Enhancing the Provability in Digital Archives by Using a Verifiable Metadata Analysis Web Service

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

Paul Christopher Johannes
provet
University Kassel
Kassel, Germany
paul.johannes@uni-kassel.de

Abstract—**In a variety of research areas the requirements for good scientific practice demand a proper long-term archiving of the data produced and handled throughout scientific processes. While the documentation was traditionally written in paper-based laboratory notebooks, the increasing amount of digital data and corresponding metadata has led to the development of electronic laboratory notebooks (ELN). To ensure the integrity and authenticity of the data special mechanisms have to be established while being stored in the ELN and digital archives for several decades. As different research areas use individual scientific tools in their processes, this paper describes a generic data verification system to enhance the provability of data and the corresponding metadata. The system was implemented using a Web service that uses multiple ingress, verification and egress modules. By using the Web service presented in this paper, the provability of scientific data in digital archives is profoundly enhanced. By automatically evaluating the probative force of the data and metadata and adding the system's digital signature as a result, the provability can be ensured by a separate third party that is trusted on the side of the ELN operator as well as the scientific community and jurisdiction.**

Keywords- ELN; evidence; digital archive; digital signature; provenance

I. INTRODUCTION

The amount of data that is generated and processed in scientific processes has been continuously growing [1]. Major drivers behind this trend are large scientific experiments, e.g., the LHC at CERN, which produces an immense volume of large-scale data. On the other hand more and more scientific institutions and funding organizations have issued guidelines for safeguarding good scientific practice that recommend a proper long-term archiving of the scientific data that led to published results. Besides protecting the bit stream, these practices also require a certain amount of compliance regarding the scientific process to ensure the long-term comprehensibility of the data. In this paper we introduce a solution that is able to enhance the provability of data being ingested in a digital archive by evaluating the metadata and checking the consistency. Results of the evaluation are logged and a model to quantify the verifiability and provability of the data (regarding the integrity and authenticity within the scientific process) is being presented in this paper.

To allow the integration with different scientific processes, tools and especially electronic laboratory notebooks (ELN) the solution was developed as a modular Web service. Different digital archives can also be connected to the Web service using specific modules. State of the art digital signatures are used to verifiably sign the evaluation results and to protect the provability in the connected archives. Besides the automatic verification and evaluation of metadata, where appropriate and possible, the system also checks digital signatures and timestamps that were generated by the scientist or archivist before being sent to our system.

In Section II, we give an overview on data in scientific processes, from the generation and processing up to publication and archival forming a scientific data lifecycle. We also give references to related work and related projects that focus on scientific long-term archiving, scientific provenance and provability of scientific records. In Section III, we describe the model that we developed to measure the probative force of scientific data and measures that we use to protect the authenticity and integrity of data while being stored in digital archives. Modules and the applied mechanisms that measure the probative force are described in detail in Section IV. The result from the data verification performed by these modules is quantified, classified and signed using the techniques described in Section V. In Section VI, we conclude our findings and discuss their impact on the scientific process as a link to future research.

II. ELECTRONIC DOCUMENTATION IN THE RESEARCH PROCESS

Documenting research digitally in ELN is not only a fad or de rigueur but state of the art practice in many fields of research [2]. Slowly ELN will replace the traditional lab journal made out of paper [3]. Their implementation into the research process presents challenges to scientist, research facilities, universities and technical staff [4]. This holds especially true for its seamless implementation into the research process and scientific data lifecycle.

A. *Scientific Data Lifecycle*

An assessment of the research process in general is difficult since it is highly individual and differs from scientific branch to branch and researcher to researcher. The quantity of different ELN available on the market [5] shows how distinct the methods of different scientific branches are. Still the

ISBN: 978-1-61208-200-4

scientific process flow may be roughly divided into five phases. At the beginning the experiments are planned on the basis of theoretical considerations and their set parameters (design). On this basis experiments are performed and then evaluated (implementation and processing). In the course this, effects are discovered and results are documented and thus serve oneself and others, possibly by subsequent publication (analysis and publication). Finally, the gathered data and results have to be archived so that they can be used for later reuse or for presentation (archiving). In relation to the origin, use, storage and reuse of the processed data we speak of the scientific data lifecycle [6].

