

Redaktionssysteme und Virtuelle Forschungsumgebungen

Veranstalter: AG Elektronisches Publizieren der Union der deutschen Akademien der Wissenschaften und Bayerische Akademie der Wissenschaften

Datum, Ort: 20.10.2014–21.10.2014, München

Bericht von: Jörg Wettlaufer, Akademie der Wissenschaften zu Göttingen

Der diesjährige Workshop der AG Elektronisches Publizieren war dem Thema „Redaktionssysteme und virtuelle Forschungsumgebungen“ gewidmet. Das Thema ist ein Dauerbrenner für die AG (und nicht nur dort), weil praktisch alle Vorhaben im Akademienprogramm Software einsetzen, die redaktionelle Aufgaben unterstützt. Was die virtuellen Forschungsumgebungen (VRE) betrifft, so ist es um diesen Begriff seit geraumer Zeit stiller geworden. Die hohen Erwartungen, die noch vor etwa 5 Jahren in diesem Bereich verbreitet waren, haben sich nicht erfüllt und der einführende Vortrag von MICHAEL PIETROWSKI (Mainz) fragte daher konsequent nach der Zukunft der Idee von der virtuellen Forschungsumgebung zwischen den beiden Polen „Werkzeugkasten“ und „monolithische“ Software. Die Idee eines Werkzeugkastens, die im Folgenden von vielen der Beiträge aufgenommen wurde, kann und muss aber weiter differenziert werden. Verstehen wir darunter modular aufgebaute Software, die, mit entsprechenden Schnittstellen ausgestattet, für die jeweiligen Aufgaben im Forschungsprozess Werkzeuge zur Verfügung stellt, die je nach Bedarf kombiniert und erweitert werden können? Oder denken wir mehr an einen Mix aus sogenannter Standardsoftware, die nicht notwendigerweise modular oder open-source daher kommt, dafür aber hohen Qualitätsansprüchen genügt und sich trotz der Heterogenität als Forschungsumgebung begreifen lässt?

Die Beiträge und Projektvorstellungen des Workshops haben am Ende beide Varianten präsentiert, wenn auch die modular gebauten CMS-Systeme in der Überzahl waren. Angelehnt an die Virtuellen Learning Environments (VLE), die ebenso wenig wie VREs bislang erfolgreich definiert werden konnten,

und am Beispiel der Erfahrung am Institut für Europäische Geschichte in Mainz mit datenbankbasierten Forschungsumgebung FUD (vergleiche dazu den Beitrag zum Projekt „Arthur Schnitzler digital“ weiter unten) formulierte Pietrowski eine Reihe von Thesen zum Charakter und zur möglichen Funktion solcher Umgebungen in Forschungsprozessen. Eine VRE sei zunächst kein Produkt, sondern eine „sozio-technische Umgebung“, so eine der zentralen Thesen bzw. Beobachtungen. Als solche muss sie nicht nur technisch, sondern auch eben auch sozial funktionieren und von den Forschern und Institutionen akzeptiert werden und deren Interessen berücksichtigen. Sodann hinterfragte er die Vorstellung vom generischen geisteswissenschaftlichen Forschungsprozess, der die Entwicklung einer ebenso generischen VRE erlauben würde. Niemand brauche Neuimplementierungen von generischen Werkzeugen – diese liegen ja in der Regel schon vor. Es käme jedenfalls auf die Auswahl der richtigen Abstraktionsebene an, um sinnvoll generische Anwendungen in den Arbeitsprozess zu integrieren, die auch für andere Projekte nutzbar wären. Schließlich komme es gerade im Kontext der Akademievorhaben auf eine langfristige Verfügbarkeit an, die die Forschungs- und Redaktionsaktivitäten nachhaltig unterstützen.

