

**Workshop der Arbeitsgruppe
Elektronisches Publizieren der Union der
deutschen Akademien der
Wissenschaften**

Veranstalter: Zentrum für Informations-, Medien- und Kommunikationstechnologie (ZIMK), Universität Trier; Sonderforschungsbereich 600 „Fremdheit und Armut“, Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften, Universität Trier; Nordrhein-Westfälische Akademie der Wissenschaften und der Künste, Düsseldorf

Datum, Ort: 04.10.2010–06.10.2010, Düsseldorf

Bericht von: Jörg Wettlaufer, Residenzenkommission der Akademie der Wissenschaften zu Göttingen

Der diesjährige Workshop der Arbeitsgruppe „Elektronisches Publizieren“ der Akademienunion war dem Thema der Repositorien gewidmet, denjenigen Behältnissen also, in denen die Bits und Bytes der Akademienprojekte bis zur nächsten Migration und vielleicht auch darüber hinaus aufbewahrt werden sollen. Das Thema ist drängend. Mit jedem auslaufenden Projekt wird der Verlust der häufig nicht nur gedruckt, sondern im zunehmenden Maße auch digital vorliegenden Forschungsergebnisse virulent. Wenn erst die Mitarbeiter nicht mehr verfügbar sind, können wichtige Metadaten oft nur noch unter großem Aufwand rekonstruiert werden. Diese sind aber zur Aufbewahrung und Speicherung der eigentlichen Daten unbedingt notwendig. Häufig wurden digitale Arbeitsinstrumente zur Bewältigung der Forschungsaufgaben in den Einzelprojekten erstellt, die nie zur (gedruckten oder digitalen) Publikation vorgesehen waren, für das Forschungsfeld des Projekts jedoch wertvollste Datensammlungen enthalten, die auch zukünftige Projekte oder Forschergenerationen gerne zur Verfügung hätten. Seitdem viele Projekte zunehmend hybrid publizieren, stehen zudem digital born publications bereit, die ein neues Zuhause und Pflege brauchen.

Das Problem der oftmals fehlenden Langzeitarchivierung ist vielerorts erkannt, allein es fehlt der Masterplan, dieser Gefahr zu be-

gegnen. Die Akademien entwickeln erst langsam die notwendige Infrastruktur, suchen strategische und operative Partner, die bei der Bewältigung der Aufgabe zur Seite stehen können und richten aus ihrem Etat Stellen für „Digitalisierungsbeauftragte“ ein, die die Vorhaben vor Ort mit Rat und Tat unterstützen. Zu lange hat man tatenlos zugesehen, vielleicht in der Hoffnung, der technische Fortschritt überholt sich eines Tages selber und die Probleme der Datensicherung könnten sich damit auflösen. Das Gegenteil ist eingetreten. Schon jetzt kämpfen Forscher und Institutionen allenthalben mit veralteten Datenformaten, proprietären Standards, gebrochenen Workflows und leider auch immer noch der Uneinsichtigkeit einzelner, die vielleicht zurecht vermuten, die Folgen ihres Zögerns nicht mehr selber verantworten zu müssen. Für Strategien und die Entwicklung zukünftiger Langzeitarchive werden Rechenzentren, Archive und Bibliotheken der Wissenschaftsstandorte, die mit den angesprochenen Problemen in vielfältiger Weise selber zu kämpfen haben, mit ihren Erfahrungen und Lösungskonzepten in Erscheinung treten.

Die Tagung war, der angesprochenen Dramaturgie Rechnung zollend, in zwei große Abschnitte aufgeteilt. Zunächst stellten einzelne Projekte ihre Situation in Bezug auf Forschungsdaten und Online-Publikationsprojekte vor. Die Bandbreite reichte hier erwartungsgemäß von umfassenden IT-Infrastrukturen, die sogar schon Überlegungen zur Langzeitarchivierung beinhalten, bis zu Redaktionen, die nach klassischem Vorbild Manuskripte an Verlage weiterreichen und noch keinen durchgängig digitalen Workflow etabliert haben. In einem zweiten Block stellten dann verschiedene Dienstleister im Bereich des Digital-Publishing ihre Angebote vor, den Arbeitsablauf der Forschungsprojekte mit standardkonformen Werkzeugen effizienter zu gestalten und die nachhaltige Nutzung der meist digital vorliegenden Arbeitsergebnisse zu ermöglichen.

