

Die Summe geisteswissenschaftlicher Methoden? Fachspezifisches Datenmanagement als Voraussetzung zukunftsorientierten Forschens

Moeller, Katrin

katrin.moeller@geschichte.uni-halle.de
Historisches Datenzentrum Sachsen-Anhalt, Martin-
Luther-Universität Halle-Wittenberg, Deutschland

#ur#o, Matej

matej.durco@oeaw.ac.at
Austrian Center for Digital Humanities, Österreichische
Akademie der Wissenschaften, Österreich

Ebert, Barbara

barbara.ebert@rfii.de
Leiterin des Rats für Informationsinfrastrukturen (RfII),
Geschäftsstelle Göttingen

Lemaire, Marina

esciences@uni-trier.de
Servicezentrum eSciences, Universität Trier, Deutschland

Rosenthaler, Lukas

lukas.rosenthaler@unibas.ch
and Service Center for the Humanities DaSCH,
Universität Basel (DHLab) und Schweizerische Akademie
der Geistes- und Sozialwissenschaften, Schweiz

Sahle, Patrick

sahle@uni-koeln.de
Data Center for the Humanities (DCH), Universität zu
Köln, Deutschland

Wuttke, Ulrike

Ulrike.Wuttke@gmx.net
Stellvertretende Vorsitzende der AG Datenzentren
des DHd, Fachbereich Informationswissenschaften,
Fachhochschule Potsdam, Deutschland

Wettlaufer, Jörg

jwettla@gwdg.de
Göttingen Centre for Digital Humanities, Georg-August
Universität Göttingen, Deutschland

Zusammenfassung des Vorhabens

Die AG Datenzentren des DHd möchte ein Panel mit insgesamt fünf Diskussionspartnern und zwei Moderatoren veranstalten, um die Herausbildung fachbezogenen Datenmanagements im deutschsprachigen Raum in den digitalen Geisteswissenschaften weiter zu fördern sowie den Stand und die Perspektiven dieses Forschungsbereichs zu diskutieren. Das Panel reiht sich damit in den fachlich übergreifenden Organisationsprozess zum Datenmanagement ein. Im Mittelpunkt der Diskussion soll hier die Veränderung der Datenkultur in der Geisteswissenschaft stehen und damit die Frage, wie eine Konsolidierung, Vernetzung und fachliche Abrundung von Diensten die Anreize zur wissenschaftlichen Nutzung des Datenmanagements erhöhen können. Dies soll mit zwei Schwerpunktsetzungen erfolgen:

A. Zunächst möchten wir anhand eines Überblicks zu bestehenden Angeboten von Datenzentren informieren und aufzeigen, welche fachbezogenen Leistungen momentan oder in naher Zukunft von Datenzentren in einem Netzwerk verteilter DH-Strukturen eingebracht werden. Dieser Block macht die Schwerpunktsetzungen, Rollen und herausgehobene fachliche Dienste einzelner Datenzentren sichtbar, die über allgemeine Anforderungen an ein Datenzentrum hinausragen. Daran anschließend kann über die mögliche Ausgestaltung kooperativer Strukturen und die Zusammenarbeit innerhalb eines Netzwerkes der DH-Datenzentren diskutiert werden.

B. Weiterhin fragt das Panel grundlegend nach zentralen Aufgabenstellungen im Bereich von Standards aus den einzelnen Fachgebieten der Geisteswissenschaften: Welche fachlichen Standards müssen Datenzentren über die zuvor gezeigten Ansätze hinaus kooperativ entwickeln, um passfähige Daten für eine möglichst große Breite von Nachnutzungen interdisziplinär abzusichern? Wo liegen Möglichkeiten und wo Grenzen von fachlicher Standardisierung und Normalisierung über die einzelnen Fachdisziplinen der Geisteswissenschaft hinweg? Hierzu sollen Erkenntnisse und Erfahrungen aus abgeschlossenen bzw. weitgehend realisierten Projekten dazu dienen, abstrahierende Anforderungen aus fachwissenschaftlicher Perspektive zu benennen, die zugleich einen Mehrwert der Realisierung besitzen. Insgesamt geht es darum, die Notwendigkeit eines geisteswissenschaftlichen sowie fachspezifischen Angebots nachdrücklich unter Beweis zu stellen, zu etablierende Standards zu benennen und Desiderate zu erkennen.

