Looking Beyond Data Through the Ages

Dennis Wehrle*, Wiebke van Ekeris** and Uli Hahn***

*Albert-Ludwigs University Freiburg, Professorship in Communication Systems
**Albert-Ludwigs University Freiburg, University Library Freiburg
***Ulm University, Communication- and Information Center

Abstract: Accessing data has always been crucial in order to get access to (cultural) knowledge. But in the new ages different challenges arise to keep complex data accessible. Concerning these challenges a linguistic use case, which has been set up in 1993, serves as an example to show that it might be hard to re-enact an old system in order to reuse its research data and processes. Re-enacting the old system showed that other crucial points emerge in the context of research data management that need attention. In order to increase the probability of gaining access to research data in the future Data Management Plans can be identified as central tool. Although preservation of and creating access to data is a global task, the scientific management of data is still strongly shaped through national regulations and practices and hence delivers a quite fragmented picture.

1 Introduction

Access to data has always been crucial in order to get access to knowledge. Through cave painting one can learn a lot about the past. This was also practiced in the ancient world where data was stored in books and scrolls with the purpose to get access to cultural knowledge and to preserve it for the future. In science the accessibility of data is important, too, not only to ensure re-examination of publications. One of the most important points, is to facilitate re-usage of research data in order to prevent duplication of effort. The European Union, with their framework programme for research and innovation, or the DFG on a national basis in Germany are driving forces to encourage re-usage. But “experiments and numerical calculations can only be repeated if all important steps are reproducible. For this purpose, they must be recorded” [For13]. Thus the accessibility to research data needs to be ensured to foster new research. This also enables interdisciplinary research. One example for this interdisciplinarity is the climate database called *tambora*[1]. For a given time and place, one gets the weather condition when a correspondent entry in the map is given. This data is based on letters, ship diaries, newspaper articles, etc. Shared with other research areas, this data could be reused for other purposes, too. But (old) research data also have value to be reused for further research, as it is the case for our use case. The research project at hand was collecting local dialects of the south-western region of Germany, which were of interest to be reused for further research again. There are

[1] Tambora, `http://www.tambora.org` [last access 27.06.2014]

different approaches to granting accessibility to this data. The approach described in this paper uses the method of emulation in order to re-enact an outdated networked research environment. However this use case shows the problem which arises from the technical progress. Nowadays data is generated and stored mostly digitally. Compared to the past, the lifetime of systems and formats is changing fast. Additionally getting access to, and understanding of, data is more complex compared to books. Thus it is not sufficient to only ensure access to the data, because without having the tools to interpret the data, knowledge will be lost.

In order to see how knowledge, respectively data, is dependent on its access to it, first a historic vision of cases in the analog age is given in section two. With these examples in mind, a shift to the digital age shows in section three, which aspects of access to data changed from the analog to the digital age. Furthermore, the methods of research data management and Emulation-as-a-Service (EaaS) are introduced. The linguistic use case shows in section four, how a functional approach is used, to grant access to old research data. Section five finally takes an abstract stance on EaaS trying to typify this method within the field of research data management in general.

2 Data Management in the Past

When looking into the past the loss of data and cultural knowledge is omnipresent. In the following, two historical examples are given in order to emphasize the problem of the loss of data due to natural or man-made incidents: cave paintings from the Stone Age and the Library of Alexandria in the ancient world. One consequence of this loss is that sustainable accessibility of knowledge and data is not possible anymore.

Stone Age Cave paintings are the oldest preserved type of man-made data. On the basis of these paintings characteristics like the actual age, the fauna and flora at that time, as well as the use of tools can be studied. In Europe the most prominent places for cave paintings are southern France and northern Spain [LB00]. The cave paintings were formed in the Palaeolithic, around 40,000 BP to 12,000 BP.

However, it is unknown how many cave paintings are lost, destroyed or not yet discovered. As a consequence, the preserved cave paintings might not be representative. Another aspect is that the accessibility of cave paintings is, in general, limited. Because the original cave paintings cannot be stored in archives like scrolls and books, copies of the paintings (in form of photographs etc.) have to be produced and stored for the purpose of providing broad access to the data. In any case, information is lost by the process of copying, because certain information like the technique of scratching or the composition of color is not passed on.

