

Bausteine Forschungsdatenmanagement
Empfehlungen und Erfahrungsberichte für die Praxis von
Forschungsdatenmanagerinnen und -managern

An der Schnittstelle von Fach- und Informationswissenschaft

Das INF-Projekt des SFB 980 "Episteme in Bewegung.
Wissenstransfers von der Alten Welt bis in die frühe Neuzeit"

Sibylle Söringⁱ Germaine Götzelmannⁱⁱ Philipp Hegelⁱⁱⁱ
Michael Krewet^{iv} Danah Tonne^v

2019

Zitiervorschlag

Söring, Sibylle, Germaine Götzelmann, Philipp Hegel, Michael Krewet und Danah Tonne. 2019. An der Schnittstelle von Fach- und Informationswissenschaft. Das INF-Projekt des SFB 980 "Episteme in Bewegung. Wissenstransfers von der Alten Welt bis in die frühe Neuzeit". *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 2/2019: S. 89-95. DOI: [10.17192/bfdm.2019.2.8083](https://doi.org/10.17192/bfdm.2019.2.8083).

Dieser Beitrag steht unter einer
[Creative Commons Namensnennung 4.0 International Lizenz \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).

ⁱFreie Universität Berlin. ORCID: [0000-0002-1698-3289](https://orcid.org/0000-0002-1698-3289)

ⁱⁱKarlsruher Institut für Technologie.

ⁱⁱⁱFreie Universität Berlin. ORCID: [0000-0001-6867-1511](https://orcid.org/0000-0001-6867-1511)

^{iv}Freie Universität Berlin. ORCID: [0000-0001-7807-6089](https://orcid.org/0000-0001-7807-6089)

^vKarlsruher Institut für Technologie.

1 Der Sonderforschungsbereich

Der an der Freien Universität Berlin angesiedelte Sonderforschungsbereich 980 "Episteme in Bewegung. Wissenstransfers von der Alten Welt bis in die frühe Neuzeit"¹ untersucht seit 2012 Prozesse des Wissenstransfers und Wissenswandels in europäischen und nicht-europäischen Kulturen vom dritten Jahrtausend v.Chr. bis ca. 1750 n.Chr. Dabei wird eine grundsätzliche Neuorientierung der wissenschaftsgeschichtlichen Forschung zur Vormoderne angestrebt. In derzeit 25 Teilprojekten mit über 80 Forschenden aus 27 Fachdisziplinen bündelt der SFB vor allem die 'Kleinen Fächer' der Freien Universität aus den Fachbereichen Philosophie und Geisteswissenschaften sowie Geschichts- und Kulturwissenschaften. So sind neben der griechischen und lateinischen Philologie, den Religions-, Kunst- und Literaturwissenschaften u.a. die Ägyptologie, Altorientalistik, Arabistik, Judaistik und Neogräzistik beteiligt. Aus dieser disziplinären Vielfalt speist sich das breite Spektrum von Forschungsgegenständen – von der altägyptischen Pyramideninschrift bis hin zum frühneuzeitlichen Sprachlehrwerk für Geschäftsreisende – und die interdisziplinäre Expertise, derer es bedarf, um historische Wissensbestände und -bewegungen aus transkultureller Perspektive zu untersuchen.

2 Das INF-Projekt: Ziele und Ressourcen

Den Sonderforschungsbereich unterstützt seit Beginn der zweiten Förderphase 2016 ein Informationsinfrastrukturprojekt (INF)², das eine Dateninfrastruktur (Repositoryum, Analysesoftware; vgl. die Abschnitte 3 und 4) für die digital arbeitenden Teilprojekte aufbaut. Die in diesem Rahmen in Kooperation der Freien Universität Berlin mit dem Karlsruher Institut für Technologie (KIT) entwickelten Methoden und Verfahren sollen perspektivisch auch projektübergreifend nutzbar sein. Durch die Datenmodellierung nach internationalen Standards und die strukturierte Vorhaltung der Forschungsdaten in einem Repositoryum soll zudem auch nach Projektende Anschlussforschung ermöglicht werden. Das INF-Teilprojekt bündelt sowohl fachwissenschaftliche als auch informationstechnologische und infrastrukturelle Expertise, um einerseits innovative Forschungsmethoden der Digital Humanities zu entwickeln und zu erproben, andererseits Forschungsdaten und -ergebnisse nachhaltig nutzbar zu machen. Die informationswissenschaftliche Personalausstattung am KIT wird an der Freien Universität durch zwei fachwissenschaftliche Stellenanteile ergänzt. Flankiert wurde das INF-Projekt von 2016 bis 2018 außerdem von dem Gastprojekt "Handschriften in Bewegung. Werkzeuge zur Dokumentation, Auswertung und Visualisierung texttopographischer Dynamiken"³ in Kooperation mit der Technischen Universität Darmstadt, in dem

¹<http://www.sfb-episteme.de/>. Letzter Zugriff am 17.10.2019.