Forschungsstand

Der digitale Wandel, das Entstehen, Archivieren, Erschließen und Nachnutzen großer Datenbestände verändert wissenschaftliche Forschung in ihren Grundlagen, Arbeitsweisen und Methodiken fundamental (OECD 2007; HRK 2014; RfII 2016; Allianz 2010; DFG 2015). Die zahlreichen Aspekte dieses Wandels beschäftigen Fach-Communities und Fachverbände in umfassender Weise mit Fragen des Datenmanagements, der Lizenzen, Rechte und des Datenschutzes, der technischen und inhaltlichen Langzeitarchivierung sowie Nachnutzung von Daten. Die intensiven Diskussionen um das Datenmanagement sind verbunden mit einem grundlegenden Nachdenken über künftige Aufgaben, Strukturen und Workflows bestehender und sich neu formierender Institutionen oder Akteure und führen zu grundsätzlichen Reflexionen über die notwendige Schaffung verteilter nationaler infrastruktureller Angebote (RfII 2016: 39f.; RfII 2017). In der jüngeren Vergangenheit sind bevorzugt technische Aspekte des Datenmanagements diskutiert worden, wie sie etwa die interoperationale Metadatenhaltung (Bibliotheksparadigma), Probleme der Speicherformate, Speichermedien und Dokumentation in der technischen Langzeitarchivierung über einen Rahmen von zehn Jahren hinaus (Archivparadigma) betreffen (Nestor 2016). Solche Lösungen und Angebote können oft in übergreifenden technischen Strategien gefunden werden, die ein zentrales Aufgabengebiet von Bibliotheken und Archiven umfassen. Wie in den Arbeitsbereichen der digitalen Wissenschaft allgegenwärtig zu beobachten ist, lässt sich dabei eine sukzessive, aber dennoch intensive Spezialisierung und Differenzierung digitaler Erfordernisse und Techniken beobachten. War beispielsweise noch das erste große deutschsprachige Überblickswerk der nestor-Arbeitsgruppe zur Langzeitarchivierung fachübergreifend technisch geprägt (Neuroth 2010), offenbaren die neueren Arbeitsergebnisse eine deutliche Hinwendung zur fachbezogenen Darstellung (Neuroth 2012). Allerdings blieben auch in dieser Hinsicht die Parameter der technischen Langzeitarchivierung aus der Sicht der Archive und Bibliotheken bestimmend.

Gleiches lässt sich mit Blick auf die Wissenschaft erkennen, wenn man etwa auf die sich konstituierenden Fachgruppen im Rahmen des Verbandes der Digital Humanities im deutschsprachigen Raum ¹ oder andere traditionelle Fachverbände mit ihren sich ausdifferenzierenden digitalen Arbeitsweisen/Methoden schaut. ² Intensiv diskutiert wurden solche Themen der fachbezogenen Binnendifferenzierung auch in der Auftaktveranstaltung zum Förderprogramm "'Mixed Methods' in den Geisteswissenschaften?" der VW-Stiftung im Frühsommer 2017, die eine Vielzahl deutschsprachiger Forschungsprojekte im Bereich der Digital Humanities versammelte. ³

Um nachhaltig wirksam zu sein, braucht Forschungsdatenmanagement eine enge Anbindung an die Wissenschaft mit ihren spezifischen Problemen und fachlichen Erfordernissen (RfII 2016: 60; RfII 2017: 1f., AG DZ 2017: 3). Nicht in erster Linie der technische Service zur Langzeitarchivierung standardisierter Daten, sondern die fachbezogene Datenkuration und die Tiefenerschließung auf der Basis von Standards sichert die Güte und Qualität von Forschungsdaten und ihre Nachnutzbarkeit für spezifische wissenschaftliche Forschungsprojekte. Qualitätsgerechtes Forschungsdatenmanagement - das auch im analogen Zeitalter den Kern jeder Forschung ausmachte - sichert den wissenschaftlichen Mehrwert von Daten, schafft vielfältige Nutzungsanreize und innerwissenschaftliche Akzeptanz (RfII 2016: 38-40; AG DZ 2017: 2). Im gleichen Maße wie daher die übergreifende Basis technischer Lösungen zur Metadatenhaltung, Lizenzierung und Langzeitarchivierung entwickelt werden muss, sind fachbezogene Standards der Datenkuration, der kontrollierten Vokabularien, der fachbasierten Quellenkritik und Annotation sowie spezialisierte Werkzeuge der fachlichen Datenverarbeitung zur Erzeugung nachhaltiger Datenstrukturen notwendig. Während erstere eher in den Verantwortungsbereich der Bibliotheken und Archive fallen, sind letztere das zentrale Aufgabenfeld der Wissenschaft mit ihren spezifischen Fachgebieten.