B. *Related Work*

The law concerning record keeping in scientific research processes is spread among many statues and is diverse. It changes from jurisdiction to jurisdiction. Specialized work into the provability of scientific records is rare. Related work usually refers to specific scientific fields. Charrow [7], for example, covers the complex laws and generally accepted procedure that relate to biomedical research in the USA. In this paper, we will discuss the legal provability of ELN in the most universal and abstract way possible.

Improving the secure long-term archival of digitally signed documents is still a challenge [8]. Solutions not only depend on the available technology but also on the law of the applicable jurisdiction [9]. To record all past versions of scientific data, versions of the database should be preserved in a continuous fashion [10]. We will expand on this research and apply its findings to stored scientific research data.

The provenance (also referred to as the audit trail, lineage, and pedigree) of electronic scientific data should contain information about the process and data used to derive it. It is documentation that is key to preserving the data, to determine its quality and authorship and to reproduce as well as validate the results [11]. These are all important parts of the scientific process and scientific metadata and can be circumstantial evidence to the authenticity and integrity of electronic research data. There are strong arguments to preserve metadata for legal evidence as a regular practice [12]. We seek to create a viable solution for (automatically) creating, archiving and utilizing metadata that is generated in the scientific data lifecycle.

The challenge is to record uniform and usable provenance metadata that meets regulatory needs while minimizing the modification burden on the scientist and the performance overhead on scientist and system [13]. Minimizing setup and maintenance costs by automating the database design, data load, and data transformation tasks helps to seamlessly integrate an ELN into the research process [14]. Current ELN offer little or no direct support for "scientist-oriented" queries of provenance information [15]. We address this with our automated weighting and classification model.

III. DESIGNING A SYSTEM TO IMPROVE THE PROBATIVE FORCE OF SCIENTIFIC DATA

Any system for the electronic documentation of the research process and its data should be designed with the goal to ensure the legal provability of its content. As opposed to the scientific provability of theories by means of experiments or empiric studies, which design and scope always depends on the subject matter and the branch of science and relies on peer review, the need for legal provability stems from the concern of the scientist to prove his results not only to other scientists but also to other people and institutions, e.g., in a court of law. Reasons for this are manifold [6]. For example, a scientist could be accused of scientific deception [16] or scientific fraud [17]. If the data he presented is questioned, he will want to show, that the data was not invented, alternated, falsified or parts of it suppressed, in order to exonerate his credibility and exculpate himself from any wrongdoing or even criminal liability.

A. *Digital Signatures and Timestamps*

To authenticate an author of electronic research data and evidence the integrity of the record, certain kinds of digital signatures can be used. These digital signatures (i.e., advanced electronic signatures) are data in electronic form, which are associated with other electronic data and serve as a method of authentication. In order to have probative value, they must be uniquely linked to the signatory, be capable of identifying him and be linked to the data to which it relates in such a manner that any subsequent change in the data is detectable. Many countries have adopted rules to the legality and evidentiary value of digital signatures (see [24] for a comprehensive but not complete list; see also [18]). Within the EU, for example, the rules have been harmonized by Directive 1999/93/EC of the European Parliament and of the Council of 13 December 1999 on a Community framework for electronic signatures. The rules therein had to be adopted by all EU Member States. They ensure that mutual legal recognition of qualified certificates and electronic signatures from third countries is applied if certain reliability conditions are met.

The Directive defines different classes of electronic signatures. The main provision of the Directive states that an advanced electronic signature based on a qualified certificate created by a secure-signature-creation device satisfies the legal requirements of a signature in relation to data in electronic form in the same manner as a handwritten signature satisfies those requirements in relation to paper-based data (for convenience this type of signature is usually called a "qualified signature"). It is also admissible as evidence in legal proceedings.

