



Separatum aus:

THEMENHEFT 12

*Elisabeth Lienert / Joachim Hamm
Albrecht Hausmann / Gabriel Viehhauser (Hrsg.)*

Digitale Mediävistik

Perspektiven der Digital Humanities für die Altgermanistik

Publiziert im November 2022.

Die BmE Themenhefte erscheinen online im BIS-Verlag der Carl von Ossietzky Universität Oldenburg unter der Creative Commons Lizenz [CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/). Die ›Beiträge zur mediävistischen Erzählforschung‹ (BmE) werden herausgegeben von PD Dr. Anja Becker (München) und Prof. Dr. Albrecht Hausmann (Oldenburg). Die inhaltliche und editorische Verantwortung für das einzelne Themenheft liegt bei den jeweiligen Heftherausgebern.

<http://www.erzaehlforschung.de> – Kontakt: herausgeber@erzaehlforschung.de
ISSN 2568-9967

Zitiervorschlag für diesen Beitrag:

Rapp, Andrea: Digitale Infrastruktur und Forschungsdatenmanagement, in: Lienert, Elisabeth/Hamm, Joachim/Hausmann, Albrecht/Viehhauser, Gabriel (Hrsg.): Digitale Mediävistik. Perspektiven der Digital Humanities für die Altgermanistik, Oldenburg 2022 (BmE Themenheft 12), S. 81–96 (online).

Andrea Rapp

Digitale Infrastruktur und Forschungsdatenmanagement

Abstract. Seit 2020 wird in Deutschland mit einer Bund-Länder-Förderung die Nationale Forschungsdateninfrastruktur (NFDI) aufgebaut. Sie hat zum Ziel, Datenbestände aller Wissenschaftsbereiche entlang der FAIR-Prinzipien und ohne Disziplinen- und Ländergrenzen zu erschließen und langfristig zu sichern. Der Beitrag stellt das NFDI-Konsortium Text+ vor, das die digitale Infrastruktur und das Forschungsdatenmanagement für Sprach- und Textdaten organisiert. Hierbei liegt der Fokus zum einen auf Anwendungsbeispielen aus der Mediävistik, zum anderen auf den Möglichkeiten der Beteiligung der Fachcommunity, der zukünftigen Weiterentwicklung sowie den bereits artikulierten konkreten Bedarfen.

1. Vorbemerkung/Einleitung

Anders als in vielen Natur- und Ingenieurwissenschaften drängt sich ein Bezug zu ›Infrastrukturen‹ für viele Geisteswissenschaften zunächst nicht auf. Dennoch sind gerade die Geistes- und Kulturwissenschaften¹, die sich mit den kulturellen Erzeugnissen des Menschen (im weiten Sinne) befassen, seit jeher auf umfangreiche analoge Infrastrukturen angewiesen und haben ihren Aufbau, ihre Pflege und ihren Betrieb seit Jahrhunderten professionalisiert. Archive, Bibliotheken und Museen und die dazugehörigen Wissenschaften und Ausbildungsberufe sind für geisteswissenschaftliche Forschung und Lehre unabdingbar. Aus der umfassenden digitalen Transformation dieser Einrichtungen und des gesamten Wissenschaftsbetriebs erklären sich die Notwendigkeit und der Anspruch der Geistes-

wissenschaften auf professionelle digitale Infrastrukturen. Zum einen liegen immer mehr – wenn auch noch lange nicht alle – Quellen und Ressourcen geisteswissenschaftlicher Forschung als digitales Surrogat vor, zum anderen wird mit digitalen Methoden geforscht. In diesem Prozess entstehen annotierte, mit Expertise und Wissen angereicherte digitale Derivate der ursprünglichen Forschungsobjekte als Ergebnisse und Grundlagen weiterer geisteswissenschaftlicher Forschung, deren Bewahrung und Pflege – das Forschungsdatenmanagement – derselben Professionalität bedürfen wie analoge Archive, Bibliotheken und Museen.

Im Folgenden möchte ich zeigen, wie sich das [NFDI-Konsortium Text+](#) der Aufgabe annimmt, die digitale Infrastruktur und das Forschungsdatenmanagement für Sprach- und Textdaten in diesem Sinne aufzubauen und zu organisieren. Ich fokussiere dabei für den Beitrag in unserem Rahmen exemplarisch auf die mediävistische Philologie. Nach einer Definition dessen, was geisteswissenschaftliche Forschungsdaten umfassen und einer kursorischen Vorstellung der Nationalen Forschungsdateninfrastruktur NFDI gehe ich vor allem auf das Konsortium Text+ ein. Hierbei liegt der Fokus auf den Möglichkeiten der Beteiligung der Fachcommunity, der zukünftigen Weiterentwicklung sowie den bereits artikulierten konkreten Bedarfen.