Hinzu treten die zahlreichen Herausforderungen des geisteswissenschaftlichen Datenmanagements im Unterschied zu natur- und sozialwissenschaftlichen Projekten. Geisteswissenschaftliche Daten bilden in vielen Anwendungsbereichen komplexe Textkorpora mit hochspezialisierten Formen der Transkription, Annotation und der Unterscheidung verschiedener Ebenen von Quellenbegriff, Lemmatisierung, Normalisierung, Standardisierung und Begriffserklärung. Ähnliche komplexe Formen der Erschließung gelten ebenso in eher auf Objekte und Gegenstände bezogene Geisteswissenschaften. Ansätze der einen Disziplin sind (bis heute) nicht ohne weiteres auf eine andere übertragbar bzw. nicht fachübergreifend bekannt. Gleichzeitig ist es Anliegen von Datenzentren übergreifende Standards bzw. Austauschformate zu definieren und zu etablieren, welche die Transparenz und Verschneidung von Daten überhaupt erst einmal ermöglichen.

Ebenso muss auf der Ebene der Datenzentren intensiv über Kooperationsformen und Vorgehensweisen für neue heterogene Datenbestände mit einem Mix aus Quellen, Daten, Methoden, Medien und Rechten nachgedacht werden (AG DZ 2017: 3).

Fragestellung und Aufbau des Panels

Die DH ist angetreten, um computergestützte Ansätze und Methoden quer über die geisteswissenschaftlichen

Fachdisziplinen zu entwickeln. Was jedoch macht die Summe geisteswissenschaftlicher Bedürfnisse aus, um Daten tatsächlich interdisziplinär und langfristig nutzbar zu machen? Die "Kritik der digitalen Vernunft" besteht für ein solches geisteswissenschaftliches Unterfangen darin, gemeinsame Anforderungen und Standards der Datenhaltung aus bestehenden Projekten und Erfahrungen zu formulieren, zusammenzutragen, zu systematisieren und zu abstrahieren. Zudem bedürfen solche Standards der Akzeptanz durch Forschende, sollen sie Wirksamkeit entfalten. Standards müssen daher einen klaren Mehrwert für wissenschaftliche Arbeit bieten.

1. Viele Datenzentren haben mittlerweile ihre reguläre fachliche Arbeit aufgenommen, projektbasiert Erfahrungen gesammelt und neben der regulären Arbeit oft spezifische Schwerpunkte gebildet. In Form einer Präsentation der wichtigsten Akteure, Dienste und Angebote möchten wir über das bestehende Leistungsangebot innerhalb der AG Datenzentren des DHd-Verbandes informieren, wobei bevorzugt die Schwerpunktsetzungen im Bereich der Standards thematisiert werden. Auf diese Weise können sich Tagungsteilnehmer einen Überblick zum Dienstleistungsspektrum von Datenzentren verschaffen, was angesichts der insgesamt unübersichtlichen (weil sich dynamisch entwickelnden) Forschungslandschaft einen erheblichen Vorteil darstellt. Diese ca. 15-20 Minuten dauernde Präsentation beruht auf einer Umfrage unter den im Verband organisierten Einrichtungen und Akteuren, die von der Organisatorin des Panels in Zusammenarbeit mit der AG Datenzentren vorstrukturiert und präsentiert wird. Bei Bedarf können weitere zentrale fachwissenschaftliche Angebote in den Geisteswissenschaften einbezogen werden. Dieser Leistungskatalog ist einerseits Grundlage der Diskussion mit dem Publikum und andererseits für die Außendarstellung der AG Datenzentren der DHd förderlich.

