

## Langzeitarchivierung in der Praxis – ein nestor/DigCurV-School-Event

*Simone Görl, Martina Kerzel, Kerstin Schröder*

Die Notwendigkeit der dauerhaften Bewahrung digitaler Dokumente steht seit vielen Jahren im Fokus von Gedächtnisorganisationen und wissenschaftlichen Einrichtungen. Nachdem in den letzten Jahren einige theoretische Grundlagen geschaffen und verschiedene Tools entwickelt wurden, treten nun vermehrt Fragestellungen zur konkreten Umsetzung der digitalen Langzeitarchivierung (LZA) in den Vordergrund.

Die notwendigen Strategien müssen, ausgerichtet auf die jeweilige digitale Kollektion, die langfristige Verfügbarkeit sowie die Nachnutzung der archivierten Ressourcen sicherstellen und dabei zukünftige Nutzungsszenarien berücksichtigen. Eine Schwierigkeit hierbei liegt im stetigen Wandel des Informationsmarktes. Dies bedeutet u.a. für die Praxis die Notwendigkeit, die Nachnutzung und Interpretierbarkeit der digitalen Objekte sicherzustellen. Dies wird bspw. durch wiederkehrende Emulations- und Migrationsszenarien sichergestellt.<sup>1</sup>

Der Aufgabe auf dem Gebiet der digitalen Langzeitarchivierung notwendige Weiter- und Fortbildungen zu entwickeln bzw. zu organisieren, haben sich u.a. das Kompetenznetzwerk nestor sowie DigCurV (Digital Curator Vocational Education Europe) verschrieben. Der nestor-MoU-Kreis „Qualifizierung“ (Memorandum of Understanding)<sup>2</sup> setzt sich für den Ausbau der curricularen Berücksichtigung des Themas Langzeitarchivierung und für die Entwicklung von E-Tutorials ein, richtet

Schools und Seminare aus und beteiligt sich an Konferenzen und Workshops. Ein primäres Ziel der Sommer und Winter Schools ist die Vermittlung wichtiger Themen und Aspekte der digitalen Langzeitarchivierung. Daneben dienen sie ebenfalls der Netzbildung und dem Austausch von Erfahrungen und Kenntnissen unter den Teilnehmern. Die Organisation und Durchführung der Winter School 2011 ist durch das EU-Projekt DigCurV unterstützt worden. Dieses internationale Projekt widmet sich der Entwicklung eines Rahmencurriculums zur digitalen Langzeitarchivierung. Dies soll die Fort- und Weiterbildung für Mitarbeiter aus Bibliotheken, Archiven und Museen im Bereich des langfristigen Managements digitaler Sammlungen verbessern.<sup>3</sup>

Die Winter School 2011 von nestor und DigCurV fand, wie schon zwei Veranstaltungen im Jahre 2007, in Gernrode im Harz statt. Vom 21. bis zum 23. November 2011 lernten und diskutierten ca. 35 Teilnehmer in den Seminarräumen des Cyriakusheims.



©Fanny Wirsing

<sup>1</sup> Liegmann/Neuroth: „Einführung“.In: *nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*, Version 2.0, 2009, hrsg.v. H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, K. Huth, S. 1:1-1:10. Hier S.1:2.

<sup>2</sup> <http://nestor.sub.uni-goettingen.de/education/index.php> (11.02.2012).

<sup>3</sup> <http://www.digcur-education.org/> (03.02.2012).

der Forschung und der hochschulischen Ausbildung geladen. Unter den Teilnehmern befanden sich Mitarbeiter aus Archiven, Museen und Bibliotheken mit teilweise umfangreichen Erfahrungen auf dem Gebiet der digitalen LZA. Weiterhin nutzten zahlreiche Studierende aus unterschiedlichen Studiengängen der Memorandum-Partner dieses Qualifizierungsangebot.<sup>4</sup> Ihr Teilnahmeentgelt konnte dank der Unterstützung durch Sponsoren<sup>5</sup> stark reduziert werden. Außerdem erhielt jeder Teilnehmer ein Exemplar des im Juni 2009 erschienenen *nestor Handbuches*.<sup>6</sup> Organisiert und moderiert wurde die Veranstaltung von der SUB Göttingen in Zusammenarbeit mit der FH Köln, unterstützt von der HTW Berlin und der HdM Stuttgart.

