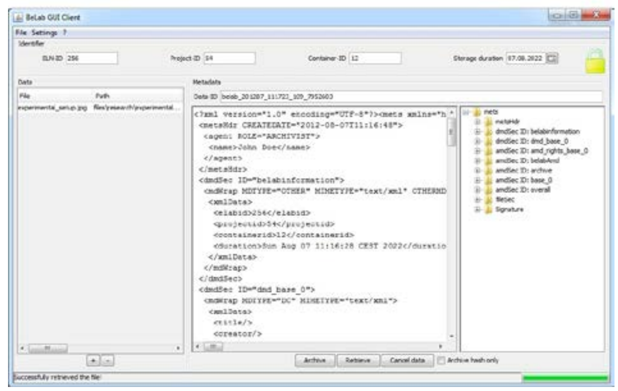

Figure 2. Grafical user interface of the BeLab client.

The generation of an archive according to the UOF that was described in the previous section can be performed by adding files to the data section. This is done with the plus symbol below the file table. After all files have been added and all needed or desired metadata has been entered, the data collection will be automatically converted to the UOF object and will be send to the BeLab system by pressing the button "archive". It should be noted that the file format is very important for proper long-term preservation. Here, all file formats will be accepted. But by using the universal object format, the possibility of file migration is given. The migration has to be supported by the archive system.

The BeLab system will convert the UOF object to the BeLab data model. After that the BeLab system will evaluate the data, as described in Section V.A., and submit it to the integrated archive system. After the data model was stored in the archive the client application will automatically retrieve and display the generated BeLab object from the archive. To retrieve the data the corresponding identifier that was returned by the BeLab system is used.

The main focus of the client application lies on the representation of the metadata. This includes the metadata that was specified by the user as well as the metadata that was generated during the BeLab process, as described in Section V.A. In addition, the unique identifier of the corresponding archive will be displayed. The metadata will be displayed in a text field as well as a tree structure. By selecting a tree node, the corresponding content will be displayed as XML in the text field, e.g. the log file which has been written during the BeLab process.

Only the part of the digital signature and the BeLab classification will be presented in a different way. The classification will be shown as a table that contains all filenames and the corresponding probative values that were calculated by the BeLab system, as described in Section V.A. The validation of the digital signature will not only be represented as a detailed list in the text field, but also as a lock symbol at the top of the window. An open lock is associated with an invalid signature whereas a closed lock stands for a valid signature. Also the background color of the text field will be green if the signature is valid.

In addition to the representation in the application, the information can be exported to a PDF report. This report contains the metadata as well as an evaluation of the documented hash values and the result of their verification. Hence the integrity of the files can be verified and demonstrated, respectively.

To retrieve an existing archive file the corresponding unique identifier has to be entered into the data id field. By clicking the button “retrieve” the data will be requested and represented after a successful authentication and authorization of the user. Then it is possible to edit the data and metadata again. For example, files can be added or removed from the file table and archive object, respectively. To update the corresponding object in the archive, the “archive” button can be clicked again. The new data will be stored with an increased version number. To update data that has already been archived previously, the old data identifier has to be entered into the data id field.

Additionally, a data object can be deleted from the archive by using the client application. It should be noted that the corresponding object will only be marked as deleted instead of being physically removed from the archive. Using a further metadata item the storage duration can be indicated. After this time the corresponding object will be automatically canceled by the archive system.

In cases in which the files cannot be submitted to an external system, e.g., to ensure data protection or confidentiality of private data or because the amount of data is too large, the client application offers the possibility to submit only the corresponding unique hash values to the BeLab system. In this case the user needs to take care of the long-term preservation of the data. It should be noted that the solution has not been designed for large scale data management, but rather for research areas with limited amount of data.

For the data transmission, the user can choose between two transfer protocols. The first option is the Simple Object Access Protocol (SOAP) [25]. SOAP supports an authentication using a password or certificate. The second alternative is the REpresentational State Transfer (REST) [26]. Here, only the password-based authentication is possible.

In research areas in which the documentation based on a laboratory notebook is not common, the presented solution can be used to structure the scientific data and finally submit it to an archive according to the rules of GSP.