Qualified signatures are recognized by procedural law in many EU-Member states and, due to their origin from a known, trusted party, which is monitored by a respective EU-Member state, of high probative value in a court of law. Therefore the integrity of an electronic document signed with a verifiable qualified signature can be proven as well as its authenticity. In the context of research data, a scientist can evidence that his electronic records originate from him and have not been altered since the last time he has signed them with his qualified signature. But in addition to that, a scientist will want to show that his data has not been altered since a definite point in time. This could, e.g., be the date of record, the day of the experiment, and the date of archival.

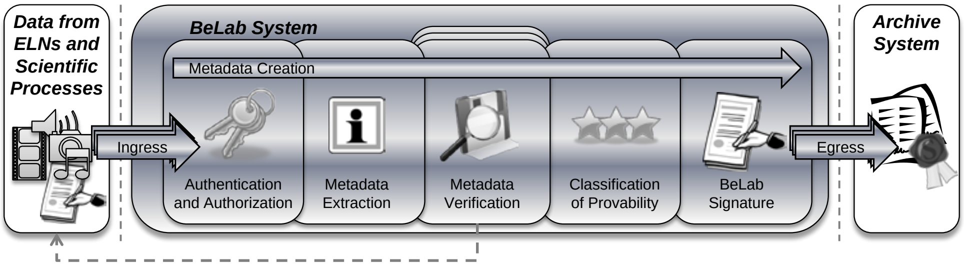

Figure 1. Modules of the BeLab Web Service

Later modification and falsification of electronic records can be ruled out by using trustworthy digital timestamps. According to the RFC 3161 standard, a trusted timestamp is a timestamp issued by a trusted third party (TTP) acting as a Time Stamping Authority (TSA). It is used to prove the existence of certain electronic data before a certain point without the possibility that the owner can backdate the timestamps. The newer ANSI X9.95 standard for trusted timestamps expands on RFC 3161 by adding data-level security requirements that can ensure data integrity against a reliable time source that is provable to any third party. Both standards can be used to create trustworthy timestamps that cannot be altered without detection and to sustain an evidentiary trail of authenticity. This holds especially true, if the TSA uses legally recognized digital signatures for his timestamps, like qualified signatures.

B. Evidentiary Value and Classification

A trail of authenticity, like a complete chain of evidence, is in the context of scientific research data a means to show that no scientific deception or fraud has been committed. To do so, the scientist wants to be able to prove that no data has been altered or falsified since recording it and no data has been suppressed. It is therefore advisable to ensure data integrity and authenticity as early as possible in the research process. The later the archival of research data, the greater is the possibility that the data has been tampered with before any kind of security measure like a digital signature or timestamp has been applied. Both, early application of digital signature and automation of the signing process are evidence for the integrity of the data [19].

But still: A general assessment of the evidentiary value of an electronic document is difficult to near impossible. The evidentiary value (i.e., probative force) always depends on the individual case: Who bears the burden of proof, i.e., which party needs to prove an assertion of fact, and whether the offered evidence is at all suitable to prove the fact [20]? In order to prepare for the eventuality of a legal dispute, the user of electronic records should assess their probative value. It is crucial to know how the authenticity and integrity of electronic documents can be proven and how this can be sustained [18]. Therefore it makes sense to give the user a comprehensible system for evaluating the authenticity, integrity and security of the dataset at hand. Such a system should classify the evidentiary value of data based on the data formats, the metadata collected and the kind of digital signatures used. Through the classification the scientist receives an indication of the degree of evidentiary value and potential risks of long-term archiving and preserving his research data or ELN.

C. Data and Metadata Model

As described in Section II, several different tools and ELN are typically used to manage data in scientific processes. Even in single scientific processes of a specific research area we found multiple ELN, and ordinary applications, e.g., Microsoft Office, being used together with custom software solutions, i.e., for data analysis. To support these data sources we implemented a Web service (called BeLab) that offers a generic input interface. Figure 1 depicts the modules and data processing to analyze and preserve the probative force of scientific data. By using HTTPS the confidentiality of the data transfer is protected.

