

Mapping the Landscape of eResearch. Text – Image – Annotation

Veranstalter: Max Planck Digital Library

Datum, Ort: 22.02.2012-23.02.2012, Berlin

Bericht von: Jörg Wettlaufer, Göttingen Centre for Digital Humanities (GCDH)/Akademie der Wissenschaften (ADWG) zu Göttingen

Der Workshop „Mapping the Landscape of eResearch. Text – Image – Annotation“ fand am 22. und 23. Februar 2012 im Harnack Haus, der Tagungsstätte der Max-Planck-Institute, in Berlin statt. Er war etwa in der Mitte der Projektlaufzeit der *Digitization Lifecycle*-Initiative der MPDL angesiedelt, die von vier geisteswissenschaftlichen Instituten, darunter die MPIs für Rechtsgeschichte in Frankfurt am Main und für Bildungsforschung in Berlin sowie die MPIs in Florenz (Kunstgeschichte) und Rom (Biblioteca Hertziana) sowie einigen assoziierten Partnern zusammen mit der Max-Planck Digital Library, getragen wird. Ziel des auf insgesamt zwei Jahre angelegten Projekts ist es, digitale Ressourcen für die Veröffentlichung von Primärdaten und Forschungsergebnissen in einer Web-basierten Umgebung für die beteiligten Institute zu entwickeln und öffentlich verfügbar zu machen. Dabei soll ein Set von Tools etabliert werden, die einen effizienten Workflow ermöglichen. Diese Tools sollen am Ende auch außerhalb der MPDL als Open Source nachnutzbar sein. Damit wird das Projekt grundsätzlich auch Elemente einer virtuellen Forschungsumgebung enthalten, z.B. Annotationsmöglichkeiten für Text und Bilder, kollaborative Arbeitsumgebungen sowie Meta-suchen über heterogene Ressourcen und deren Verlinkung.

Der Workshop war in die drei im Titel genannten Teilaspekte aufgefächert. Die geplante Einführung in das Gesamtprojekt durch KARL HÄRTER (MPI Rechtsgeschichte, Frankfurt am Main) nach der Eröffnung des Workshops durch Andrea Kulas, der Projektkoordinatorin, musste aufgrund des Streiks am Frankfurter Flughafen entfallen. Ebenso konnte Christoph Ringstetter aus München aus familiären Gründen nicht wie geplant anreisen, so dass sein Vortrag über

„Improving Access to Historical Documents. Special Lexica and Ressources“ leider ausfallen musste. Alle Vorträge und die Diskussionen wurden in englischer Sprache gehalten. MALTE DREYER, Leiter der Abteilung Forschung und Entwicklung der MPDL, sowie JAN SIMANE vom MPI in Florenz präsentierten zunächst das *Digitization-Lifecycle*-Projekt in seinen technischen und wissenschaftlichen Umrissen. Neben der Etablierung von Tools zur Metadatenauszeichnung, dem Vergleich und der gemeinsamen Nutzung von Texten und Bildern im Rahmen eines Rechte-Managements und von Werkzeugen zur kollaborativen Zusammenarbeit sind auch Metasuchsysteme sowie Indexwerkzeuge geplant, mit der heterogene Sammlungen verbunden und gemeinsam nutzbar gemacht werden. Außer der Entwicklung neuer Werkzeuge will das Projekt jedoch auch Richtlinien für die Digitalisierung innerhalb der MPG zur Verfügung stellen, die in Anlehnung an die DFG-Praxisregeln „Digitalisierung“ (Stand: April 2009) einzelnen Projekten die Arbeit erleichtern sollen.

SEBASTIAN RAHTZ (Oxford) von den Oxford University Computer Services teilte im eröffnenden Referat im Stil einer keynote seine Befürchtungen und Hoffnungen hinsichtlich der Interoperabilität und Austauschbarkeit des Textauszeichnungsstandards der Text Encoding Initiative (TEI) mit. Aufgrund seiner langjährigen Erfahrung in diesem Bereich stellte er die grundsätzlich Verwendbarkeit des Standards für Textsammlungen heterogenen (Kodierungs-)Ursprungs in Frage, da die TEI inzwischen eine Vielzahl von Kodierungsmöglichkeiten für ähnliche Textbestandteile zur Verfügung stellt, die eine Verwendung der Texte bzw. Auszeichnungen über Einzelprojekte hinaus erschweren und damit eines der Hauptanliegen der Standardisierung unterlaufen. Deshalb plädierte er für eine Lösung des Problems über die Verwendung von Schemata im ODD (One Document does it all) Format, in dem Informationen über die jeweils verwendete Teilmenge des Markups zur Verfügung gestellt werden. In der sich anschließenden Sektion über Bilder in digitalen Forschungsumgebungen stellte nach der Einführung durch Anette Creutzburg zunächst UTE DERCKS (Flo-

