

Gute Forschungsdaten, bessere Forschung: wie Forschung durch Forschungsdatenmanagement unterstützt wird

Mache, Beata

mache@sub.uni-goettingen.de
Staats- und Universitätsbibliothek Göttingen

Trippel, Thorsten

thorsten.trippel@uni-tuebingen.de
Eberhard Karls Universität Tübingen, Deutschland

Effinger, Maria

effinger@ub.uni-heidelberg.de
Universitätsbibliothek Heidelberg

Gradl, Tobias

tobias.gradl@uni-bamberg.de
Otto-Friedrich-Universität Bamberg

Haaf, Susanne

haaf@bbaw.de
Berlin-Brandenburgische Akademie der Wissenschaften

Hinrichs, Erhard

erhard.hinrichs@uni-tuebingen.de
Eberhard Karls Universität Tübingen, Deutschland

Horstmann, Wolfram

horstmann@sub.uni-goettingen.de
Staats- und Universitätsbibliothek Göttingen

Müller, Lydia

lydia@informatik.uni-leipzig.de
Universität Leipzig

Schrade, Torsten

torsten.schrade@adwmainz.de
Akademie der Wissenschaften und der Literatur Mainz

Teich, Elke

e.teich@mx.uni-saarland.de
Universität des Saarlandes

Kurzzusammenfassung

In diesem Panel geht es um die Förderung der geisteswissenschaftlichen Forschung durch eine planvolle Erhebung, Archivierung, Veröffentlichung und die dadurch ermöglichte Nachnutzung von Forschungsdaten, die sowohl zur Qualitätssicherung in der Forschung beitragen als auch nicht zuletzt neue Fragestellungen erlauben. Aus unterschiedlichen Perspektiven soll in dem Panel beleuchtet werden, welchen Mehrwert das Datenmanagement für die Forschung in den digitalen Geisteswissenschaften hat, wie man diesen Mehrwert erreicht und auch die Veröffentlichung der Forschungsdaten als ein selbstverständliches Element der Dissemination der Forschungsergebnisse etabliert und wie man gleichzeitig den Aufwand für die Forschung abschätzen kann.

Ziele des Panels

Vor dem Hintergrund der notwendigen Diskussion zum Management, zur Interoperabilität, Sicherung, Publikation und Zitation aber auch zur Auswahl und Kassation von Forschungsdaten wird das Panel Experten aus der Forschung, Entwicklung, aus wissenschaftsgeleiteten Forschungsinfrastrukturen sowie dem Bibliotheks- und Universitäts-Verlagswesen zusammenbringen. Im Fokus steht die kritische oder auch kontroverse Auseinandersetzung, welche Aufgaben Forschende selbst übernehmen können und wollen, welche Aufgaben bei digitalen Infrastrukturen anzusiedeln sind, welche gemeinsamen Strategien aller Mitwirkenden nötig sind, wo eine Parallelisierung von Entwicklungen vermieden werden muss und wo eine Vielfalt von Ansätzen wünschenswert wäre.

Dazu werden die Teilnehmenden des Panels anhand von vier Leitfragen aus ihrer jeweiligen Perspektive kurz Stellung zum Thema nehmen.

Leitfragen

- Welche spezifischen Anforderungen aus der geisteswissenschaftlichen Praxis beeinflussen das Datenmanagement?
- Wie beeinflusst die Datenveröffentlichung herkömmliche Publikationen und welche Datenmanagementoptionen sind für die Publikationen von Daten relevant?
- Wieweit können Forschungsdatenmanagementpläne die Forschenden dabei unterstützen, schon vom Moment der Formulierung der Forschungsfrage und der Beantragung eines Projektes an für die sichere forschungsbegleitende und forschungsabschließende Veröffentlichung der gesammelten, erhobenen und

erstellten Daten – möglichst den FAIR-Prinzipien konform – zu sorgen?

