

Datenzentren für die nachhaltige Forschung in den Digital Humanities

Sahle, Patrick

sahle@uni-koeln.de

Data Center for the Humanities, Universität zu Köln

Trippel, Thorsten

thorsten.trippel@uni-tuebingen.de

CLARIN-D, Universität Tübingen

Neumann, Gerald

gneumann@bbaw.de

Berlin-Brandenburger Akademie der Wissenschaften

Engelhardt, Claudia

claudia.engelhardt@sub.uni-goettingen.de

Humanities Data Center, Staats- und

Universitätsbibliothek Göttingen

Kurzawe, Daniel

kurzawe@sub.uni-goettingen.de

Humanities Data Center, Staats- und

Universitätsbibliothek Göttingen

Schäfer, Felix

felix.schaefer@dainst.de

IANUS, Deutsches Archäologisches Institut

Wörner, Kai

kai.woerner@uni-hamburg.de

Geisteswissenschaftliche Infrastruktur für Nachhaltigkeit (gwin), Universität Hamburg

Durch die Digitalisierung der Forschung und die damit verbundenen gestiegenen Anforderungen an das Forschungsdatenmanagement ergibt sich ein zunehmender Bedarf an umfassender und forschungsorientierter Datenexpertise. Während die Geisteswissenschaften in diesem Prozess einen großen methodischen Fundus an Analysemethoden entwickeln, mangelt es noch an der Infrastruktur, den Institutionen und stabilen Praktiken, die notwendig sind, um Forschungsdaten für künftige Forschergenerationen zu bewahren und auf geeignete Weise zur Verfügung zu stellen. Datenzentren spielen im Idealfall nicht nur zum Zeitpunkt der Aufnahme abgeschlossener Projekte, also bei der Übernahme, Archivierung und ggf. Einbettung in eigene Repositorien eine wichtige Rolle. Vielmehr werden

die Weichen für die optimale Archivfähigkeit und Nachnutzbarkeit von Forschungsdaten bereits vorher gestellt und erfordern ggf. eine Beratung von Anfang an. Auf der anderen Seite ist die Kuratierung von Daten und Anwendungen kein einmaliger Akt, sondern eine andauernde Aufgabe. Es stellt sich deshalb die Frage, welche Funktionen geisteswissenschaftliche Datenzentren und Forschungsinfrastrukturen für Forschende in den Digital Humanities insgesamt haben können und sollen. Wenn ein Mehrwert für die Forschung durch die dauerhafte Anschlussfähigkeit von Daten und Anwendungen an den ganzen Forschungsprozess entstehen soll, dann erfordert dies anscheinend einen ganzen Strauß an Diensten und Angeboten, die sich von der koordinierten Unterstützung durch Fachberatung, der Vermittlung von Hardware, der Beantragung von Mitteln, dem Betrieb von Repositorien bis zur anhaltenden Pflege von Daten und Anwendungen erstrecken.

Dieses Panel bringt Vertreter von geisteswissenschaftlichen Datenzentren und den zentralen DH-Infrastrukturprojekten zusammen. Gemeinsam wird diskutiert, wie die Forschung in ihrer Arbeit und in der Erzeugung und Nachnutzung von Daten unterstützt werden kann, wie erste Erfahrungen mit der Kuratierung von Daten aussehen und wie zielgerichtete zusätzliche Angebote gestaltet werden können.

Insbesondere soll dabei aufgezeigt werden, was heute bereits an Angeboten verfügbar ist und auf welche Weise diese in der Forschung genutzt werden. Für Forschende in den Digital Humanities bietet das Panel die Möglichkeit, über weiterhin bestehende Hürden und Herausforderungen zu diskutieren und so den Dialog zwischen der Forschung und den Infrastrukturen voranzutreiben. Die Themen reichen von konkreten Beispielen aus der Forschungspraxis bis zu Modellen und momentanen Fragestellungen zur institutionellen Organisation geisteswissenschaftlicher Forschungsdatenzentren.