Ancient World The Library of Alexandria was the most famous and largest library in the ancient world [Mac00]. More than 400,000 mixed and 90,000 single scrolls are said to have been stored in the main library, as well as about 40,000 books in the satellite-daughter library. Also the first subject catalogue of the world with six sections and around 120,000 scrolls of classical poetry and prose was build up there. The scrolls were stored in linen or leather jackets and were kept in racks in the hall or in the cloisters and lasted for at least 300 years until a great fire burnt most of the library down. Unfortunately, most of the scrolls and books were lost. The exact date and the reason for the fire remains unknown. Three incidents or a combination of them are probable, though: 48 BC during the fighting between Julius Caesar and Cleopatra, 272 AD after an attack of Aurelian, the emperor, or 391 AD, while Theophilus (the Christian patriarch of Alexandria) tried to turn one of the temples in the city into a church [Mac00].

However, the story of the Library of Alexandria shows that due to disasters – natural or man-made – the loss of data and cultural knowledge can happen at any time.

3 Data Management Today

Nowadays, in the age of digital data, natural catastrophes like fires are not the main threats with respect to the loss of knowledge and data, compared to the Stone Age and the Ancient World. The short lifetime of data storage media and formats, their diversity and complexity have to be considered as modern threats with respect to archiving. Another aspect is the permanent access to and the interpretability of data, which has become a new challenge in the digital age.

According to Neuroth et al. (2010) the lifetime of digital storage media is – with around 100 years – considerably shorter than the lifetime of ancient scrolls or cave paintings, which are still preserved since the Paleolithic and the ancient times [NOS$^+$10]. Not only the lifetime of storage media, but also the lifetime of data formats is limited, because data formats are in flux. In the analog age, different media were building around the data. For example, since the 20th century libraries and archives have increased the diversity of media to be stored: in addition to books and journals, microfilms, microfiches, disc records, audio and video tapes, and finally digital media were considered [PRSU11]. In the digital age, the digital medium becomes predominant and replaces the various analog media step-by-step. Additionally, the complexity of digital data has increased. For example, climate data and models cannot be described only by text. Therefore, the access to data has changed, because complete structures of these complex systems have to be represented. As a consequence, the access to knowledge has changed, too, because it is no longer possible or useful to copy all data into books. One functional approach to enable access to knowledge and data is emulation, which is described later on in this section and is the basis for our use case in section four. In order to provide permanent access to, and ensure interpretability of, digital data, the performance of research data management is of general interest, while in the case of emulation it is of explicit interest.

3.1 Research Data Management

The access to data has generally changed over times: In the ancient world, the traditional libraries were responsible for providing access to scrolls and books for their respective audience. Nowadays, their role has changed. The main task of libraries still is to provide access to books, articles, and other text-based publications. Additionally the fields of digital data, research data, and e-Science have become more and more important. These new fields come along with new scopes of duties like licencing, data management, long-term preservation of digital data, retrodigitalisation, and open access [BW11]. In order to cope with these new fields and to provide broad access to research data, a worldwide repository landscape has shaped. At the moment this landscape is still fragmented. There are many isolated solutions depending on the different institutions and the special fields of science, because the requirements with regard to technology, content, and archiving differ tremendously [BHM11].

Platforms like re3data[2] or Databib[3] resemble powerful search engines that bring an organizational structure into the scattered repository landscape. Their existence is proof enough that an individual overview over the repositories is not possible anymore. Apart from the infrastructural side, there is another aspect worth mentioning in connection with research data management. In their "Guideline on Research Data Management" (Leitfaden zum Forschungsdaten-Management), Jens Ludwig and Harry Enke describe the management of research data as measures which guarantee that digital research data are usable. What measures are necessary depends on the purpose of the use. Ludwig and Enke define four cases: (1) using the research data as a working copy for the scientific process, (2) re-using the research data for future research, (3) archiving documents for complying with the good scientific practice and (4) archiving the data for complying with legal or other non-scientific demands [LE13]. Research data management does not only imply the handling of data but also the process of data generation. Therefore both the scientist's side as well as the motivation to publish the generated data has to be considered. Most common may be the publishing of data due to regulations by external funding agencies, or because it is common practice in the scientist's respective research field. The priority initiative 'Digital Information' by the Alliance of German Science Organisations has postulated principles for the handling of research data. The principles are formulated very cautiously: While long-term preservation of, and the principle of open access to, data is supported by the alliance, at the same time this principle "shall be balanced against the scientific and legal interests of the researchers." [4] External funding organizations like the German Research Foundation (DFG), who are part of the Alliance of German Science Foundations, ask for compliance with the principles of good scientific practice, including the storage of and the access to all primary research data for at least ten years [For13]. Here, a divide between publishing and storing research data becomes obvious. While it is easy to formulate clear guidelines in order to maintain the quality of German research as such, it is not so easy to impose open-access-publishing for research data upon scientists. Although the Alliance

