

Themenkreis „Fokus Management, Marketing, Innovationen“

Vereinbarkeit von Forschungsprozess und Datenmanagement in den Geisteswissenschaften – Forschungsdatenmanagement nüchtern betrachtet

Marina Lemaire, Universität Trier, Servicezentrum eSciences

Zusammenfassung

In den Geisteswissenschaften stößt der Terminus Forschungsdatenmanagement auf viel Unverständnis und erscheint den Forschenden unvereinbar mit ihrer alltäglichen Arbeit. Der hiesige Beitrag zeigt anhand von Anwendungsbeispielen mit der Virtuellen Forschungsumgebung FuD, dass Datenmanagement dem Forschungsprozess innewohnt. Bei der Überführung der traditionellen Forschungsarbeit ins Digitale handelt es sich um eine Neuinterpretation des Bekannten, wobei jedoch viele Aspekte bereits in der Planungsphase expliziter gemacht werden müssen, was allerdings der Schärfung des Forschungsprozesses zuträglich ist. Zudem verfügen die Forschenden der Geisteswissenschaften für diesen Transformationsprozess aufgrund ihrer Ausbildung bereits über viele der dafür notwendigen Fähigkeiten.

Summary

In the humanities, the term research data management encounters much incomprehension and seems to researchers incompatible with their daily work. Based on practical examples with the virtual research environment FuD, this article shows that data management is inherent in the research process. The transfer of traditional research processes to the digital domain is merely a reinterpretation, although many aspects have to be defined more explicitly already in the planning phase. However, this is conducive to sharpening the research process. Furthermore, the humanities scientists have most of the necessary skills for this transformation process due to their education.

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/2018H4S237-247>

Autorinnenidentifikation: Lemaire, Marina: GND 1137528877,

ORCID: <https://orcid.org/0000-0003-4726-2481>

Schlagwörter: Forschungsdaten; Datenmanagement; Forschungsprozess; Geisteswissenschaften

1. Stand des Forschungsdatenmanagements in den Geisteswissenschaften

In den Geisteswissenschaften hat es den Anschein, dass die Forschenden die Termini Forschungsdatenmanagement (nachfolgend: FDM), Forschungsdaten und Datenmanagement mit der alltäglichen Forschungsarbeit als nicht miteinander vereinbar wahrnehmen. Dabei hat die Digitalisierung schon längst in den Forschungsalltag Einzug gehalten. Doch wie wir in unserer täglichen Beratungspraxis immer wieder feststellen, ist sie in vielen Disziplinen noch nicht so weit vorangeschritten, dass die Potenziale der digital erarbeiteten Forschungsdaten und -ergebnisse ausgeschöpft werden, da

bislang der planvolle digitale Umgang mit der Forschungsgrundlage eher ausbleibt. Obwohl bereits viele Forschungsdaten und -ergebnisse digital vorliegen, sind sie weder auffindbar, noch nachweisbar qualitätsgesichert und damit zur Nachnutzung geeignet. Deshalb fordern Drittmittelgebende und Wissenschaftsorganisationen zunehmend zum FDM auf und appellieren, sich um die Forschungsdaten dem technologischen Wandel angemessen zu kümmern und die Forschungsergebnisse auf ein an die digitalen Bedingungen angepasstes Fundament zu stellen, u.a. um sie im Sinne von Open Science und Open Access möglichst breit zugänglich und nachnutzbar zu machen.¹ Doch (nicht nur) die Forschenden der Geisteswissenschaften sehen sich vor eine scheinbar unlösbare Aufgabe gestellt: Aus ihrer Sicht müssen sie für das FDM Zeit und Kosten im großen Umfang investieren, ohne dass dieses besondere Engagement momentan mit entsprechend steigender wissenschaftlicher Reputation angemessen honoriert werden würde. Ihnen fehlt häufig auch die Vorstellungskraft, dass ihre Forschungsdaten für andere von Interesse sein könnten, bzw. welchen Erkenntnisgewinn die Anwendung von digitalen Methoden auf ihre Forschungsdaten bringen könnten.² Sie scheinen zu befürchten, dass Schwächen bei der Erhebung, Analyse und Interpretation der Primärdaten offensichtlich werden könnten, die mit der Publikation ihrer Forschungsergebnisse öffentlich werden. Sie haben das Gefühl, sich durch die Offenlegung der Daten angreifbarer zu machen. Auch werden die Aufforderungen/Vorgaben zur Bereitstellung von Forschungsdaten oft als Angriff auf die Wissenschaftsfreiheit aufgefasst. Ebenso fällt es ihnen (zu Recht) schwer, die rechtlichen Aspekte, die im Zusammenhang mit ihren Forschungsdaten relevant sind, einzuschätzen.³ Sie meinen zudem, nicht über die notwendigen technischen Kompetenzen zu verfügen. Schlussendlich scheinen sie der Auffassung zu sein, dass ihre zumeist hermeneutisch geprägte, oft rekursive Forschungsarbeit und der Erkenntnisgewinn unvereinbar sind mit einem scheinbar deterministisch starren Datenmanagementprozess.⁴ Das dem nicht so ist, soll an den nachfolgenden Praxisbeispielen, wie wir sie am