renz) das Projekt CENOBIUM (Multimediale Darstellung romanischer Kreuzgangkapitel im Mittelmeerraum) vor, in dem hochauflösende 3D-Ansichten romanischer Kapelle über das Internet zur Verfügung gestellt werden. Die beeindruckende Life-Präsentation des Portals verdeutlichte die Möglichkeiten aber auch zugleich die Kosten einer solchen Darstellungsweise, da die Herstellung der für die 3D-Animationen notwendigen Photographien und Laservermessungen sehr teuer ist. Anschließend berichtete MARTIN WARNEKE (Lüneburg) über die Projekte *hyperimage.eu* und *meta-image.de*, wobei letzteres im engeren Umfeld der Prometheus Bilddatenbank realisiert wurde. Es handelt sich um ein Tool zur kollaborativen Zusammenarbeit bei der Identifikation von Motiven und Bildern und deren Verlinkung sowie der Beifügung von Metadaten zu Bildern. Erneut zum Thema Text sprach am nächsten Tag GEORG VOGELER (Graz) am Beispiel des Projekts *monasterium.net*, in dem eine Vielzahl von europäischen Archiven ihre digitalisierten Bestände anbieten. Darüber hinaus verwendet *monasterium.net* eine eigene virtuelle Forschungsumgebung für historische Urkunden und Akten, in der über einen online XML-Editor (Daniel Ebner) Transkriptionen von Urkunden unterstützt werden sowie über ein Bild-tool (André Streicher) Metadaten einzelnen Bildabschnitten zugewiesen werden können. CHRISTIAN THOMAS von der Berlin-Brandenburgischen Akademie der Wissenschaften stellte anschließend das Deutsche Text Archiv (DTA) vor, dessen Fokus auf der linguistischen Analyse von gedruckten deutschsprachigen Texten aus der Zeit zwischen 1650 und 1900 liegt. Das DTA promotet ein eigenes XML-basiertes Basisformat, das aus 80 Elementen der TEI P5 besteht und auf diese Weise den Problemen begegnet, die Sebastian Rahtz in seinem einführenden Vortrag thematisiert hatte. Zum eigentlichen Textarchiv mit seinen Analysetools gibt es zwei interessante Erweiterungen, DTAE und DTAQ. Die Akronyme stehen zum einen für ein Erweiterungsprojekt des Korpus und zum andern für ein Tool zur Qualitätskontrolle der Erfassungsgenauigkeit der Texte.

Die thematisch dritte Sektion des Workshops beschäftigte sich mit verschiedenen

Annotationsmöglichkeiten an Texten, Bildern bzw. digitalen Medien im Allgemeinen. Im inzwischen abgeschlossenen Projekt *Sanddrart.net* (Laufzeit von 2007 bis März 2012), das eine digitale Edition von Joachim von Sandrarts kunsttheoretischen Schriften (Teutsche Academie der Bau-, Bild- und Mahlerey-Künste) des späten 17. Jahrhunderts online bereitstellt, wurden von CARSTEN BLÜM (Frankfurt am Main) Annotationen im Kontext von Metadatenanreicherung und tiefe Verlinkung von Text im Rahmen dieses Projekts thematisiert. Dieses breite, generische Verständnis von Annotation findet auch in den von den von ERHARD HINRICHS und KATHRIN BECK (beide Tübingen) vorgestellten WebLicht-Tools zur webgestützten linguistischen Textanalyse und Annotation Anwendung. Die verteilt angelegte und weit entwickelte computerlinguistische Tool-Sammlung – *WebLicht: Web-basierte LRT Services* – ermöglicht die eigene Zusammenstellung von Workflows zu einem Text, in denen bspw. ein Tokenizer, ein Lemmatizer, ein Named Entity Tool und viele weitere Werkzeuge zur Anreicherung mit Metadaten und semantischen Auszeichnungen angewendet werden können. Eine Ausweitung von Annotationen auf verschiedene Arten von Medien, die auch Ton und Videodokumente umfassen können, stellte anschließend RAINER SIMON (Wien) vom Österreichischen Institut für Technologie in Wien vor. Das YUMA (Universal Media Annotator) genannte Werkzeug, dessen Prototyp im Rahmen des EU-finanzierten EuropeanaConnect Projekts entstand und sich momentan auf dem Weg zu einem Open-Source Programm befindet, umfasst im wesentlichen JavaScript Bibliotheken, die einfach in eigene Webseiten eingebunden und dort zur Annotation von verschiedenen Medientypen eingesetzt werden können. Der letzte Vortrag der Sektion von GEORG SCHELBERT (Berlin) präsentierte am Beispiel der Projekte *Lineamenta* (Eine Forschungsdatenbank für Architekturzeichnungen) und *Cipro* (eine online-Präsentation historischer Pläne der Stadt Rom) die Möglichkeiten der Georeferenzierung im kultur- und kunsthistorischen Kontext. Mit dem Projekt *Zuccaro* sei hier zudem eine virtuelle Vernetzung der Informationsressourcen im Rahmen eines Zo-