Der Titel der Winter School lautete „Langzeitarchivierung in der Praxis“. Auch dieses Mal wurde, wie bereits 2010, auf eine Einführung in das Thema verzichtet und stattdessen die Schwerpunkte auf das praktische Vorgehen in der LZA und auf die technische Umsetzung gelegt. In insgesamt vier Lektionen mit anschließenden praktischen Übungen und einem abschließenden Vortrag wurden praxisnahe Themen behandelt und diskutiert.

Das Konzept der Veranstaltung sah eine theoretische Einführung durch einen Fachexperten vor, dem ein angeleiteter Praxisteil folgte. Zu diesem Zweck wurden die Teilnehmer in Gruppen aufgeteilt, zu denen jeweils auch

---

4 Entsprechend den MoU-Vereinbarungen konnte manchen studentischen Teilnehmern die Teilnahme an der School in Verbindung mit anderen Aktivitäten als Studienleistung anerkannt werden.

5 Vgl. [http://nestor.sub.uni-goettingen.de/winter\\_school\\_2011/index.php](http://nestor.sub.uni-goettingen.de/winter_school_2011/index.php) (11.2.2012)

6 *Nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*, Version 2.0, hrsg. v. H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, K. Huth. Online-Version unter: <http://nestor.sub.uni-goettingen.de/handbuch/> (12.02.2012).

mindestens ein Fachexperte gehörte, der die Gruppe im Bedarfsfall unterstützen konnte. Die während dieser Gruppenarbeitsphase erreichten Ergebnisse wurden im Anschluss jeweils dem gesamten Teilnehmerkreis vorgestellt und diskutiert. Den Abschluss bildete eine Zusammenfassung der Ergebnisse durch den jeweiligen Dozenten.

Da in der Praxis die Planung der Übernahme von Objekten zu Projektbeginn einer der wichtigsten Schritte ist, so startete auch die Winter School mit diesem Thema. „Wege ins Archiv – Wie plant man die Objektübernahme?“ war folglich der Titel des Vortrages von Jens Ludwig aus der Niedersächsischen Staats- und Universitätsbibliothek (SUB) Göttingen. Anhand des von der AG Standard



©Fanny Wirsing

des Kompetenznetzwerkes *nestor* erarbeiteten Leitfadens wurden die einzelnen Schritte vorgestellt. Der Prozess besteht aus drei Hauptstufen: der Definition des zu archivierenden Objektes und seiner Beschreibung mittels Metadaten, dem Prozess, der die Bildung von Transferpaketen, die Prüfung der Validität sowie den Transfer ins System umfasst, sowie dem Management, also der Klärung vertraglicher und rechtlicher Aspekte der Informationsübernahme, der Finanzierung und der Qualitätskontrolle etc.

Im Anschluss wurde der Leitfaden mit der WissGrid-Checkliste, einem Planungsinstrument für das Management von Forschungsdaten verglichen. Mit seiner Hilfe können alle Aspekte von der Auswahl und Bewertung der

Objekte über die Infrastruktur des Archivierungssystems bis zum Zugriff durch Nutzer im Vorfeld erarbeitet werden. Die WissGrid-Checkliste stellt somit ein Werkzeug zur Erstellung eines Datenmanagementplans dar. Im Gegensatz dazu wurde der nestor-Leitfaden nur für den Teilaspekt der Übernahme von Objekten entwickelt. Der vorgenommene Vergleich zeigt widererwarten, dass der nestor-Leitfaden viele Themen aus dem kompletten Life Cycle enthält. Dies verwundert jedoch nicht mehr, wenn man bedenkt, dass im Rahmen der Übernahme der Objektewichtige Entscheidungen getroffen werden müssen. Zahlreiche Aspekte, die im Prozess der LZA zu einem späteren Zeitpunkt zum Tragen kommen, müssen vor dem tatsächlichen Transfer der Daten in das System geklärt werden.

In der anschließenden Übung sollten Übergabevereinbarungen zwischen dem Produzenten



der Daten und dem Betreiber des Langzeitarchivs in Rollenspielen ausgehandelt werden. Hier zeigte sich eines der Hauptprobleme von Verhandlungen. Dies ist der unterschiedliche Gebrauch von Begriffen in

der Fachwelt und im allgemeinen Sprachgebrauch. Die Teilnehmer konnten erfahren, wie schwierig die Gespräche sind, da die mitgebrachten Kenntnisse im Bereich der LZA teilweise sehr unterschiedlich waren. Diese Szenarien waren auch deshalb sehr realitätsnah, da solche Verhandlungen in Zukunft auf zahlreiche Bibliotheken, universitäre Rechenzentren oder auch IT-Abteilungen von Zeitschriftenverlagen etc. zukommen werden.