2. Was sind Forschungsdaten in den Geisteswissenschaften?

Ganz vereinfacht gesagt, sind alle Daten, die im Rahmen des (geistes-)wissenschaftlichen Forschungsprozesses erzeugt und benötigt werden, Forschungsdaten. Mit dem Vorschlag von Johanna Drucker, solche Daten nicht als »data« (something given), sondern als »capta« (something actively taken) zu sehen, erhalten sie ihre epistemische Qualität (Drucker 2011, S. 3). Sowohl der [Rat für Informationsinfrastrukturen](#) (2016, Anhang, S. A–13) als auch das [DARIAH-Stakeholdergremium](#) »Sammlungen«

(Oltersdorf/Schmunk 2016, S. 3) haben Definitionen für Forschungsdaten erarbeitet.

Im Hinblick auf den Vorgang der Datenerhebung umfassen geisteswissenschaftliche Forschungsdaten demnach sämtliche digitale Daten, die bei Quellenerschließungen, Transkriptionen, Editionen, Erhebungen, Grabungen, Experimenten, Messungen, Simulationen, Aufnahmen oder Umfragen usw. entstehen oder deren Ergebnis sind. So vielfältig wie die Erhebungsmethoden sind auch ihre Formen, Medien und Formate, sie erscheinen als Texte, Bilder, mehrdimensionale Modelle, Audio, Video, Tabellen, Datenbanken, Dokumentationen, Software, Verzeichnisse usw. Sie liegen disziplinspezifisch in unterschiedlichen Aggregationsstufen vor und verwenden zumeist verschiedene digitale Formate. Sie sind eingebunden in den Forschungsdatenlebenszyklus, d. h. Ergebnis und Ausdruck einzelner Schritte des Forschungsprozesses bzw. Grundlage und Ausgangspunkt weiterer Schritte (vgl. forschungsdaten.info).

3. Die Geisteswissenschaften in der Nationalen Forschungsdateninfrastruktur

Die drängende Notwendigkeit, den nachhaltigen Betrieb digitaler Infrastrukturen für das Forschungsdatenmanagement für die gesamte Wissenschaftslandschaft in Deutschland zu organisieren, führte 2018 zu einer Bund- und Ländervereinbarung zum zunächst auf zweimal fünf Jahre befristeten Aufbau einer Nationalen Forschungsdateninfrastruktur ([Bund-Länder-Vereinbarung](#)). Als Ziel wurde die Erschließung und langfristige Sicherung von Datenbeständen entlang der FAIR-Prinzipien und ohne Disziplinen- und Ländergrenzen formuliert. Dafür sollten 30 Konsortien für eine breite Abdeckung der Wissenschaft eingerichtet werden.² Zentral für Akzeptanz und Bedarfsorientierung ist dabei, dass »[d]ie NFDI [...] in einem aus der Wissenschaft getriebenen Prozess als vernetzte Struktur eigeninitiativ agierender Konsortien aufgebaut [wird].« ([Deutsche Forschungs-](#)

[gemeinschaft 2020](#)). Das bedeutet, dass die Konsortien (und mithin die NFDI) von den Wissenschaftler:innen selbst in Kooperation mit passenden Infrastruktureinrichtungen geformt, betrieben und gesteuert werden. Durch diese Konstruktion wird gewährleistet, dass die Infrastruktur bedarfs- und wissenschaftsgerecht angelegt wird, sie erfordert jedoch auch von den Forschenden einen hohen Aufwand an Ressourcen und Zeit und nicht zuletzt an Kenntnissen in einem Maße, das vermutlich doch deutlich über das auch vorhandene Engagement für analoge Infrastrukturen hinausgeht. Die Nachhaltigkeit des Konstrukts wird sich ebenfalls noch erweisen müssen, denn als Förderung sind zunächst zweimal fünf Jahre vorgesehen, evaluiert wird der NFDI-Prozess durch die DFG und den Wissenschaftsrat (vgl. [Deutsche Forschungsgemeinschaft 2018](#)).

Für die Geisteswissenschaften stellen die Vielfalt und Heterogenität ihrer Disziplinen einerseits und die Überschneidungen bei den Quellen und Ressourcen bei gleichzeitiger Heterogenität der Zugriffsmethoden andererseits eine große Herausforderung dar. Dem sind sie durch einen intensiven Vernetzungs- und Abstimmungsprozess begegnet, der in einem Memorandum of Understanding dokumentiert ist, das für die erste Antragsrunde 2019 publiziert und anschließend für die weiteren Runden aktualisiert wurde (vgl. Brünger-Weilandt [u. a.] 2020 und [forschungsinfrastrukturen.de](#)). Dieses Memorandum legt den Grundstein für einen gemeinsamen Aufbau einer NFDI in den Geistes- und Kulturwissenschaften, indem sich die Initiativen zur Kooperation verpflichten und Verantwortlichkeiten und die Form ihrer Zusammenarbeit definieren.