- 1 Vgl. u.a. Steuerungsgremium der Schwerpunkttinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen: Den digitalen Wandel in der Wissenschaft gestalten. Die Schwerpunkttinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen. Leitbild 2018 - 2022, Dezember 2017, <<https://doi.org/10.2312/allianzoa.015>>; EC Research & Innovation, European Commission - Directorate-General for Research & Innovation: H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020, 26.07.2016. Online: <http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf>, Stand: 08.06.2018; DFG, Deutsche Forschungsgemeinschaft: Leitlinien zum Umgang mit Forschungsdaten, 30.09.2015. Online: <http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf>, Stand: 06.08.2018.
- 2 Vgl. Tenopir, Carol; Allard, Suzie; Douglass, Kimberly u. a.: Data Sharing by Scientists: Practices and Perceptions, in: PLOS ONE 6 (6), 29.06.2011, S. e21101, <<https://doi.org/10.1371/journal.pone.0021101>>.
- 3 Zu dieser Problematik ist kürzlich vom Projekt DataJus des Bundesministeriums für Bildung und Forschung die Kurzfassung eines Gutachtens zu den rechtlichen Rahmenbedingungen des FDM erschienen. s. Lauber-Rönsberg, Anne; Krahn, Philipp; Baumann, Paul: Gutachten zu den rechtlichen Rahmenbedingungen des Forschungsdatenmanagements (Kurzfassung), 12.07.2018. Online: <https://tu-dresden.de/gsw/jura/igewem/jfbimd13/ressourcen/dateien/publikationen/DataJus_Kurzfassung_Gutachten_12-07-18.pdf?lang=de&set_language=de>, Stand: 06.08.2018.
- 4 Zum vorhergehenden Abschnitt vgl. Kaden, Ben: Warum Forschungsdaten nicht publiziert werden, LIB-REAS.Dokumente, LIBREAS.Projektberichte, 13.03.2018, <<https://libreas.wordpress.com/2018/03/13/forschungsdatenpublikationen/>>, Stand: 06.08.2018; Lemaire, Marina; Rommelfanger, Yvonne; Ludwig, Jan u. a.: Umgang mit Forschungsdaten und deren Archivierung. Bericht zur Online-Bedarfserhebung an der Universität Trier, eSciences Working Papers 02, Universität Trier, Trier 2016, S. 27-30. Online: <<http://nbn-resolving.de/urn:nbn:de:hbz:385-10156>>; Klein, Olivier; Hardwicke, Tom E.; Aust, Frederik u. a.: A practical guide for transparency in psychological science, psyarxiv.com (preprint), 25.03.2018, <<https://doi.org/10.17605/OSF.IO/RTYGM>>; Eynden, Veerle van den; Bishop, Libby: Sowing the seed: Incentives and motivations for sharing research data, a researcher's perspective, Knowledge Exchange Report, UK Data Archive, University of Essex, November 2014. Online: <http://repository.jisc.ac.uk/5662/1/KE_report-incentives-for-sharing-researchdata.pdf>, Stand: 06.08.2018; Tenopir/Allard/Douglass u.a.: Data Sharing by Scientists, 2011, Abb. 12.

Servicezentrum eSciences (SeS | esciences.uni-trier.de) mit der Virtuellen Forschungsumgebung FuD (VFU FuD | fud.uni-trier.de) durchführen, exemplarisch gezeigt werden.

2. Forschungsdatenmanagement mit FuD anhand von Praxisbeispielen

2.1. Was ist FuD?

FuD ist ein Softwaresystem, das den gesamten wissenschaftlichen Arbeitsprozess in seinen verschiedenen Forschungsphasen unterstützt und die orts- und zeitunabhängige Zusammenarbeit ermöglicht. Es ist für die qualitativ forschenden Geistes- und Sozialwissenschaften konzipiert und besteht aus drei Teilkomponenten: der Arbeitsumgebung für die Datenerfassung, -bearbeitung, -analyse und -auswertung; der Publikationsumgebung für die online-Präsentation digitaler Bestände; das Datenrepositorium ViDa⁵ für die Erstellung von Datensammlungen und deren Anreicherung mit Archivierungs- metadaten für eine mindestens zehnjährige Aufbewahrung oder deren langfristige Bereitstellung.

Die *FuD-Arbeitsumgebung* ist individuell konfigurierbar, um zum einen die Bearbeitung von Forschungsdaten unterschiedlichen Dokumenttyps (archivalisches Schriftgut, Briefe, Tagebücher, Personen-, Ortsinformationen, Bilddaten, Interviews, Blogbeiträge u.v.m.) zu ermöglichen; zum anderen bietet sie individuell an die Forschungsmethode adaptierbare Analysewerkzeuge. Dadurch wird die strukturierte und – sofern von den Projekten gewollt – an internationalen Auszeichnungs- und Metadatenstandards orientierte Datenerfassung, -bearbeitung und -analyse unterstützt. Auf diese Weise kann mit ihr die Grundlage für die langfristige Datenverfügbarkeit und -nachnutzung geschaffen werden. Das System ist nicht nur für die Erstellung von digitalen Publikationen geeignet, sondern es kann als Werkzeug für das FDM eingesetzt werden, um über mehrere Forschungsprojekte hinweg die Forschungsdaten einheitlich zu erfassen, zu verwalten, nachzunutzen, zu teilen und ggf. gemeinsam zu bearbeiten. FuD wird an zahlreichen Forschungseinrichtungen in unterschiedlichen Anwendungsbereichen eingesetzt: Es entstehen und entstanden Editionen sowohl im Druck als auch als digitale Ausgabe für Briefe, Werke und andere Quellenmaterialien. Die Aufbereitung und Bereitstellung unerschlossener Archivbestände ist ein weiterer Anwendungsfall. Ebenso werden linguistische Texterschließungen, wie z.B. historisch semantische Untersuchungen, Wortbildungsartenbestimmungen oder Diskursanalysen damit durchgeführt. Des Weiteren können Daten für historisch-geographische Analysen erfasst, aufbereitet und analysiert werden.⁶