Der zweite Veranstaltungstag widmete sich in theoretischer und praktischer Hinsicht einerseits den Werkzeugen, die für eine digitale Langzeitarchivierung zur Verfügung stehen

und andererseits dem Thema

„Langzeitarchivierung von E-Mails“.

Stefan Funk aus der SUB Göttingen stellte die



©Fanny Wirsing

LZA-Werkzeuge koLibRI (kopal Library for Retrieval and Ingest)<sup>7</sup> und FITS (File Information Tool Set)<sup>8</sup> vor, mit deren Hilfe verschiedene technische Schritte innerhalb eines LZA-Workflows, z.B. während der Interaktion mit einem Archivsystem, bewältigt werden können. Funk beschäftigt sich seit 2001 mit den verschiedenen technischen Aspekten der LZA innerhalb von Projekten wie etwa kopal<sup>9</sup> oder DP4lib (Digital Preservation for Libraries)<sup>10</sup> und kann daher auf umfangreiche Erfahrungen im Thema zurückgreifen.

Nach der Erläuterung und Veranschaulichung, welche Schritte zu einer technischen Realisierung der Langzeitarchivierung eines digitalen Objekts gehören, bevor ein Datenpaket physisch in ein Archivsystem eingespeist werden kann (Ingest-Policy<sup>11</sup>), folgte der Selbstversuch in angeleiteten Gruppen. Die Aufgabe bestand darin, mehrere vorgefertigte Dateiodner mit unterschiedlichem Dateninhalt mit Hilfe der soeben vorgestellten Werkzeuge zu verpacken

7 [http://dp4lib.langzeitarchivierung.de/index\\_koLibRI.php.de](http://dp4lib.langzeitarchivierung.de/index_koLibRI.php.de) (03.02.2012).

8 <http://code.google.com/p/fits/> (03.02.2012).

9 <http://kopal.langzeitarchivierung.de/index.php.de> (03.02.2012).

10 <http://dp4lib.langzeitarchivierung.de/> (03.02.2012).

11 Die Ingest-Policy von DP4Lib umfasst z.B. neben einer Beschreibung von Schnittstellen auch die Beschreibung der Transferpakete des Daten-Lieferanten. Vgl. dazu

[http://dp4lib.langzeitarchivierung.de/downloads/DP4lib-Ingest-Spezifikation\\_v2.0.pdf](http://dp4lib.langzeitarchivierung.de/downloads/DP4lib-Ingest-Spezifikation_v2.0.pdf) (03.02.2012).

(koLibRI) und zu validieren (JHOVE<sup>12</sup> bzw. FITS). Eine der Herausforderungen stellte dabei die Arbeit mit der Kommandozeile dar, die einigen z.B. noch aus DOS-Zeiten bekannt, anderen Teilnehmern aber nicht geläufig war. Das Packen und anschließende Untersuchen der SIPs (Submission Information Package)<sup>13</sup> führte zu interessanten Ergebnissen: offenbar waren einige der Dateien nicht für die Langzeitarchivierung geeignet, da sie nur wenige oder nur wenig aussagekräftige Metadaten besaßen. Dies konnte bereits bei der Erstellung eines SIPs mit verschiedenen TIFF-Dateien festgestellt werden. Die für die Verpackung geeigneten Dateien wurden dann in einem zweiten Schritt mit Hilfe des JHOVE-Tools untersucht. Bei genauer Betrachtung der ausgegebenen Log-Files konnten vielfältige Unstimmigkeiten festgestellt werden. So muss man bei den Voreinstellungen das passende Modul auswählen (z.B. das Modul, das zur Prüfung von PDF-Dateien eingesetzt werden kann), weil sonst die Prüfergebnisse nicht valide sind und eine Wohlgeformtheit nicht festgestellt werden kann, da der Bezugsrahmen für Wohlgeformtheit fehlt. Dieses Erfolgserlebnis, das in den vier Gruppen in unterschiedlicher Ausprägung stattfand, war der ausgezeichneten Vorbereitung der Veranstalter zu verdanken und ihrem Gespür für die Balance zwischen schrittgenauer Anleitung und Raum für eigene Erfahrungen der Teilnehmer.