Folgende Themenschwerpunkte und Fragestellungen bieten eine Orientierung zu möglichen Inhalten:

- Ausgangspunkte: Datengetriebene Forschung in den Geisteswissenschaften - was wird beforscht, welche Methoden werden angewendet, welche Daten entstehen und wie wird mit diesen Daten umgegangen? Welche Anforderungen ergeben sich daraus an die Langzeitarchivierung, Bereitstellung und Nachnutzung der Daten? Wie können die geisteswissenschaftliche Forschung und die langfristige Sicherung und Bereitstellung ihrer Ergebnisse durch die Datenzentren unterstützt werden? Welche Lücken in den momentanen Angeboten zeigen sich dabei? Konkrete Projekte, Beispiele oder Methoden können demonstriert und diskutiert werden.
- Zugänglichkeit: Wie lassen sich die bestehenden Angebote nutzen und wer ist für "meine Forschung" zuständig? Mit wem lassen sich Projekte planen, wer hostet mein Projekt und wer übernimmt die Daten und Anwendungen nach der Projektlaufzeit? Fragen

dieser Art können direkt zwischen den Forschenden und Zentren diskutiert werden.

- Zielgruppen: An wen richten sich welche Angebote? Nicht ein Schuh passt allen. Wie finden Forscher zu den passenden Angeboten? Wie können Angebote für ein breites Anforderungsspektrum mit vielen heterogenen Forschungsfragen und Forschungsmethoden gestaltet werden? Wie stellen die Forschungsdatenzentren sicher, sich nicht an der Zielgruppe vorbei zu entwickeln?
- Zuständigkeit: Die aktuelle Landschaft an Forschungsdateninfrastrukturen lässt sich anhand mehrerer Dimensionen kartieren: regionaler bzw. standortgebundener gegenüber übergreifendem Auftrag; Spezialisierungen auf Datentypen, Disziplinen, Methoden; Art, Schichtung und Umfang der Dienstleistung: Beratung - umfassende Begleitung - technische Dienstleistungen, Übernahme von Daten, Anwendungskonservierung. Wo sind bestehende Initiativen zuzuordnen und wo ergeben sich Lücken? In welchem Verhältnis stehen die lokalen Datenzentren zu den übergreifenden Infrastrukturprojekten CLARIN-D und DARIAH-DE? Wie lassen sich die Aktivitäten der verschiedenen geisteswissenschaftlichen Datenzentren sinnvoll koordinieren und aufeinander abstimmen; welche Modelle der Kooperation sind denkbar?
- Beratung von Anfang an: Wie können z. B. fachspezifische Datenzentren IT-Empfehlungen zum nachhaltigen Umgang mit digitalen Daten in bestimmten Bereichen geben, um auf die Problematik der langfristigen Nutzbarkeit von digitalen Forschungsdaten aufmerksam zu machen und um Forschende mit praktischen Informationen zu Fragen des Forschungsdatenmanagements zu unterstützen?
- Kuratierung komplexer Forschungsdaten: Wie können Daten und Anwendungen aus beendeten Projekten übernommen und in übergreifende Strukturen der Langfristverfügbarkeit eingebracht werden? Hier liegen bei einigen Datenzentren inzwischen erste praktische Erfahrungen vor, über die zu diskutieren ist.
- Vertrauenswürdige Repositorien: Wie lassen sich projektübergreifende und stabile "generische" Lösungen für die Vorhaltung von Forschungsdaten aufbauen? Welchen Abdeckungsgrad können Repositorien hinsichtlich der Vielfalt der Forschungsprojekte erreichen? Wieso sind Repositorien nicht in allen Fällen die Lösung und was kann in solchen Fällen getan werden?
- Angepasste Dienste für die Fachgemeinschaft: Wie können Werkzeuge zur Kuratierung, Nutzung und Bearbeitung von Datenzentren oder größeren Forschungsinfrastrukturen entwickelt und dauerhaft vorgehalten werden, um die Reproduzierbarkeit von Forschungsergebnissen und eine kontinuierliche Nutzung und Weiterentwicklung zu gewährleisten?
- Verteilte Datenhaltung und virtuelle Kollektionen: Wie lässt sich eine verteilte Datenhaltung und übergreifende Nutzung in den DH organisieren? Wie kann die

Zusammenarbeit von Datenzentren technisch durch Schnittstellen umgesetzt werden, die z. B. gemeinsame Datenmodelle, Formate und PID-Systeme verwenden?

- In welchem Maße erfordern technisch übergreifende Lösungen auch eine institutionelle Förderierung?
- Institutionelle Stabilität: Können nur Datenzentren als dauerhafte Einrichtungen die nachhaltige Bereitstellung von Forschungsdaten und Anwendungen gewährleisten oder gibt es andere Optionen?
- Fachbereichsspezifische Kompetenzen: Wie können an Datenzentren und Forschungsinfrastrukturen dauerhafte Kompetenzen und Strukturen zur Beratung von Wissenschaftlerinnen und Wissenschaftlern etabliert werden?