Die *Publikationsumgebung* bietet zum einen unterschiedliche Suchwerkzeuge in Form von Volltext-, feldspezifischer und Filtersuche, um sowohl dem Laien einen einfachen Zugang zu den Materialien zu gewähren, als auch den Koryphäen die Möglichkeit zu geben, komplexere Suchanfragen zu formulieren. Zum anderen werden die Dokumente je nach projektspezifischer Anforderung mit ihren Metadaten, Volltexten und ggf. Annotationen in Form von z.B. Personen- und Ortsauszeichnungen, textkritischen Kommentaren etc. dargestellt.

5 www.vida.uni-trier.de (Prototyp; befindet sich im Relaunch).

6 Vgl. Liste der FuD-Anwenderprojekte <www.fud.uni-trier.de/referenzen>.

Die *Archivierungsumgebung* ViDa dient der Aufbereitung und Aufbewahrung ausgewählter Daten für deren Archivierung und ggf. Bereitstellung. Sie verfügt über ein Modul zur Erstellung einer Datensammlung, um die Forschungsdaten⁷ zusammenzustellen und sie mit Archivierungsmetadaten (z.B. Informationen zum Forschungskontext, Datengeber, Datenarten, Zeitraum der Datenerhebung, Zugriffsbedingungen [Lizenzen] etc.) anzureichern.

Bei der (Weiter-)Entwicklung von FuD ist es von Beginn an ein zentrales Anliegen gewesen, die Forschenden soweit wie möglich nicht mit den (daten-)technischen Aspekten des FDMs zu konfrontieren. Daher verfügt die FuD-Arbeitsumgebung über eine Benutzeroberfläche, die die Forschenden davon entlastet, sich mit der XML-Auszeichnung ihrer Daten auseinandersetzen zu müssen. Die TEI-konforme XML-Ausgabe der Forschungsdaten wird über eine Exportschnittstelle gewährleistet, um sicherzustellen, dass die Daten unabhängig von der FuD-Software nachgenutzt werden können.

Zur FuD-Philosophie gehört, neben dem Betrieb und der Entwicklung der Software, dass die Projekte, sofern gewünscht, von der Konzeptionsphase bis zur Archivierung von der fachwissenschaftlichen FuD-Koordination begleitet werden, um sie bei der Entwicklung eines digitalen Forschungskonzepts zu unterstützen, das das FDM und dessen Umsetzung im Forschungsalltag möglichst effizient gestaltet. Dabei werden Schritt für Schritt das projektspezifische Datenmodell und die zugehörigen Workflows gemeinsam entwickelt.⁸ Wie sich dies in der Praxis ausgestalten kann, wird im nachfolgenden Kapitel exemplarisch dargestellt.

2.2. Entwicklung eines digitalen Forschungskonzeptes am Beispiel des Einsatzes der FuD-Software

Um ein digitales Forschungskonzept zu entwickeln, wird der Forschungsprozess zu Grunde gelegt, der sich im Arbeitsprogramm eines Projektes widerspiegeln sollte. Dabei ist das Arbeitsprogramm ausgerichtet auf die Ziele und Fragestellungen, die Untersuchungsmaterialien und die anzuwendenden Methoden. Diese Aspekte, die Gegenstand eines Exposés für ein Forschungsvorhaben oder eines Drittmittelantrages sein können, sind für die fachwissenschaftliche FuD-Koordination von Interesse, um evaluieren zu können, wie FuD im Forschungsprojekt eingesetzt werden kann: Zunächst wird der Textabschnitt *Ziele und Fragestellungen* konsultiert, weil digitale Projekte stärker vom Ende her geplant werden müssen, als es die Forschenden aus der analogen Welt gewohnt sind. Denn es werden durch die Datenerfassung und -bearbeitung bereits entscheidende Wege eingeschlagen, die nur noch mit teilweise immensem Aufwand wieder verlassen oder umgeleitet werden können. Wird also eine Information nicht in der Art und Weise erfasst, wie sie für das anvisierte Ziel gebraucht wird, so ist dieses Ziel nicht ohne Mehraufwand realisierbar. Ein häufig vorkommender Fall ist z.B. die chronologische Sortierung von Daten. Sofern jedes Informationsobjekt mit einem eindeutigen

7 Hier können auch externe Daten (von Speichermedien, andere Datenbanken), die nicht mit der FuD-Arbeitsumgebung erzeugt wurden, in das Datenrepositorium übernommen werden. Das System wurde im Rahmen des SFB 600 „Fremdheit und Armut“ als Prototyp entwickelt und wird derzeit zu einem generischen, Disziplinen übergreifenden Datenrepositorium ausgebaut.