Der Nachmittag war der LZA von E-Mails gewidmet. Karin Schwarz von der Fachhochschule Potsdam führte in das Thema ein und behandelte dabei Fragen wie etwa rechtliche Rahmenbedingungen, insbesondere im Verwaltungssektor privatwirtschaftlicher und

---

12 <http://hul.harvard.edu/jhove> (06.02.2012).

13 Vgl. N. Brübach: „Das Referenzmodell OAIS“. In: *nestorHandbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*, Version 2.3, 2010, hrsg. v. H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, K. Huth, S. 4:3-4:14. Hier S.4:7.



öffentlicher Einrichtungen, den Zusammenhang zwischen Problemen bei E-Mail-Management und E-Mail-Archivierung, Metadaten, Speicher- und Aussonderungsstrategien. Die erste problematische Hürde beginnt demnach unmittelbar beim Versand der Nachricht, die in der Regel aus mehreren Elementen besteht: Header (u.a. mit Metadaten zu Absender, Empfänger, benutzten Zwischenstationen), Body (der eigentliche Nachrichtentext) und Anhänge. Abhängig vom Gesamtsystem des Empfängers kann bereits die nach dem Versand empfangene und geöffnete E-Mail ihre Authentizität verlieren, indem sie, anders als vom Sender vorgesehen, rezipiert wird (z.B. ohne Formatierungen und ohne Bildmaterial). Die rechtskonforme Aufbewahrungspflicht von E-Mails (im öffentlichen Dienst zur Nachvollziehbarkeit des verwaltungsrechtlichen Handelns eine Notwendigkeit) betrifft einen großen Teil nicht nur der in öffentlichen Einrichtungen arbeitenden Personen.

Der Praxisteil stellte die vier Arbeitsgruppen vor unterschiedliche Herausforderungen: Eine Gruppe hatte die Aufgabe, die Vor- und Nachteile des LZA-Programms Mailstore<sup>14</sup> für den Arbeitsalltag eines Professors herauszuarbeiten. Die zweite Gruppe diskutierte, welche Metadaten für die LZA der E-Mails von Bedeutung sind. Die dritte Gruppe hatte den Auftrag, Überlegungen über die übermittelten Informationen einer E-Mail anzustellen, welche Informationen benötigt werden und wel-

---

14 <http://www.mailstore.com/de/> (03.02.2012).

che auch zukünftig von Interesse sein könnten. Die vierte Gruppe hatte die Aufgabe, aus einem E-Mail-System zwei verschiedene E-Mails für die LZA vorzubereiten und dabei folgende Fragen zu beantworten: Welche der angegebenen Export- und Speichermöglichkeiten werden gewählt, um ein möglichst archivfähiges Objekt übernehmen zu können?

Alle Ergebnisse der vier verschiedenen Gruppen hatten Eines gemeinsam: Überraschenderweise schien es nie die perfekte Lösung, Antwort oder Strategie zu geben, sondern immer stand einer Sammlung von Vorteilen der einen Methode oder Vorgehensweise eine Reihe von Nachteilen gegenüber, die es galt, gegeneinander abzuwägen und eine Entscheidung zu treffen.

Am Ende dieses zweiten Veranstaltungstages waren die Teilnehmer sowohl sensibilisiert für die vielfältigen Fragen und Probleme bei der Archivierung von E-Mails als auch für die technischen Schritte bei der Vorbereitung von Daten für deren Übergabe an ein Langzeitarchiv. Am Abend hatten die Teilnehmer Zeit, die Erkenntnisse und Ergebnisse der vergan-



©Fanny Wirsing

genen zwei Tage bei einem gemütlichen Zusammensein im Cyriakuskeller zu diskutieren.

Der dritte und zugleich

letzte Veranstaltungstag legte den Fokus zunächst auf die Webarchivierung. In der vierten und abschließenden Lektion mit dem Titel „Web(site)archivierung: Warum, Was und Wie?“ gab Anna Kugler, Bayerische Staatsbibliothek (BSB), einen umfassenden Überblick über Hintergründe, Ziele, Strategien, Techniken und Herausforderungen der Webarchivie-

rung im nationalen und internationalen Kontext. Dazu erfolgte in einem einführenden Teil eine Erläuterung der Anfänge, die bis in die 1990er Jahre mit der