Gemäß der Anlage der NFDI-Struktur als Community-getragene Institution haben sich auch die Geisteswissenschaften an den Bedarfen spezifischer Communitys organisiert. Das seit 2020 geförderte Konsortium NFDI4Culture adressiert die Architektur, die Kunstgeschichte, die Musik-, Theater- und Tanzwissenschaft sowie die Medienwissenschaften, während sich das seit 2021 geförderte Konsortium Text+ an alle sprach- und textbasierten Disziplinen wendet, insbesondere an die Linguistik, die Alt- und

Neuphilologien, die Literaturwissenschaft, die Philosophie, die Anthropologie sowie die nichteuropäischen Kulturen und Sprachen. Die aktuell (Stand Mai 2022) in der Begutachtung befindlichen Konsortien NFDI4-memory und NFDI4objects richten sich auf die Geschichte, die Wirtschafts- und Sozialgeschichte, die Area Studies, die Religionswissenschaft, die historische Philosophie sowie auf Gedächtnisinstitutionen bzw. auf die Archäologie, die Archäometrie, die Archeological Sciences, die Anthropologie, die Antiken Architekturen, die Cultural Heritage Studies und die Conservation Science. Als gemeinsame Querschnittsbereiche wurden Metadaten, Normdaten, Terminologien, Provenienz, Recht und Ethik sowie Data Literacy identifiziert.

Sprache und Text spielt in zahlreichen Bereichen eine wichtige, ja eine zentrale Rolle, nicht allein in den Geisteswissenschaften. Der Fokus von Text+ lässt sich dennoch gut eingrenzen durch das auf Sprache und Text selbst gerichtete Erkenntnisinteresse, das seine Community eint.

4. Das NFDI-Konsortium Text+

Wie oben bereits benannt, baut Text+ eine auf Sprach- und Textdaten ausgerichtete Forschungsdateninfrastruktur auf. Das Konsortium orientiert sich dabei in seiner Struktur dezidiert an Datendomänen, nicht an Fach-Disziplinen, um der Vielfalt und Heterogenität der Fächer und ihrer Nutzungscommunitys gerecht zu werden. Der aktuelle Fokus der ersten Förderphase liegt auf den Datendomänen digitale Sammlungen/Korpora, lexikalische Ressourcen und Editionen. Sie haben eine für die Disziplinen häufig konstitutive Funktion, eine lange Forschungstradition und sind mit ausgereiften methodologischen Paradigmen verknüpft, die charakteristische, aber auch bereichsübergreifende Praktiken der Erzeugung, Kuratierung und des Managements von Daten erfordern. Bevor diese Datendomänen gleich ausführlicher vorgestellt werden, soll kurz erläutert werden, wie bereits im Vorfeld bzw. bei der Profilierung des Konsortiums und

der Ausarbeitung des Antrags Bedarfe der Forschenden erhoben wurden. Ihre Berücksichtigung ist durch demokratische Einbindung in die Governance-Struktur und im Monitoring durch gewählte Wissenschaftler:innen fest in Text+ verankert.³

Im Sommer 2020 wurde ein initialer und thematisch offener Call for User Stories gestartet, mit denen eine große Bandbreite an Disziplinen, Datendomänen und Forschungsfragen erfasst werden sollte. Dieser Call erbrachte innerhalb kurzer Zeit 120 Rückmeldungen mit ausgearbeiteten User Stories erbrachte. Aufgrund der Vorgaben, möglichst eine inhaltliche Struktur einzuhalten, die die Bereiche 1. Motivation, 2. Ziele, 3. Herausforderungen, 4. Lösung, 5. Evaluation durch die Community umfasste, konnten die Stories sehr gut verglichen und ausgewertet werden und sind in großem Umfang in den Antrag eingegangen. Alle **User Stories** wurden nach Datendomänen sowie nach DFG-Fachsystematik klassifiziert und unter den Namen und mit Zustimmung der Autor:innen auf der Text+-Homepage veröffentlicht. Die Auswertung, die klare Prioritäten der Nutzenden ergeben hat (z. B. Interoperabilität und Nachnutzbarkeit sowie Zugänglichkeit von Daten), sowie die Daten für die Analyse sind ebenfalls veröffentlicht (Rißler-Pipka [u. a.] 2021). Diese Calls sollen in regelmäßigen Abständen wiederholt werden, um auf aktuelle Entwicklungen und Bedarfe eingehen zu können und die Nutzungscommunity erweitern zu können.

Im Folgenden wird nun beschrieben, welche Datentypen und -angebote zu den drei Datendomänen gerechnet werden, im Anschluss werden einige User Stories aus der Mediävistik exemplarisch vorgestellt und eingeordnet.

