

Was sind und was sollen Datenzentren in den Geisteswissenschaften?

Panel der AG Datenzentren im Verband DHd

Aus Gründen der "guten wissenschaftlichen Praxis", wegen der Nachnutzbarkeit von Forschungsdaten und hinsichtlich eines fortlaufenden wissenschaftlichen Diskurses, besteht für Forscher ein dringender Bedarf an der nachhaltigen Zugänglichkeit vertrauenswürdiger geisteswissenschaftlicher Forschungsdaten. Aus der Sicht der Forschungsförderer geht es bei der Sicherung und anhaltenden Bereitstellung von Projektergebnissen um die Effizienz von Projektfinanzierungen. Datenzentren scheinen die institutionelle Antwort auf die anstehenden Aufgaben zu sein. Hier gibt es bereits einige erfahrene Vorreiter und viele aktuelle Projekte, dennoch fehlt ein übergreifendes Bild und ein gemeinsames Verständnis, was ein Datenzentrum eigentlich ausmacht. Die AG *Datenzentrum* des DHd-Verbandes stimmt ein solches gemeinsames Verständnis der Ziele und Aufgaben von Datenzentren ab, und identifiziert offene Entwicklungspotentiale. Mitglieder der AG Datenzentren sind Wissenschaftler, die ihr Forschungsdatenmanagement bisher schon selbst durchgeführt haben, sowie bestehende und entstehende Datenzentren aus Deutschland, Österreich und der Schweiz.

Datenzentren gewährleisten die Zugänglichkeit und die Nachnutzbarkeit geisteswissenschaftlicher Forschungsdaten jenseits der aktiven Projektlaufzeit. Aufgaben der Datenkuration (z.B. Dokumentation des Projektkontextes, Migration von Datenformaten) werden in enger Zusammenarbeit mit Wissenschaftlern durchgeführt. Existierende Werkzeuge der Langzeitarchivierung decken dabei nicht immer die spezifischen Anforderungen geisteswissenschaftlicher Forschungsdaten ab; so entstehen z.B. erst langsam Ansätze für die Langzeitverfügbarkeit von Datenbanken, Präsentationssystemen und interaktiven Visualisierungen. Insgesamt scheint es auch erhebliche Unterschiede zwischen Modellen der Research Life Cycles in den Naturwissenschaften und den Geisteswissenschaften zu geben, so dass Lösungsvorstellungen aus dem einen Bereich nicht ohne weiteres in den anderen zu übertragen sind.

Der Aufbau von Datenzentren hat nicht nur wissenschaftliche und technische Aspekte, sondern vor allem auch organisatorische Stabilität und finanzielle Nachhaltigkeit müssen gesichert sein. So war unter den Vorreitern geisteswissenschaftlicher Datenzentren das AHDS, das mit den AHDS-Spezialzentren zu Archäologie, Geschichte, Literatur, darstellende und bildende Künste hunderte Kollektionen mit Millionen an Datensätzen und Datenobjekten beherbergt hat. Nach 10-jährigem Betrieb stellte der Förderer (AHRC, und schließlich auch JISC) die Finanzierung für das AHDS im Jahr 2008 kurzfristig ein, trotz eines Aufschreis der Geisteswissenschaften weltweit.

Ein Datenzentrum deckt insgesamt viele Bereiche ab: von technischen Fragen, über konzeptionelle Herausforderungen bis hin zu organisatorisch-institutionellen Problemen. Es bewegt sich dabei zwischen teils widersprüchlichen Anforderungen, um die langfristige Zugänglichkeit und Nachnutzbarkeit von Forschungsdaten mit möglichst geringen Kosten und Aufwänden zu ermöglichen.

Antworten auf diese Herausforderungen müssen in lokalen Initiativen gefunden, und an die jeweiligen Kontexte und Zielgruppen angepasst werden. Gleichzeitig sind sie relevant für die gesamte DH-Community, denn Vertrauen in das Angebot eines Datenzentrums kann nur auf Basis eines gemeinsamen Verständnisses und gemeinsamer Werte entstehen. Letztlich kann und muss die gesamte DH-Community zu Entwicklung, Betrieb und Nachhaltigkeit der Angebote von Datenzentren beitragen. Dies impliziert insbesondere auch die Notwendigkeit einer stärkeren Vernetzung oder gar Föderation zwischen den Datenzentren, die zu einer gemeinsamen Konzeptentwicklung, einer besseren Lastverteilung und einer insgesamt stabileren Forschungsdatenlandschaft beiträgt.