©Fanny Wirsing

Gründung des Internet Archive (1996<sup>15</sup>) zurückreichen, sowie eine Schilderung der Beweggründe für die Archivierung von Internetseiten wie z.B. die Auffindbarkeit und Referenzierbarkeit von Inhalten, was gerade für die Wissenschaft dauerhaft ermöglicht werden sollte. Im Rahmen einer anschließenden Definition der wichtigsten Fachtermini, die für eine inhaltliche Abgrenzung als notwendig erachtet werden, lernten die Teilnehmer beispielsweise den Unterschied zwischen Websites und Webpages sowie den Maßnahmen zu deren Langzeitarchivierung kennen. Weiterhin wurden die Kriterien bzw. Strategien vorgestellt, nach denen die Archivierung von Websites erfolgt und die vorab in einer Policy der verantwortlichen Institutionen definiert werden müssen. Einige wichtige Anhaltspunkte sind beispielsweise die Frequenz der Archivierung oder signifikante Eigenschaften („significant properties“) von Websites wie Inhalt oder Design. Nach einem kurzen Blick auf die rechtlichen Rahmenbedingungen, stand der Einsatz möglicher Technologien zur Websitearchivierung im Zentrum der Präsentation. Dazu zählen sowohl die sog. Crawl-Technologie und integrierte Softwaresysteme wie z.B. das WebCuratorTool<sup>16</sup>, welches auch in der BSB eingesetzt wird, als auch Software zur Präsentation der archivierten Websites. Im letzten Teil des Vortrags wurden verschiedene

15 Internet Archive, <http://www.archive.org/> (03.02.2012).

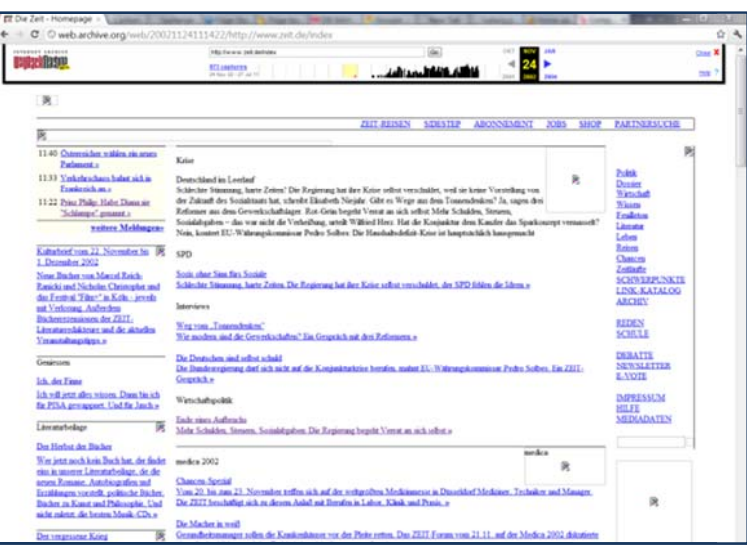
16 <http://webcurator.sourceforge.net/> (03.02.2012).

internationale Projekte wie das IIPC (International Internet Preservation Consortium<sup>17</sup>) zur Webarchivierung beschrieben und einige nationale Lösungsansätze in ihren unterschiedlichen Zielsetzungen präsentiert. Dazu zählen etwa das Web-Archiv des Deutschen Bundestags<sup>18</sup> oder das Webseitenarchiv des Zentralarchivs zur Erforschung der Geschichte der Juden in Deutschland.<sup>19</sup> Abschließend wurde anhand des Webarchivierungsprojekts an der BSB ein Umsetzungsweg in der Praxis im Detail vorgestellt.

Nach der Vermittlung der theoretischen Grundlagen beschäftigten sich die Teilnehmer in der anschließenden Gruppenübung mit den Webarchivierungsaktivitäten verschiedener Institutionen im In- und Ausland in der Praxis. Dabei versuchten die Kursteilnehmer im Rahmen einer Bestandsaufnahme sich ein Bild von der Landschaft der Webarchivierung zu erarbeiten, in dem ausgewählte Webarchive von Bibliotheken und Archiven auf Vollständigkeit, Sammlungsauftrag, Zugänglichkeit und Rechtsverhältnisse überprüft wurden. Das Ergebnis fiel dabei eher ernüchternd aus. Die

in der Diskussion gefallenen Formulierungen wie „Flickenteppich“ oder „Schweizer Käse“ beschreiben die aktuelle Situation wohl am treffendsten. Im Ergebnis konnte festgestellt werden, dass es keinen einheitlichen Ansatz oder eine einheitliche Erschließungsstrategie gibt. Zudem wurde in der Diskussion deutlich, dass die Teilnehmer eine grundlegende Verständigung auf die Inhalte der Policies, vor allem auf die Sammlungstiefe, -objekte und -häufigkeit, wünschen, damit insgesamt systematischer und einheitlicher archiviert wird. Die Aufgabe der ersten Erarbeitung eines Vorschlags sollte bei der Deutschen Nationalbibliothek (DNB) liegen, da sie durch ihren gesetzlichen Auftrag eine zentrale Rolle in der LZA einnimmt.