2. Nach dieser Präsentation leitet die Moderation in den Diskussionsteil über. Eingeladen werden Vertreter und Vertreterinnen geisteswissenschaftlicher Datenzentren mit unterschiedlichen Schwerpunktsetzungen und Funktionen, die Impulsreferate zu notwendigen Anforderungen fachwissenschaftlicher Standards im Datenmanagement zu geben. Diese Statements werden vor der Tagung allen Diskussionspartnern zur Verfügung gestellt, um eine angeregte Diskussion über fachliche Belange und Standards für das Datenmanagement der Geisteswissenschaften zu ermöglichen und auch Kritik zu formulieren. Die Panelorganisatorin und Moderatoren sichten die Beiträge zuvor redaktionell-moderierend. Hier sollen vor allem Möglichkeiten und Grenzen von bestehenden innerfachlichen Standards (z.B. existierende Ansätze wie GND, XML, TEI, Transkriptionsregeln; neue Entwicklungen wie z.B. "Ontologie historischer Berufe" etc.) benannt und diskutiert werden. An Einzelbeispielen soll kontrovers diskutiert werden, welche Standards die Geisteswissenschaft und das Datenmanagement tatsächlich voranbringen. Sind es eher technische Aspekte und Rahmenbedingungen (Tools,

Formate, Annotationswerkzeuge), die formale Aspekte des Datenmanagements vereinheitlichen oder thematisch übergreifende Normansätze, Ontologien oder kontrollierte Vokabulare, die häufig aber sehr voraussetzungsvoll und arbeitsintensiv sind? Zielgerichtet sollen so Ansätze, Möglichkeiten und Grenzen diskutiert werden. Fachwissenschaftliche Aspekte sollen zusätzlich durch die Öffnung der Diskussion für das Publikum einbezogen werden. Gleichzeitig kann das Publikum Hindernisse und Desiderate des Forschungsdatenmanagements formulieren. Insgesamt sollen zudem mögliche Kooperationen und Organisationsstrukturen zwischen verschiedenen Datenzentren angerissen werden (Wer darf Standards entwickeln? Wer muss sie nutzen?). Nach Möglichkeiten werden dazu Beiträge des Publikums bzw. die Use Cases der Diskussion aufgegriffen. Diese Diskussion soll die fachliche Positionierung und Ausdifferenzierung der Datenzentren schärfen.

Es ist die Aufgabe der Moderation, entsprechend der Schwerpunktsetzungen in der Diskussion abstrahierende Aussagen zur Ansetzung von Standards zu finden, die aber zuvor innerhalb der AG und auch der eingeladenen Diskutanten bereits diskutiert wurden, um ein konzentriertes und fachbezogen-klares Fazit der Diskussion zu ermöglichen.

Zusammensetzung des Panels

Organisation und Präsentation:

Dr. Katrin Moeller, Historisches Datenzentrum Sachsen-Anhalt (Hist-Data), Martin-Luther-Universität Halle-Wittenberg, Deutschland

Moderation:

Dr. Ulrike Wuttke, Stellvertretende Vorsitzende der AG Datenzentren des DHd, Fachbereich Informationswissenschaften, Fachhochschule Potsdam, Deutschland

Dr. Jörg Wettlaufer, Göttingen Centre for Digital Humanities, Georg-August Universität Göttingen, Deutschland

Datenzentren:

Marina Lemaire, M.A., Servicezentrum eSciences, Universität Trier, Deutschland

DI Matej #ur#o, Austrian Center for Digital Humanities, Österreichische Akademie der Wissenschaften, Österreich

Dr. Barbara Ebert, Leiterin des Rats für Informationsinfrastrukturen (RfII), Geschäftsstelle Göttingen

Prof. Dr. Lukas Rosenthaler, Data and Service Center for the Humanities DaSCH, Universität Basel (DHLab) und Schweizerische Akademie der Geistes- und Sozialwissenschaften, Schweiz

apl. Prof. Dr. Patrick Sahle, Data Center for the Humanities (DCH), Universität zu Köln, Deutschland

