

Forschungsinfrastrukturen in den digitalen Geisteswissenschaften.

Wie verändern digitale Infrastrukturen die Praxis
der Geisteswissenschaften?

Martin Huber, Sybille Krämer, Claus Pias
Symposienreihe „Digitalität in den Geisteswissenschaften“

Gefördert durch

DFG Deutsche
Forschungsgemeinschaft

IMPRESSUM

HERAUSGEBER

Martin Huber, Sybille Krämer, Claus Pias

KONTAKT

Julia Menzel

Digitalität in den Geisteswissenschaften

DFG-geförderte Symposienreihe

Universität Bayreuth

Universitätsstr. 30

95447 Bayreuth

www.digitalitaet.dfg@uni-bayreuth.de

1. Auflage November 2019

Wir danken der Deutschen Forschungsgemeinschaft (DFG Projekt Projektnummer 287972711) für die Förderung.

Eine gemeinsame geisteswissenschaftliche Forschungsdatenkultur: Utopie oder Dystopie?

Heike Neuroth, Ulrike Wuttke (Potsdam)

1. Wissenschafts- und förderpolitischer Rahmen

1.1 Einleitung

Durch die digitale Transformation der Forschungsprozesse ist ein starker Anstieg der Produktion, Aggregation und Nachnutzung von digitalen Forschungsdaten in allen Wissenschaftsdisziplinen zu verzeichnen.¹ Die digitale Transformation und die damit zusammenhängende Datenflut (*data deluge*) betrifft nicht nur die Wissenschaft, sondern auch die Wirtschaft, die Politik, die Verwaltung und die breitere Gesellschaft. Im Rahmen der förderpolitischen Debatte ist die sachgemäße Aufbewahrung und Bereitstellung von Forschungsdaten in Deutschland zum Beispiel in die *Vorschläge zur Sicherung guter wissenschaftlicher Praxis*² der DFG aufgenommen bzw. wird in den *Leitlinien zum Umgang mit Forschungsdaten*³ der DFG erläutert. Einzelne Fachdisziplinen haben nach dem Vorbild der DFG-Richtlinien eigene fachspezifische Empfehlungen veröffentlicht, darunter z. B. die Biodiversität, Soziologie, Sprachkorpora etc.⁴ Neben dem Argument, dass die Ergebnisse öffentlich geförderter Forschung auch der breiten Öffentlichkeit zur Verfügung stehen sollten, werden diesbezüglich u. a. die Sicherung der wissenschaftlichen Qualität und der Nachvollziehbarkeit der Forschung anhand der unmittelbaren Datengrundlage als Argumente angeführt. Gerade in Zeiten von *fake journals* und dem Zurückziehen von Artikeln spielt die Prüfung der Datengrundlage eine zunehmende Rolle. So hat z. B. die Fachzeitschrift *JAMA* 2018 dreizehn Artikel des weltweit anerkannten Ernährungspsychologen Brian Wansink (Cornell University) zurückgezogen, dessen Studien 20.000mal zitiert wurden, weil die Ergebnisse nicht validierbar sind, u. a. auch, weil kein Zugriff auf die Originaldaten möglich war.⁵

2017 hat sich die Initiative [Collections as Data](#) in den USA gegründet, die von dem [Institute of Museum and Library Services](#) initiiert wurde. Hier wurden 10 Prinzipien formuliert, die quasi parallel zu anderen Initiativen (s.u.) inhaltlich und strategisch in die gleiche Richtung gehen. So lautet zum Beispiel ein Prinzip: „Collections as data development aims to encourage computational use of digitized and born digital collections“ oder: „Collections as Data should be made openly accessible by default, except in cases where ethical or legal obligations preclude it.“ Allein diese beiden Prinzipien zielen darauf ab, dass die Sammlungen aus den sogenannten

¹ vgl. Borgman, Christine L.: *Big Data, Little Data, No Data. Scholarship in the Networked World*. Cambridge (Mass.), London: MIT Press 2015, S. 3-16.

² DFG - Deutsche Forschungsgemeinschaft: *Vorschläge zur Sicherung guter wissenschaftlicher Praxis / Proposals for Safeguarding Good Scientific Practice*, Denkschrift / Memorandum, Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“ / Recommendations of the Commission on Professional Self Regulation in Science. Bonn: Wiley-VCH 2013, S. 21-22.

³ DFG - Deutsche Forschungsgemeinschaft: *Leitlinien zum Umgang mit Forschungsdaten*. 2015. https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf.