Abgerundet wurde die Veranstaltung durch einen Fachvortrag von Andreas Lange, dem Leiter des 1997 in Berlin gegründeten Computerspielmuseums.<sup>20</sup> Lange ist zugleich Projektleiter des Computerspielmuseums für das EU-Forschungsprojekt KEEP (Keeping Emulation Environments Portable<sup>21</sup>) und Sprecher der AG Emulation von nestor. In seinem spannenden und bei einigen Teilnehmern sicher nostalgische Gefühle auslösenden Vortrag mit dem Titel „Zugang und Absturz. Über die Chancen und Herausforderungen der digitalen Bewahrung anhand des Beispiels von Computerspielen“ weitete er den Blick auf das zu bewahrende kulturelle Erbe, zu dem neben traditionellem Bibliotheks-, Archiv- und Museumsgut auch interaktive Medien wie Computerspiele bzw. Spielesoftware gehören. Während bei Texten vornehmlich die Migration im Vordergrund steht, bewarb er die Emulation als die geeignetste Methode zur Bewahrung des „Look and Feel“ von Computerspielen, da dadurch die Komplexität digitaler Artefakte



17 <http://netpreserve.org/about/index.php> (03.02.2012).  
 18 <http://webarchiv.bundestag.de/cgi/archive.php> (03.02.2012).  
 19 <http://www.zentralarchiv.uni-hd.de/> (03.02.2012).

20 <http://www.computerspielmuseum.de/> (03.02.2012).  
 21 <http://www.keep-project.eu/ezpub2/index.php> (06.02.2012).

am besten repräsentiert wird. Nach einer aktuellen Situationsbeschreibung, welche durch eine extrem dynamische Marktveränderung gekennzeichnet ist, führte Lange einige – mittlerweile historische – Computerspiele mit Emulatoren vor und betonte dabei schließlich die wichtige Rolle von Fans bzw. Gamern, die als erste den Anstoß gaben, auch retrospektiv über PC-Spiele nachzudenken.

Als abschließendes Fazit lässt sich festhalten, dass den Teilnehmern im Laufe der Veranstaltung deutlich wurde, wie weit respektive kurz gefasst die LZA verstanden und interpretiert wird. Die unterschiedliche Interpretation des Begriffs führte, unterstützt durch die verschiedenen beruflichen Schwerpunkte (Archiv, Bibliothek, Museum) teilweise zu längeren Diskussionen, die eher um eine Interpretation des Begriffs bemüht waren. Eine Einführung in das Thema hätte diese Diskussionen womöglich vorweggenommen und in den einzelnen Sessions zu einer stärkeren Konzentration auf

das eigentliche Thema geführt. Jedoch ist dies der einzige Kritikpunkt an der Winter School.

Durch die praxisnahen Vorträge und Übungen wurde die nestor Winter School zur LZA in der Praxis von allen Beteiligten als große Hilfe empfunden, bisheriges theoretisches Wissen auch praktisch anwenden zu können. Eine wichtige Rolle spielte dabei sicherlich – neben der angenehmen Atmosphäre – auch die Zusammensetzung der Dozenten sowie der Teilnehmer, die ein großes Spektrum des kulturellen Erbes abdeckten. Der intensive Erfahrungs- und Wissensaustausch unter den Kollegen war in diesen idealen Rahmenbedingungen auch hervorragend dazu geeignet, neue Kontakte zu knüpfen und/oder bestehende zu vertiefen.

Die Website der nestor/DigCurV Winter School 2011 ist unter [http://nestor.sub.uni-goettingen.de/winter\\_school\\_2011/index.php](http://nestor.sub.uni-goettingen.de/winter_school_2011/index.php) erreichbar.



Die Autorinnen danken Frau Fanny Wirsing, Herrn Prof. Dr. Achim Oßwald und Herrn Stefan Strathmann für die Bereitstellung des Fotomaterials.