8 Für detailliertere Informationen zum Geschäfts- und Organisationsmodell von FuD vgl. Minn, Gisela; Burch, Thomas; Lemaire, Marina u. a.: FuD2015 – Eine virtuelle Forschungsumgebung für die Geistes- und Sozialwissenschaften auf dem Weg in den Regelbetrieb, eSciences Working Papers 01, Universität Trier, Trier 2016. Online: <<http://nbn-resolving.de/urn:nbn:de:hbz:385-10103>>.

Datum (Tag, Monat Jahr) versehen werden kann, ist dies unproblematisch. Doch gerade bei historischen Daten fällt es schwer, ein so präzises Datum anzugeben. Bislang hilft man sich in der analogen Welt mit Bezeichnungen wie „um 1500“ oder „zwischen 1450 und 1530“. Das sind keine Angaben, die ein Computer korrekt zu sortieren vermag: Sortiert man ein solches Dokument vor oder hinter den 1.1.1500? Daher ist es in einem solchen Fall bspw. sinnvoll, zusätzlich ein maschinenlesbares Datum den Informationsobjekten hinzuzufügen, damit später die Daten korrekt chronologisch sortiert werden können. Dies direkt bei der Erschließung der Dokumente in einem dafür vorgesehenen Sortierungsfeld zu erfassen, ist weniger aufwendig, als im Nachhinein das Datenmodell und die Daten selbst nochmals überarbeiten zu müssen.

Der Abschnitt *Materialien und Methoden* bildet die Grundlage für die Entwicklung des Datenmodells. Die verwendeten Untersuchungsmaterialien müssen mit Metadaten beschrieben werden, um sie voneinander zu unterscheiden und ihre Provenienz zu verzeichnen. Die angewendeten Methoden müssen ebenfalls im Datenmodell Berücksichtigung finden. Dies kann bedeuten, dass das Metadatenchema der Materialien entsprechend erweitert werden muss oder ein eigenes Datenmodell für die Analyse entwickelt wird.

Schließlich zeigt das *Arbeitsprogramm* auf, wie die Arbeitsschritte im Forschungsprozess inhaltlich logisch aufeinander aufbauen und somit Einfluss auf den Einsatz des Softwaresystems (Workflows) oder auch auf die Definition des Datenmodells haben. Hierin konkretisiert sich auch die methodische Vorgehensweise. Die Entwicklung des Datenmodells und der Workflows gehen also Hand in Hand.

2.2.1. Entwicklung eines Datenmodells und der Workflows für die Datenerfassung und -analyse

Um ein Datenmodell z.B. für eine digitale Briefedition zu entwickeln, wird zunächst das Metadatenmodell für die Briefe entworfen. Dabei wird klassisch nach den Informationen gefragt, die notwendig sind, um einen Brief eindeutig zu identifizieren, also zitierbar zu machen und ihn zu beschreiben. Dazu kann sich bspw. an dem auf dem TEI-basierenden Auszeichnungsschema *correspDesc*⁹ orientiert werden. Dazu gehören Informationen wie Absender, Empfänger, Absendedatum, -ort, ggf. Empfängerdatum und -ort, Angaben zum Fundort (Archiv, Signatur), ggf. zu Druckorten etc. Weitere Felder können hinzukommen, wenn bspw. das Incipit, der Überlieferungszustand (z.B. fragmentiert, Wasserschaden), eine Kategorisierung von Briefftypen (z.B. Geburtstagsgrüße, Amtsschreiben, Einladungen) oder unterschiedliche Arten von Beziehungen zwischen Briefen hergestellt werden müssen, z.B. wenn die Korrespondenzkette nur durch die fachwissenschaftliche Expertise ermittelt werden kann. Je nach Art der Untersuchungsmaterialien, deren Überlieferung sowie deren geplanter Weiterverarbeitung/Darstellung wird das Metadatenmodell der Dokumente entwickelt.

Neben der rein inhaltlichen Definition, welche Metadatenfelder gebraucht werden, muss in diesem Schritt auch festgelegt werden, wie die Felder zu befüllen sind. So ist es bspw. wichtig, Personen die Verfassende und/oder Adressat sind, so zu erfassen, dass sie zum einen identifizierbar sind, um z.B. Namensgleichheiten bzw. Schreibvarianten abzufangen oder Tippfehler zu minimieren. Daher ist es

9 Vgl. P5: Richtlinien für die Auszeichnung und den Austausch elektronischer Texte, <correspDesc>, Version 3.4.0, tei-c.org, 23.07.2018, <<http://www.tei-c.org/release/doc/tei-p5-doc/de/html/ref-correspDesc.html>>.

sinnvoll, für jeden Brief neben der Originalschreibweise eine standardisierte Personenbezeichnung festzulegen und in einem Personenregister zu verwalten, um die eindeutige Zuordnung der richtigen Person zu gewährleisten.