pe/Plone Frameworks realisiert worden. Am Ende des Workshops gab MALTE DREYER (München) eine kurze Zusammenfassung der Ergebnisse und einen Ausblick auf die geplanten Schwerpunkte der *Digitization Lifecycle*-Initiative.

Man wird auf die Ergebnisse des Projekts gespannt sein dürfen. Die PPT-Präsentationen sind inzwischen auf der Webseite des Workshops verfügbar (<http://www.mpdl.mpg.de/main/landscapeofresearch.htm>¹). Die geplante Bereitstellung als Open Source ist vorbildlich und ermöglicht die Nachnutzung von Tools auch außerhalb der MPG. Auf der anderen Seite zeigte die Motivation des Workshops, dass es inzwischen schon rein quantitativ nicht mehr einfach ist, einen Überblick zu den existierenden Tools und Projekten zu erhalten, die in den Geisteswissenschaften in den letzten Jahren entwickelt wurden. Eine aktuelle, interdisziplinäre und möglichst auch kollaborativ verwaltete informative Übersicht, die auch Brückenfächer – z.B. die Archäologie – umfassen und möglichst international sein sollte, wäre daher wünschenswert.²

Konferenzübersicht:

Andrea Kulas (Max Planck Digital Library): Welcome

Introducing Digitization Lifecycle (DLC)

Malte Dreyer (Max Planck Digital Library): Technical Implications

Jan Simane (Kunsthistorisches Institut in Florenz, Max-Planck-Institut): Scientific Implications

Introduction: Text in DLC (Klaus E. Werner, Rom)

Sebastian Rahtz (Oxford University Computing Services): The TEI - private and public concerns

Introduction: Images in DLC (Anette Creutzburg, Florenz)

Ute Dercks (Photo Library of the Kunsthistorisches Instituts in Florenz, Max-Planck-Institut): Cutting-edge technology meets the Middle Ages. CENOBIUM - A Project for the Multimedia Representation of Romanesque Cloister Capitals in the Mediterranean Region

Martin Warnke (Leuphana University of Lüneburg, Institute for Culture and Aesthetics of Digital Media): Observations on Images

Introduction: Synthesis First Day, Text in DLC (Ingo Caesar, Frankfurt a. M.)

Georg Vogeler (University of Graz): Lessons from Monasterium.net: More Efficient Cooperation between Science and Cultural Heritage Institutions through Online Collaboration

Christian Thomas (Berlin-Brandenburg Academy of Sciences and Humanities): DTAE - Enlarging the Reference Corpus of the Deutsches Textarchiv (DTA) - Production, Conversion and Interchange of XML/TEI Encoded Full Text

Introduction: Annotations in DLC (Malte Dreyer, München)

Carsten Blüm (Goethe University Frankfurt): Sandrart.net - An Enriched Online Edition of a 17th Century Text

Erhard Hinrichs, Kathrin Beck (Eberhard Karls University Tübingen): Web-Based Linguistic Annotation - Current Practice and Future Directions

Rainer Simon (Austrian Institute of Technology): Collaborative Media Annotation with YUMA

Georg Schelbert (Humboldt-University Berlin): The Topography of Knowledge. On Georeferencing of Cultural History Data

Malte Dreyer (Max Planck Digital Library): Final Remarks and Farewell

Tagungsbericht *Mapping the Landscape of eResearch. Text – Image – Annotation.*

¹Die Links zu Präsentationen sind auf der Seite http://colab.mpdl.mpg.de/mediawiki/Digitization_Lifecycle_Mapping_the_Landscape_of_eResearch#Mapping_the_Landscape_of_eResearch unterhalb der Abstracts zu finden.

²Eine Sammlung von vor allem im anglophonen Sprachraum entstandenen Projekten und Tools findet sich unter der Adresse <http://www.arts-humanities.net>. Speziell für Deutschland existiert eine Projektübersicht unter <http://www.kulturerbe-digital.de>, die vor allem den Zeitraum bis 2008/2009 gut abdeckt. Weitere Informationen finden sich im Digital Research Tools Wiki (DiRT) und unter <http://www.projectbamboo.org>.

22.02.2012-23.02.2012, Berlin, in: H-Soz-u-
Kult 02.04.2012.