[2] Registry of Research Data Repositories, http://re3data.org [last access 27.06.2014]

[3] Databib, http://databib.org [last access 27.06.2014]

[4] Principles for the Handling of Research Data, http://www.allianzinitiative.de/en/core_activities/research_data/principles/ [last access 27.06.2014]

argues that publicly funded research needs to be made accessible publicly, the legal situation is far more complicated, especially concerning cooperations with private partners. Yet, in both cases – be it to follow the DFG's recommendations for safeguarding good scientific practice, or to follow the alliance's ideal of publishing research data as open access – data management plans should stand at the beginning of every research project that is collecting data. "A data management plan is a document outlining how the research data collected or generated will be handled during a research project, and after it is completed, describing what data will be collected / generated and following what methodology and standards, whether and how this data will be shared and / or made open, and how it will be curated and preserved ..." [Com13a] Data management plans are already common practice in the US or the UK with whole institutions forming around the idea of preserving digital data.[5] With data management planning tools like DMPonline from the DCC[6], where scientists get help with formulating their data management plans according to their national and their funder's demands, the discrepancies between the UK and the US on the one side, and countries like Germany on the other side become apparent. In Germany such a DMP-tool that helps scientists in the funding stage of their project to create a data management plan only is offered at Bielefeld University. Of course, there is also a lot of activity when it comes to advocating the necessity of caring about research data, for example the NESTOR-initiative taking care of long term preservation and the above mentioned DFG formulating guidelines for scientists and various projects like the WissGrid initiative seek to improve the situation in Germany. In the UK the Digital Curation Center (DCC) puts significant effort into the sensitization of scientists but they also give lots of practical information that helps to improve the scientists' workflows, something that in Germany is not practiced to this extend. With the advent of Horizon 2020, data management plans might get more popular in Germany as well. Research projects that participated in the Open Research Data Pilot already contained data management plans. The optional article 29.3 of the Model Grant Agreement in Horizon 2020 now provides the legal requirements for such projects to integrate data management plans in the new framework program. Article 29.3 consists of two parts: Part a) is addressing the compulsory tagging of the research data with metadata, necessary for open access or long-term publishing in a repository. Part b) describes what is needed apart from the metadata in order to maintain the chance of re-using the concerned data sets: "Regarding the digital research data generated in the action ('data'), the beneficiaries must [...] provide information – via the repository – about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and – where possible – provide the tools and instruments themselves)."[Com13b] As shown in section five, a data management plan does not infringe the author's copyright. Yet, tagging the data, and giving a full subset of the used programs, hardware and software, will remain the responsibility of the scientist. By demanding data management plans from the scientists, a whole subset of positive effects could be achieved. In order to render long term preservation a valuable action, data has to be kept interpretable and tools to access the data should not be forgotten. How important this is can especially be seen when looking closely at preservation (or re-enacting) methods like Emulation-as-a-Service.

[5]RDM guidance webpages, `http://www.dcc.ac.uk/resources/policy-and-legal/rdm-guidance-webpages/rdm-guidance-webpages` [last access 27.06.2014]

[6]DMPonline, `https://dmponline.dcc.ac.uk/` [last access 27.06.2014]

3.2 Emulation-as-a-Service (EaaS)

Emulation is becoming a key strategy for digital preservation and securing access to digital artifacts. By digitally recreating computer systems, emulation ensures that digital objects can be rendered in their native environments and thus maintain their original ”look and feel.” In many cases it is only feasible to use the original environment in order to reuse research data and its processes, because a migration to a newer system might not be successful. Nevertheless, deploying operational emulation setups is a complex and laborious task requiring specific technical expertise and knowledge of how to maintain obsolete computer systems. Thus up to now emulation has been perceived as a domain reserved for technical experts only. In order to address both technical and organizational challenges, a scalable emulation-based service model has been developed: Emulation-as-a-Service (EaaS). EaaS is focusing on a functional approach. Accordingly the goal is to make digital artifacts sustainably accessible, which in turn means that the original system may be (slightly) changed without loosing the correctness and completeness of the digital artifact.