⁴ vgl. DFG - Deutsche Forschungsgemeinschaft: *Umgang mit Forschungsdaten*. https://www.dfg.de/foerderung/antrag_gutachter_gremien/antragstellende/nachnutzung_forschungsdaten/.

⁵ vgl. z. B. Lingenhöhl, Daniel: *Der tiefe Fall eines Ernährungspapstes*. Spektrum.de News 06.03.2019. <https://www.spektrum.de/news/der-tiefe-fall-eines-ernaehrungspapstes/1593004>.

Gedächtnisinstitutionen wie Bibliotheken, Museen und Archive als Daten-Grundlage für digitale (automatische) Verfahren und Methoden dienen und ohne weitere Mühen für die Nachnutzung zur Verfügung stehen sollten.

Im Zusammenhang mit der allgemeinen Forderung nach der Bereitstellung von Forschungsdaten als [Open Research Data](#) (ORD) im Sinne der Open Science-Bewegung, steht in jüngster Zeit verstärkt ihre Bereitstellung gemäß den sogenannten [FAIR-Prinzipien](#)⁶ im Vordergrund und rückt somit die Eignung der Daten für die Nachnutzung in den Mittelpunkt des Interesses. FAIR steht für „Findable, Accessible, Interoperable, Reuseable“ und gilt für die intellektuelle Nachnutzung durch Menschen, aber besonders für die maschinelle Nachnutzung (*machine-actionability*), d.h. „the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention“⁷. Die Europäische Kommission hat im November im Rahmen der [EOSC](#) (European Open Science Cloud) u. v. a. a. den Bericht *Turning FAIR into reality*⁸ veröffentlicht, der die breite Spanne an notwendigen Änderungen beschreibt, um die sogenannte FAIRness von Daten sicherzustellen.

Durch diese Entwicklungen werden Forschungsdatenmanagement und aktive Forschungsdatenmanagementpläne (DMPs)⁹ immer wichtiger und auch die Geisteswissenschaften stehen nun vor der Herausforderung, ein grundlegendes Verständnis dafür zu entwickeln:

- was geisteswissenschaftliche Forschungsdaten sind,
- wie ihr Umgang systematisch geregelt werden kann,
- welche Kernkompetenzen in der Lehre bzw. in der Weiterbildung vermittelt werden sollten und
- welche systemischen Veränderungen (Stichworte Förder- und Evaluationsprozesse) notwendig sind, um die zunehmende Digitalisierung der Geisteswissenschaften (Stichwort Digital Humanities) zu unterstützen.

Hierbei gilt es zunächst, den Forschungsdatendiskurs in den Geisteswissenschaften überdisziplinär zu öffnen, um einen allgemeinen definitorischen Konsens zu erreichen, der auch Erfahrungen aus bereits etablierten Forschungsdatenkulturen einbezieht. Erst dann kann eine Verfeinerung von disziplinspezifischen Fragestellungen vorgenommen werden. Im Folgenden werden Handlungsbedarfe und mögliche Lösungsszenarien bei der Etablierung einer gemeinsamen Forschungsdatenkultur in den Geisteswissenschaften vor dem Hintergrund aktueller Entwicklungen und grundsätzlicher Prämissen skizziert.

1.2 Aktuelle Entwicklungen

Neben den bereits oben genannten europäischen Veröffentlichungen sind in letzter Zeit eine Reihe wichtiger Veröffentlichungen auf nationaler Ebene erschienen, die einen großen Einfluss auf die Praktiken der Wissenschaftler*innen und die förderpolitischen Rahmenbedingungen haben werden. So hat der Deutsche Bundestag im September 2018 *Die Hightech-Strategie 2025 – Forschung und Innovation für die Menschen* veröffentlicht, in der es sehr konkret heißt:

⁶ vgl. Wilkinson, Mark D. et al.: The FAIR Guiding Principles for Scientific Data Management and Stewardship. in: Scientific Data 3 2016.

⁷ GO FAIR: FAIR Principles. <https://www.go-fair.org/fair-principles/>.

⁸ European Commission Expert Group on FAIR Data: Turning FAIR into Reality. Final Report and Action Plan from the European Commission Expert Group on FAIR Data. European Commission: Brüssel 2018.

⁹ vgl. z. B. Neuroth, Heike / Engelhardt, Claudia / Klar, Jochen / Enke, Harry: Aktives Forschungsdatenmanagement - Research Data Management Organiser. In: ABI Technik 38,1 (2018), S. 55-64.