- Wie bestimmen wir die technischen Kriterien für die Sicherung und Interoperabilität der heterogenen Forschungsdaten, damit sie gesichert, gefunden und nachgenutzt werden können? Reichen die Kriterien zur Abschätzung des Aufwands aus?

Hintergrund

Geisteswissenschaftliche Forschung basiert darauf, Ideen und Konzepte anderer Forschender zu betrachten, zu kommentieren und weiter zu entwickeln. Im Rahmen des Paradigmenwechsels (siehe etwa Berry, 2011; Baum and Stäcker 2015 ; Thiel, 2012), der sich gerade auch in den digitalen Geisteswissenschaften durch teilautomatische Analysen (siehe bereits Busa, R., 1951),

Visualisierungen und Verknüpfungen unterschiedlicher Datentypen manifestiert, ist die Nachnutzung von Daten, Informationen und Konzepten von Anfang an ein integraler Bestandteil der Forschung gewesen (siehe etwa den Impact von Busa beschrieben von Winter, 1999; die Daten von Busa unter <http://www.corpusthomicum.org> arbeiten von Antonio Zampolli, 1973; etc.)

Die Veröffentlichung von Forschungsdaten selbst ist kein Novum - in Anhängen, in Tabellen, in Abbildungen wurden "Daten" analog als Merkmale der Wissenschaftlichkeit einer Publikation veröffentlicht, und selbst Busa (ebd.) referenziert existierende Konkordanzen und Indizes, die allerdings noch nicht elektronisch vorlagen. Die elektronische Publikation und Weitergabe ist auch nicht neu: Bereits in den 1990er Jahren fanden sich elektronische Beilagen auf CD-ROM in Verlagspublikationen und auch bei Hausarbeiten; Webseiten enthielten schon früh Forschungsdaten¹, die – zumindest während der Projektlaufzeit – diese auch für andere Forschende verfügbar machten.

Die Publikation über öffentlich zugängliche Datenrepositorien ist dagegen recht neu, d.h. in zentralen Einrichtungen, die auch über Projektlaufzeiten hinaus Forschungsdaten vorhalten können, dafür aber eine Übergabe durch Datenbereitsteller nach bestimmten Qualitäts- und Beschreibungsrichtlinien verlangen. Dabei erlauben Repositorien, extrem große Mengen von (vernetzten) Daten zu veröffentlichen: alle erhobenen Daten können publiziert werden. Darunter können auch Daten sein, die für die eigene Forschungsfrage letztendlich nicht relevant waren, aber den Forschungsweg und die Ergebnisse nachvollziehbar machen, also sogenannte "null-results". Innerhalb von Enhanced Publications können sie referenziert werden und z.B. mittels Linked Data als dynamisch vernetzende Daten aufeinander verweisen. Ihre Nachnutzung erlaubt es, Analysen, die in der Vergangenheit nur als Gedankenexperimente möglich waren, praktisch durchzuführen.

Auch deswegen setzen Forschungsförderer zunehmend voraus, dass ein Konzept zur Nachnutzung der Daten, auf denen Forschungsergebnisse beruhen, zusammen mit Projektanträgen eingereicht wird, und dass zudem bereits existierende Daten und Werkzeuge gegebenenfalls nachgenutzt werden (siehe Senat der DFG, 2015; Deutsche Forschungsgemeinschaft, 2013; Deutsche Forschungsgemeinschaft, 2009; Deutsche Forschungsgemeinschaft, undatiert; H2020 Programme, 2013; Allianz der deutschen Wissenschaftsorganisationen, 2010). Gleichzeitig wird durch die Attribution der Datenaufbereitung die Leistung derjenigen sichtbar, die die Daten bereitgestellt haben, so dass die Veröffentlichung "infrastrukturbezogene[r] wissenschaftliche[r] Leistungen" und Forschungsdaten "in Qualifikationsverfahren ergänzend anerkannt werden" können (vgl. Wissenschaftsrat, 2011). Die Bereitstellung von Forschungsdaten folgt dabei den sogenannten FAIR-Prinzipien², ist Teil der sich entwickelnden Open-Science-Landschaft und wird etwa von wissenschaftsgeleiteten Forschungsinfrastrukturen wie CLARIN-D (siehe Hinrichs und Trippel, 2017) und DARIAH-DE (Gradl und Henrich, 2016) unterstützt.