The fundamental building blocks of the EaaS architecture are abstract emulation components to standardize deployment and to hide individual emulator complexity. Each emulation component encapsulates a specific emulator type as an abstract component with a unified set of software interfaces for data I/O and technical as well as user-machine interaction. This way, different classes of emulators become interoperable. This means that various emulation technologies like qemu[7] and different virtualization technologies, for instance Oracles VirtualBox[8], could be used simultaneously. Currently, emulation components for all major past and present desktop systems, e.g. PPC, Sparc, M68k, Intel-based x86 and major operating systems, e.g. OS/2, MS Windows, Mac OS 7 and Linux are available for deployment. A detailed technical description of the EaaS framework and its workflows can be found in earlier work [RVvL12]. With EaaS it is not only possible to preserve a single system but to preserve a complete networked environment, too. Thus, preservation of a scientific desktop including its research data and processes which have dependencies on other machines, is feasible. After the system is successfully preserved, end users have access to the preserved system remotely through standard (web-)clients (e.g. web browsers).

Although emulation is certainly a key strategy in digital preservation still it does not seem to have found its place in the field of research data management. The “Handbook on Research Data Management” [BHM11], for example, does not list emulation per se; nonetheless emulation is hidden behind the scenes. When looking at the Curation-Lifecycle-Model[9] developed in 2010 by the DCC, it gets obvious that EaaS also fits perfectly into this lifecycle. After appraising and selecting what to preserve and ingest, which could be summarized as preparation step of EaaS, the emulation itself takes place covering the steps of preservation action and storage, ultimately providing for access, use, and reuse of the research data.

[7] QEMU open source processor emulator, `http://www.qemu.org` [last access 27.06.2014]]

[8] VirtualBox, `http://www.virtualbox.org` [last access 27.06.2014]

[9] DCC Curation Lifecycle Model, `http://www.dcc.ac.uk/resources/curation-lifecycle-model` [last access 27.06.2014]

4 Use Case: Linguistics

In the linguistics department at the University of Freiburg, a long running research project got recently shut down. Since the 1970s, the project had been collecting local dialects of the south-western region of Germany.[10] Later on it became one of the first projects in this field of the humanities to move from laborious paper based evaluation to a computerized environment. All collected data got added to a database. The data then was used to create customized dialect maps depending on various input parameters which got published in numerous publications and theses. Several specific workflows and even custom font and symbol shapes were created to produce Postscript output from a TeX file source. Many researchers and PhD students added to the project and refined the workflows over time. Unfortunately, no actual user of the system has a full understanding of how the system is set up and configured. As the contained research data has value to be reused for new research and cannot easily be extracted from the system, we took the system as a use case for preserving full scientific original environments.

In order to be able to reuse this research data, it is not only vital to preserve this data but to preserve the complete system, too, as it provides different procedures to process the data. With the EaaS approach, it is feasible to preserve not only a single system but a complete environment (e.g. a Client-Server infrastructure). In order to preserve a system an analysis and a preparation step, ideally carried out while the machine is still operational, is required. This includes an assessment of hardware, software (e.g. operating system, memory, network configuration, etc.) as well as its configuration to be able to migrate the machine to an emulated environment. In a second step, an imaging process of the physical hard disk is required [WRCv12]. Furthermore, to ensure proper re-enactment of the old system, it is not only vital to know explicit hardware requirements but also to determine explicit and implicit dependencies and their functional expectations. Such dependencies and expectations could be, for instance, a system with a hard coded address to connect to a database system within the local network, or, like in this use case, the system could expect network drives which will be initialized at startup. Thus, these network resources have to be available. Finally, the resulting re-enactment must be checked with regard to whether the expectations are met and the system is functionally working.

4.1 System Analysis

The system got set up in 1993 and consisted of one server machine running on OS/2 driving an IBM DB2 database and five client machines, offering access to the database over LAN. By the end of the project the server was fully functional and at least three clients were completely, two more partially, working (two to three hard disks were more or less unreadable, in another machine the floppy drive was failing or the system was not booting because of mainboard failure). The LAN infrastructure was running both TCP/IPv4 and NetBUI on a physical TokenRing infrastructure. All machines were x86

[10]The project was relevant enough to get added to the permanent exhibit of the Freiburg University Museum, `http://www.uniseum.de` [last access 27.06.2014]

hardware, featuring 486DX2 for the clients and a Pentium/Overdrive CPU (upgraded from 486) in the server. The machines were equipped with 8 to 16 MByte of RAM. All machines were connected to 15" CRT monitors powered by CirrusLogic VGA graphic adaptors with 1 MByte of RAM. The server hard disk was a SCSI disk of 1.1 GByte, the client hard disks of 240 MByte were using IDE adaptors. All clients were identically configured and the installation on the clients were originally identical. The system description illustrates how the technical configuration has changed over the last 20 years compared to today's computers. On basis of the derived information, a proper virtual machine was configured. As a next step to re-enact the system, images of the hard disks have to be created.