²<http://www.sfb-episteme.de/teilprojekte/informationsinfrastruktur/index.html>. Letzter Zugriff am 17.10.2019.

³<http://www.sfb-episteme.de/teilprojekte/gastprojekt/index.html>. Letzter Zugriff am 17.10.2019.

insbesondere innere und äußere Wissenstransfers in Manuskriptüberlieferungen behandelt wurden. Die Freie Universität Berlin unterstützt das INF-Projekt zudem durch eine am Center für Digitale Systeme / Universitätsbibliothek angesiedelte, haushaltsgeförderte Stelle zum Infrastruktur-Auf- und Ausbau an der Heimatinstitution sowie zur Erarbeitung eines Nachhaltigkeitskonzepts⁴.

3 Gemeinsame Forschung und Entwicklung

In bewusst enger Verzahnung von Fach- und Informationswissenschaft entwickelt und implementiert das INF-Projekt in der gegenwärtigen Förderphase unter dem Titel "Bücher auf Reisen. Informationstechnologische Erschließung von Wissensbewegungen in vormodernen Kulturen" zum einen ein zentrales Forschungsdatenrepositorium für den SFB und betreibt zum anderen gemeinsame Forschung und Entwicklung mit derzeit vier ausgewählten Pilotprojekten. Der vorliegende Beitrag skizziert, wie die Kollaboration von Geisteswissenschaften, Computerphilologie und Informatik sowohl innerhalb des INF-Projekts als auch im Zusammenspiel mit den Pilotprojekten ausgestaltet wird, welche Herausforderungen bei gemeinsamer Forschung entstehen und welche Rolle das INF-Projekt eines großen geisteswissenschaftlichen Verbundes in den Nachhaltigkeitsüberlegungen der Heimatinstitution spielen kann.

In den knapp drei Jahren seines Bestehens hat das INF-Projekt verschiedene Dienste und Werkzeuge entwickelt und erprobt, die auf Open Source-Lösungen und Standards (z.B. Web Annotation Data Model⁵) aufbauen und in spezifischen Anpassungen die Anforderungen der jeweiligen Teilprojekte bedienen. Dazu zählt neben dem Datenrepositorium unter anderem auch ein Web Annotation Protocol Server⁶. In der Infrastruktur werden nicht nur die statischen, unveränderlichen Metadaten zugänglich gespeichert, sondern auch die ‚dynamischen‘ Metadaten, die bei Annotation von digitalisiertem Quellenmaterial entstehen.

Die gemeinsame Arbeit von Fach- und Informationswissenschaftlern an Forschung und Entwicklung steht bei dem INF-Projekt im Vordergrund, zum Beispiel im Kontext dynamischer Metadaten. Werkzeuge, Dienste und Datenmodelle werden unmittelbar und forschungsgetrieben in den jeweiligen Arbeitsprozess der Pilotprojekte eingegliedert. Wie das Beispiel der weiter unten dargestellten Untersuchungen zu Wissenstransfers zeigt, kann dabei die jeweilige geisteswissenschaftliche Forschungsfrage im gegebenen Zusammenhang nicht oder nicht zureichend ohne die im INF-Projekt entwickelten Methoden, Verfahren und Infrastrukturen beantwortet werden.

⁴<https://www.cedis.fu-berlin.de/services/projektentwicklung/aktuell/dh-infrastruktur/index.html>.
Letzter Zugriff am 17.10.2019.

⁵vgl. <https://www.w3.org/TR/annotation-model>. Letzter Zugriff am 17.10.2019.

⁶vgl. <https://www.w3.org/TR/annotation-protocol> (letzter Zugriff am 17.10.2019.) sowie Tonne, Danah et al.: Ein Web Annotation Protocol Server zur Untersuchung vormoderner Wissensbestände. In: DHd 2019 Digital Humanities: multimedial & multimodal. Konferenzabstracts, Frankfurt am Main, S. 285-288 (978-3-00-062166-6), <https://doi.org/10.5281/zenodo.2596095>.