Sammlungen

Zu dieser umfangreichen, sehr offen gestalteten und vielfältigen Datendomäne zählen Sprach- und textbasierte Sammlungen aller Art. Sie umfasst Sammlungen geschriebener, gesprochener oder gebärdeter Sprache und Texte sowie sprach- und textbezogene Experimental- oder Messdaten, die auf Grundlage wissenschaftlicher Kriterien gesammelt wurden, z. B.

Textsammlungen, mono- und multimodale Aufnahmen, Sensordaten, Befragungen, Verzeichnisse, Kataloge usw. Zu dieser Domäne gehören 61 User Stories, sie stellt damit mit über 50 Prozent Anteil den weitaus größten Bereich und zeigt den immensen Bedarf am Aufbau, insbesondere an der nachhaltigen und offenen Verfügbarkeit digitaler Sammlungen aller Art für die geisteswissenschaftliche Forschung auf. Drei Beispiele aus der Mediävistik demonstrieren zum einen die Bandbreite der Datentypen und der damit verbundenen Anforderungen, zum anderen nicht zuletzt die Interdisziplinarität der ›Mediävistik‹ im Sinne eines *umbrella term*.

Die User Story ›Reference Corpus Middle Low German/Low Rhenish (1200–1650) (ReN)‹⁴ beschreibt eine in der Forschung gängige Situation, wie Korpora mit viel Aufwand aufgebaut und mit erschließenden Annotationen angereichert werden. Eine Verknüpfung mit weiteren Korpora und niedrigschwelligen Abfrage- und Nutzungsmöglichkeiten einerseits könnten den Wert für die philologische Forschung noch deutlich erhöhen, eine nachhaltige Bereitstellung durch technologische Aktualisierungen andererseits schonte Ressourcen und sorgte für wissenschaftliche Transparenz und Kontinuität. Projektförmige Strukturen und Personalmangel und -fluktuation führen häufig dazu, dass solche wertvollen Ressourcen verwaisen – eine zentrale Infrastruktur könnte hier Abhilfe schaffen.

Die Sammlung ›Roman Seals and Inscriptions‹⁵ adressiert eine Quellengruppe, die nur selten in regionalen Inschriftenkorpora aufgenommen wird bzw. dort nachgewiesen ist, dadurch wird die Forschung erschwert. Insbesondere der philologischen Forschung sind wertvolle Quellen nur schwer zugänglich und damit kaum mit anderen Inschriftenquellen vergleichbar, so dass hier kein Gesamtüberblick erreichbar ist. Eine Koordination der weltweit verstreuten Quellen und Interoperabilität der Erschließungsinstrumente würde die Forschung also entscheidend erleichtern.

Die ›Bibliotheca Arabica‹⁶ erschließt die Geschichte der Arabischen Literatur von den Anfängen bis ins 19. Jahrhundert und mit ihr eine reichhaltige und vielfältige Handschriftenkultur. Eine besondere Herausforderung

ung liegt hier in der Verarbeitung der nichtlateinischen Schrift. Auch hier werden klar Unterstützung und Koordination beim Thema Interoperabilität und Standardisierung adressiert.

Diese (und weitere) User Stories im Feld ›Sammlungen‹ eint, dass zum einen zentrale Anlaufstellen und Orte für nachhaltige Sicherung und Pflege von Daten gefordert werden, weil Einzelforschende oder zeitlich befristete Projekte diese Nachhaltigkeit nicht leisten können, zum anderen der Wunsch nach Standardisierung und Interoperabilität – auch auf internationalem Level.

Lexikalische Ressourcen

Unter Lexikalische Ressourcen fassen wir Daten, die die Verwendung von Wörtern in Sätzen, Texten und multimodaler Kommunikation beschreiben, z. B. Wörterbücher und Enzyklopädien, Normdaten, terminologische Datenbanken, Wortnetze, Ontologien, Wortkarten und linguistische Atlanten. Obwohl die Wörterbucharstellung und die Lexikographie seit langem und umfassend digital transformiert sind und Normdaten und Ontologien zu den Schlüsseltechnologien eines FAIRen Forschungsdatenmanagements gehören, sind hier mit 19 User Stories die wenigsten Rückmeldungen eingegangen. Dies ist darauf zurückzuführen, dass beispielsweise Wörterbucharbeit in vergleichsweise wenigen Vorhaben und Arbeitsstellen gebündelt ist, anders als Sammlungen und Korpora, die fast flächendeckend in den Wissenschaften entstehen.