VI. Conclusion and Future Work

In this paper we introduced a simple client application that allows preserving GSP requirements as defined, i.e., by the German Research Foundation (DFG) [2] or corresponding regulations issued by scientific institutions (e.g., [1]). The client connects to a web service (the BeLab system as described in [23]) that evaluates the data and metadata submitted by the client regarding its probative force and suitability for long-term storage. After the evaluation, the data is digitally signed to protect the integrity and authenticity of the data while being stored in a long-term archive. Results of the evaluation are delivered back to the client. The client can be used later on to prove the probative force, consistency and integrity of the scientific data using the embedded digital signature. Compared to existing information management solutions and specific scientific data management tools like ELN and LIMS, the client offers a simplified way to ensure integrity and authenticity of scientific data. It can be easily integrated into every scientific process or workflow that produces or processes data in form of files. Moreover most of the existing tools that support the data management in scientific workflows do not support the protection of integrity and authenticity of the data or its long-term interpretability. As existing metadata standards are used, the long-term interpretability of the results of the evaluation (and the data) is addressed.

Together with other scientific institutions the project currently identifies an integration of the client and the BeLab web service into existing scientific processes. A future enhancement of the client will therefore focus on the automatic usage of the client. One example could be to automatically submit files to the BeLab system using predefined rules as they are being created in a directory monitored by the client. Furthermore, lab equipment could be connected to the client using custom or industry standard interfaces. This way additional metadata context could be provided to allow further evaluation of the probative force in the BeLab system. To be used as a generic tool in a variety of scientific processes and research areas, the support for specific data formats and interfaces of lab equipment have to be enhanced in future versions. If the client collects metadata throughout the entire research process, the integrity of the workflow could be proven later on by verifying the digital signature attached to the evaluation results stored in the archive. NoSQL databases for the storage of scientific data, as described, e.g., in [28], could also be a promising option. Especially the document-oriented types are well suited to store scientific data and metadata. Extensions to evaluate the probative force, to store the digital signature, as described for our solution, and to retrieve and check the integrity have to be developed to ensure GSP with NoSQL databases.

Another field for future research could be the usage of the stored data and metadata as evidence during trials. In Germany a specific procedure is being developed to receive evidence records for PDF files using e-mail and digital signatures. The client, while being used by an auditor or reviewer, could automatically send the signed result of the verified metadata to an e-mail address that can be used by the judges or attorneys during the trial. The probative force of the data and metadata can be enhanced even further by including identification, authentication and digital signature mechanisms of digital passports carried by scientists.

REFERENCES

[1] Karlsruher Institute of Technology (KIT), "Regeln zur Sicherung guter wissenschaftlicher Praxis im Karlsruher Institut für Technologie (KIT)," http://www.kit.edu/downloads/K_OBP_XX_RI_01_05-10.pdf 04.12.2012.

[2] Deutsche Forschungsgemeinschaft, "Recommendations of the Commission on Professional Self Regulation in Science -Proposals for Safeguarding Good Scientific Practice," January, 1998, http://www.dfg.de/en/research_funding/legal_conditions/good_scientific_practice/ 04.12.2012.

[3] Max-Planck-Gesellschaft (MPG), "Rules of Good Scientific Practice," 2000, http://imprs.ice.mpg.de/ext/fileadmin/imprs/pictures-files/Good_Scientific_Practice_MPG.pdf 04.12.2012.

[4] OECD, "OECD Principles on Good Laboratory Practice," Paris, 1998, http://search.oecd.org/officialdocuments/displaydocumentpdf/?doclanguage=en&cote=env/mc/chem%2898%2917 04.12.2012.

[5] A. Bryaman, "Social Research Methods," Oxford, 2012.

[6] U. Degenhardt, "Requirements for Repository Systems," eScience Seminar, 2009, http://colab.mpdl.mpg.de/mediawiki/images/e/ea/ESci_09_Sem_2_Requirements_from_MPI_f_Dynamics_and_Self-Organization_Degenhardt.pdf 04.12.2012.

[7] L.M. Chan and M.L. Zeng, "Metadata Interoperability and Standardization - A Study of Methodology, Part I," D-Lib Magazine, Vol. 12, Nº. 6, 2006.

[8] Bundesamt für Sicherheit in der Informationstechnik (BSI), "BSI Technical Guideline 03125: Preservation of Evidence of Cryptograohically Signed Documents v.1.1," 2011, https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Publications/TechGuidelines/TG03125/TG-03125_main.pdf 04.12.2012.

[9] T. Steinke, “The Universal Object Format – An Archiving and Exchange Format for Digital Objects,” in Research and Advanced Technology for Digital Libraries, Springer Berlin, 2006, pp. 552–554.

[10] M. Feijen, "What researchers want," SURFfoundation, 2011, http://www.surf.nl/nl/publicaties/documents/what_researchers_want.pdf 04.12.2012.