- Die umfassende Nutzung von Forschungsdaten ermöglicht die Beantwortung neuer, auch multidisziplinärer wissenschaftlicher Fragen.
- Für die Nutzung von Daten müssen international anschlussfähige Lösungen gefunden werden. Um unsere Wettbewerbsfähigkeit in einer datengetriebenen Wissenschaft und Wirtschaft zu sichern, ergänzen wir unsere Förderung durch übergreifende Infrastruktur-Ansätze wie den Aufbau einer nationalen Forschungsdateninfrastruktur sowie einer European Open Science Cloud.¹⁰

Konsequenterweise folgte im November 2018 die *Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI)*, die in ihrer Präambel auf die Notwendigkeit von potentiellen NFDI-Konsortien als „wissenschaftlich breit nutzbare Datensätze mit gesellschaftlichem Mehrwert“¹¹ hinweist. In dem zugrunde liegenden Förderprogramm wird darauf hingewiesen, dass „ein stimmiges Konzept zu Datennutzung und -zugang sowie Auffindbarkeit und Nachnutzbarkeit der Daten, welches entlang der FAIR-Prinzipien ausgerichtet ist“¹², vorliegen muss. Dafür stehen in den nächsten zehn Jahren bis zu 90 Millionen Euro pro Jahr im Endausbau für insgesamt ca. 30 sogenannte fachspezifische NFDI-Konsortien zur Verfügung. Im Rahmen der RDA-DE Tagung im Februar 2019 am GFZ Potsdam¹³ haben sich bereits einige potenzielle fachspezifische NFDI-Konsortien vorgestellt¹⁴, die in den Geisteswissenschaften verankert sind (u. a. NFDI4Text+, NFDI4Memory und NFDI4Culture). Anfang 2019 hat sich zudem das [Geisteswissenschaftliche Forum NDFI](#) für den Informationsaustausch gegründet, an dem sich bereits über 20 geisteswissenschaftliche Fachverbände bzw. Fachgesellschaften beteiligen.

2. Geisteswissenschaftlicher Kontext

Die bisherige Forschungsdatendiskussion in den Geisteswissenschaften ist von zwei Tendenzen geprägt: der besonderen Spezifik geisteswissenschaftlicher Forschungsdaten¹⁵ sowie der Notwendigkeit der Abgrenzung einzelner Disziplinen. Zum einen haben Geisteswissenschaftler*innen ein eher ambivalentes Verhältnis zum Datenbegriff.¹⁶ Zum anderen sind auf Grund der disziplinären Vielfalt innerhalb der Geisteswissenschaften und weil potenziell eine breite Spanne von Materialien als geisteswissenschaftlicher Untersuchungsgegenstand in Frage

¹⁰ BMBF - Bundesministerium für Bildung und Forschung: Forschung und Innovation für die Menschen: Die Hightech-Strategie 2025. Berlin 2018, S. 55.

¹¹ GWK - Gemeinsame Wissenschaftskonferenz: Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI) vom 26. November 2018 (BAnz AT 21.12.2018 B10), S. 1.

¹² GWK - Gemeinsame Wissenschaftskonferenz: Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI) vom 26. November 2018 (BAnz AT 21.12.2018 B10), Paragraph 5, Abschnitt g.

¹³ vgl. RDA DE: RDA Deutschland Tagung 2019. <https://www.rda-deutschland.de/events/tagung-2019>.

¹⁴ Einige haben dazu die Möglichkeit einer Kurzvorstellung im Plenum genutzt, andere das Posterformat gewählt.

¹⁵ vgl. u. a. Sahle, Patrick / Kronenwett, Simone: Jenseits der Daten. Überlegungen zu Datenzentren für die Geisteswissenschaften am Beispiel des Kölner ‚Data Center for the Humanities‘. In: LIBREAS. Library Ideas, 23 (2013), S. 76–96, Buddenbohm, Stefan / Engelhardt, Claudia / Wuttke, Ulrike: Angebotsgenese für ein geisteswissenschaftliches Forschungsdatenzentrum. In: Zeitschrift für digitale Geisteswissenschaften 2016,1. Open Research Data Task Force: Realising the Potential. Final Report of the Open Research Data Task Force“. Open Research Data Task Force, 2018, S. 6.