Die User Story ›Old High German Dictionary: Supplements‹⁷ spricht ein klassisches, langlaufendes Akademiewörterbuch an, das als Sprachstadienwörterbuch und Thesaurus die Aufgabe hat, das Althochdeutsche gesamthaft zu erfassen und zu dokumentieren. Die digitale Version ermöglicht zwar leicht die Aufnahme von Neufunden und Korrigenda, dennoch muss der Konnex zum gedruckten Wörterbuch und damit die Zitierfähigkeit gewahrt werden. Neben Datenhaltung und Versionierung tut sich damit die

Anforderung auf, Datenaggregation zu steuern und die Daten insgesamt persistent zu adressieren.

Dagegen geht es bei der User Story ›Word family database of historical German⁸ um ein Vorhaben, das auf den vorhandenen historischen Wörterbüchern und Referenzkorpora zum Deutschen aufbaut und diese Daten nachnutzt. Bedarf besteht insbesondere im Hinblick auf ein zentrales Repository sowie bei der Harmonisierung vor allem der Metadaten, aber auch bei der Entwicklung von Tools.

Die Verbindung von gedruckten und digitalen Ergebnispublikationen und das dazugehörige Spannungsfeld von statischer, aber recht nachhaltiger Buchpublikation und dynamischer Webpublikation ist ein Aspekt vieler zentraler Nachschlagewerke und Wörterbücher – allein aufgrund des Umfangs des dort bearbeiteten Materials und der damit zusammenhängenden langen Laufzeit von Vorhaben, die zum Teil ja noch vor der Digitalen Transformation begonnen wurden. Zugleich wird in Feld Lexikalische Ressourcen besonders deutlich, welche Potenziale in der Aggregation, Interoperabilität, Nachnutzung und übergreifenden Auswertungsmöglichkeit dieser Daten liegen.

Editionen

Editionen sind kritische Repräsentationen historischer Dokumente. Sie bestehen aus der zuverlässigen methodengeleiteten Bewahrung, Präsentation und Kommentierung aller Arten von Texten in verschiedenen Sprachen und Schriftsystemen. Sie liefern damit unabdingbare Grundlagen für (geistes-)wissenschaftliche Forschungen. Die Relevanz zeigt sich bereits durch einfache Recherchen in Forschungsinformationssystemen; so liefert die Datenbank [GEPRI](#)S rund 1500 ›Editionen‹, das Informationssystem der Akademien [AGATE](#) meldet 111 mit dem Schlagwort ›Edition‹ getaggte Vorhaben.⁹ In den 23 User Stories dieser Datendomäne wird der Bedarf nach nachhaltigen Infrastrukturen für die Sicherung der Ergebnisse ganz besonders deutlich.

«Editing a Medieval Hispanic Poetry Corpus with TextGrid»¹⁰ befasst sich mit der Edition mittelalterlicher Literatur, wobei die Forschenden an ihrem Standort in Argentinien nicht auf ein Umfeld gleichartiger Vorhaben vor Ort bauen konnten, sondern Kooperation im internationalen Bereich suchten. Die Nutzung von TextGrid-Komponenten war aufgrund der Offenheit zwar leicht möglich, dennoch ergaben sich im Detail Schwierigkeiten bei der Nutzung eigener TEI-Schemata und avancierterer X-Technologien. Auch eine Diversifizierung von Dokumentationssprachen (Multilingualität von Software und ihrer Dokumentation) wird angemahnt.

Die User Story »Digital Humanities and Medieval Literary Studies«¹¹ geht generisch auf die Problematik ein, dass für die Erstellung von Editionen neben philologischen auch technologische Kompetenzen erforderlich sind. Daher besteht ein hoher Bedarf an Beratung, Weiterbildung und Training bzw. spezifischer Kooperation. Bei Editionen entstehen wertvolle Forschungsdaten als »Zwischen- und Neben-Ergebnis« wie etwa Transkriptionen, für die ein zentraler Nachweis ebenso fehlt wie ein zuverlässiger Speicherort. Und nicht zuletzt sind das Interface einer Edition und seine Funktionalität ein zentrales Ergebnis, das ebenso wie die »(Roh-)Daten« gesichert werden muss.

»Inscriptions in Germany from the Middle Ages to Early Modern Times (Die Deutschen Inschriften)«¹² verdeutlicht die Besonderheit der Quellengattung Inschriften, die u. a. darin besteht, dass die Verbindung von Schrift, Schrifträger und Aufstellungsort nicht gelöst werden kann, zugleich sind diese Objekte ungeheuer vielfältig (Gebäude, Skulpturen, Grabsteine, Glocken, liturgische Objekte usw.). Die Erstellung von kontrollierten Vokabularen, das Aggregieren von Daten und die Aushandlung und Propagierung von Standards sind daher unabdingbar für die Inschriftenforschung und könnten über verschiedene geisteswissenschaftliche Konsortien hinweg adressiert werden.

