



Forschungsdaten in bester Lage: Nutzungsszenarien und Perspektiven digitaler Forschungsinfrastrukturen. DFG-Projekt „Digitaler Wissensspeicher“, Berlin-Brandenburgische Akademie der Wissenschaften, 05.04.2016–06.04.2016.

Reviewed by Ulrike Wuttke

Published on H-Soz-u-Kult (July, 2016)

Forschungsdaten in bester Lage: Nutzungsszenarien und Perspektiven digitaler Forschungsinfrastrukturen

Vom 5.-6. April 2016 fand in Berlin der Workshop „Forschungsdaten in bester Lage: Nutzungsszenarien und Perspektiven digitaler Forschungsinfrastrukturen“ statt. Er wurde ausgerichtet vom DFG-Projekt „Digitaler Wissensspeicher“ an der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW). In zwölf kurzen Vorträgen mit anschließenden Plenumsdiskussionen wurden die bisherigen Ergebnisse des Wissensspeicherprojekts sowie Zukunftsszenarien vorgestellt. Es waren Vertreterinnen und Vertreter thematisch relevanter Projekte und Infrastruktureinrichtungen als Sprecher geladen, um zentrale Fragen, die sich im Rahmen digitaler Forschungsinfrastrukturen stellen, im breiteren Kontext zu diskutieren: zur Bereitstellung, Integration und Vernetzung von großen heterogen und dezentral vorliegenden Datenbeständen sowie nach deren langfristiger Verfügbarkeit als zentrale Herausforderungen der Digitalisierung der Geisteswissenschaften und Digital Humanities.

Vier Themenbereiche prägten den Workshop: Erstens Forschungsinfrastrukturen, Forschungsdatenrepositorien und Metadatensuchmaschinen, zweitens die Nachhaltigkeit digitaler geisteswissenschaftlicher Forschung, insbesondere der Forschungsergebnisse und Infrastrukturen, bzw. die Aufgaben und Herausforderungen der Datenkuration, drittens Möglichkeiten der Präsentation und Visualisierung von Forschungsdaten in den Geisteswissenschaften und viertens rechtliche Fragen. Ä

Den ersten Schwerpunkt des Workshops bildete die Vorstellung von Forschungsinfrastrukturen, Repositorien und Metadatensuchmaschinen sowie der Erfahrungsaustausch über deren Rolle und die zu lösenden Herausforderungen. MAXI KINDLING (Berlin) stellte das „Registry of Research Data Repositories“ (re3data.org (05.06.2016).) vor. Bei diesem Service von DataCite (05.06.2016). handelt es sich um das weltweit umfangreichste Verzeichnis von Forschungsdatenrepositorien aus allen Forschungsbereichen. Zum Zeitpunkt des Vortrags waren rund ein Zehntel der insgesamt rund 1.500 verzeichneten Repositorien den Geisteswissenschaften zuzuordnen (davon schätzte Kindling den Anteil wirklich disziplinspezifischer Forschungsdatenzentren auf rund 50 Prozent, da Mehrfachnennungen möglich sind). Kindling stellte definitorische Schwierigkeiten bezüglich des Begriffs „Forschungsdatenrepositorium“ als kleinsten gemeinsamen Nenner habe man sich laut Kindling innerhalb von re3data.org auf folgende Definition geeinigt: „A research data repository is a subtype of a sustainable information infrastructure which provides long-term storage and access to research data [...]“. Die Begriffe Forschungsdaten und Langfristigkeit wurden im Plenum kontrovers diskutiert, sowie die trotz Standardisierungsbestrebungen auf nationaler und internationaler Ebene vorherrschende Vielfalt von Datentypen, Zugangsberechtigungen, verwendeten Datenlizenzen, Zertifikaten und Schnittstellen heraus. Besonders problematisch seien die unterschiedliche Meta-

datenqualität, der fehlende Zugang zu den in den Repositorien enthaltenen Daten und die unzulängliche Verwendung von PIDs.