¹⁶ vgl. u. a. Cremer, Fabian / Klaffki, Lisa / Steyer, Timo: Der Chimäre auf der Spur: Forschungsdaten in den Geisteswissenschaften. In: O-Bib 5,2 (2018), S. 142-162.

kommt, geisteswissenschaftliche Forschungsdaten und die für ihre Interpretation benötigte Expertise sehr heterogen. Die Natur geisteswissenschaftlicher Forschungsdaten kann berechtigterweise nicht nur gestellt werden als „Was sind geisteswissenschaftliche Daten?“ sondern auch als „Wann sind es geisteswissenschaftliche Daten?“:

The humanities are in a position similar to the social sciences in determining what or when something might be used as data. [...] Several characteristics of the humanities distinguish them from the sciences and social sciences with respect to data scholarship. One is their focus on interpretation and reinterpretation. [...] A second characteristic is the vast variety of data sources and the expertise to interpret them. [...] As a result of these two qualities, data sources can be reused repeatedly, often becoming more valuable as they cumulate over time. Conversely, because the same objects can be represented and interpreted in so many ways, it is difficult to apply classification mechanisms that transcend the array of possible uses of collections.¹⁷

Geisteswissenschaftliche Forschungsdaten können sowohl analoger als auch digitaler Natur sein und in Form von Texten, Bildern, Audio-, Multimedia- oder 3D-Daten vorliegen. Diese Daten können darüber hinaus in verschiedenen (Meta-)Datenformaten in unterschiedlichen Qualitätsstufen, z. B. im Fall von mittels OCR erstellten Volltexten oder Transkriptionen im Double-Keying-Verfahren, vorliegen. In vielen Fällen liegen die (Meta-)Daten mehrsprachig und zum Teil in nicht-lateinischen Schriften vor. Oftmals sind die Daten sehr stark untereinander vernetzt, da sie mehrere Kommentar- bzw. Interpretationsschichten reflektieren. Gerade der letzte Punkt, die Subjektivität geisteswissenschaftlicher Forschung und der daraus resultierenden Forschungsdaten, stellt eine weitere Herausforderung dar.¹⁸

Im Rahmen von [DARIAH-DE](#) (Digitale Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften) wurde sich auf die folgende Definition für Forschungsdaten geeinigt:

Unter digitalen geistes- und kulturwissenschaftlichen Forschungsdaten werden innerhalb von DARIAH-DE all jene Quellen/Materialien und Ergebnisse verstanden, die im Kontext einer geistes- und kulturwissenschaftlichen Forschungsfrage gesammelt, erzeugt, beschrieben und/oder ausgewertet werden und in maschinenlesbarer Form zum Zwecke der Archivierung, Zitierbarkeit und zur weiteren Verarbeitung aufbewahrt werden können.¹⁹

Die oben angerissenen epistemologischen Fragestellungen zur Natur geisteswissenschaftlicher Daten können in diesem Rahmen nicht im Detail ausgeführt werden. Es soll jedoch festgehalten werden, dass geisteswissenschaftliche Forschungsdaten, im Gegensatz zu vielen naturwissenschaftlichen Forschungsdaten, in der Regel besonders langlebig sind. Ein historisches Korpus z. B. in Form einer Datenbank oder einer digitalen Edition, Vorhaben, an denen oft lange Jahre gearbeitet wurden, ist für lange Zeit eine autoritative Quelle. Dazu kommt, dass es trotz gewisser Gemeinsamkeiten, die im Folgenden eruiert werden, trügerisch ist, von DEN Geisteswissenschaften oder DEN Digital Humanities zu sprechen, da sich hinter diesen

¹⁷ Borgman 2015, S. 166.

¹⁸ vgl. u. a. Raspe, Martin / Schelbert, Georg: Genau, wahrscheinlich, eher nicht. Beziehungsprobleme in einem kunsthistorischen Wissensgraph. In: Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. Hg. von Andreas Kuczera / Thorsten Wübbena / Thomas Kollatz. 2019 (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 4).

¹⁹ DARIAH-DE: Weiterführende Informationen. <https://de.dariah.eu/weiterfuehrende-informationen>.

Begriffen sehr unterschiedliche Daten, Methoden und Verfahren, Geschwindigkeiten der Digitalisierung sowie Anforderungen und Bedarfe verbergen.