Gerade diese Frage nach der langfristigen Perspektive der Technologien und Softwarearchitekturen bereitete den Teilnehmerinnen und Teilnehmern des Workshops die größten Sorgen, denn weiterhin ist dergleichen nicht in Sicht. Im Gegenteil. THORSTEN SCHRADE (Mainz) und KollegInnen wiesen in ihrem Beitrag zu „XML basiert edieren, Content Management gestützt publizieren: Integrierte Forschungs- und Editionsumgebungen in den Projekten ‚Deutsche Inschriften Online‘ und ‚Digitales Familiennamenwörterbuch Deutschlands‘“ auf die kurzen Produktzyklen bei heutiger Software hin, die sichere Aussagen in der Regel nur für zwei bis drei Jahre zuließen. Vor der Prämisse von mittleren Laufzeiten von Akademieprojekten zwischen 12 und 25 Jahren führt dies immer wieder zu Problemen und Unsicherheiten bei der Planung der Softwarearchitekturen. Niemand kann garantieren, dass die heute verwendeten

CMS oder Frameworks auch noch in 10 Jahren existieren und von einer Community gepflegt werden. Auf der anderen Seite erweisen sich gerade oft Ein-Personen Software-Projekte als recht ausdauernd und langzeitverfügbar, allerdings zum Preis der permanenten Unsicherheit und Sorge über den Gesundheitszustand dieser Person, mit der die Software in der Regel steht oder fällt. In der Digitalen Akademie in Mainz setzt man jedenfalls auf eine ganze Reihe von Werkzeugen, um dem Titel entsprechend XML basiert zu editieren und CMS gestützt zu publizieren. Für letzteres kommt im Mainz das weit verbreitete System typo3 zum Einsatz, für das in der Mainzer Digitalen Akademie schon eine Reihe von Extensionen programmiert wurden, die durchaus auch generisch sind und in anderen typo3 gestützten Umgebungen zum Einsatz kommen. Für die Publikation in XML nutzt man hingegen unter anderem die beliebte individualisierbare Werkzeugleiste des Oxygen XML Editors aus der BBAW.

ANDREAS KUCZERA (Gießen) und STEFAN DUMONT (Berlin) stellten anschließend die Digitale Arbeitsumgebung „Ediarum“ und ihre Anwendung im Akademieprojekt „Regesta Imperii“ vor. In diesem TELOTA-Projekt wird ebenfalls die eXist XML Datenbank für die Speicherung verwendet. Die Ausgabe erfolgt entweder als Webseite via REST, XQuery und XSLT oder gedruckt als PDF über ConTeXt. UTE RECKER-HAMM (Mainz) beschloss die erste Sektion mit einem Vortrag über das Redaktionssystem des Mittelhochdeutschen Wörterbuchs. Dieses Wörterbuch benutzt einen selber entwickelten Editor bzw. Redaktionssystem (TAReS), der Ergebnis eines DFG geförderten Projekts ist. Das in diesem Projekt entwickelte internetbasierte Redaktionssystem unterstützt alle Phasen der Arbeit an einem historischen Belegwörterbuch und ist als Clients-Server Architektur konzipiert, in dessen Zentrum eine MySQL Datenbank steht. Es beinhaltet Funktionen für die elektronische Publikation des Wörterbuchs und seiner Materialien (Belegarchiv, Quellenverzeichnis usw.) im Internet.¹

Die zweite Sektion begann mit einem Beitrag von ULRICH SCHMID (Münster) über den virtuellen Handschriftenlesesaal des Instituts für neutestamentliche Textforschung.

Dort werden griechische Manuskripte des neuen Testaments virtuell präsentiert und sind unter einer einheitlichen Oberfläche konsultierbar. Damit wird eine *editio critica maior* des neuen Testaments angeboten, die zur Zeit ca. 5.600 Handschriften umfasst. Das Projekt kann auf eine 150jährige Geschichte zurückblicken. Technologisch wird die Portalsoftware Liferay eingesetzt, die auch dem OpenSocial Standard entspricht und Crowdsourcing bei der Transkription der Manuskripte erlaubt. Weniger eine Präsentationssoftware als vielmehr ein Verwaltungsinstrument stellte anschließend JÜRGEN HEROLD (Greifswald/Göttingen) mit dem Redaktionssystem Epigra 3 des Projekts „Deutsche Inschriften Online“ vor. Diese als Client-Server System selber konzipierte System erlaubt passgenau die Aufnahme der Inschriften in die MySQL Datenbank, die im Backend wirkt. Das speziell entwickelte System entspricht damit genau den Bedürfnissen der Inschriftenforscher.