Auch wenn der Mehrwert des Forschungsdatenmanagements in diesem Bereich dokumentiert ist, als Anforderung formuliert wird und positiv konnotiert ist (siehe Bargheer u.a. 2017), erscheint es einigen Forschenden trotzdem eher als aufgezwungene, lästige Notwendigkeit denn als Mehrwert für die Forschung in den DH. Auch deshalb sollen in dem Panel die Vertreter der Forschungsinfrastrukturen und die WissenschaftlerInnen reflektieren, ob die bisher an die Forschenden herangetragenen Angebote ausreichen, um ihre Anforderungen zu erfüllen und ihr Vertrauen zu gewinnen, und, was die Forschungsinfrastrukturen tun können, um hier weitere Fortschritte zu erzielen.

Panel Format

Das Panel besteht aus drei Teilen:

- Kurze Impulsvorträge von max. 5 Minuten von den Panel-Teilnehmenden unter Einnahme ihrer jeweiligen Perspektive
- Diskussion der Leitfragen zwischen den Panelteilnehmenden moderiert aus der Anwenderperspektive
- Öffnung der Diskussion für die Zuhörenden durch die Moderation

Panel-Teilnehmende:

Moderatorin

Elke Teich (Saarbrücken): Als Sprecherin eines interdisziplinären SFB verfügt Prof. Teich über

eine hervorragende Kenntnis der Erfordernisse und Hindernisse im Forschungsdatenmanagement in großen Forschungsverbänden. Sie wird das Panel aus der spezifischen Perspektive der Geisteswissenschaften moderieren.

Teilnehmende auf dem Panel und Impulsvorträge

- Lydia Müller (Mind Research Repository, Universität Leipzig): In den Kognitionswissenschaften ist die Verwaltung von Daten im Zusammenhang mit der Publikation von Artikeln zusammen mit den zugrundeliegenden Daten teilweise gelebte Praxis. Lydia Müller ist als Maintainerin des Mind Research Repository (<http://openscience.uni-leipzig.de/>) an einer Schlüsselstelle, da dieses Verzeichnis die Integration von Daten und Publikationen ermöglicht. Vor dem Hintergrund der dort gehosteten Daten und Publikationen kennt sie die Fallstricke bei der Integration von Daten und Publikationen und weiß, wie die Daten nachhaltig und gesichert abgelegt werden können und hat Erfahrungen damit, auf diese Daten zu verweisen. Sie steht damit für einen Aspekt der FAIR-Prinzipien für Forschungsdaten im Rahmen des Forschungsdatenzyklus in den Geisteswissenschaften: der Aufbewahrung und Zugänglichmachung von Daten.
- Susanne Haaf (BBAW) ist am Aufbau und Betrieb des Deutschen Textarchivs (DTA) beteiligt und hat die Spezifikation des standardisierten DTA-Basisformats maßgeblich mit beeinflusst, das Eingang in die einschlägigen DFG-Richtlinien gefunden hat. Damit war sie an der Erstellung von geisteswissenschaftlichen Forschungsdaten selbst beteiligt und entwickelte aktiv Standards und Richtlinien, die für große Datenmengen eingesetzt wurden. Daneben vermittelte sie die zugrundeliegenden Workflows, Richtlinien und Technologien im Rahmen von Workshops und Lehrveranstaltungen innerhalb der DH-Nutzercommunity. Das DTA vertritt mehr als 2000 Nutzende, die zur Weiterentwicklung seines bislang einzigartigen Korpus interoperabler Forschungsdaten des historischen Deutschen für die Geisteswissenschaften beitragen.
- Torsten Schrade (Leiter der Digitalen Akademie der Akademie der Wissenschaften und der Literatur Mainz, Professor für Digital Humanities an der Hochschule Mainz): Torsten Schrade ist ausgewiesener Experte für Forschungsdatenmanagement und Webtechnologien für die geisteswissenschaftliche Grundlagenforschung, für historische Fachinformationssysteme, webbasierte Arbeitsumgebungen, digitale Editionen, Webservices für geisteswissenschaftliche Fachdaten und Semantic Web Anwendungen.
- Thorsten Trippel: Im Rahmen von CLARIN-D und dem europäischen Projekt Parthenos beschäftigt sich Thorsten Trippel mit der Erstellung