[11] M. Rubacha, A.K. Rattan, and S.C. Hosselet, "A Review of Electronic Laboratory Notebooks available in the market today," JALA, vol. 16, Feb.2011, pp. 90–98 , doi:10.1016/j.jala.2009.01.002.

[12] M. Dreyer, N. Bulatovic, U. Tschida, and M. Razum, "eSciDoc – a Scholarly Information and Communication Platform for the Max Planck Society," Proc. German e-Science Conference, 2007.

[13] T. Schlauch and A. Schreiber, "DataFinder – A Scientific Data Management Solution," Proc. PV 2007, 2007.

[14] R. Altenhöner, "Data for the future: The German project ‘Co-operative development of a long-term digital information archive’ (kopal)," Library Hi Tech, Vol. 24 Iss: 4, 2006, pp. 574–582, doi:10.1108/07378830610715437.

[15] B.A. Weber, H. Yarandi, M.A. Rowe, and J.P. Weber, "A Comparison Study: Paper-based Versus Web-based Data Collection and Management," in Applied Nursing Research, Vol. 18 Iss: 3, 2005, pp. 182–185.

[16] S. Hackel, P.C. Johannes, M. Madiesh, J. Potthoff, and S. Rieger, "Scientific Data Lifecycle – Beweiswerterhaltung und Technologien," Proc. 12. Deutscher IT-Sicherheitskongress (BSI-IT-SEC 2011), SecuMedia, 2011, pp. 403–418.

[17] M.H. Elliott, "Electronic Laboratory Notebooks: A Foundation for Scientific Knowledge Management Edition III," Atrium Research & Consulting LLC, Wilton, CT USA.

[18] P. Boogaard and P. Pijanowski, "Electronic Laboratory Notebooks (ELN) Mean Many Things to Many People," Feb. 09, 2012, http://www.laboratory-journal.com/science/information-technology-it/electronic-laboratory-notebooks-eln-mean-many-things-many-people/ 04.12.2012.

[19] D. Morris, "LIMS vs ELN arch enemies or best of friends?," DDW, 2009, http://www.ddw-online.com/informatics/p146753-lims-vs-elns-arch-enemies-or-best-of-friends-summer-09.html 04.12.2012.

[20] T. Kuipers and J. v. d. Hoeven, "Insight into digital preservation of research output in Europe - Survey report," 2009, http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf 04.12.2012.

[21] F. Rudolphi and L.J. Goossen, "Electronic Laboratory Notebook: The Academic Point of View," J. Chem. Inf. Model., 2012, 52 (2), pp 293–301 DOI: 10.1021/ci2003895.

[22] S. Fischer-Diskau, "Das elektronisch signierte Dokument als Mittel zur Beweissicherung," Nomos, 2006.

[23] J. Potthoff, S. Rieger, and P.C. Johannes, "Enhancing the Provability in Digital Archives by Using a Verifiable Metadata Analysis Web Service," Proc. 7th ICIW 2012, 2012, pp. 112–117.

[24] Digital Library Federation, "<METS> Metadata Encoding and Transmission Standard: Primer and Reference Manual," Version 1.6 Revised, 2010, http://www.loc.gov/ standards/mets/ 04.12.2012.

[25] M. Gudgin, M. Hadley, N. Mendelsohn, J.J. Moreau, H. Frystyk Nielsen, A. Karmarkar, and Y. Lafon, "SOAP Version 1.2 Part 1: Messaging Framework (Second Edition)," 2007, http://www.w3.org/TR/soap12-part1/ 04.12.2012.

[26] A. Rodriguez, "RESTful Web services: The basics," IBM, developerWorks, 2008, http://www.ibm.com/developerworks/webservices/library/ws-restful/ 04.12.2012.

[27] M. Ney, G. Kloss, and A. Schreiber, "Using Provenance to support Good Laboratory Practice in Grid Environments," in Data Provenance and Data Management in eScience, Springer, 2011.

[28] A. Szalay, "Extreme Data-Intensive Scientific Computing," in Computing in Science Engineering, vol.13 no.6, 2011, pp.34-41.

[29] J. Silbermann, S. Weinert, C. Wernicke, and M. Frohme, "Quality and information management in the laboratory," Logistics and Industrial Informatics (LINDI), 2011, pp. 93-98.

[30] K.T. Taylor, "Evolution of electronic laboratory notebooks," in Collaborative Computational Technologies for Biomedical Research, Wiley, 2011, pp. 303-320.