4 Anforderungen aus den Fachwissenschaften: Mehrwerte für die Kodikologie im Kontext der Erforschung von Wissenstransfers

Im Projekt "Prozesse der Traditionsbildung in der *De interpretatione*-Kommentierung der Spätantike"⁷ werden Wissenstransferprozesse in Verbindung mit der Überlieferung der Aristotelischen Schrift *De interpretatione* erforscht. Diese Überlieferung stellt einen konkreten Fall für einen hoch komplexen Wissenstransfer dar, der in der Fachliteratur als innerhalb eines Menschenlebens kaum erforschbar gilt. Zu Beginn des INF-Projekts lagen in diesem Pilotprojekt bereits äußerst umfangreiche Digitalisatdaten vor. Zusätzlich wurden deskriptive Metadaten zu den Digitalisaten fachwissenschaftlich aggregiert, modelliert und inhaltlich angereichert. Entsprechend den lizenzrechtlichen Möglichkeiten können die Digitalisate mit ihren Metadaten in einem zentralen Repositorium abgelegt werden. Über Annotationstools können Forschende unmittelbar am jeweiligen Digitalisat z.B. Transkriptionen von überlieferten Texten und Diagrammen vornehmen. Diese Transkriptionen und Beschreibungen dienen der Erforschung von Verwandtschaften einzelner Handschriften. Sie können perspektivisch durchsucht, quantitativ analysiert und auch in Verbindung mit deskriptiven Metadaten zu den Manuskripten (z.B. zur Provenienz, Datierung usw.) ausgewertet werden. In einem ersten Schritt können Verwandtschaften von Handschriften untereinander erforscht werden, um anschließend tiefere Einsichten in die Modalitäten von Wissenstransferprozessen der Vormoderne zu gewinnen. Ein Datenrepositorium sowie spezifische digitale Werkzeuge bilden wesentliche Bestandteile der täglichen Forschungspraxis. Für die fachwissenschaftliche Forschung erweisen sich beide – Datenrepositorium und digitale Werkzeuge – als methodische Erweiterungen, die neue Forschungserkenntnisse ermöglichen.

Repositorium, Dienste und Tools bilden eine Arbeitsinfrastruktur, mit der Forschende kollaborativ und dezentral an Wissenstransferprozessen arbeiten können. Digitale Tools wie z.B. Annotations- und Auswertungstools können über REST-Schnittstellen mit Funktionalitäten zum Erzeugen, Auslesen und Aktualisieren von Datensätzen direkt an das Repositorium angebunden werden. Neben den Daten müssen auch diese Tools und Dienste technisch nachhaltig gepflegt und bereitgestellt werden, um sie weiterhin im Sinne der FAIR-Prinzipien⁸ fachwissenschaftlich nutzen zu können. Die Freie Universität Berlin als Heimatinstitution des SFB nimmt hierbei eine zentrale Rolle ein.

Die intensive Nutzung der Forschungssoftware durch die Fachwissenschaft ermöglicht es, durch ein kontinuierliches Feedback die Software fortlaufend an die jeweiligen fachwissenschaftlichen Anforderungen anzupassen. Eine solche wissenschaftsge-

⁷<http://www.sfb-episteme.de/teilprojekte/sagen/A04/index.html>. Letzter Zugriff am 17.10.2019.

⁸https://www.go-fair.org/wp-content/cache/wp-rocket/www.go-fair.org/fair-principles/index.html/_gzip. Letzter Zugriff am 17.10.2019.

triebene Softwareentwicklung trägt wesentlich zu technischen wie fachwissenschaftlichen Fortschritten in Forschungsfragen bei.

5 Herausforderungen und Lösungsansätze

5.1 Lessons learned

Gestartet mit der zweiten Förderphase (2016-20), musste zu Beginn des INF-Projekts 2016 der überwiegende Teil der Forschungsdaten zunächst noch erhoben und modelliert werden. Die zentralen Infrastrukturkomponenten konnten nach Jahresfrist prototypisch vorgelegt werden; im Vordergrund der Arbeit der INF-Teilprojekte stand in der zweiten Förderphase die Generierung und Aufbereitung der Materialien für die digitale Speicherung und Auswertung. Nicht immer ist der damit verbundene zeitliche und personelle Aufwand für die fachwissenschaftlichen Teilprojekte im Voraus – und zumal für einen Zeitraum von mehreren Jahren – kalkulierbar: Wie komplex sich z.B. eine entsprechende Quellenerschließung darstellt, die auf die digitale, quantitative Analyse ausgerichtet ist, zeigt sich häufig erst im Kontext der Erhebung selbst. Auch erweist sich die Datengrundlage nicht in allen Fällen als ergiebiger "use case", welcher weiter zu verfolgend und für die Weiterentwicklung von Analysesoftware nachnutzbar ist.

Solche Lernprozesse sind für die weitere Planung äußerst wertvoll und werden nach Möglichkeit für die Beantragung einer weiteren Förderphase nutzbar gemacht. Bei der Konzeption eines vergleichbaren INF-Projekts wäre daher die dezidierte Einplanung von Personalressourcen auch innerhalb der jeweiligen fachwissenschaftlichen Teilprojekte für die Aufbereitung von (analogen wie digitalen) Materialien für die digitale Auswertung dringend anzuraten.