und Umsetzung von Datenmanagementpläne für die geisteswissenschaftliche Forschung. Zusammen mit Antragstellern entwickelt er dazu Datenmanagementpläne, wie sie von Drittmittelgebern zunehmend gefordert sind, und begleitet die Projekte über alle Phasen der geisteswissenschaftlichen Forschung - vom Erstellen oder Auffinden von Forschungsdaten, über die Aufbereitung, Analyse, Bereitstellung bis zur Archivierung, um die FAIR-Prinzipien umzusetzen.

- Tobias Gradl: Im Rahmen seiner Forschung hat sich Tobias Gradl innerhalb von DARIAH mit der nachhaltigen Aufbereitung von Forschungsdaten beschäftigt, besonders mit Daten, die nicht standardkonform aufbereitet wurden und im Rahmen spezifischer Forschungsfragestellungen anfallen. Daher bearbeitet er Modelle, um Daten in einer nachhaltig erschlossenen Form bereitzustellen.
- Maria Effinger (UB Heidelberg): Im Rahmen des DFG-Projekts OA-Monos, des Universitätsverlags "Heidelberg University Publishing" und der Fachinformationsdienste "arthistoricum.net" und "Proyplaeum" hat Maria Effinger XML-basiertes Publizieren implementiert und in der Universitätsbibliothek in Heidelberg umgesetzt. Die Veröffentlichungsmöglichkeiten über universitätseigene Verlage nehmen für offene Publikationen und damit verbundene Daten eine wichtige Funktion ein, indem sie in einer engen Zusammenarbeit mit den Forschenden die prinzipielle Nachvollziehbarkeit und Überprüfbarkeit wissenschaftlicher Ergebnisse sichern. Diesen Aspekt wird Frau Effinger im Panel vertreten.

Fußnoten

1. Zum Beispiel gab es schon im Jahr 2000 auf der Webseite <http://www.homes.uni-bielefeld.de/gibbon/EGA/> Sprachdokumentationsdaten der Sprache Ega in Côte d'Ivoire. Diese Seite kann als typisches Beispiel für Projektwebseiten mit Daten angesehen werden, deren Nachhaltigkeit durch das Engagement des betreibenden Einzelwissenschaftlers bestimmt wird.
2. Findable, Accessible, Interoperable, und Re-usable, siehe <https://www.force11.org/group/fairgroup/fairprinciples>

Bibliographie

Allianz der deutschen Wissenschaftsorganisationen (2010): "Grundsätze zum Umgang mit Forschungsdaten". <http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsätze.html> [letzter Zugriff 15. Januar 2017].

Bargheer, Margo, / Zeki, Mustafa Dogan / Horstmann, Wolfram / Mertens, Mike / Rapp, Andrea

(2017): "Unlocking the digital potential of scholarly monographs in 21st century research," in: *Liber Quarterly* 27(2).