Des Weiteren sollte für jedes Feld überlegt werden, ob es sinnvoll sein könnte, mit Listen zu arbeiten, um ein einheitliches Vokabular, z.B. für die Bezeichnung von verschiedenen Dokumenttypen, zu verwenden oder um die Eingabe zu erleichtern, weil dann bei der Bearbeitung nicht jedes Mal das Wort erneut eingegeben werden muss; dadurch können Schreibfehler vermieden bzw. leichter korrigiert werden. Bei der Erarbeitung von Listen muss überlegt werden, ob Hierarchien notwendig und sinnvoll sind. Dies gilt z.B. für die Sachverschlagnwortung mittels einer hierarchisierten Sachsystematik. Wenn hierbei nicht auf eine bereits bestehende Systematik zurückgegriffen wird, so obliegt es den Projektbeteiligten festzulegen, nach welchen Kriterien ein solcher Index aufgebaut wird, wie viele Ebenen er umfassen soll, wie neue Begriffe hinzugefügt werden und schlussendlich auch, nach welchen Kriterien sie vergeben werden. Dies sind rein vom Forschungsgegenstand her inhaltlich zu begründende Festlegungen des Projektes. D.h. ein Datenmodell wird immer entsprechend der fachwissenschaftlichen Festlegungen modelliert.

Wenn inhaltliche Analysen an Volltexten oder Bildern vorgesehen sind, müssen für deren Aufnahme ebenfalls entsprechend Felder im Datenmodell vorgesehen und die Verfahren festgelegt werden, wie die Volltexte/Bilddaten in die Datenbank gelangen, z.B. durch Importe. Am Ende dieser Planungen für die Forschungsphase der Datenerhebung steht das Metadatenmodell für die Untersuchungsmaterialien.

Im nächsten Schritt werden die Methoden der Datenanalyse in das Datenmodell überführt. Dafür ist es notwendig, die einzelnen methodischen Analyseschritte präzise in ihren logischen Zusammenhängen zu beschreiben, sodass sie in ein Datenmodell überführt werden können. D.h. nicht, dass alle inhaltlichen Analysekategorien bereits bekannt sein müssen, sondern lediglich die methodische Vorgehensweise und die Verfahrensweise bei der Bildung der Analysekategorien feststehen muss: Zur Veranschaulichung ziehe ich hier das Projekt „Wortbildung in Mündlichkeit und Schriftlichkeit. Beziehungen zwischen Wort- und Textbildung im geschriebenen und gesprochenen Deutsch“ heran.¹⁰ Es vergleicht schriftliche und mündliche Kommunikation im Hinblick auf ihre Wortbildungstypik und -spezifik. Für den Vergleich musste zunächst in den Dokumentmetadaten zusätzlich festgehalten werden, ob es sich um gesprochenen oder geschriebenen Text handelt. Das Datenmodell für die Analyse der einzelnen Wortbildungen¹¹ in den Texten wurde gemeinsam mit dem Wissenschaftler entwickelt, der seine methodische Vorgehensweise anhand von Beispielen detailliert erläuterte: Es wurde dann festgelegt, dass bei seiner Vorgehensweise der Ausgangspunkt der Analyse einer Wortbildung die Bestimmung der Wortbildungsart ist. Daher wurde ein Analyseindex mit den Wortbildungsarten angelegt. Das bedeutet in der Praxis, dass er eine Wortbildung im Volltext markiert

10 Habilitationsprojekt von Dr. Sören Stumpf, vgl. Stumpf, Sören: Textsortenorientierte Wortbildungsforschung. Desiderate, Perspektiven und Beispielanalysen, in: Zeitschrift für Wortbildung 2 (1), 2018, S. 165–194.

11 Als Wortbildungen bezeichnet man solche Wörter einer Sprache, die auf der Grundlage und mithilfe vorhandenen Sprachmaterials gebildet sind, wie *Haus*tür, *zurück*geben, aber auch *nerv*ig.

und ihr die entsprechende Analysekatgorie (Wortbildungsart, z.B. Suffixderivat) zuweist.¹² Da alle anderen Analysekatgorien (z.B. Wortart, Lexikalisierung, Grundform) abhängig von der jeweiligen Analyse einer Wortbildung sind, wurde das Metadatenschema des Analyseobjektes¹³ um diese Analysekatgorien in Form von weiteren Feldern erweitert.¹⁴ Auf diese Weise können die verschiedenen Ausprägungen von einzelnen Wortbildungsarten (z.B. hinsichtlich ihrer Wortart) strukturiert erfasst und später quantitativ ausgewertet werden.

