

Humanities Data Centre – grundlegende Überlegungen in der Designphase eines geisteswissenschaftlichen Forschungsdatenzentrums

Stefan Buddenbohm¹, Claudia Engelhardt², Ulrike Wuttke³

¹Max-Planck-Institut zur Erforschung multireligiöser und multiethnischer Gesellschaften, ²Staats- und Universitätsbibliothek Göttingen, ³Akademie der Wissenschaften zu Göttingen
buddenbohm@mmg.mpg.de, claudia.engelhardt@sub.uni-goettingen.de, uwuttke@gwdg.de

Schlagwörter: Forschungsdatenzentrum, Langzeitarchivierung, Forschungsdatenmanagement

Zusammenfassung: Forschungsdaten sind sowohl Ergebnis von Forschung als auch Grundlage für neue Forschungsfragen. Die zunehmende Nutzung digitaler Ressourcen und Methoden in der Forschung widerspiegelt sich sowohl im wachsenden Umfang als auch in der zunehmenden Komplexität von digitalen Forschungsdaten, sowohl in den Geisteswissenschaften wie auch in anderen Disziplinen. Aus verschiedenen Gründen ist die Erhaltung dieser Forschungsdaten notwendig: Dokumentationszwecke beispielsweise für Förderer oder aufgrund rechtlicher Bestimmungen, Nachvollziehbarkeit und Reproduzierbarkeit von Forschungsergebnissen, aber vor allem auch die Möglichkeit der Nachnutzung für neue Forschungsvorhaben. Die Herausforderungen hinsichtlich des Forschungsdatenmanagement und der Langzeitarchivierung können jedoch nur mit einem umfassenden Verständnis ihrer Entstehungs- und Nutzungsbedingungen gemeistert werden. Da diese von Infrastrukturanbietern nur im engen Austausch mit den Fachdisziplinen eruiert werden können, scheinen disziplinspezifische Forschungsdatenzentren am besten geeignet, die damit verbundenen Aufgaben zu übernehmen.

Während der Designphase des Humanities Data Centres (HDC, 2014-2016) werden daher im Dialog mit der Wissenschaft und Infrastruktureinrichtungen die Grundlagen für den Aufbau eines Forschungsdatenzentrums für die Geisteswissenschaften geschaffen. Das Projektkonsortium besteht neben geisteswissenschaftlichen Forschungseinrichtungen aus Rechenzentren und einer Universitätsbibliothek.

Grundsätzlich lässt sich die Langzeitarchivierung von Forschungsdaten entlang von drei aufeinander aufbauenden Ebenen strukturieren:

- Bitstream Preservation: Der physische Erhalt des gespeicherten Datenobjekts (Bitstream) auf einem entsprechenden Speichermedium,
- Technische Nachnutzbarkeit: Sicherstellung der Zugänglichkeit der Forschungsdaten auch bei veränderten technischen Bedingungen,
- Intellektuelle Nachnutzbarkeit: Sicherstellung der vollständigen Nutzbarkeit und Interpretierbarkeit des intellektuellen Gehalts der Forschungsdaten, beispielsweise durch Metadaten und die Dokumentation von Kontextinformationen, die das ursprüngliche Forschungsszenario nachvollziehbar machen.

Darüber hinaus hängt die Nachhaltigkeit von Forschungsdaten stark von einem stabilen, organisatorischen Rahmen ab, innerhalb dessen die entsprechenden Umgebungen und Werkzeuge bereitgestellt werden können. Nicht zuletzt ist aber der beständige Austausch mit den wissenschaftlichen Nutzern von großer Bedeutung, um mit dem Angebot (dem Forschungsdatenzentrum) den Anforderungen der Wissenschaft zu entsprechen beziehungsweise dieses neuen Entwicklungen und Bedürfnissen anzupassen.

Vor diesem Hintergrund stellen sich bei der Konzeption eines geisteswissenschaftlichen Forschungsdatenzentrums, das sowohl die Langzeitarchivierung als auch die Bereitstellung der Forschungsdaten für die Nachnutzung sicherstellen soll, verschiedene Fragen:

- Was sind Forschungsdaten in den Geisteswissenschaften und welche Forschungsdatentypen sollen vom Angebot des Datenzentrums berücksichtigt werden? Wie können geeignete Objektmodelle für die Bereitstellung und Archivierung dieser Forschungsdaten aussehen? Wie kann mit Forschungsdaten umgegangen werden, die nicht dokumentenbasiert sind, sondern beispielsweise aus Datenbanken bestehen?

- Wie kann die Zusammenarbeit zwischen Wissenschaft und Forschungsdatenzentrum erfolgreich sein? Welche Angebote hinsichtlich Beratung und Schulung sind besonders geeignet, um der Bedeutung des Forschungsdatenmanagements gerecht zu werden? Welche Implikationen hat das für mögliche Organisationsformen (-einheiten) eines Forschungsdatenzentrums?
- Die langfristige Nachhaltigkeit und Nachnutzbarkeit von Forschungsdaten ist nicht nur ein technisches, sondern vor allem ein organisatorisches Thema. Bestimmte geisteswissenschaftliche Forschungsdaten (zum Beispiel Editionen, Korpora, Wörterbücher) behalten über einen längeren Zeitraum ihre Forschungsrelevanz. Wie lässt sich diese Anforderung in organisatorischer und infrastruktureller Hinsicht umsetzen?
- Welche bestehenden und zukünftigen Standards für ein Forschungsdatenzentrum sind zu beachten, um Interoperabilität und Kooperation zwischen Forschungsdatenzentren zu fördern? Wie kann dies in Einklang mit der Anforderung der Skalierbarkeit gebracht werden?
- Ein Forschungsdatenzentrum muss über einen längeren Zeitraum lernen und sein Angebot anpassen. Gleichzeitig ist aber die Stabilität der konkreten (technischen) Angebote wichtig: für die technische Infrastruktur zum stabilen Aufbau der technischen Dienste; für Nutzer des Forschungsdatenzentrums um bereits zu Projektbeginn die Angebote in ihr Datenmanagement einplanen zu können und auch zu Projektende noch darauf vertrauen zu können. Stabilität und Innovation sind dabei zwar keine Gegensätze, müssen aber gegeneinander abgewogen werden. Wie leistet ein Forschungsdatenzentrum diesen Ausgleich zwischen Erneuerung und Stabilität des Angebotes?