Baum, Constanze / Stäcker, Thomas (2015): "Methoden – Theorien – Projekte," in: *Zeitschrift für digitale Geisteswissenschaften*, 1(Sonderband: Grenzen und Möglichkeiten der Digital Humanities), 4-12. doi:DOI 10.17175/sb001_023

Berry, David M (2011): "The Computational Turn: Thinking about the Digital Humanities", in: *Cultural Machine*, 12, 1-22. <http://www.culturemachine.net/index.php/cm/article/view/440>

Busa, Roberto (1951): *Sancti Thomae Aquinatis Hymnorum Ritualium Varia Specimina Concordantiarum*. Milano: Fratelli Bocca.

Deutsche Forschungsgemeinschaft (2013): "Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“. *Sicherung guter wissenschaftlicher Praxis*. http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf [letzter Zugriff 15. Januar 2017]

Deutsche Forschungsgemeinschaft (DFG): "Umgang mit Forschungsdaten: DFG-Leitlinien zum Umgang mit Forschungsdaten". http://www.dfg.de/foerderung/antrag_gutachter_gremien/antragstellende/antragstellung/nachnutzung_forschungsdaten/ [letzter Zugriff 15. Januar 2017]

Deutsche Forschungsgemeinschaft (DFG) Ausschuss für wissenschaftliche Bibliotheken und Informationssysteme (Unterausschuss für Informationsmanagement) (2009) "Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten. Bonn: Deutsche Forschungsgemeinschaft. http://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf [letzter Zugriff 15. Januar 2017]

Geyken, Alexander (2011): "Wege zu einem historischen Referenzkorpus des Deutschen: das Projekt Deutsches Textarchiv," in: Perspektiven einer corpusbasierten historischen Linguistik. Internationale Tagung des Akademienvorhabens „Altägyptisches Wörterbuch“, Berlin, Germany.

Geyken, Alexander / Haaf, Susanne / Boenig, Matthias / Thomas, Christian / Wiegand, Frank (seit 2007): Deutsches Text Archiv (DTA). <http://www.deutschestextarchiv.de/> [letzter Zugriff 15. Januar 2017]

Grabl, Tobias / Henrich, Andreas (2016): "Die DARIAH-DE-Föderationsarchitektur – Datenintegration im Spannungsfeld forschungsspezifischer und domänenübergreifender Anforderungen," in: *Bibliothek - Forschung und Praxis*, 40(2), 222-228. doi:10.1515/bfp-2016-0027

Hinrichs, Erhard / Trippel, Thorsten (2017): "CLARIN-D: eine Forschungsinfrastruktur für die sprachbasierte Forschung in den Geistes- und

Sozialwissenschaften," in: *Bibliothek - Forschung und Praxis*, 1(41).

Senat der Deutschen Forschungsgemeinschaft (DFG) (2015): "Leitlinien zum Umgang mit Forschungsdaten. <https://www.wissenschaftsrat.de/download/archiv/10466-11.pdf> [letzter Zugriff 15. Januar 2017].

Thiel, Thomas (2012): "Digital Humanities: Eine empirische Wende für die Geisteswissenschaften?", in: *Frankfurter Allgemeine Zeitung*, 24.07.2012. <http://www.faz.net/aktuell/feuilleton/forschung-und-lehre/digital-humanities-eine-empirische-wende-fuer-die-geisteswissenschaften-11830514.html> [letzter Zugriff 15. Januar 2017].

Winter, Thomas Nelson (1999): "Roberto Busa, S.J., and the Invention of the Machine-Generated Concordance," in: *Faculty Publications, Classics and Religious Studies Department*. <http://digitalcommons.unl.edu/classicsfacpub/70/> [letzter Zugriff 15. Januar 2017].

Wissenschaftsrat (2011): *Übergreifende Empfehlungen zu Informationsinfrastrukturen* Vol. Drs. 10466-11. <https://www.wissenschaftsrat.de/download/archiv/10466-11.pdf> [letzter Zugriff 15. Januar 2017].

Zampolli, Antonio (1973): "Humanities Computing in Italy," in: *Computers and the Humanities*, VII(6), 343-360. doi:10.1007/BF02395110