Auch außerhalb der Geisteswissenschaften gibt es bisher keine standardisierte Definition von Forschungsdaten in Deutschland. Alle bisherigen Beschreibungen wurden bewusst offen formuliert.²⁰ In einigen naturwissenschaftlichen Fächern (wie z. B. der Astrophysik²¹ oder der Klimaforschung²²) wurde jedoch bereits ein Konsens darüber erreicht, was Forschungsdaten sind und wie sie beschrieben und aufbewahrt werden. Ein solcher Lösungsansatz (*Good Practice*) aus einer etablierten Forschungsdatenkultur kommt aus der Astrophysik. Da in dieser Fachdisziplin die benötigten Großgeräte sehr kostspielig sind, beteiligen sich in der Regel zahlreiche Länder an ihrer Entwicklung sowie den Investitions- und Betriebskosten. Dies führt dazu, dass eine Kultur der internationalen Zusammenarbeit entwickelt wurde bzw. werden musste, d. h. die Wissenschaftler*innen teilen sich die Beobachtungszeit und in der Regel auch die Forschungsdaten. Hieraus entwickelte sich eine gemeinsame Forschungsdatenkultur, die zu einer Reihe weiterer Standardisierungen bezüglich Werkzeuge, Dienste, Datenformate etc. geführt hat. Als ein Beispiel sei hier der [Sloan Digital Sky Survey](#) (SDSS) angeführt, der im Jahr 2000 den Startschuss für die „modernste Kartierung des Himmels“²³ eröffnete. Mit der mittlerweile 15. Datenpublikation (*data release*) bildeten diese Daten bisher die Grundlage für mehr als 8.000 wissenschaftliche Publikationen und es entstand ein Katalog von mehr als 200 Millionen Galaxien sowie Millionen von Sternen und Quasaren. In einem Hinweis zur Datenzitierung heißt es: „If you have used public SDSS data in your paper, please cite the following papers describing the instruments, survey, and data analysis as appropriate.“²⁴

In den Geisteswissenschaften fehlt dagegen weitestgehend ein Konsens darüber, was Forschungsdaten sind und vor allem, wie sie beschrieben und aufbewahrt werden sollten, sowohl in den einzelnen Fachdisziplinen als auch darüber hinaus. So produziert das [Deutsche Textarchiv](#) (DTA) sehr viele Forschungsdaten(-sammlungen), z. B. hochauflösende Digitalisate, Transkriptionen als Volltexte, Annotationen, TEI-Auszeichnungen, zahlreiche Exportformate und unterschiedliche Metadaten sowie Dokumentationen. Was davon sind Forschungsdaten? Was muss wie aufbewahrt werden? Welche, auch zukünftigen, Nutzungsszenarien sind vorstellbar? Ist eine Unterscheidung zwischen aktiver und passiver Nutzung von Forschungsdaten sinnvoll, weil manche Forschungsdatensammlungen eher den Anspruch haben, nur langzeitarchiviert zu werden, um der guten wissenschaftlichen Praxis zu dienen?²⁵ Welche Metadatenstandards sind verbindlich und international interoperabel? Welcher Status kommt Forschungssoftware zu? Wem gehören die Forschungsdaten in den Geisteswissenschaften?

²⁰ So z. B. die Definition der Arbeitsgruppe Forschungsdaten der Allianz der deutschen Wissenschaftsorganisationen (vgl. Schwerpunktinitiative „Digitale Informationen“: Forschungsdaten. <https://www.allianzinitiative.de/archiv/forschungsdaten/>).

²¹ vgl. z. B. NASA Science: Astrophysics Data Centers. <https://science.nasa.gov/astrophysics/astrophysics-data-centers>.

²² vgl. z. B. NOAA - National Oceanic and Atmospheric Administration - Earth System Research Laboratory (ESRL) - Physical Sciences Division (PSD): Data and Imagery. <https://www.esrl.noaa.gov/psd/data/>.

²³ Max-Planck-Institut für Astronomie: Startschuss für modernste Kartierung des Himmels. Webseite.

²⁴ SDSS - Sloan Digital Sky Survey. How to Cite SDSS. <https://www.sdss.org/collaboration/citing-sdss/>.

²⁵ vgl. Pronk, Tessa E.: The Time Efficiency Gain in Sharing and Reuse of Research Data, In: Data Science Journal 18,1 (2019), S. 10: „Moreover, results suggest overall efficiency can be increased if not too much effort is put into sharing datasets with low probability of reuse.“ Vgl. auch DFG – Deutsche Forschungsgemeinschaft: Leitlinien zum Umgang mit Forschungsdaten. 2015, S. 1 „Den Regeln der Guten Wissenschaftlichen Praxis folgend sollen Forschungsdaten in der eigenen Einrichtung oder in einer fachlich einschlägigen, überregionalen Infrastruktur für mindestens 10 Jahre archiviert werden.“