After the authentication and authorization of the data transfer originating from the ELN (based on the user of the ELN or the ELN as a whole), metadata is extracted from the received data and stored in a container that is used during the subsequent verification and classification of the probative force. We chose an existing implementation that offers an extendible (XML-based) metadata container encoding & transmission mets standard. The ingress modules build up a TAR file of the data to be archived. By using a common standard the ELN can use a unified way to describe metadata information for the archived data.

The result of the execution of the verification modules, being stored in the model, is quantified in the classification module, which is explained in Section V. To protect the classification of the probative force the system signs the classification result using a digital signature. Finally the BeLab system offers a generic output interface, which supports different egress modules to digital archive systems that preserve the bit stream and the long-term interpretability of the data.

D. Archive systems

Multiple archive systems can be connected to the BeLab system by implementing specific egress modules. The implementation of the BeLab system also allows the combination of multiple archives, i.e., to ensure redundancy and fault tolerance across the archives. Typically these measures against bit stream errors or manipulation of the stored data are already addressed within the long-term archive system. An example for a standard that implements such a mechanism that uses digital signatures and digitally signed timestamps is described in RFC 4998 as long-term archive and notary service. Archive systems that are receiving data from the BeLab system can verify the digital signature of the BeLab system in their ingress interface or during a cyclic refresh of the long-term archived data. The egress modules of the BeLab system serialize the data being stored in the BeLab model as described in the previous section and transmit the data to the archive.

IV. AUTOMATIC WEIGHTING MODULES

As described in Section III.C, the probative force of the data being submitted to the BeLab system is processed and stored using a unified model inside the BeLab system. Ingress modules supply the data to be archived and associated metadata using a TAR file that includes a mets XML file containing the metadata. To enhance the quality of the automatic verification of the data and metadata carried out by the BeLab system, an additional metadata extraction is performed before the execution of the verification modules. This also allows the selection of the proper verification modules that support the specific data or metadata format or content. Furthermore by comparing the results of the metadata extraction to the supplied metadata an initial consistency check of the submitted data is performed.

A. Integrity by Checksums

The metadata contains unsigned hash value of each file in the archive file, as described in Section III. To ensure the integrity of the submitted data the checksum module calculates the hash value of each file again and compares it to the value that was received from the ingress module. In order to use the same algorithm to calculate the hash value the algorithm name is specified in the mets container.

B. Integrity and Authenticity by Digital Signatures

The checksum module as described in the previous section does not suffice to ensure the integrity and authenticity of electronic documents. As described in Section III.A, digital signatures, i.e., based on a X.509 certificate, are needed to ensure the authenticity of the document. Some computer programs, e.g., Microsoft Office, OpenOffice and Acrobat, offer the opportunity to sign the corresponding document. In these examples the signature is integrated in the file format. Other programs, e.g., Cryptonit, offer the possibility to save the signature in a separate file. In this way it is possible to sign electronic documents which do not offer the integration of signatures. Before archiving signed data, the electronic signature and corresponding certificates should be verified. Therefore the BeLab system implements specific modules which are designed to support different signature standards, i.e., PKCS#7 and XML-DSig.

1) Embedded Digital Signatures

The digital camera from the company Kappa optronics GmbH [25] which can be used to collect data samples in scientific processes, offers the opportunity of electronically signed images. Therefore, the image will be signed before the data will be transferred to the computer that the camera is attached to. Using the software included with the camera different file formats can be selected, e.g., jpg, gif or bmp. The signed images have a data format that was developed especially by Kappa optronics GmbH. The verification module that makes it possible to check the signature has to understand this specific format.

Hence a specific (Kappa) module was implemented for the BeLab system. The file format is based on three parts which contain the image data with header information and custom content, the signature and the public key to decrypt the signature. In the first step the Kappa module calculates the hash value based on defined data range. After the signature decryption both hash values will be compared to verify the signature.

In contrast, the integration of signatures in PDF is based on PKCS#7. So, a corresponding weighting module can use frameworks which are already developed such as Bouncy Castle Crypto API [23]. With this framework it is possible to verify digital signatures based on PKCS#7 which can be used in the PDF weighting module. This module also allows an automatic verification of the integrity of the data, validity of embedded certificates and optionally contained signed timestamp [21].