RAFFAEL BALL (Universitätsbibliothek Regensburg) verkündete in seinem Einführungsvortrag das Ende des gedruckten Buches und forderte die Anpassung der Bibliotheken an die neue Aufgabe als „Pro-

blemlösungsportale“. In den folgenden Beiträgen zeigte sich wieder einmal, dass gerade die „exotischen“ Projekte aufgrund besonderer Anforderungen an Zeichensätze und Zeichencodierungen in der Umstellung ihrer Arbeitsprozesse besonders weit fortgeschritten sind. Nicht zufällig stellte WOLFGANG MEIER (Heidelberger Akademie der Wissenschaften, HAW), der an der Entwicklung des inzwischen so beliebten XML Datenbanksystems eXist entscheidend beteiligt ist, zusammen mit CLAUDIA WENZEL (HAW) das Projekt der Erfassung Buddhistischer Steininschriften in China der HAW vor. Ob Keilschrift, Hieroglyphen oder das Arabisch des Korans – alle diese Schriften erfordern spezielle Zeichensätze und haben damit schon früh die Umstellung auf XML TEI sowie Unicode vollzogen. Die Plattform „Deutsche Inschriften Online“¹ wurde von TORSTEN SCHRADER und ALINE DEICKE (beide Akademie der Wissenschaften und der Literatur Mainz) vorgestellt; sie ist ein Beispiel für eine Lösung, bei der mit zwei Mitarbeiterstellen und studentischen Hilfskräften eine Infrastruktur, die „Mainzer Digitale Akademie“, geschaffen wurde, die einerseits eine Nachnutzung durch verschiedene Projekte gewährleistet, zugleich aber auch anderen Akademien eine Betreuung und Umsetzung digitaler Vorhaben ermöglicht. So wurde die Digitalisierung der bislang nur gedruckt vorliegenden Bände in einer ersten Phase durch die Göttinger Akademie finanziert, deren digitales Akademieportal erst im Aufbau begriffen ist. Auch bei den Wörterbüchern und Editionen ist die Lage ernst. Zwar setzen viele auf den XML/TEI P5 Standard, doch eine Strategie für die langfristige Archivierung der Daten oder die Erhaltung lauffähiger Portale über das Ende der Projektlaufzeit hinaus konnte kaum jemand vorweisen. In einem Statement forderte der Generalsekretär der Akademiunion Dieter Herrmann die Projektverantwortlichen auf, schon frühzeitig an die Konvertierung der Daten in standardkonforme Formate zu denken und dabei eng mit verwandten Projekten zusammen zu arbeiten. Dabei sollte das Rad nicht jeweils neu erfunden werden, das heißt, die Orientierung an erfolgreichen Strategien und Standards sollte

im Vordergrund der Bemühungen stehen.

Problematisch ist bei vielen Projekten noch die Doppelfunktion der Repositorien als Datenspeicher einerseits und als Portal zur Verbreitung der Forschungsergebnisse andererseits. Konservierung und Langzeitarchivierung stellen grundsätzlich andere Anforderungen an die zu verwendenden Standards und Datenformate als die ansprechende Präsentation der Ergebnisse. Die Präsentation beansprucht häufig viele Ressourcen, die dazu verwendeten Lösungen sind aber extrem zeitsensibel, veralten also eben so schnell wie z.B. das Webseitendesign oder unterliegen Wandlungsprozessen wie die best practice der usability. Hinsichtlich der stark zeitgebundenen Komponenten der Datenpräsentation bietet sich insbesondere die Emulation als Möglichkeit an, veraltende Forschungsrepositorien mit der dazugehörigen Software für einige Zeit zu konservieren. Allerdings sind hier aufgrund der verwendeten Virtualisierungstechnik teilweise Anpassungen der Software notwendig, um z.B. Schreibprozesse zu unterbinden und so eine Archivfunktion zu ermöglichen. THOMAS BURCH und MARINA MÜLLER (Sonderforschungsbereich 600, Universität Trier) und PETER LEINEN (Rechenzentrum, Universität Trier) stellten eine derartige Lösung für das dort eingesetzte „Forschungs- und Dokumentationssystem FuD“ vor, dessen Inhalte und Funktionalität auf diese Weise auch nach Ende des SFB interessierten Wissenschaftlerinnen und Wissenschaftlern weiter zur Verfügung stehen soll. Einen anderen Weg beschreitet das Deutsche Rechtswörterbuch, das als Projekt der Heidelberger Akademie der Wissenschaften schon seit Ende der 1990er-Jahre eine Online-Präsenz mit open access policy anbietet. In diesem Jahr neu hinzugekommen ist ein Textarchiv „Deutsche Rechtsquellen in digitaler Edition“², das juristische Literatur des 15. und 16. Jahrhunderts online präsentiert und ebenso wie das Rechtswörterbuch auf einer angepassten Implementation einer Faust Datenbank basiert. Probleme der Langzeitarchivierung sind hier aufgrund der Laufzeit des Projekts bis 2035 noch nicht akut, aber schon vorhersehbar. Hier wird zu entscheiden sein,