Auch im Feld Editionen zeigt sich die spezifische Vielfalt der Anforderungen und Bedarfe: zum einen gibt es sehr viel Abstimmungsbedarf in

einem in vielfältiger Hinsicht und aus Forschungsgründen auch zurecht heterogenen Bereich, zum anderen muss nach organisatorischen und technischen Lösungen für die nachhaltige Sicherung von Editionen, insbesondere für die Roh- und ›Peripheriedaten‹ sowie für die Interfaces gesucht werden.

5. FAIR-Prinzipien und Vernetzung

Obwohl die Aufgliederung in die genannten drei Datendomänen aufgrund ihrer je spezifischen Anforderungen im Datenmanagement notwendig und sinnvoll ist, gibt es auch viele Berührungs- und Überschneidungspunkte, die nicht aus dem Blick geraten dürfen, sondern die erst eine echte wechselseitige Erschließung erlauben. Daher ist der systematische Ausbau von Verlinkungen zwischen den Datendomänen ein zentrales Ziel von Text+-. Editionen und Korpora liefern gesicherte Belege für lexikalische Ressourcen, während wiederum Lexikalische Ressourcen Editionen und Korpora strukturiert erschließen. Korpora liefern Rohdaten und Grundlagen für Editionen, während Editionen ›korpusfähig‹ und in Korpora integriert werden. Grundlage dafür sind zum einen persistente Identifikatoren und normierte Vokabulare bzw. allgemeine, standardisierte Normdaten wie die Gemeinsame Normdatei (GND), GeoNames sowie beispielsweise der Vokabular-Server DANTE. Hier finden sich persistent adressierbare Normdaten zu Personen, Körperschaften, Geografika, Ereignissen, Werken und Sachbegriffen, deren Verlinkung – z. B. in den Apparaten von Editionen – den wissenschaftlichen Kommentar von Routineangaben entlastet. Für die GND gilt beispielsweise, dass durch die Verbindung mit Thesauri und Fachdatenbanken der verschiedenen wissenschaftlichen Disziplinen einerseits und weiteren offenen Normdaten andererseits zu jeder GND-Entität nach Bedarf dann auch Daten aus der Linked Data Cloud wie Wikidata, Wikipedia, VIAF usw. ›automatisch‹ verlinkt werden können. Durch solche Verlinkungen werden Wissensbasen auf- und ausgebaut, so dass nicht

allein die Auffindbarkeit von Forschungsdaten (findability), sondern auch die Möglichkeiten von automatisierten Verfahren wie Text and Data Mining (reusability) entscheidend befördert werden. In Deutschland wird die GND als kollaboratives zentrales Tool breit disziplinübergreifend genutzt und kooperativ ausgebaut und gepflegt. Darüber hinaus ist sie international mit Normdaten (authority files) weiterer Nationalbibliotheken verlinkt und damit ein wichtiger Hub in der Linked Open Data Cloud, die auch in Text+ eine entscheidende Rolle spielt.¹³ Die GND soll daher für sprach- und textbasierte Forschungsdaten weiterentwickelt und ausgebaut werden, was nicht allein den Textdisziplinen zugutekommen wird, sondern als Beitrag zu einem domänenübergreifenden Wissensgraphen der NFDI insgesamt sowie weiten Bereichen der Kultur und Wissenschaft zu sehen ist. Im Rahmen von Text+ wird daher eine entsprechende GND-Agentur eingerichtet, um niedrigschwellige qualitätsgesicherte Beteiligungsmöglichkeiten für Forschende zu schaffen und zugleich den Vernetzungsgrad der GND auch durch Terminologie-Mappings zu erweitern.

Diese effiziente, das Wissensnetz erweiternde und verdichtende Erschließung und Sicherung unserer Quellen gelingt jedoch nur, wenn möglichst alle wissenschaftlichen Forschungsdaten im Sinne der **FAIR-** und **CARE-**Prinzipien verfügbar gemacht werden (Blumesberger 2021). Wenn wir erwarten, dass wir für unsere Forschung auf Daten des Kulturellen Erbes und geisteswissenschaftliche Forschungsergebnisse dauerhaft zugreifen können, müssen wir nicht allein Aspekte von Individualität/Besonderheit und Standardisierung aushandeln, sondern auch und vor allem unsere eigenen Daten und Ergebnisse zur Verfügung stellen. Es ist jedoch auch klar, dass dies nur in einer vertrauenswürdigen und nachhaltig abgesicherten digitalen Forschungsinfrastruktur erfolgen kann – insofern muss die Chance NFDI von den Forschenden genutzt und mitgestaltet werden.

6. Zusammenfassung: Community-Beteiligung erwünscht!

Text+ bietet verschiedene und vielfältige Möglichkeiten der Partizipation und der Mitgestaltung. Da ist zum einen das Angebot, das Text+ in Form von Daten, Tools und Services an die Community macht, zum anderen die Einladung zum Engagement und zur Mitbestimmung im Konsortium. Hier greife ich nur einige Dinge exemplarisch heraus, über die Website und die Kommunikationsaktivitäten z. B. in den Sozialen Medien oder auch direkte Anfragen kann man sich über alle Angebote informieren und sich einbringen.