Auf eine schon etablierte (Fach-)spezifische Software setzt hingegen das Projekt „Kleine und fragmentarische Historiker der Spätantike“, das anschließend von JOHNATAN GROSS (Düsseldorf/Göttingen) vorgestellt wurde. Hier kommt der in der Community beliebte „Classical Text Editor“ zum Einsatz, der seit vielen Jahren von einem Entwickler gepflegt und vertrieben wird. Damit erhebt sich die für viele Projekte erhebliche Frage nach Langzeitverfügbarkeit der verwendeten Tools und Systeme und der Abhängigkeit von einzelnen Personen, deren Verlust nicht ersetzbar wäre.

Im Abendvortrag stellte ECKHART ARNOLD (München) von der gastgebenden Akademie Architekturen und Trends virtueller Forschungsumgebungen und Redaktionssysteme vor. Er sieht die Entwicklung vom Desktop zum Web und von der monolithischen Forschungsumgebung hin zum Werkzeugkasten verlaufend. Das Hinterfragen des Referenzcharakters des Gedruckten für die digitalen Projekte am Ende zog eine Diskussion über die Zukunft des Buches nach sich,

¹ DFG-gefördertes Projekt „Internetbasiertes Artikelredaktionssystem für die Produktion und Publikation von Wörterbüchern in dezentralen Arbeitsstellen“, <<http://www.mhdwb.uni-trier.de/TAReS/>> (25.11.2014)

in der die unterschiedlichen Standpunkte sowie die immer noch verbreitete und auch berechnete Skepsis gegenüber der langfristigen Verfügbarkeit von Ressourcen auf dem Internet deutlich wurden.

In der dritten Sektion am nächsten Morgen wurde von ALEXANDER GEYKEN (Berlin) das Redaktionssystem des „Digitalen Wörterbuchs der deutschen Sprache“ (DWDS) vorgestellt, das maßgeblich auf der Entwicklung einer studentischen Hilfskraft dieses Projekts beruht. Es verwendet die XML-Datenbank eXist 2.2 RC2 sowie den Editor Oxygen als Eingabetool. Im Projekt werden damit 425 Lemmata und 1,8 Millionen Belege verwaltet. Das Trierer Wörterbuchnetz und das DWDS nutzen dieselben XML Quellen, stellen sie aber anders dar. Als backup-tool sowie als Repositorium und die Webansicht wird SVN 1.3.3 verwendet. Als Format für die Artikel wird ein eigenes XML-Schema eingesetzt, das aber relativ kompatibel mit TEI sein soll.

STEFAN DÖRR (Heidelberg) präsentierte anschließend das Redaktionssystem des „Etymologischen Wörterbuchs des Altfranzösischen“ (DEAF). Im Hintergrund arbeitet hier eine MySQL Datenbank mit entsprechendem Schema. Die Lemmata werden mit XML Markup versehen und manuell von handschriftlichen Zetteln erstellt. Als Lemmatisierer für Altfranzösisch kommt das selber entwickelte regelbasierte Programm „Lemming“ zum Einsatz, dessen Ergebnisse dann mit den Lemmata des Tobler-Lommatzsch abgeglichen werden. Es gibt im Projekt außerdem eine Rückkopplung mit französischen Urkunden und diesbezüglich eine Kooperation mit der Universität Zürich. Korpuslinguistische Informationen werden speziell farblich ausgezeichnet. Das System wird als sehr benutzerfreundlich beschrieben und soll in Zukunft auch in anderen Projekten der Heidelberger Akademie Anwendung finden (z.B. FEW = Französisches Etymologisches Wörterbuch).

Zum Abschluss der Sektion stellten NIKOLAI SOLMSDORF und OLIVER VON CRIEGERN (München) das Redaktionssystem des Wörterbuchs der tibetischen Schriftsprache vor. Bislang wurde dort eine Desktopanwendung verwendet, die auf Clearlook basiert. In Zukunft soll allerdings auf ein webbasiertes System umgestellt werden, das zu Zeit mit

dem Web2py Framework entwickelt wird. Eine Herausforderung sind dabei die Besonderheiten der Tibetischen Sprache.