4.2 Challenges and Issues to Re-enact the System

There are several methods available to create an image of a hard disk: (1) in-system, non-intrusive method, which means to boot a networked mini linux operating system on the original system (preservation target) or (2) an intrusive method, which involves removing the hard disk for connecting and dumping it to a suitable imaging system. Such a dumping system has to provide appropriate connectors (e.g. SCSI or IDE), either internally (e.g. via mainboard) or externally (like USB-connectors). Because there was no suitable mini linux and no suitable network connection available, the intrusive method was applied. Therefore, the machines were disassembled to remove the drives. In order to create an image of the SCSI server hard disk and the IDE client hard disk, proper connectors had to be found. Both connectors are outdated today and are not used on new machines anymore. A working controller was found for SCSI, whereas different controllers were used for IDE. Interestingly the IDE system imaging procedure was more difficult than imaging the SCSI disk, because the logical layer changed over time. Therefore the produced disk image was not correct or the original disk was not even recognized. Finally, a controller could be found with which a correct image of the disk could be produced. Despite the imaging problem, there were other problems when starting the system for the first time. Because of the use of a virtualized environment the machines' expectations of the system were not fulfilled. This means that on a virtualized environment there are not exactly the same hardware components (e.g. graphic card, network card, etc.). Accordingly, we had to solve these issues too, by changing configuration parameters and installing new drivers manually.

4.3 Result

After solving the aforementioned problems, the two systems could finally be preserved and re-enacted. This includes the server machine which hosts the database for the linguistic data and a client machine which queries the database and processes the data (e.g. create a custom map). Despite the fact that the system could be preserved, knowledge is still lost (maybe just temporarily) as the location and the knowledge of how to use various tools is unknown. One possibility to process the data is to use a tool named `ListOASRV.cmd`, which produces a custom map. But the application of this command is not intuitive as several steps have to be done, for example a key combination of ALT+Z, in order to produce the final result. Without the knowledge of how to use the tools, one might not be able to produce the desired result.

The use case clearly shows that it is possible to re-enact an old system for re-usage of research data. But it also illustrates that it might be hard to re-enact old systems properly as knowledge of the respective system is needed. Additionally, the IDE-Controller problem demonstrates that also out-dated hardware might be necessary to properly migrate the system.

5 Taking Care of the Future: The Necessity of Data Management Plans

This paper started out by stating that the loss of knowledge is – and has always been – omnipresent, be it in the form of paintings, writings or data. While in the analog age the loss of data could be linked to either man-made or natural causes, these boundaries are blurred in the digital age. Digital data, as well as its analog pendant, certainly suffer from natural decay. But, as the use case demonstrates, the decay of digital data may also have (unintended) man-made causes. Scientists can certainly not consider the various problems with which one is confronted by the preservation process of a system, which resembles a blindness of what is to come. This blindness calls for certain preventive steps when handling digital data that were started to be articulated with the practice of writing data management plans. The German Research Foundation introduced such data management plans in their 2010 proposal preparation instructions for project proposals, yet made those plans only mandatory for projects where "research data or information will be systematically produced."[11] The central question to such plans is, how scientific data can be kept accessible and, if the data is not accessible anymore, how accessibility to the scientifically relevant data can be achieved again. It becomes apparent that science cannot be reduced to text or data anymore, but that for achieving full accessibility (e.g. to a single system, but also to whole research environments like in the demonstrated use case) the tools that are used to make the data interpretable proof equally important.

As the use case shows, Emulation-as-a-Service can definitely serve as a solution to make old research data reusable again. At the same time, however, EaaS will certainly not be a solution for every outdated system. Although EaaS contains a certain "out-of-the-box"-component that results in a highly automated workflow, this only becomes possible if enough information about the system is available. In order for EaaS to work fluidly, data management plans that provide all relevant data in order to maintain the system are certainly helpful. Missing information on software, hardware, and data are the main reason, why re-enacting old systems is harder, than it will be with systems for which data management plans have been written. It may also be helpful to use (centrally) managed

[11] Leitfaden für die Antragstellung, http://www.dfg.de/formulare/54_01/54_01_de.pdf [last access 27.06.2014]

(virtual) research environments, where a standardized or partly standardized system will be used within a working group. If such environments are planed before or at the start of the project, it will be easier to migrate the system later or even just move the virtual environment into EaaS.