Gegenwärtig scheinen Fragen aus drei Bereichen zu diskutieren zu sein ...

- **Grundbegriffe, Paradigmen und Definitionen** im Forschungsdatendiskurs. Was sind Daten und Forschungsdaten? Gibt es spezifische Datentypen in den Geisteswissenschaften? Wie sieht der Research Data Life Cycle in diesem Bereich aus? Müssen wir zwischen "Daten" und "Ressourcen" unterscheiden? Was bedeutet "Archivierung" und "Kuratierung"? Was sind "Datenzentren" - als konzeptioneller und als institutioneller Begriff? Wie werden allgemeine Begriffsdiskussionen in lokale Konzepte übersetzt und welche Auswirkungen auf das jeweilige Angebot und die organisatorische Nachhaltigkeit haben sie? Welche Profile von Datenzentren lassen sich hinsichtlich ihrer Ausrichtungen unterscheiden? Wie lässt sich das Anforderungsprofil an Datenzentren und das Angebotsspektrum von Datenzentren sinnvoll strukturieren?
- **Lokale Umsetzung und Organisation der Nachhaltigkeit** geisteswissenschaftlicher Daten. Welche Strategien zur Sicherstellung von Nachhaltigkeit können erfolgreich sein? Wie unterscheiden sich Datenzentren in Bezug auf ihre Struktur und Zielgruppe (z.B. institutionelle, disziplinspezifische, nationale Datenzentren). Wie können Angebote von Datenzentren strukturiert werden

(z.B. nach Gigabyte, Datentypen, Projekten, Forschungsmethoden, etc.), so dass Vergleichbarkeit hergestellt werden kann? Wie können die Sicherungsprozesse im Einzelnen organisiert werden: die Übernahme der Daten und Ressourcen, ihre Beschreibung und Dokumentation durch Metadaten, sowie ihre Langzeitarchivierung und ggf. der dauerhafter Betrieb von einzelnen Systemen. Braucht man jenseits der Archivierung auch transparente Lösch-Prozesse für ungenutzte Daten? Wie lassen sich für diese Prozesse und Leistungen geeignete langfristige Finanzierungsmodelle denken?

- **Zusammenarbeit mit den FachwissenschaftlerInnen.** Müssen Forschungsprojekte bereits ab Antragstellung durch erfahrene Datenkuratoren beraten und begleitet werden, um das Datenmanagement bereits von Beginn an auf Nachhaltigkeit auszurichten und um gleichzeitig den Gesamtaufwand von Wissenschaftlern für Datenmanagement-Aufgaben möglichst gering zu halten? Wie kann eine gute Balance zwischen Standardisierung und innovativen projektspezifischen Lösungen gefunden werden? Wie können Standardisierungsbemühungen in Fachcommunities unterstützt werden, die gleichzeitig übergreifende Sicherungsprozesse in den Datenzentren erst realistisch machen? Wie kann DH-weit die Publikation von kuratierten Daten aus der Forschung als wissenschaftliche Leistung gefördert werden, geleitet von der Vision eines Impact Factors für solche veröffentlichten Ressourcen?

Das Panel wird in drei "Runden" verlaufen. Zunächst werden einige Vertreter der verschiedenen Institutionen und Projekte kurz über den jeweiligen Entwicklungsstand und Ziele berichten. Danach sollen die oben genannten Themenfelder vorgestellt und thesenartig beleuchtet werden. Schließlich soll die Diskussion der aufgeworfenen Fragen mit dem Publikum geführt werden.

Als Teilnehmer des Panels sind vorgesehen: Matej Durco; Daniel Kurzawe; Lukas Rosenthaler; Patrick Sahle (Convenor der AG); Johannes Stigler und ggf. weitere AG-Mitglieder