Additionally a module was implemented that allows users to verify XML-DSig signatures [21]. This kind of signature is used, for example, in the OpenOffice file format. To verify the signature the Bouncy Castle Crypto API was used. The OpenOffice file format is based on a zip archive file. Hence the archive needs to be unpacked to verify the signature. The OpenOffice module unpacks the archive with the given structure and verifies the signature. Even though the verification of signatures based on PKCS or XML standard can be handled in a uniform manner individual modules are needed.

2) External Electronic Signatures

External digital signatures are typically based on a uniform standard (PKCS#7). Therefore a single module was implemented to check for each file whether a corresponding signature file has been supplied. The signature file has to match the file name and must be placed in the same folder as the associated file [21]. If an external signature has been found the module starts the verification, as described in Section IV.B.1).

C. Sequence Detection and Workflow Verification

Digital cameras produce images that are usually named on the basis of a sequential number. These numbers can be used to automatically verify the completeness of a received series of images. One or more missing files can be interpreted in two ways: The owner might have forgotten to submit all files or he might have been trying to alter the series of images [21]. For whatever reason the files are missing, the proprietor of the ELN should be informed about it, as this might have an impact on the probative force of the data. Therefore the corresponding weighting module should take the sequence detection into account [19].

The implemented module works in two steps: First it checks whether or not a series exists. Therefore all numbers in the provided filename will be removed. After that the remaining part of the filename will be compared with the other results from previous filenames. If there are more than x identical designations a new series of files, e.g., images, is found. The value x can be set individual by the user. If a series of images is found the module checks the completeness in a second step. First, the lowest and highest number of the series has to be determined. Next the series can be analyzed whether or not an image is missing. All steps of the sequence detection will be noted in a log [21].

V. Classification Modules Based on Logging Mechanism

Not all implemented verification and weighting modules are suitable for each file format. The right modules must be selected based on the corresponding file type. Therefore a specific module, that is able to determine the data format, was implemented [21]. As described, there are dependencies between different modules, meaning that some modules need the result of another module and should be executed in the right order. The module for determining the file format should be executed first.

For each module a value (index) can be defined which indicates the order of modules. In addition, there are other conditions for selecting a module, e.g., dealing with file archives or separate files. These conditions can also be defined in the weighting module definition.

A. The Verification Result as Log

Due to the modular approach of the BeLab system, the result of a weighting module should be flexible [22]. Therefore, the result model of the verification was implemented as a log which can include different results [21]. Each entry consists of three values: key, content and type of content. The key element defines the validation being executed, the content element includes the result of this validation, and the type element declares the data type of the result. For each module a set of keys is defined. For example, the module for embedded signatures, as described in Section IV, defines the keys: “signature found”, “signature verified” and “signer”. In this case the type of “signature found” and “signature verified” is Boolean and the type of “signer” is String.

Each weighting module can produces any number of log entries. More than one module can analyze one file. The result is a list of logs which is stored as a tuple of filename and result. The complete verification and logging process is shown in Figure 2.

B. Log Analysis to Determine the Probative Force

The coordination of the verification process is performed by a validation controller. First, the controller chooses the right weighting modules for the evaluation of the whole archive, e.g., the sequence detection module, as presented in Section IV. Appropriate modules are chosen by the corresponding file format after that. When the validation process is finished, the coordination of the classification process is performed by a separate controller. This controller chooses the classification modules the same way as the validation controller does it with the weighting modules. This ensures that each result of the verification is taken into account in the classification process. The goal is to map the results of the verification regarding the degree of probative force, the suitability for long-term preservation and the degree of secure (i.e., consistent) data generation.

The probative force is primarily determined by the use of digital signatures and digital timestamps. For example, if the embedded signature module, as described in Section IV, has found a signature, the corresponding classification module would check its validity based on the result of the embedded signature module. If it is valid the module checks the specific kind of signature in a second step. The result is a value from B1 standing for “no signature” to B6 the highest qualification based on the qualified electronic signature, as described in Section III [22].