Tools alleine ohne Dokumentationen, Tutorials und Schulungen haben kaum Chancen auf Nutzung, daher sind Aktivitäten im Bereich Education and Training auch für Text+ zentral. Es werden eigene Workshops und Tagungen entwickelt, aber auch Beiträge zu anderen Workshops, Verbandstagungen und Summerschools usw. geleistet. Die Struktur und damit die Bündelung von Expertise nach Datendomänen ist besonders gut geeignet, um adressatenorientierte und zielgruppenspezifische Angebote zu entwickeln. Darüber hinaus wird die Community (mindestens) jährlich zu Vollversammlungen eingeladen, das Text+ Plenary 2022 findet am Montag, den 12. September 2022 in Mannheim statt.

Text+ verfügt über die Möglichkeit, das Portfolio zu erweitern, indem neue Partner und Angebote aufgenommen werden. Im April 2022 erfolgte eine erste Ausschreibung zur Einreichung von Kooperationsprojekten zur weiteren Integration von Daten und Services in die Text+-Infrastruktur, die über die sogenannten Flexfonds des Antrags ermöglicht werden und deren Auswahl von den Mitgliedern der SCCs vorgenommen wird.¹⁴ Nicht zuletzt soll noch einmal darauf hingewiesen werden, dass gerade die SCCs und das OCC als zentrale Mitbestimmungsorgane des Konsortiums vom Engagement der Community leben. Bei der Vollversammlung im September 2022 werden die Vertreter*innen für die nächsten Jahre gewählt (Weimer 2022).

Der oben beschriebene Call for User Stories sowie die Ausschreibung für Kooperationsprojekte werden regelmäßig wiederholt.

Es liegt also auch an den Mediävist*innen, ihrer Disziplin und ihren Anliegen in der NFDI eine starke Stimme zu geben.

Anmerkungen

- 1 Ich folge in unserem Zusammenhang einer weiten Definition von Geistes- und Kulturwissenschaften und werde im Folgenden Geisteswissenschaften als eben- solchen weit gefassten Terminus verwenden und nicht weiter ausdifferenzieren; vgl. abouthumanities.sagw.ch/02-was-geisteswissenschaften.html. Alle im Bei- trag angegebenen URLs wurden zuletzt eingesehen am 26.05.2022.
- 2 Stand Mai 2022 werden 19 Konsortien gefördert (www.nfdi.de/konsortien/); weitere 15 bewerben sich in der dritten Ausschreibungsrunde um Förderung. In- formationen zum Prozess hat aktuell www.dfg.de/foerderung/programme/nfdi/.
- 3 Siehe dazu die Governancessstruktur mit Scientific Coordination Committees und dem Operations Coordination Committee, die den Task Areas zugeordnet und deren Mitglieder u. a. von einschlägigen Fachverbänden nominiert und gewählt sind; www.text-plus.org/ueber-uns/governance/.
- 4 Ingrid Schröder, Robert Peters: www.text-plus.org/en/research-data/user-story-339/.
- 5 Marietta Horster: www.text-plus.org/en/research-data/user-story-308/.
- 6 Boris Liebrecht: www.text-plus.org/en/research-data/user-story-346/.
- 7 Uwe Kretschmer, Brigitte Bulitta: www.text-plus.org/en/research-data/user-story-517/.
- 8 Jost Gippert, Sarah Ihden, Ralf Plate, Ingrid Schröder: www.text-plus.org/en/research-data/user-story-516/.
- 9 Auswertungen Stand Mai 2020.
- 10 Gimena del Rio Riande: www.text-plus.org/en/research-data/user-story-407/.
- 11 Gabriel Viehhauser: www.text-plus.org/en/research-data/user-story-409/.
- 12 Arbeitsstelle Inschriften, Nordrhein-Westfälische Akademie der Wissenschaften und der Künste (Universität Bonn): www.text-plus.org/en/research-data/user-story-420/.
- 13 Vgl. die Grafik Folie 9 Semantische Interoperabilität in: Hinrichs [u. a.] 2020.
- 14 www.text-plus.org/forschungsdaten/kooperationsprojekte/.