Hinzu kommt, dass auch organisatorische und strukturelle Fragen in den Geisteswissenschaften noch unzureichend geklärt sind. Während zahlreiche naturwissenschaftliche Disziplinen, großteils in engen internationalen Kooperationen, eigene Forschungsdatenzentren aufgebaut haben, stehen ähnliche Bestrebungen in den Geisteswissenschaften noch am Anfang. Die im Vergleich sehr junge Klimaforschung hat es zum Beispiel sehr früh verstanden, sowohl in Deutschland als auch auf internationaler Ebene, eng zu kooperieren und gemeinsame Standards und Werkzeuge zu entwickeln. Obwohl hierfür alleine in Deutschland die Zusammenarbeit zwischen mehr als 100 universitären und außeruniversitären Forschungseinrichtungen organisiert werden musste, steht mit verschiedenen Förderinstrumenten und zahlreichen Fachgremien eine tragfähige Struktur zur Verfügung, die das breite und interdisziplinäre Feld der Klimaforschung (z. B. Meteorologie, Ozeanographie, Politologie, Ökonomie, Soziologie etc.) organisieren hilft. Mit dem [Deutschen Klimarechenzentrum](#) (DKRZ) existiert in Deutschland auch ein für die Archivierung und Zitierung der Forschungsdaten verantwortlicher Partner.

In Analogie hierzu ist in den Geisteswissenschaften vieles noch unklar bzw. nicht geregelt. Welche Rolle kommt Gedächtnisinstitutionen wie Bibliotheken, Archiven und Museen zu, die eine große Menge potenzieller geisteswissenschaftlicher Forschungsdaten produzieren, in dem sie ihre Sammlungen digitalisieren und zur Verfügung stellen? Wie sollten einheitliche Regelungen für die Nachnutzungsbedingungen für diese digitalen Sammlungen aussehen, so dass Wissenschaftler*innen rechtssicher darauf zugreifen und diese nachnutzen können? Allein der momentan geringe und wenig systematische Stand der Digitalisierung von geisteswissenschaftlichen Quellen bzw. Sammlungen in den Gedächtnisinstitutionen stellt ein Problem für datengetriebene Forschungsfragen dar. Mangels genauen Berechnungen sei hier auf die Schätzung von Michael Knoche verwiesen, laut der „erst ein niedriger zweistelliger Prozentsatz der nationalen Buchproduktion in elektronischer Form zur Verfügung“ steht, weil Ressourcen zur Digitalisierung fehlen und das restriktive deutsche Urheberrecht eine erhebliche Barriere darstellt.²⁶ Hinzu kommt die Heterogenität der Beschreibungsformate (Metadaten), der Dateiformate, Auszeichnungssprachen etc., die maschinelle Auswertungsverfahren über mehrere digitale Sammlungen (Korpora) hinweg mit einem vertretbaren Zeitaufwand nahezu unmöglich macht.

Ohne einen disziplinübergreifenden Forschungsdatendiskurs in den Geisteswissenschaften unter Einbeziehung der Gedächtnisinstitutionen wird es kaum möglich sein, auf nationaler Ebene ein gemeinsames Verständnis einer geisteswissenschaftlichen Forschungsdatenkultur zu etablieren. Diese ist jedoch die Voraussetzung:

- für nationale und internationale interdisziplinäre Kooperationen und damit einhergehende Standardisierungen,
- um vor allem zukünftigen Anforderungen der Förderer (bzw. auch Politik und Gesellschaft) im Kontext der FAIR-Prinzipien zu begegnen,
- um eine gemeinsame gute datenwissenschaftliche Praxis zu etablieren,
- um gemeinsam den Herausforderungen der digitalen Transformation zu begegnen mit abgestimmten Notwendigkeiten in Entwicklungen, Förderinstrumenten etc.,

²⁶ vgl. Knoche, Michael: Die Idee der Bibliothek und ihre Zukunft. Göttingen: Wallstein 2018, S. 67-73, Zitat S. 69. Auch Klaffki, Lisa / Schmunk, Stefan / Stäcker, Thomas: Stand der Kulturgutdigitalisierung in Deutschland: Eine Analyse und Handlungsvorschläge des DARIAH-DE Stakeholdergremiums „Wissenschaftliche Sammlungen“, Göttingen 2018 (DARIAH-DE Working Papers, 26), S. 6-17 verzeichnet einen ähnlich spärlichen Umfang der Kulturgutdigitalisierung in Deutschland.

- um koordiniert (Mindest)Standards, Policies etc. zu entwickeln, die jeweils fachspezifisch ausdifferenziert werden zu können und
- für die Entwicklung der notwendigen Lehr-, Weiterbildungs-, und Schulungskonzepte.