The suitability for long-term preservation is based on the file format. Because there is no unambiguous approach to define the qualification there are some roles which can be used. For example, the complexity of the file format, the usage of open standards in an index for the suitability of a data format. In this case the result of the classification is L1 (inappropriate), L2 (appropriate), L3 (recommended) and can be understood as a recommendation, meaning that the BeLab system does not require specific file formats to support individual formats used in scientific processes [22]. However, the data container described in Section III must be used.

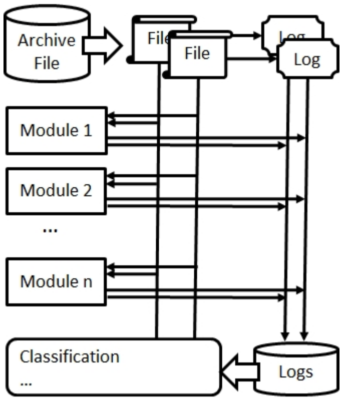

Figure 2. Verification process according to [21]

The degree of the secure data generation can be associated with the completeness of data [21], as shown in Section IV.D, or with the knowledge about the weighting modules used in the analysis phase. The completeness of data, for example, a series of digital pictures, indicates the integrity of this data and the scientific process. A stronger indication are electronic signatures which were used in the data generation phase [19]. For these signatures a corresponding weighting module, as shown in Section IV.B and classification module can be implemented. If the signature is valid, a secure data generation can be assumed. The BeLab system distinguishes the values S1 (insecure data generation) and S2 (secure data generation).

In fact, more than one module classifies each file resulting in different weighting modules being involved. Hence a mechanism to merge the classification results was needed. Initially, the final classification result is set to undefined (u). So, the classification for one file and the three categories, as described above, starts with a tuple <u, u, u>. If a module classifies a file, the result is another tuple, for example <u, B6, u>, meaning that the degree of probative force is set to the highest level of security and the other categories are undefined. These two results are merged by comparing each value of the three categories separately. If one of the values is undefined the defined value is preserved. In case both values are set, the lower value is taken. This way, a manipulation to a higher classification by a user is prevented.

All results of the weighting modules, the result of the classification process and eventually detected errors during the manufacturing process, are documented in a copy of the metadata. Finally, the metadata file and hence all included data (whether it was already signed or not) is signed by the BeLab system to ensure the integrity and traceability of the BeLab process. In addition, if multiple digital signatures for different documents were used, the verification is centralized. To do so, first the checksums have to be verified by the checksum module, as described in Section IV. Subsequently the signature of the metadata has to be verified, to ensure the integrity of submitted data.

VI. Conclusion and Future Work

By using ELN for the documentation of the scientific process, all data can be administrated centrally. Measures for the integrity and authenticity for paper-based laboratory notebooks cannot be transferred to the electronic documentation easily. By using the BeLab system, it is possible to ensure the integrity and authenticity of electronic data before the archive process. Using the developed weighting modules the needed data analysis can be executed automatically and without disturbing the scientist during his work. Additionally, the scientist gets useful information from the BeLab system during the archive process, i.e., about the suitability of the used file format. As the results from the evaluation of the BeLab system are stored in the attached digital archive, they can be used subsequently to prove the authenticity and integrity of the data. In Section IV.B.1), the example of a digital camera being used in scientific processes was given. One of the next requirements will be to support more measuring instruments and the validation of their data, e.g., using existing Web services.

References

[1] J. Gray, D.T. Liu, M.N. Santisteban, A. Szalay, D.J. DeWitt, and G. Heber, “Scientific data management in the coming decade,“ ACM SIGMOD Record, vol. 34, Dec. 2005, pp. 34-41, doi:10.1145/1107499.1107503.

[2] M. Kihlén, “Electronic lab notebooks – do they work in reality?,” DDT, vol. 10, Sep. 2005, pp. 1205-1207.

[3] M. Kihlén and M. Waligorski, “Electronic lab notebooks – a crossroad is passed,” DDT, vol. 8, Nov. 2003, pp. 1007-1009.

[4] D.J. Drake, “ELN implementation challenges,” DDT, vol. 12, Aug. 2007, pp. 647-649.