5.2 Management und Ergebnispräsentation nicht publizierbarer Forschungsdaten

Im Kontext der Publikation bzw. Zugänglichmachung von Forschungsdaten und ihrer digitalen Auswertung stellt lizenzrechtlich geschütztes Material das INF-Projekt des SFB 980 vor Herausforderungen. Hierzu zählt u.a. auch die Frage, wie Forschungsergebnisse an derartigem Material, z.B. den Digitalisaten im weiter oben bereits genannten Teilprojekt zur *De interpretatione*-Kommentierung online präsentiert werden können. Dies gilt etwa für die Auswertung von Annotationen zu nicht gemeinfreiem Bildmaterial: Diese Annotationen sind Resultate wissenschaftlicher Arbeit, aber ohne öffentlich zugängliches digitales Objekt nur schwer sinnvoll zu publizieren. Das Repositorium mit seinen z.T. bereits im Produktivbetrieb befindlichen Werkzeugen zur Annotation von Bild und Text sowie zur Variantenerkennung wird es ermöglichen, ursprünglich am Digitalisat vorgenommene Annotationen in der Präsentationsschicht

mit einem gemeinfreien und im Fall von *De interpretatione* kanonisch gezählten Text zu verbinden und zu visualisieren.

5.3 Technische, fachwissenschaftliche und infrastrukturelle Nachhaltigkeit

Die technische Nachhaltigkeit umfasst nicht nur die Speicherung, Zugänglichkeit und permanente Referenzierbarkeit von Forschungsdaten und -ergebnissen, sondern auch die zur Generierung, Analyse und Visualisierung von Daten entwickelten Werkzeuge. Insbesondere für die Werkzeuge und Dienste sind außerdem Konzepte zur fachwissenschaftlichen Nachhaltigkeit zu entwickeln, welche die Nutzung von im Verbundvorhaben entwickelten Infrastrukturen, Werkzeugen und Diensten sowie der gewonnenen Daten in Forschung und Lehre z.B. mittels Dokumentationen und Tutorien auch über das Projekt hinaus ermöglichen, begleiten und befördern.

Hinsichtlich der Entwicklung von Nachhaltigkeitskonzepten für die entwickelte Infrastruktur und ihre Funktionalitäten liegen die Herausforderungen auf institutioneller Ebene vor allem im Umstand begründet, dass technologische, organisatorische und strukturelle Infrastrukturen für geisteswissenschaftliche Forschungsdaten an der Freien Universität Berlin gegenwärtig noch im Aufbau begriffen sind. So liegen z.B. bislang zwar verbindliche Erklärungen für die Publikation im Open Access, jedoch keine universitären Leitlinien, Policies oder Services für das Forschungsdatenmanagement oder die Verwendung von Metadatenstandards vor. Die Anforderungen, aber auch die bereits erfolgten Arbeiten des SFB können für die Konzeption solcher Services wertvolle Hinweise liefern.

Auch das institutionelle Forschungsdaten- und Publikationsrepositorium "Refubium"⁹ steht derzeit noch nicht im vollen Umfang zur Verfügung: So ist z.B. keine Zugriffsbeschränkung für publizierte Forschungsdaten möglich; auch sind keine Werkzeuge für die Analyse und/oder Visualisierung von Forschungsdaten eingebunden. Für dynamische Datensätze, die Veränderungen unterliegen, ist keine generische Lösung verfügbar.

Insgesamt erweisen sich die personellen und technischen Ressourcen des Infrastrukturprojekts nach lediglich drei Jahren als noch nicht so weit skalierbar, dass sie als stabile Infrastruktur generisch und dauerhaft zur Verfügung gestellt werden könnten. Das Hosting wird daher derzeit und bis auf Weiteres durch das KIT gewährleistet. Vorgehen ist perspektivisch jedoch eine Nutzbarkeit sämtlicher entwickelter Werkzeuge und Dienste auf institutioneller Ebene sowie als Open Source-Publikation. Um aktuelle Entwicklungen auf regionaler sowie auf Bund-/Länderebene mitzuverfolgen und ggf. einbeziehen zu können, engagieren sich die Mitglieder des SFB in den einschlägigen

⁹<https://refubium.fu-berlin.de/>. Letzter Zugriff am 17.10.2019.

Gremien und Verbundvorhaben (Berlin University Alliance¹⁰, DARIAH-DE¹¹, Research Data Alliance (RDA)¹², Nationale Forschungsdateninfrastruktur (NFDI)¹³).

¹⁰<https://www.berlin-university-alliance.de/>. Letzter Zugriff am 17.10.2019.

¹¹<https://de.dariah.eu/en/>. Letzter Zugriff am 17.10.2019.

¹²<https://www.rda-deutschland.de/>. Letzter Zugriff am 17.10.2019.

¹³<https://www.dfg.de/foerderung/programme/nfdi/index.html>. Letzter Zugriff am 17.10.2019.