SASCHA GRABSCH (Berlin), MARCO JÄRGENS (Berlin) und NIELS-OLIVER WALKOWSKI (Berlin) stellten in ihren Vorträgen den Digitalen Wissensspeicher (05.06.2016) der BBAW unter unterschiedlichen Gesichtspunkten vor. Sascha Grabsch gab eine allgemeine Einführung zum Wissensspeicher, der neben einem zentralen Zugang zu allen digitalen Ressourcen der BBAW auch eine semantische Verknüpfung dieser heterogenen Bestände anbietet. Momentan sind über eine Million digitale Ressourcen, die im Kontext von 170 verschiedenen Projekten, wie Akademienvorhaben, Drittmittelprojekten und Initiativen entstanden sind, indexiert und durch Metadaten beschrieben. Neben technischen Herausforderungen, wie der nachträglichen Entwicklung von Schnittstellen, stellte Grabsch insbesondere organisatorische und administrative Aspekte heraus. So waren mangels Dokumentation für die Erschließung der laufenden und beendeten digitalen Projekte der Akademie umfangreiche Gespräche mit (ehemaligen) Projektmitarbeiterinnen und -mitarbeitern erforderlich. Dazu kamen ressourcenintensive Umbaumaßnahmen an (laufenden) Projekten im Rahmen der Erschließung.

Auf die grundlegenden Technologien des Wissensspeichers als Infrastruktur zur Integration von zunächst heterogenen digitalen Ressourcen ging Marco Jürgens ein. Zu technischen Details siehe: (05.06.2016). Besonders hervorzuheben sind die eine Volltextsuche weit übersteigenden Such- und Präsentationsfunktionen, die u.a. auf Apache Lucene, Donatus (05.06.2016) und DBPedia Spotlight (05.06.2016) beruhen. Weitere Entwicklungsarbeit ist hinsichtlich der Automatisierung und der Workflows notwendig. Niels-Oliver Walkowski hinterfragte in seinem Vortrag den Wissensbegriff im Kontext des Wissensspeichers (Ist Wissen die Repräsentation eines Diskurses durch die Verknüpfung digitaler Ressourcen?) und gab einen Ausblick auf die weitere Entwicklung des Projekts im Rahmen der Modellierung geisteswissenschaftlicher Forschungsprozesse und der Dokumentation von Wissenschaftsprozessen (SDM Siehe u.a. Steffen Hennicke (et al.), D3.4: Research Report on DH Scholarly Primitives, 2015 (05.06.2016) und NeMO Siehe (05.06.2016) und Luise Borek (et al.) TaDiRAH: a Case Study in Pragmatic Classification, DHQ 10 (2016):1). In den Diskussionen spielten die Rolle des Wissensspeichers als Endbaustein einer Veröffentlichungskette, d.h. als Teil der wissenschaft-

lichen Kommunikation, und für die Vermittlung von Best Practices eine wichtige Rolle, sowie die komplexen Frage nach dem wissenschaftlichen Mehrwert des Wissensspeichers und seiner Zielgruppe.

Die technische Umsetzung sowie der wissenschaftliche Mehrwert einer Metasuchmaschine standen auch im Mittelpunkt der Vorstellung der Judaica-Suchmaschine durch HARALD LORDICK (Essen). (05.06.2016). Lordick argumentierte, dass konfigurierbare nutzerzentrierte Suchmöglichkeiten aus den vorhandenen Suchinfrastrukturen heraus relativ einfach zu entwickeln seien. Der wissenschaftliche Mehrwert bestehe im Vergleich zu kommerziellen Anbietern wie Google in der Transparenz über den Datenraum und der Kuration durch Fachwissenschaftlerinnen und -wissenschaftler, die zu einer weitaus höheren Relevanz der Treffer führe. Das Konzept wird momentan im Rahmen von DARIAH-DE als Branded Search Siehe u.a. Tobias Gradl, Andreas Henrich, Christoph Plutte, Heterogene Daten in den Digital Humanities: Eine Architektur zur forschungsorientierten Federation von Kollektionen. In: Grenzen und Möglichkeiten der Digital Humanities. Hg. von Constanze Baum / Thomas Stäcker. 2015 (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 1). DOI: (05.06.2016). weiterentwickelt. JÄRG LEHMANN (Berlin, Bern) stellte die *lessons learned* bezüglich der Datenakquise und der technischen Umsetzung beim Aufbau von CENDARI (05.06.2016) vor, einer virtuellen Forschungsinfrastruktur für die europaweite Recherche in Archivbeständen zum Mittelalter und zum Ersten Weltkrieg. Siehe Jakub Beneš (et al.), The Cendari White Book of Archives, 2016 <http://www.cendari.eu/thematic-research-guides/white-book-archives> (05.06.2016). LISA DIECKMANN (Köln) widmete ihren Beitrag den Herausforderungen bei der Zusammenführung heterogener kunst- und kulturhistorischer Daten am Beispiel des digitalen Bildarchivs prometheus. (05.06.2016). Im Mittelpunkt der Diskussion stand vor allem die heterogene Metadatenqualität bzw. deren Verbesserung durch die verstärkte Einbindung von Verantwortlichen und Normdaten, z.B. der GND.