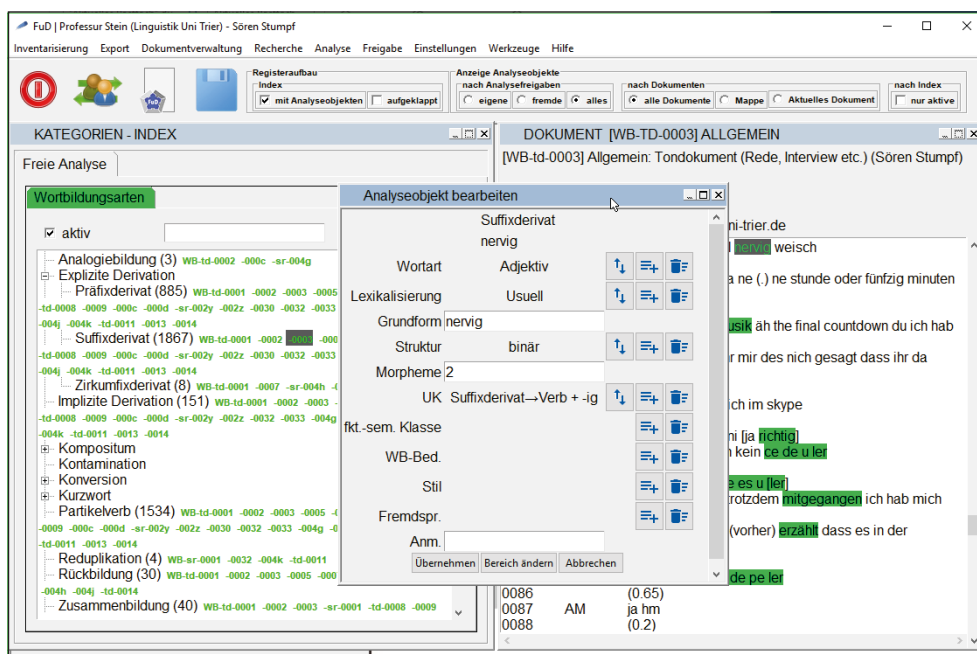


Abb. 1: Bildschirmabzug der Analyseumgebung des Habilitationsprojekts von Dr. Sören Stumpf, Germanistische Linguistik, Universität Trier.

Je nach Forschungsmethode kann es notwendig sein, auch für die Analysekatgorien zusätzliche Informationen, wie z.B. den GND-Link oder synonyme Bezeichnungen, im Datenmodell zu berücksichtigen. Auch hierfür können zusätzliche Felder in der FuD-Arbeitsumgebung konfiguriert werden.

Die Entwicklung des Datenmodells und der zugehörigen Workflows ist ein iterativer Prozess, bei dem die Forschenden relativ kleinteilig ihren Forschungsprozess und ihre Methoden beschreiben müssen und das FuD-Team diese in das Datenmodell übersetzt, das die Grundlage für die Systemkonfiguration

12 Siehe Abb. 1 im linken Fenster Analyseindex „Wortbildungsarten“.

13 Analyseobjekte sind in der FuD-Terminologie annotierte Textpassagen bzw. Bildausschnitte. Im abgebildeten Beispiel ist das der im rechten Fenster dunkelgrau markierte Textabschnitt „nervig“. Siehe Abb. 1.

14 Siehe Popup-Fenster „Analyseobjekt bearbeiten“ (mittig) in Abb. 1.

bildet. Hierfür ist ein sehr intensiver Austausch über das methodische Vorgehen und die Arbeitsweisen erforderlich, um die Arbeitsabläufe zu operationalisieren. Während dieser detaillierten Analyse werden manchmal die Grenzen der datentechnischen Umsetzung der gewählten Forschungsmethode deutlich bzw. wird offensichtlich, wo entweder rein fachwissenschaftlich das methodische Vorgehen spezifiziert werden muss oder zunächst erste Befunde aus der Forschungsarbeit abgewartet werden müssen, um die abschließende Ausgestaltung des Datenmodells vorzunehmen. Dieser Operationalisierungsschritt schärft den Blick der Forschenden dafür, worauf bei der Erhebung und Auswertung der Untersuchungsmaterialien bzgl. der methodischen Vorgehensweise insbesondere zu achten ist und wo noch Unschärfen im Forschungskonzept konkretisiert werden müssen. FDM ist also nicht nur ein dem Forschungsprozess inhärenter Prozess, sondern kann sogar ein Instrument zur Überprüfung des Forschungskonzeptes sein.

3. Forschungsdatenmanagement nüchtern betrachtet: Die Neuinterpretation eines traditionellen Rezeptes

Das vorhergehende Kapitel zeigte, wie die Übertragung des Forschungsprozesses in ein digitales Konzept die Auseinandersetzung mit dem Forschungsgegenstand und der Methode unterstützt und ggf. den Blick für noch zu füllende Leerstellen schärft, die sich möglicherweise erst durch die tiefere Einarbeitung in die Untersuchungsmaterialien ergeben. Diese Vorgehensweise behindert auch nicht den, meist iterativen gar rekursiven, geisteswissenschaftlichen Forschungsprozess, weil es an sich um die Festlegung/Operationalisierung der methodischen Vorgehensweise geht und nicht um die Inhalte oder gar Ergebnisse. Die Forschenden können sich weiterhin bei der Analyse von ihren Materialien überraschen lassen und weitere Analysekatoren entsprechend ihrer vorab entworfenen Methodik ergänzen. Der planvolle Einsatz digitaler Werkzeuge wird (jedenfalls derzeit noch nicht) die intelligente und findige Interpretation der Analyseergebnisse ersetzen; sie aber nachvollziehbarer und transparenter machen.