Literaturverzeichnis

Sekundärliteratur

- Blumesberger, Susanne: Forschungsdaten in den Geisteswissenschaften. Bereits selbstverständlich oder doch noch etwas exotisch? In: o-bib. Das offene Bibliotheksjournal 8–4 (2021) ([online](#)).
- Brünger-Weilandt, Sabine/Bruhn, Kai-Christian/Busch, Alexandra W./Hinrichs, Erhard/Maier, Gerald/Paulmann, Johannes/Rapp, Andrea/von Rummel, Philipp/Schlotheluber, Eva/Schmidt, Dörte/Schrade, Torsten/Simon, Holger/Stein, Regine/Teich, Elke: Memorandum of Understanding by NFDI Initiatives from the Humanities and Cultural Studies, Version 28.09.2020 ([online](#)).
- Bund-Ländervereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI) vom 26. November 2018, in: Bundesanzeiger Amtlicher Teil, 21.12.2018 B10 ([online](#)).
- Deutsche Forschungsgemeinschaft: Nationale Forschungsdateninfrastruktur (NFDI): DFG übernimmt Auswahl und Evaluation der Konsortien, in: Pressemitteilung Nr. 58 (07. 12.2018) ([online](#)).
- Deutsche Forschungsgemeinschaft: Nationale Forschungsdateninfrastruktur – Ausschreibung 2020 für die Förderung von Konsortien, in: Information für die Wissenschaft Nr. 29 (25.05.2020) ([online](#)).
- Drucker, Johanna: Humanities Approaches to Graphical Display. In: DHQ: Digital Humanities Quarterly 5.1 (2011) ([online](#)).
- Hinrichs, Erhard/Henrich, Andreas/Rapp, Andrea/Stein, Regine: Text+: Sprach- und Textbasierte Forschungsdateninfrastruktur. Text+ Präsentation bei der NFDI-Konferenz 2020. Leibniz-Institut für Deutsche Sprache (IDS) Mannheim 2020 ([online](#)).
- Oltersdorf, Jenny/Schmunk, Stefan: Von Forschungsdaten und wissenschaftlichen Sammlungen. Zur Arbeit des Stakeholdergremiums »Wissenschaftliche Sammlungen« in DARIAH-DE. Bibliothek Forschung und Praxis 40 (2016), S. 179-185 ([online](#)).
- Rat für Informationsinfrastrukturen: Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland. Göttingen 2016 ([online](#)).
- Rißler-Pipka, Nanette/Barthauer, Raisa/Buddenbohm, Stefan/Calvo Tello, José/Friedrichs, Sonja/Weimar, Lukas: Community Involvement in Research Infrastructures: The User Story Call for Text+. 2021 ([online](#)).
- Weimer, Lukas: Bitte mitmachen! DHD-Blog 31. Januar 2022 ([online](#)).

Online-Ressourcen

About Humanities. Was sind Geisteswissenschaften?:

<https://abouthumanities.sagw.ch/02-was-geisteswissenschaften.html>.

AGATE. A European Gateway for the Academies of Sciences and Humanities:

<https://agate.academy/>.

Bund-Ländervereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI) vom 26. November 2018:

[https://www.gwk-](https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Papers/NFDI.pdf)

[bonn.de/fileadmin/Redaktion/Dokumente/Papers/NFDI.pdf](https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Papers/NFDI.pdf).

DANTE (DATendrehscheibe für Normdaten und TERminologien):

<https://dante.gbv.de/search>.

DARIAH-DE (Digital Research Infrastructure for the Arts and Humanities):

<http://dx.doi.org/10.20375/0000-000E-67ED-4>.

Forschungsdaten.info: <https://www.forschungsdaten.info/themen/informieren-und-planen/datenlebenszyklus/>.

Geistes- und kulturwissenschaftliche Forschungsinfrastrukturen e. V.:

<http://forschungsinfrastrukturen.de/>.

GeoNames: <http://www.geonames.org/>.

GIDA (Global Indigenous Data Alliance): <https://www.gida-global.org/care>.

GND (Gemeinsame Normdatei):

https://gnd.network/Webs/gnd/DE/Home/home_node.html.

GO FAIR: <https://www.go-fair.org/fair-principles/>.

NFDI (Nationale Forschungsdateninfrastruktur): <https://www.nfdi.de/>.

RfII (Rat für Informationsinfrastrukturen): <https://rfii.de/de/start/>

TextGrid: <https://textgrid.de/>.

Text+: <https://www.text-plus.org/>

[https://www.text-plus.org/ueber-uns/governance/;](https://www.text-plus.org/ueber-uns/governance/)

[https://www.text-plus.org/forschungsdaten/kooperationsprojekte/;](https://www.text-plus.org/forschungsdaten/kooperationsprojekte/)

[https://www.text-plus.org/forschungsdaten/user-stories/.](https://www.text-plus.org/forschungsdaten/user-stories/)

Anschrift der Autorin:

Prof. Dr. Andrea Rapp

Technische Universität Darmstadt

Institut für Sprach- und Literaturwissenschaft

Residenzschloss, Marktplatz 15

64283 Darmstadt

E-Mail: andrea.rapp@tu-darmstadt.de