¹ DIO <<http://www.inschriften.net>> (06.12.2010).

² <<http://www.drqed.it.de>> (06.12.2010).

ob die Daten in einer Emulation oder, wahrscheinlicher, mit einer Migration in ein offenes Format in einer längerfristigen Perspektive zugänglich gehalten werden sollen. Besonderen Wert auf eine ansprechende Präsentation der Daten legt derzeit das bereits genannte Portal „Deutsche Inschriften Online“, das Gemeinschaftsprojekt der Inschriftenarbeitsstellen an den deutschen Akademien. Hier sind sehr plastisch die synergetischen Effekte zu beobachten, die das know how verschiedener Arbeitsstellen zu einem modernen XML-basierten System zusammenführt, das zu 100 Prozent über Open Source Software realisiert wurde. Es ist abzusehen, dass in diesem Lösungsansatz das größte Potential steckt, nämlich die Zusammenarbeit von thematisch und strukturell ähnlichen Projekte zu organisieren und hier gemeinsame Lösungen zu erarbeiten, die von allen Beteiligten genutzt und in gegebenenfalls leicht angepasster Form betrieben werden können. Die inhaltlichen Anforderungen an Editionsprojekte auf der einen und einem Wörterbuchprojekt auf der anderen Seite sind einfach zu unterschiedlich, um mit ein und derselben Softwarelösung optimal abgebildet zu werden. Verwandte Projekte sollten sich also in kleinen Gruppen zusammenfinden und zusammen mit starken Partnern (Bibliotheken und Archiven) eine für sie optimale Lösung erarbeiten.

Ein Vertreter eines solchen „starken Partners“, der Direktor des Freiburger Rechenzentrums GERHARD SCHNEIDER, hielt den öffentlichen Abendvortrag am zweiten Tag zu dem Thema „Repositorien – Digitalmüll der Zukunft oder Zukunft digitaler Bibliotheken?“ Dieser kritische Aspekt, nämlich die Gefahr der Produktion von digitalem Müll, der niemanden interessiert bzw. der von niemandem mehr verwendet werden kann, wurde insgesamt auf der Tagung nur selten angesprochen. Vor allem aber legte Schneider den Finger in die Wunde des Standardisierungsproblems, das er zugleich als Machtproblem thematisierte. Wer den Standard setzt, der hat zugleich auch Macht über Daten und ihre Zukunft. In der digitalen Welt sind daher Standards und die Kompatibilität von Systemen und Softwarearchitekturen entscheidende Kriterien für die Zukunftssicherheit von Daten und ihren Präsentationen. XML

nach dem Standard der TEI in der Version P5 stellte sich für eigentlich alle Projekte als kleinster gemeinsamer Nenner für textbasierte Datenrepositorien heraus. Im multimedialen Bereich sieht es schon schwieriger aus. Die Standards bei Digitalisierungsprojekten, METS und MODS im Bereich der digitalen Bibliotheken, gehören zurzeit sicher ebenso zu einer zukunftsweisenden Metadaten-speicherung wie spezielle XML-Standards für Noten/Musik oder für Inschriften. Zumindest im Bereich der Akademieprojekte sollte man sich verbindlich auf diese Standards verständigen, die ja auch die Voraussetzung für DFG-Förderung einzelner Digitalisierungsvorhaben darstellen.