Es wurde dargelegt, dass für den qualitätsgesicherten und nachhaltigen Umgang mit Forschungsdaten die „klassische“ wissenschaftliche Vorgehensweise mit den heutigen digitalen Arbeitsweisen in Einklang gebracht werden muss, indem ein digitales Forschungskonzept erstellt wird. Das Datenmanagement ist somit ein den Forschungsprozess unterstützendes Instrument in Zeiten der Digitalisierung und wohnt ihm quasi inne. Dafür müssen jedoch zusätzliche Arbeitsschritte für die Generierung qualitätsgesicherter Datenbestände im Arbeitsprogramm berücksichtigt und gleichzeitig die notwendigen finanziellen Mittel für Software sowie andere IT-Dienste im ausreichenden Maße eingeplant werden. Daher ist es unabdingbar, mit dem Entwurf des *digitalen* Forschungskonzeptes bereits in der Planungsphase zu beginnen.

Die Datenmanagementplanung beginnt also mit der Erstellung des Forschungskonzeptes, das neben der Entwicklung der Fragestellung, der Auswahl der Untersuchungsmaterialien sowie der benötigten Methoden auch die Erstellung eines Arbeitsprogramms vorsieht. Darin legen die Forschenden fest, welche Arbeitsschritte notwendig sind, um die Forschungsfrage zu beantworten respektive die Ziele zu erreichen. Das Arbeitsprogramm orientiert sich weiterhin klassisch am Forschungsprozess:

Für jede Phase werden bislang häufig die Techniken und Werkzeuge mal intuitiv und unbewusst vorausgesetzt (z.B. Recherchen in Datenbanken, Texteditorssoftware für die Erstellung der Publikation), mal bewusst (z.B. spezifische Messapparate; Audioaufnahmegeräte) festgelegt. Das FDM führt nun dazu, dass die Forschenden vieles, was sie bislang intuitiv und unbewusst vollzogen haben, nun explizit darstellen müssen, wenn sie den FDM-Anforderungen gerecht werden wollen. Daher ist für jeden geplanten Arbeitsschritt notwendig, die Frage zu stellen, ob digitale Daten entstehen respektive digitale Werkzeuge eingesetzt werden und wenn ja, welche Auswirkung dies auf den entsprechenden Arbeitsschritt im Forschungsprozess selbst, aber auch auf nachfolgende hat. Daher verläuft der Datenmanagementprozess nicht nur parallel zum Forschungsprozess, sondern hat ggf. auch Einfluss auf ihn, weil bspw. die Verwendung eines bestimmten Softwaretools es erfordert, einen Arbeitsschritt zu einem anderen Zeitpunkt durchzuführen, oder es hat gar Auswirkungen auf die Forschungsmethode selbst.

Das *Neue* an der Digitalisierung des Forschungsprozesses ist also, dass die Forschenden ihren Forschungsprozess zu einem früheren Zeitpunkt detailgenauer planen, am besten in einem Datenmanagementplan dokumentieren und Arbeitsschritte bzgl. der methodischen Vorgehensweise expliziter beschreiben müssten, die momentan eher implizit erfolgen. Dies ist notwendig, um Arbeitsabläufe zu konzipieren, die von Computern verstanden werden können.

Für die Operationalisierung des Forschungsprozesses sind nicht grundlegend neue Fähigkeiten gefragt: Denn die Entwicklung eines digitalen Forschungskonzeptes ist keine völlig neue Aufgabe im Forschungsprozess, sondern eine ihm immanente. Schließlich müssen Forschende stets die Kontextbedingungen ihrer Forschungsarbeit kritisch reflektieren: So wie Physiker/innen bei einem Experiment auf die Umgebungstemperatur achten müssen, Statistiker/innen den Algorithmus für die Bereitstellung von Zufallszahlen kennen müssen, so müssen bspw. Historiker/innen sich die Überlieferungs- und Entstehungsgeschichte ihrer Quellen vergegenwärtigen, um die Vollständigkeit, Aussagekraft und Glaubhaftigkeit beurteilen zu können. Für diese Einschätzung werden relativ intuitiv Kriterien angewendet, die nicht nur Historiker/innen, sondern alle qualitativ forschenden Wissenschaftler/innen im Laufe ihrer Ausbildung im Umgang mit ihren Untersuchungsmaterialien erlernen, während ihrer Forschungstätigkeit quasi automatisieren und nicht mehr so viel darüber nachdenken. Es ist also ein anderes, im Augenblick noch nicht verinnerlichtes, Nachdenken über den reflektierten Einsatz digitaler Werkzeuge in der geisteswissenschaftlichen Forschung notwendig. Dies wird durch wiederholte Anwendung und durch die Erweiterung der propädeutischen Curricula in den nächsten Jahren zur Selbstverständlichkeit werden, genauso wie es für Geschichtswissenschaftler/innen nach Leopold von Ranke selbstverständlich geworden ist, die äußere und innere Quellenkritik anzuwenden. Die Erstellung eines *digitalen* Forschungskonzeptes ist also nicht mehr als die Neuinterpretation des *traditionellen*.