[5] 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.

[6] 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.

[7] R.P. Charrow, Law in the Laboratory. Chicago, IL: UCh. Press, 2010.

[8] C. Troncoso, D. De Cock, and B. Preneel, “Improving secure long-term archival of digitally signed documents,” Proc. 4th ACM international workshop on Storage security and survivability (StorageSS 08), ACM, 2008, pp. 27-36, doi:10.1145/1456469.1456476.

[9] W. Zimmer, T. Langkabel, and C. Hentrich, "ArchiSafe: Legally Compliant Electronic Storage," IT Professional, vol. 10, Jul.-Aug. 2008, pp. 26-33, doi:10.1109/MITP.2008.82.

[10] P. Buneman, S. Khanna, K. Tajima, and W. Tan, „Archiving scientific data,“ ACM TODS, vol. 29, Mar. 2004, pp. 2-42, doi:10.1145/974750.974752.

[11] S.B. Davidson and J. Freire, “Provenance and scientific workflows: challenges and opportunities,” Proc. ACM SIGMOD international conference on Management of data (SIGMOD 08). New York, NY:ACM, 2008, pp. 1345-1350, doi:10.1145/1376616.1376772.

[12] W.L. Wescott II, „The Increasing Importance of Metadata in Electronic Discovery,“, Rich.J.o.L.T-, vol. 14, Article 10, http://law.richmond.edu/jolt/v14i3/article10.pdf <retrieved: March 12, 2012>.

[13] B. Howe, K. Tanna, P. Turner, and D. Maier, “Emergent Semantics: Towards Self-Organizing Scientific Metadata,“ in Semantics of a Networked World, M. Bouzeghoub, C. Goble, V. Kashyap and S. Spaccapietra, Eds. Berlin, DE: Springer 2004, pp. 177-198, doi:10.1007/978-3-540-30145-5_11.

[14] S. Bowers, T. McPhillips, B. Ludäscher, S. Cohen, and S.B. Davidson, “A Model for User-Oriented Data Provenance in Pipelined Scientific Workflows,” in Provenance and Annotation of Data, L. Moreau and Ian Foster, Eds. Chicago, IL: Springer, 2006, pp. 133-147, doi:10.1007/11890850_15.

[15] Y.L. Simmhan, B. Plale, and D. Gannon, “A Framework for Collecting Provenance in Data-Centric Scientific Workflows,” in Proc. International Conference on Web Services (ICWS 06), IEEE Press, Sep. 2006, pp. 427-436, doi:10.1109/ICWS.2006.5.

[16] L. Grayson, Scientific Deception: an overview and guide to the literature of misconduct and fraud in scientific research. London, UK: British Library, 1990.

[17] H. Ottemann, Wissenschaftsbetrug und Strafrecht. Hamburg, DE: Dr. Kovač 2006.

[18] S. Mason (Ed.), International Electronic Evidence, London, UK: BIICL 2008.

[19] J. Potthoff, S. Rieger, P.C. Johannes, and M. Madiesh, “Elektronisch signierende Endgeräte im Forschungsprozess,” Proc. D-A-CH Security 2011, syssec, 2011, pp. 44-55.

[20] E. Schneider, Die Klage im Zivilprozess. Cologne, DE: Otto Schmidt 2007, pp. 477ff.

[21] F. Ellmer, “Automatische Metadatenanalyse zur beweiswerterhaltenden Langzeitarchivierung im Forschungsprozess,” Karlsruhe Institute of Technology, bachelor thesis.

[22] M. Madiesh, P.C. Johannes, and J. Potthoff, “Beweissichere elektronische Labor-, Patienten- und Fallakten,” Proc. perspegtive 2011, in press.

[23] Bouncy Castle, http://www.bouncycastle.org <retrieved: March 12, 2012>.

[24] eSignatureLegalWiki.org contributors, “Main Page,” eSignatureLegalWiki.org, http://www.esignaturelegalwiki.org <retrieved: March 12, 2012>

[25] Kappa optronics GmbH, http://www.kappa.de/ <retrieved: March 12, 2012>.