Verschiedenste Konzepte und Entwicklungen wurden im zweiten Teil der Tagung vorgestellt, die aus Platzgründen hier nur stichwortartig Erwähnung finden können. Neben der BSB in München, deren Konzept GREGOR HORSTKEMPER vorstellte und bei der besonders große Datenmengen zu archivieren bzw. bereitzustellen sind, stellte HEIKE NEUROTH von der SUB Göttingen die Text-Grid Initiative als mögliches Repitorium auch für Daten aus Akademieprojekten vor. MICHAEL HOLLMANN vom Bundesarchiv in Koblenz führte in die Problematik der Archivierung digitaler Überlieferung ein – ein Problemfeld, das noch wesentlich umfangreicher erscheint, als es sich für die Akademien momentan darstellt. Die Max Planck Digital Library wurde von MALTE DREYER (München) in Bezug auf die Möglichkeit zur Verwendung als Repitorium vorgestellt, und GERALD NEUMANN sowie TORS- TEN ROEDER präsentierten die Fortschritte des Personendaten-Repositoriums der Berlin-Brandenburgischen Akademie der Wissenschaften – sozusagen ein Nachklang zur letzten Tagung der Arbeitsgruppe in Leipzig, die ja den Möglichkeiten der Vernetzung über die Personennormdatei (PND) der Deutschen Nationalbibliothek (DNB) gewidmet war. SACHA ZALA (Bern) stellte sodann die Situation der Digitalisierung von Akademievorhaben in der Schweiz dar und ROLAND KAMZELAK (Marbach) die Langzeitarchivierung von Literatur im Deutschen Literaturarchiv. Schließlich wurde von NATASCHA SCHUMANN (Frankfurt am Main,

DNB) und STEFAN STRATHMANN (Göttingen, SUB) die Langzeitarchivierung aus der Sicht von Bibliotheken und hier konkret der SUB Göttingen thematisiert. Den Abschlussvortrag hielt im Anschluss ELMAR MITTLER von der SUB Göttingen über das Dreieck Autor-Verlag-Bibliothek und beleuchtete dabei die Seite der Verlage, die sich in der Vergangenheit immer wieder mit ihren kommerziellen Interessen als Hindernis für eine open access Präsentation von Arbeitsergebnissen positioniert hatten. Obwohl hier inzwischen eine größere Offenheit auf Seiten der Verlage herrscht, bleibt das Grundproblem der doppelten Belastung öffentlicher Haushalte im klassischen Verlagsmodell bestehen und wird sich wohl nur durch einen radikalen Wandel in Hinblick auf digitale Dienstleistungen lösen lassen. Im Hintergrund stehen hier wie so oft Fragen der Nutzungsrechte, die inzwischen Forschung massiv behindern würden. Der zu dieser Thematik vorgesehene Referent TOBIAS HILLENGEIST (Landgericht Lüneburg) war leider verhindert, aber das Abstrakt seines Vortrags wie auch das aller übrigen Vorträge stehen online zur Verfügung.³

Im Fazit bleibt festzuhalten, dass die Akademievorhaben aufgefördert sind, sich bei den Datenformaten ihrer Arbeitsergebnisse an offenen Standards zu orientieren, sich für die Langzeitarchivierung und die Präsentation ihrer Arbeit starke Partner (in erster Linie Bibliotheken, Archive und Rechenzentren) zu suchen und vor allem auch synergetische Effekte zu nutzen. Diese schienen an einigen Stellen auch in Düsseldorf auf, als z.B. eine in Python geschriebene Lösung zum Auslesen von Allegrodatenbanken auf großes Interesse bei anderen Projekten stieß, die eben doch häufig mit ganz ähnlichen Problemen zu kämpfen haben.