Der zweite Themenbereich widmete sich der Nachhaltigkeit von Forschungsinfrastrukturen und der Forderung nach Institutionen oder Dienstleistern, die sich nach Ablauf der Entwicklungsphase um die Pflege und den Betrieb der Anwendungen kümmern, bzw. die Sicherstellung der langfristigen Verfügbarkeit der digitalen Forschungsdaten über die Projektlaufzeit hinaus. JOHANNES STIGLER (Graz) stellte hierzu vier Thesen auf: Ers-

tens muss die Datenkuration von Anfang an mitbedacht werden; zweitens sollte die Repräsentationsschicht nicht programmiert, sondern modelliert werden; drittens muss eine technische Standardisierung erfolgen, z. B. durch die Etablierung von Workflows und viertens müssen Lösungen für die Langzeitarchivierung gefunden werden. Auch wenn einige dieser Thesen wie Zukunftsmusik klingen mögen und im Plenum kritisch diskutiert wurden – vor allem die „on the fly“-Modellierung der Ansichten in GAMS (05.06.2016). GAMS steht als österreichischer Beitrag zu DARIAH als Open Source Lösung zur Nachnutzung zur Verfügung (05.06.2016). stieß bei den Fachwissenschaftlerinnen und -wissenschaftlern auf Widerstand – legen sie dennoch den Finger in die richtige Wunde. Die in den digitalen Geisteswissenschaften noch immer vorherrschenden proprietären Lösungen können nur durch institutionelle Strategien und einen möglichst breiten Erfahrungsaustausch aller relevanter Stakeholder, einschließlich der Fachcommunity, in nachhaltigere Lösungen und Strukturen überführt werden. Über Strategien zur Langzeitarchivierung und Datenkuration referierten auch FELIX SCHÄFER (Berlin), der die Arbeitsschritte für die Übernahme von Forschungsdaten in IANUS (05.06.2016) vorstellte, und PATRICK SAHLE (Köln) für das Kölner Data Center for the Humanities (DCH) (05.06.2016). Da IANUS davon ausgeht, dass es als disziplinspezifisches, zentrales Forschungsdatenzentrum im Gegensatz zu institutionellen Forschungsdatenzentren wahrscheinlich erst sehr spät kontaktiert wird, wird großer Wert auf die Bereitstellungen von Informationen zur Sicherung der Nachhaltigkeit der Forschungsdaten gelegt, die weit über die engere IANUS-Community als repräsentativ gelten dürften. (05.06.2016). Neben einer allgemeinen Vorstellung des Schichtenmodells des DCH stellte Sahle das „Resource Description Schema“ des DCH vor, wobei es im Plenum zu einer angeregten Diskussion zur Definition des Terminus „Resource“ kam. Die Folien zu diesem Vortrag finden sich unter: (05.06.2016). Sahle argumentierte auch, dass die oftmals geforderten zehn Jahre Langzeitarchivierungsfrist für Forschungsdaten eher ein naturwissenschaftliches Paradigma seien. In den Geisteswissenschaften, in denen Forschungsdaten eine andere Rolle spielen, gelten andere zeitliche Paradigmen (wenn auch das einmal von Klaus Graf „until judgment day“ formulierte Paradigma etwas weit geht). Gerade die Ergebnisse geisteswissenschaftlicher Grundlagenforschung (wie Editionen, Wörterbücher, Personen-, Orte-, und Sachdatenbanken) haben eine sehr lange Halbwertszeit.