Fußnoten

1. Siehe die Arbeitsgruppen des Verbands DHd: <https://dig-hum.de/dhd-ags>.
2. Vergleiche hier etwa das CfP und die daraus resultierenden Ergebnisse der Tagung: Quellen und Methoden der Geschichtswissenschaft im digitalen Zeitalter. Neue Zugänge für eine etablierte Disziplin, DIGIMET 2017, 25./26.09.2017 in Berlin, URL: <http://welt-der-kinder.gei.de/wp-content/uploads/2017/05/CfP-DH-Abschluss-tagung-WdK-02.05.2017.pdf>.
3. Dabei nahmen zahlreiche Vertreter der DH-Community in den Geisteswissenschaften teil, die auch über das eigentliche Förderprogramm "Mixed Methods" in den Geisteswissenschaften Projekte durchführen: https://www.volkswagenstiftung.de/fileadmin/downloads/publikationen/20161221_Mixed_Methods_Volkswagen_Stiftung_Bewilligungen.pdf.

Bibliographie

AG Datenzentren des DHd (Hg.) (2017): Stellungnahme der DHd AG Datenzentren und des DHd-Verbands zur Nationalen Forschungsdateninfrastruktur (NFDI), URL: http://dig-hum.de/sites/dig-hum.de/files/DHd_NFDI_Stellungnahme_2017-07-31.pdf [letzter Zugriff 22. September 2017].

Allianz der deutschen Wissenschaftsorganisationen (Hg.) (2010): Grundsätze zum Umgang mit Forschungsdaten, URL: http://www.allianzinitiative.de/fileadmin/user_upload/www.allianzinitiative.de/Grundsaeetze_Forschungsdaten_2010.pdf [letzter Zugriff 22. September 2017].

DFG (Hg.) (2015): DFG-Leitlinien zum Umgang mit Forschungsdaten, Bonn 2015, http://www.dfg.de/foerderung/antragstellung_begutachtung_entscheidung/antragstellende/antragstellung/nachnutzung_forschungsdaten/index.html [letzter Zugriff 22. September 2017].

Nestor (Hg.) (2016): Standardisierung, 2016, URL: <https://wiki.dnb.de/display/NESTOR/Standardisierung> [letzter Zugriff 22. September 2017].

Neuroth, Heike / Oßwald, Achim / Scheffel, Regine / Strathmann, Stefan / Huth, Karsten (Hg.): nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3, Göttingen 2010, URL: http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch_23.pdf [letzter Zugriff 22. September 2017].

Neuroth, Heike / Strathmann, Stefan / Oßwald, Achim / Scheffel, Regine / Klump, Jens / Ludwig, Jens (Hg.) (2012): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. Boizenburg / Göttingen.

OECD (Hg.) (2007): OECD Principles and Guidelines for Access to Research Data from Public Funding, Paris 2007, URL: <http://www.oecd.org/sti/>

[sci-tech/38500813.pdf](http://www.oecd.org/sti/sci-tech/38500813.pdf) [letzter Zugriff 22. September 2017]. **OECD (Hg.)** (2016):, "Research Ethics and New Forms of Data for Social and Economic Research", *OECD Science, Technology and Industry Policy Papers*, No. 34, OECD Publishing, Paris, URL: <http://dx.doi.org/10.1787/5jln7vnpxs32-en> [letzter Zugriff 22. September 2017].

Hochschulrektorenkonferenz (HRK) (Hg.) (2014): Management von Forschungsdaten als strategische Aufgabe der Hochschulleitungen, 14. Mai 2014, URL: <https://www.hrk.de/positionen/beschluss/detail/management-von-forschungsdaten-eine-zentrale-strategische-herausforderung-fuer-hochschulleitungen/> [letzter Zugriff 22. September 2017].

Rat für Informationsinfrastrukturen (RFII) (Hg.) (2016): Leistung aus Vielfalt, Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland, Göttingen, URL: <http://www.rfii.de/de/category/dokumente/> [letzter Zugriff 22. September 2017].

Rat für Informationsinfrastrukturen (RFII) (Hg.) (2017): Schritt für Schritt - oder: Was bringt wer mit? Ein Diskussionsimpuls zur Zielstellung und Voraussetzungen für den Einstieg in die Nationale Forschungsdateninfrastruktur (NFDI), URL: <http://www.rfii.de/de/category/dokumente/> [letzter Zugriff 22. September 2017].