Konferenzübersicht:

Eröffnungsvortrag

Rafael Ball (Direktor der UB Regensburg): Die Wissenschaftskommunikation der Zukunft - Aspekte von Science 2.0

Beiträge aus den Akademien: Alte Geschichte und Inschriftenforschung

Gernot Wilhelm (AWLM): Das Hethitologie

Portal

Jan-Peter Graeff (AWG): Vom Zettelarchiv zur Datenbank: Die Inschriften des Tempels von Edfu in Oberägypten

Frank Grieshaber (HAW): Die epigraphische Datenbank Heidelberg (EDH)

Torsten Schrade / Aline Deicke (AWLM): Die deutschen Inschriften online

Beiträge aus den Akademien: Repositorien zur Überlieferung heiliger Schriften

Holger Strutwolf / Volker Krüger (Universität Münster): Die Überlieferung des Neuen Testaments (NRWAWK, DFG)

Michael Marx (BBAW): Die Überlieferung des Korans (Corpus Coranicum)

Claudia Wenzel / Wolfgang Meier (HAW): Buddhistische Steininschriften in China

Beiträge aus den Akademien: Wörterbücher und Editionen

Andreas Deutsch (HAW): Das Deutsche Rechtswörterbuch - ein Repitorium?

Matthias Schulz (BBAW): Vom Buch zum „Deutschen Text-Archiv“ (DFG)

Stephen Dörr (HAW): Zwischen Datenbank und Wörterbuch im -Dictionnaire étymologique de l'ancien français

Brigitte Bulitta (SAW): Stand und Folgen der Retrodigitalisierung des Althochdeutschen Wörterbuchs

Beiträge aus den Akademien: Digitale Bibliotheken und Informationssysteme

Ulrich Marder / Claudia Leuker (NRWAWK): Digitale Bibliotheken am Deutschen Referenzzentrum für Ethik in den Biowissenschaften

Matthias Reinert (BAdW): Die Neue Deutsche Biographie (NDB) - von der Digitalisierung zur digitalen Dienstleistung

Markus Wesche (BAdW): Akademien in Not? Überlegungen der BAdW und der BSB zur Langzeitarchivierung ihrer Forschungsdaten

Konzepte und Entwicklungen I

³ <<http://www.akademienworkshop.de>> (06.12.2010).

Thomas Burch (Kompetenzzentrum, Universität Trier), Marina Müller (SFB 600, Universität Trier): Forschungsdaten in interdisziplinären Langfristvorhaben: Bereitstellung und Nachnutzung (DFG)

Peter Leinen (Zentrum für Informations-, Medien- und Kommunikationstechnologie (ZIMK) der Universität Trier): Perspektiven der Langzeitarchivierung der Universität Trier

Gregor Horstkemper (BSB): Perspektiven der Langzeitarchivierung der BSB

Konzepte und Entwicklungen II

Heike Neuroth (SUB Göttingen): TextGrid - Virtuelle Forschungsumgebung, Forschungsdaten und Langzeitarchivierung (BMBF)

Michael Hollmann (Bundesarchiv Koblenz): Digitale Quellen sichern und verfügbar halten: Die Lösung des Bundesarchivs

Gerald Neumann, Torsten Roeder (BBAW): Das Personendaten-Repository der BBAW (DFG)

Malte Dreyer (MPG, Max Planck Digital Library): Repositorieninfrastruktur der MPDL

Öffentlicher Vortrag

Gerhard Schneider (Direktor des Rechenzentrums der Universität Freiburg): Repositorien - Digitalmüll der Zukunft oder Zukunft digitaler Bibliotheken

Einzelaspekte und Kooperationen

Tobias Hillegeist (Landgericht Lüneburg): Rechtliche Probleme der digitalen Langzeitarchivierung wissenschaftlicher Primärdaten (ausgefallen)

Sacha Zala (SAGW): Digitalisierungsstrategien in der SAGW

Roland Kamzelak (DLA): Langzeitarchivierung von Literatur im DLA

Natascha Schumann (DNB), Stefan Strathmann (SUB Göttingen): Repositorien und Langzeitarchivierung aus der Sicht der Bibliotheken

Abschlussvortrag

Elmar Mittler (SUB Göttingen, GAW): Geisteswissenschaftliche Forschungskommunikation

der Zukunft im Dreieck Autor - Verlag - Bibliothek

Schlusswort: Wolfgang Raible (HAW, AG EP)

Tagungsbericht *Workshop der Arbeitsgruppe Elektronisches Publizieren der Union der deutschen Akademien der Wissenschaften*. 04.10.2010–06.10.2010, Düsseldorf, in: H-Soz-Kult 15.12.2010.