Konsequenterweise heißt es auf der Webseite der Gemeinsamen Wissenschaftskonferenz (GWK) zur NFDI:

Derzeit oft dezentral, projektförmig und temporär gelagerte Datenbestände von Wissenschaft und Forschung sollen im Rahmen der NFDI für das gesamte deutsche Wissenschaftssystem systematisch erschlossen werden. Die NFDI soll Standards im Datenmanagement setzen und als digitaler, regional verteilter und vernetzter Wissensspeicher Forschungsdaten nachhaltig sichern und nutzbar machen [...]. Auf diese Weise wird eine unverzichtbare Voraussetzung dafür geschaffen, neue wissenschaftliche Erkenntnisse zu gewinnen und Innovationen in Forschung und Gesellschaft zu ermöglichen.²⁷

3. Potentielle Roadmap Forschungsdateninfrastruktur digitale Geisteswissenschaften

3.1 Gemeinsame Forschungsdatenkultur

Ausgangsbasis für die nachfolgenden Überlegungen stellen die in der *Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI)* vom 26. November 2018 formulierte Ziele dar.²⁸ Diese betonen sehr stark Aspekte der Gemeinsamkeit, Koordination, Anbindung an nationale sowie internationale Entwicklungen, Zusammenarbeit, Standardisierung sowie fachübergreifende Prozesse und Strukturen. Hieraus ergibt sich die Notwendigkeit für die Geisteswissenschaften, sich gemeinsam auf den (langen) Weg zu begeben und übergeordnete Diskussionen und Lösungswege anzustoßen. Zu kleinteilige Fokussierungen auf die jeweiligen Fachdisziplinen mit ihren spezifischen Forschungsfragen werden in der Zukunft nicht mehr zielführend sein.

Wie könnte eine gemeinsame geisteswissenschaftliche Forschungsdatenkultur aussehen? Ein erster Schritt der Annäherung an diese Frage ist eine grobe Unterscheidung der Entstehungskontexte geisteswissenschaftlicher Forschungsdaten, diese entstehen meist:

- im Kontext von konkreten Forschungsfragen, in den meisten Fällen im Rahmen (kleinerer) Drittmittel-basierter Projekte sowie
- beim Aufbau von digitalen (größeren) Sammlungen, sozusagen als Grundlagenforschung, analog zu den Großgeräten in den Naturwissenschaften und als Basis für die Erforschung breiter, zukünftiger Forschungsfragen.²⁹

Allerdings scheinen die Einzelvorhaben oder kleinen Projektverbünde in den Geisteswissenschaften zu dominieren. So sind über die letzten Jahrzehnte hinweg zahlreiche digitale Daten-Silos entstanden, zwischen denen kaum ein Austausch, sei es auf der eigentlichen Datenebene

²⁷ GWK - Gemeinsame Wissenschaftskonferenz: Informationsinfrastrukturen / NFDI. <https://www.gwk-bonn.de/themen/weitere-arbeitsgebiete/informationsinfrastrukturen-nfdi/>.

²⁸ vgl. GWK 2018, Paragraph 1.

²⁹ vgl. Schöch, Christof: Aufbau von Datensammlungen. In: Jannidis, Fotis / Hubertus Kohle / Malte Rehbein (Hg). Digital Humanities. Eine Einführung. Stuttgart: J.B. Metzler, 2017, S. 223-233. Schöch unterscheidet drei Typen von Datensammlungen (Projektbezogene Datensammlungen, Gemeinschafts-Datensammlungen, Referenz-Datensammlungen).

oder auch nur auf der Ebene der Metadaten, möglich ist. Dazu kommen Nachhaltigkeitsfragen: Wie kann sichergestellt werden, dass diese Daten auch in Zukunft gefunden und zugänglich gemacht werden können, wenn sie in den meisten Fällen auf Projekthomepages liegen (dies betrifft schon den Bereich „Findable“ und „Accessible“ der FAIR-Prinzipien)? Das wichtigste Ziel der direkten Nachnutzbarkeit („Reusable“) wird bisher noch so gut wie gar nicht in den Projektvorhaben adressiert. Hier fehlen Strukturen und Prozesse für die Überführung in vertrauenswürdige Datenzentren.³⁰ Diesen Entscheidungen müssen natürlich Absprachen über die Aufbereitung der Daten, geeignete Standards, Schnittstellen und Formate vorausgehen, an die sich alle Datenproduzenten halten müssen.