Für die Übertragung des analogen Forschungsprozesses ins Digitale können die Forschenden zudem auf ihre disziplinspezifischen Fähigkeiten zurückgreifen: Denn im Grunde genommen geht es um eine Art Übersetzung, für die die Geisteswissenschaften über Methoden und analytisches Repertoire verfügen, die ihnen die notwendigen Kompetenzen und Prozeduren für diesen Transfer zur Verfügung stellen. So bringen sie für die Entwicklung von Metadatenschemata z.B. die Methode der inneren und

äußeren Quellenkritik mit. Sie erfordert nach einem durch das Untersuchungsmaterial vorgegebenen Kriterienkatalog die Quelle in ihren Entstehungskontext einzuordnen und ihre Aussagekraft zu bewerten. Die dafür heranzuziehenden Kriterien repräsentieren bereits einen Teil der Metadaten. Des Weiteren begünstigt die (vielleicht auch nur deutsche) Detailverliebtheit für die vielfältigen Literaturzitierweisen die Metadatenmodellierung. Durch sie sind die Forschenden geübt darin, einzelne Informationseinheiten (z.B. Autorenschaft, Titel, Provenienz) inhaltlich voneinander zu trennen, um sie dann in einer festgelegten Systematik für das Zitat wieder zusammenzufügen. Dies ist ebenfalls eine relevante Fähigkeit, um logisch aufeinander aufbauende Workflows für die Datengenerierung und -verarbeitung zu entwickeln. Zudem können Geisteswissenschaftler/innen für die Workflowentwicklung auf ihre Fähigkeiten zur multiperspektivischen Rekonstruktion von komplexen Systemen zurückgreifen. Denn dafür haben sie Methoden und Kompetenzen entwickelt, wie sie trotz fehlender Überlieferungen Leerstellen logisch argumentativ schließen, indem sie Regelmäßigkeiten ermitteln bzw. heranziehen, sodass sich dennoch ein stimmiges Gesamtbild des Untersuchungsgegenstandes ergibt. Zudem verfolgen sie bei ihren Untersuchungen eher einen ganzheitlichen Ansatz, der möglichst alle relevanten Faktoren in den Blick nimmt und ihren Einfluss auf das Gesamtgefüge einschätzt. All diese Fähigkeiten sind im Digitalisierungskontext sehr nützlich und es gilt sie jetzt auf den eigenen Forschungsprozess anzuwenden, um ihn der digitalen Welt anzupassen.

Literaturverzeichnis

- Steuerungsgremium der Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen: Den digitalen Wandel in der Wissenschaft gestalten. Die Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen. Leitbild 2018 - 2022, Dezember 2017. Online: <<https://doi.org/10.2312/allianz0a.015>>.
- DFG, Deutsche Forschungsgemeinschaft: Leitlinien zum Umgang mit Forschungsdaten, 30.09.2015. Online: <http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf>, Stand: 06.08.2018.
- EC Research & Innovation, European Commission - Directorate-General for Research & Innovation: H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020, 26.07.2016. Online: <http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf>, Stand: 08.06.2018.
- Eynden, Veerle van den; Bishop, Libby: Sowing the seed: Incentives and motivations for sharing research data, a researcher's perspective, Knowledge Exchange Report, UK Data Archive, University of Essex, November 2014. Online: <http://repository.jisc.ac.uk/5662/1/KE_report-incentives-for-sharing-researchdata.pdf>, Stand: 06.08.2018.
- Kaden, Ben: Warum Forschungsdaten nicht publiziert werden, LIBREAS.Dokumente, LIBREAS.Projektberichte, 13.03.2018, <<https://libreas.wordpress.com/2018/03/13/forschungsdatenpublikationen/>>, Stand: 06.08.2018.

- Klein, Olivier; Hardwicke, Tom E.; Aust, Frederik u. a.: A practical guide for transparency in psychological science, psyarxiv.com (preprint), 25.03.2018. Online: <<https://doi.org/10.17605/OSF.IO/RTYGM>>.
- Lauber-Rönsberg, Anne; Krahn, Philipp; Baumann, Paul: Gutachten zu den rechtlichen Rahmenbedingungen des Forschungsdatenmanagements (Kurzfassung), 12.07.2018. Online: <https://tu-dresden.de/gsw/jura/igewem/jfbimd13/ressourcen/dateien/publikationen/DataJus_Kurzfassung_Gutachten_12-07-18.pdf?lang=de&set_language=de>, Stand: 06.08.2018.
- Lemaire, Marina; Rommelfanger, Yvonne; Ludwig, Jan u. a.: Umgang mit Forschungsdaten und deren Archivierung. Bericht zur Online-Bedarfserhebung an der Universität Trier, eSciences Working Papers 02, Universität Trier, Trier 2016. Online: <<http://nbn-resolving.de/urn:nbn:de:hbz:385-10156>>.
- Minn, Gisela; Burch, Thomas; Lemaire, Marina u. a.: FuD2015 – Eine virtuelle Forschungs-umgebung für die Geistes- und Sozialwissenschaften auf dem Weg in den Regelbetrieb, eSciences Working Papers 01, Universität Trier, Trier 2016. Online: <<http://nbn-resolving.de/urn:nbn:de:hbz:385-10103>>.
- Stumpf, Sören: Textsortenorientierte Wortbildungsforschung. Desiderate, Perspektiven und Beispielanalysen, in: Zeitschrift für Wortbildung 2 (1), 2018, S. 165–194.
- Tenopir, Carol; Allard, Suzie; Douglass, Kimberly u. a.: Data Sharing by Scientists: Practices and Perceptions, in: PLOS ONE 6 (6), 29.06.2011, S. e21101. Online: <<https://doi.org/10.1371/journal.pone.0021101>>.