As EaaS needs a preparation step for the emulation to work properly, the first step to emulation will always be a decision making process: Is the re-enactment of the scientific data so relevant that the effort is paying off? The question which scientific data should be kept and if data is kept, for how long, is raised in other parts of the scientific domain as well. But while the preservation of scientific data seems to be fairly simple, the maintaining of its accessibility is much more complex. Repositories for scientific data are confronted with the same central questions as EaaS. Yet, while repositories can pertain a policy of "store first, ask later", EaaS works the other way round. It is essential to first consider the effort of re-enacting a system before taking action. This necessity proves that keeping data accessible does not only involve technical aspects but also entails a management aspect. By providing data management plans before or during scientific projects, EaaS can be implemented easier than stated in the use case above.

6 Conclusion

In this paper, the necessity of research data management was outlined by analyzing the ancient and modern threats with respect to the loss of cultural knowledge and data as well as the permanent access to data. Research data management is currently converted into a prerequisite for the transfer of benefits through the stipulation of data management plans by funding agencies like the DFG.

When looking into the future, an interconnection between (specialized) repositories and data centers combined with integrated EaaS becomes thinkable. Then a customized data curation, including permanent access and a sustainable long-term preservation, would be achieved. Technical and administrative interfaces are needed in order to setup such a repository network, e.g. OAI[12], service level agreements, and standardized workflows. Furthermore, local contact points for research data management and specialized IT infrastructure are required.

The project "Coordinated Structures for Verification and Efficient Reusability of Research Data Throughout the State of Baden-Württemberg" has the goal to realize the scenario mentioned above and to serve as a model for other institutions in the State of Baden-Württemberg. It is currently executed by the University of Freiburg's IT-Center and University Library, as well as by the Ulm Communication- and Information Center.

[12]Open Archives Initiative, `http://www.openarchives.org/` [last access 27.06.2014]

Acknowledgments

The work presented in this publication is part of the project "Coordinated Structures for Verification and Efficient Reusability of Research Data Throughout the State of Baden-Württemberg" ("Landesweit koordinierte Strukturen für Nachweis und effiziente Nachnutzung von Forschungsdaten") sponsored by the Ministry for Science, Research and Arts of the federal state of Baden-Württemberg, Germany.

References

[BHM11] Stephan Büttner, Hans-Christoph Hobohm, and Lars Müller. *Handbuch Forschungsdatenmanagement*. Bock+ Herchen, 2011.

[BW11] Sabine Brünger-Weilandt. *Gesamtkonzept für die Informationsinfrastruktur in Deutschland: Empfehlungen der Kommission Zukunft der Informationsinfrastruktur im Auftrag der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder*. Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz e. V., 2011.

[Com13a] European Commission. Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, 2013.

[Com13b] European Commission. Multi-beneficiary General Model Grant Agreement, 2013.

[For13] Deutsche Forschungsgemeinschaft. *Proposals for Safeguarding Good Scientific Practice*. Wiley-VCH, 2013.

[LB00] Michel Lorblanchet and Gerhard Bosinski. *Höhlenmalerei*. Thorbecke, 2000.

[LE13] Jens Ludwig and Harry Enke. *Leitfaden zum Forschungsdaten-Management: Handreichungen aus dem WissGrid-Projekt*. Verlag Werner Hülsbusch, 2013.

[Mac00] Roy MacLeod. *The Library of Alexandria: Centre of Learning in the Ancient World*. IB Tauris, 2000.

[NOS+10] Heike Neuroth, Achim Oßwald, Regine Scheffel, Stefan Strathmann, and Karsten Huth. *Nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. Verlag Werner Hülsbusch, 2010.

[PRSU11] Engelbert Plassmann, Hermann Rösch, Jürgen Seefeldt, and Konrad Umlauf. *Bibliotheken und Informationsgesellschaft in Deutschland: Eine Einführung*. Harrassowitz, 2011.

[RVvL12] Klaus Rechert, Isgandar Valizada, Dirk von Suchodoletz, and Johann Latocha. bwFLA – A Functional Approach to Digital Preservation. *PIK – Praxis der Informationsverarbeitung und Kommunikation*, 35(4):259–267, 2012.

[WRCv12] Ian Welch, Niklas Rehfeld, Euan Cochrane, and Dirk von Suchodoletz. A Practical Approach to System Preservation Workflows. *PIK – Praxis der Informationsverarbeitung und Kommunikation*, 35(4):269–280, 2012.