Die IV. Sektion begann mit einem Vortrag zur digitalen Belegpraxis des Deutschen Rechtswörterbuch durch ALMUTH BEDENBENDER (Heidelberg). Sie konzentrierte sich dabei auf die Darstellung der Nutzung digitaler Ressourcen für die Erstellung des Deutschen Rechtswörterbuchs, insbesondere des Quellenkorpus mit ca. 8.500 Siglen. Die Verlinkung zu den Belegen wurde mit dem eingesetzten FAUST Datenbanksystem in einem beeindruckenden Umfang realisiert. Ca. 500.000 Belege insgesamt, von denen mehr als 300.000 mit einem Digitalisat verknüpft sind (Textarchiv, Faksimile oder URL-Verknüpfung), werden bereitgestellt. Zurück in den Bereich der Webportale führte der Beitrag von INES GRUND und JOACHIM BERGER (Mainz), die „Europäische Geschichte Online: Einsatz und Weiterentwicklung eines Content Management Systems für eine interdisziplinäre Geschichte Europas“ (Ieg-ego.eu) präsentierten. Im Hintergrund des Portals, das mit Plone/Zope realisiert wird, steht Thomas Burch, der selber aber diesmal nicht an der Tagung teilnahm. Bislang finden sich auf dem Portal ca. 200 Originalbeiträge und 150 Übersetzungen. Die Oberfläche steht durchgängig auf Deutsch und Englisch zur Verfügung. Besonders interessant hierbei könnte auch für andere Projekte eine selber entwickelte Fußnotenfunktion für den TinyMCE Editor sein, die es ermöglicht, Fußnoten aus eingehenden Manuskripten weitgehend automatisiert in das CMS zu übernehmen. Vom Redaktionsteam in Mainz wird eine intensive Verlinkung der Beiträge innerhalb des Systems geleistet.

Die fünfte und abschließende Sektion stellte am Ende nochmal drei im weiteren Sinne „virtuelle Forschungsumgebungen“ vor, die generisch eine Reihe von Arbeitsprozessen im Umfeld von Projekten abdecken und auch kollaborativ angelegt sind. Es handelt sich zum einen um FUD, ein Datenbanksystem des schon erwähnten Thomas Burch aus Trier, dann SADE von der Berlin-Brandenburgische Akademie der Wissenschaften (BBAW) und schließlich WissKI aus Erlangen/Nürnberg. Der Beitrag von Alexander Czmil zu Sade

fiel allerdings aufgrund einer Erkrankung aus, so dass GERALD NEUMANN (Berlin) stattdessen kurz die Strategie der BBAW in Hinblick auf VREs und Redaktionssysteme erläuterte. VIVIEN FRIEDRICH und FRANK QUEENS (Wuppertal / Trier) stellten zunächst das Projekt „Arthur Schnitzler digital“ vor und erläuterten die dortigen Arbeitsprozesse und Werkzeuge. Der zu erfassende Nachlass befindet sich an der Universitätsbibliothek Cambridge, die auch die langfristige Bereitstellung der digitalen Edition zugesagt hat. Neben dem schon erwähnten FUD, das ursprünglich für den SFB 600 „Fremdheit und Armut“ in Trier entwickelt worden war, präsentierten die Mitarbeiter auch die Transkriptionsumgebung „Transcribo“, ein Eclipse-Projekt mit einem sehr graphisch gestalteten User-Interface. Es beinhaltet auch ein Kollationierungswerkzeug sowie ein TEI-Vergleichstool mit einem zweistufigen, halbautomatischen Verfahren auf der Basis von CollateX. Die Transkriptionsumgebung ist schon als Prototyp vorhanden und soll demnächst produktiv eingesetzt werden.

Zum Abschluss stellten GEORG HOHMANN (München) und JÖRG WETTLAUER (Göttingen) die wissenschaftliche Informations- und Kommunikationsumgebung (WissKI) vor und erläuterten eine der Anwendungsmöglichkeiten des Systems anhand des Projekts „Semantic Blumenbach“ der Akademie der Wissenschaften zu Göttingen. WissKI basiert auf dem CMS Drupal und stellt über Module eine Arbeitsumgebung auf Basis von Semantic Web Technologien bereit, die bislang vor allem im Kontext von Museumsprojekten eingesetzt wird. Als upper-level Ontologie wird in Erlangen CRM eingesetzt, eine DL-OWL Variante des CIDOC Conceptual Reference Models. In Semantic Blumenbach wurde versucht, die Funktionalität des Systems auf die Verwaltung größerer Mengen von TEI-Texten zu erweitern und damit die Beziehung zwischen Texten und Objekten in dem Material von Blumenbach-Online abzubilden.