Mit den Möglichkeiten der Präsentation und Visualisierung von Forschungsdaten in den Geisteswissenschaften wurde im Verlauf des Workshops ein zentrales Thema der diesjährigen Tagung des DHD-Verbandes in Leipzig (05.06.2016) aufgegriffen. Insbesondere der Beitrag von MARIAN DÄRK (Potsdam) war Visualisierungen als Mittel zur Analyse und Interpretation wachsender Informationsräume in den Geisteswissenschaften und ihrem Potential während des Forschungsprozesses gewidmet. Ziel von DÄRKs Projekt „Visualising cultural collections“ (05.06.2016) ist – wie nicht zuletzt des Digitalen Wissensspeichers auch – die Erforschung der Möglichkeiten der explorativen Sichtung digitaler Sammlungen bei großer Transparenz des Datenraums. Im Mittelpunkt der Diskussion standen Fragen der Nachhaltigkeit von Visualisierungstools und der vorgestellten Visualisierungen sowie die Entwicklung von Komponenten des Nutzertrackings, d.h. der Nachvollziehbarkeit beliebiger Wege durch digitale Sammlungen.

Wie in allen Diskussionen zur Verfügbarmachung von Forschungsdaten in den digitalen Geisteswissenschaften spielten auch die rechtlichen Rahmenbedingungen eine wichtige Rolle. Das fachliche „Know-how“ lieferte PAUL KLIMPEL (Berlin) mit einem kurzen historischen Abriss über die Entwicklung des Urheberrechts und einem Plädoyer für die großzügige Verwendung von CC-0 Lizenzen zumindest für Metadaten und Beschreibungstexte, um die Nachnutzung von Forschungsdaten zu erleichtern. Siehe z.B. John H. Weitzmann und Paul Klimpel, Handreichung Recht (1.2), 2016 (05.06.2016). Des Weiteren wies Klimpel darauf hin, dass die Regeln des Urheberrechts und die Verwendung offener Lizenzen in keinem intrinsischen Zusammenhang mit dem Schutz des geistigen Eigentums im wissenschaftlichen Sinne stehen. Diese werden durch die gute wissenschaftliche Praxis, das heißt der Pflicht zur Kennzeichnung einer Quelle zur Genüge abgedeckt.

Zum Abschluss wurden durch die Teilnehmerinnen und Teilnehmer des Workshops an fünf thematischen Fragen zur Datenkuration, Visualisierungen, Technik/Infrastruktur/Frameworks, Langzeitarchivierung sowie Infrastruktureinheiten und Forschungsdaten in kleineren Runden diskutiert. Die im Plenum vorgestellten Ergebnisse der Diskussionen überlappten sich teilweise oder ergänzten sich. So wurden beispielsweise bei der Diskussion über die Aufgaben der Datenkuration und des Berufsbilds des „Data Curators“ Parallelen zu den Ergebnissen der thematischen „Langzeitarchivierung“ und „Infrastruktureinheiten“ gezogen. Die Diskussion am Thementisch „Visualisierung“ setzte sich kri-

tisch mit den Potenzialen und Risiken der Visualisierung von Forschungsdaten auseinander. Der Thementisch âTechnik/Infrastruktur/Frameworksâ stellte technische Schwierigkeiten sowie Erfolgskriterien f r die Technikentwicklung vor, wobei sich  berschneidungen mit den Diskussionen an den Thementischen âLangzeitarchivierungâ und âInfrastruktureinheitenâ ergaben. Ein wichtiger roter Faden, war die Notwendigkeit der Sensibilisierung der Wissenschaftlerinnen und Wissenschaftler f r die Nachhaltigkeit ihrer digitalen Forschung und das Bewusstsein Infrastrukturprojekte, aber auch Bibliotheken und Rechenzentren, als kompetente Partner in diesem Bereich wahrzunehmen.

Weitere wichtiges Themen des Workshops waren die bisher ungel ste Frage der Kassation (Was wollen wir aufheben und wie lange?) und die ungewisse Zukunft vieler (auf Projektbasis gef rdeter) geisteswissenschaftlicher Infrastrukturen und Datenzentren, die nicht dazu beitr gt, das Vertrauen der Fachwissenschaftlerinnen und âwissenschaftler in die Zukunftstr chtigkeit der digitalen Forschung zu best rken. Eine wichtige Aufgabe f r die Zukunft ist somit die Gew hrleistung der Nachhaltigkeit digitaler geisteswissenschaftlicher Infrastrukturen und Datenzentren. Wenn dieses Problem ungel st bleibt, werden die Repositorien, Metasuchmaschinen oder avancierten Forschungsumgebungen wie der Digitale Wissensspeicher nur kurz in die Vergangenheit reichen oder komplett verschwinden.