Die Tagung zeigte ein weiteres Mal eindrücklich, dass auch heute noch an verschiedenen Stellen Spezialsoftware für Forschungsprojekte im Einsatz ist, obwohl schon längst Open-Source Lösungen mit ähnlicher

Funktionalität existieren. Es sind auch immer noch Parallelentwicklungen, vor allem im Bereich der Redaktions- und Präsentationsumgebungen zu beobachten. Obwohl weitgehend Einigkeit darüber bestand, dass die klassische VRE nur mehr in der Form des Werkzeugkastens eine Zukunft besitzt, ist die Zahl der Speziallösungen beeindruckend. Umso deutlicher drängt sich die Notwendigkeit des Austauschs und der Kooperation der verschiedenen Akademien auf, damit in Zukunft Parallelentwicklungen vermieden und Synergien besser ausgenutzt werden können. Als Thema für das nächste Treffen in Heidelberg wurde „Historische Semantik und Korpus-/Wörterbucharbeit“ ausgewählt.²

Konferenzübersicht:

I. Sektion

Michael Pietrowski (Mainz), Virtuelle Forschungsumgebungen: Standardsoftware oder Werkzeugkasten?

Franziska Horn / Sandra Denzer (Darmstadt), Dominik Kasper / Max Grüntgens / Torsten Schrade (Mainz), XML basiert editieren, Content Management gestützt publizieren: Integrierte Forschungs- und Editionssumgebungen in den Projekten „Deutsche Inschriften Online“ und „Digitales Familiennamenwörterbuch Deutschlands“

Andreas Kuczera (Gießen) / Stefan Dumont (Berlin), Digitale Arbeitsumgebung „ediarum“ am Beispiel der Regesten Kaiser Friedrichs III.

Ute Recker-Hamm (Mainz), Das Redaktionssystem für das Mittelhochdeutsche Wörterbuch: Erfahrungen und Wünsche nach 12 Jahren elektronischer Artikellarbeit

II. Sektion

Ulrich Schmid (Münster), Virtueller Handschriftenlesensaal

Jürgen Herold (Greifswald/Göttingen), Epigrafi: Redaktionssystem für „Die Deutschen Inschriften“

²Eine Übersicht zur Tagung und einige Folien der Beiträge können über die Adresse: <http://www.gcdh.de/agep_ws2014/doku.php?id=wiki:welcome> (25.11.2014) abgerufen werden.

Jonathan Groß (Düsseldorf/Göttingen), Kleine und fragmentarische Historiker der Spätantike: Redaktionssystem

Abendvortrag

Eckhart Arnold (München), Web-basierte Redaktionssysteme für die Geisteswissenschaften. Architekturen und Trends

III. Sektion

Alexander Geyken (Berlin), Verwaltung, Bearbeitung und Publikation mehrerer Wörterbuchsubstanzen in Einem: das Redaktionssystem des Digitalen Wörterbuchs der deutschen Sprache

Stephen Dörr (Heidelberg), Etymologisches Wörterbuch des Altfranzösischen: Redaktionssystem

Nikolai Solmsdorf / Oliver von Criegern (München), Wörterbuch der tibetischen Schriftsprache: Redaktionssystem

IV. Sektion

Almuth Bedenbender (Heidelberg), Die Nutzung digitaler Ressourcen für die Erstellung des Deutschen Rechtswörterbuchs

Ines Grund / Joachim Berger (Mainz), Europäische Geschichte Online: Einsatz und Weiterentwicklung eines Content Management Systems für eine interdisziplinäre Geschichte Europas

V. Sektion

Vivien Friedrich / Frank Queens (Wuppertal / Trier), Arthur Schnitzler digital: Arbeitsprozesse und Werkzeuge

Georg Hohmann (München) / Jörg Wettlaufer (Göttingen), Semantische VRE WissKI

Tagungsbericht *Redaktionssysteme und Virtuelle Forschungsumgebungen.* 20.10.2014–21.10.2014, München, in: H-Soz-Kult 04.12.2014.