Vorhandene kommerzielle Suchmaschinen k nnen (und wollen) wissenschaftliche Forschungsdaten nicht den wissenschaftlichen Bed rfnissen entsprechend abbilden und zug nglich machen. Eine Zusammenf hrung und Vernetzung von Forschungsdaten durch die nachhaltige Einrichtung von institutionell gest tzten Forschungsinfrastrukturen ist daher unabdingbar. Dabei gilt es ein besonderes Augenmerk auf die Datenkuratation zu legen, um den Zugriff, die Auffindbarkeit und damit die Sichtbarkeit der digitalen Ressourcen langfristig zu gew hrleisten. Hier hilft der Einsatz von technischen Standards und standardisierten Workflows, aber auch die attraktive Pr sentation mittels innovativer Visualisierungsmethoden. Die kuratierten Daten und die entwickelten Softwarekomponenten der Forschungsinfrastrukturen m ssen frei zur Verf gung stehen und ohne H rden nachnutzbar sein, um eine m glichst breite Akzeptanz innerhalb der wissenschaftlichen Community zu erlangen.

Konferenz bersicht:

Maxi Kindling (Institut f r Bibliotheks- und Informationswissenschaft, Humboldt-Universit t zu Berlin),  berblick  ber die Landschaft der Forschungsdaten-Repositorien auf Basis von re3data.org

Sascha Grabsch (Digitaler Wissensspeicher, BBAW), Der Digitale Wissensspeicher als Baustein in der Publikationskette digitaler Forschungsdaten

Johannes Hubert Stigler (Zentrum f r Informationsmodellierung, Austrian Centre for Digital Humanities, Universit t Graz), Die Sicherstellung der Verf gbarkeit von Forschungsdaten als Aufgabe von Langzeitarchivierung am Beispiel eines FEDORA-basierten Repositoriums

Felix Sch fer (IANUS, Deutsches Arch ologisches Institut Berlin), Vom Produzenten ins Archiv. Arbeitsschritte zur Datenkuratierung bei IANUS

Patrick Sahle (Cologne Center for eHumanities, Universit t K ln), Daten bernahme und âResource Description Schemaâ im K lner Data Center for the Humanities (DCH)

Marco J rgens (Digitaler Wissensspeicher, BBAW), Lessons Learned: Erfahrungen aus der Entwicklung des Digitalen Wissensspeichers

Paul Klimpel (iRights-law), Wissenschaftliche Freiheit und rechtliche Fu ngeln

Harald Lordick (Salomon Ludwig Steinheim-Institut f r deutsch-j dische Geschichte, Universit t Duisburg-Essen), M glichkeiten und Grenzen einer Judaica-Suchmaschine im digitalen Forschungsumfeld

J rg Lehmann (FU Berlin, Universit t Bern), Das Rezept der CENDARI Data Soup (CENDARI White Book of Archives)

Niels-Oliver Walkowski (TELOTA, BBAW), Integration, Interaktion, Evaluation: Neue M glichkeiten f r die Analyse digital unterst tzter Forschungsaktivit ten

Marian D rk (Institut f r Angewandte Forschung Urbane Zukunft, FH Potsdam), Von Repr sentation zu Interpretation: Visualisierung in den Geisteswissenschaften

Lisa Dieckmann (Kunsthistorisches Institut, Universit t K ln), Von Malerei bis Performancekunst â  ber M glichkeiten und Herausforderungen bei der Zusammenf hrung heterogener kunst- und kulturhisto-

rischer Daten Technik/Infrastruktur/Frameworks (Josef Willenborg)
Diskussion an Thementischen: Langzeitarchivierung (Hubert Stigler)
Data Curation (Alexander Czmiel) Infrastruktureinheiten/Archive/Bibliotheken und
Vernetzung Kunstgeschichte/Visualisierung (Lisa Dieckmann) Forschungsdaten (Markus Schnäpf)

If there is additional discussion of this review, you may access it through the network, at:

<http://hsozkult.geschichte.hu-berlin.de/>

Citation: Ulrike Wuttke. Review of , *Forschungsdaten in bester Lage: Nutzungsszenarien und Perspektiven digitaler Forschungsinfrastrukturen*. H-Soz-u-Kult, H-Net Reviews. July, 2016.

URL: <http://www.h-net.org/reviews/showrev.php?id=47409>

Copyright © 2016 by H-Net, Clio-online, and the author, all rights reserved. This work may be copied and redistributed for non-commercial, educational purposes, if permission is granted by the author and usage right holders. For permission please contact H-SOZ-U-KULT@H-NET.MSU.EDU.