Auch der Bereich des breiter angelegten Aufbaus größerer digitaler Sammlungen ist in Deutschland wenig koordiniert. So gibt es zahlreiche gelungene Einzelbeispiele wie zum Beispiel das DTA als „Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache“³¹ oder die Digitale Bibliothek in [TextGrid](#) als „umfangreiche Sammlung XML/TEI-erschlossener Texte aus Belletristik und Sachliteratur vom Anfang des Buchdrucks bis zu den ersten Jahrzehnten des 20. Jahrhunderts“³². Allerdings gibt es in Deutschland keinen nationalen Ansatz, alle digitalen Sammlungen basierend auf gemeinsamen standardisierten Vorgaben zu digitalisieren, zu erschließen und für die Nachnutzung nachhaltig als digitalen Wissensspeicher zur Verfügung zu stellen. Als Vorbild oder Inspiration kann [HathiTrust](#) in den USA dienen, das seit 2008 aufgebaut wird. Als Ziel wird Folgendes formuliert: „provides long-term preservation and access services to digitized content from a variety of sources, including Google, the Internet Archive, Microsoft, and in-house member institution initiatives“³³. Mit Bezug zu den Zielen der FAIR-Prinzipien, die Maschinenlesbarkeit der (Meta-)daten zu erhöhen, betont das HathiTrust Research Center (HTRC) das Ziel „computational analysis of the HathiTrust corpus“ zu ermöglichen³⁴ und stellt dafür eine technische Infrastruktur mit einer Reihe von geeigneten digitalen Werkzeugen und Diensten zur Verfügung.

Es ist eine schöne Vorstellung, wenn die digitalen Bestände z. B. der Bayerischen Staatsbibliothek mit dem [Deutschen Textarchiv](#), der Digitalen Bibliothek in [TextGrid](#) sowie den im [Zentralen Verzeichnis Digitalisierter Drucke](#) (ZVDD) nachgewiesenen Digitalisaten in einem deutschen Wissensspeicher zugänglich und nachnutzbar wären, versehen mit klaren rechtlichen Nutzungsbedingungen, die sich nach den Parametern von Open Science richten. Besonders wünschenswert wäre es, wenn in diesem Kontext gleich Angaben über verschiedene Stufen der Wieder- und Weiterverwendung inklusive Qualitätslevel der digitalen Daten mitgeliefert werden würden³⁵, z. B.:

- **Digitalisat:** Mindestanforderungen für verschiedene Nutzungsszenarien. In einigen Fällen braucht es hochauflösende Digitalisate, in anderen Anwendungsfällen mögen auch niedrigere Auflösungen ausreichend sein.

³⁰ vgl. DHd AG Datenzentren: Geisteswissenschaftliche Datenzentren im deutschsprachigen Raum - Grundsatzpapier zur Sicherung der langfristigen Verfügbarkeit von Forschungsdaten (Version 1.0). DHd AG Datenzentren, 2018.

³¹ DTA - Deutsches Textarchiv: DTA - Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache. <http://www.deustextarchiv.de/doku/referenzkorpus> .

³² TextGrid: Die Digitale Bibliothek bei TextGrid. <https://textgrid.de/digitale-bibliothek>.

³³ HathiTrust: Our Digital Library. https://www.hathitrust.org/digital_library.

³⁴ HathiTrust: Our Research Center. <https://www.hathitrust.org/htrc> .

³⁵ vgl. Klaffki / Schmunk / Stäcker 2018. Dort werden verschiedene Digitalisierungsklassen danach unterschieden, inwieweit sie die Nutzung durch digitale Forschungsvorhaben unterstützen.

- **Volltext:** Mindestanforderungen für verschiedene Nutzungsszenarien. Die Erstellung von Volltexten kann in vielen Fällen (automatisiert) mittels OCR-Verfahren erfolgen, in einigen Anwendungsgebieten braucht es aufwändigere Double-Keying-Methoden. Auch hier ist das Ausgangsmaterial entscheidend.
- **Metadaten:** Analog zu dem [Europeana Data Model](#) oder dem HathiTrust Metadata Management System [Zephir](#) braucht es ein Mindest-Kernset an Metadatenelementen, auf die sich die Datenproduzenten gemeinsam mit allen geisteswissenschaftlichen Fachdisziplinen einigen. Dieses Kernset sollte dabei weitgehend bereits etablierte Metadatenschema wie [DataCite](#), [Dublin Core](#) etc. nachnutzen, um die Interoperabilität und damit die Nachnutzbarkeit für interdisziplinäre Fragestellungen maximal zu unterstützen.
- **Kontrollierte Vokabularsysteme:** Für die intellektuelle und automatisierte, maschinenlesbare Nachnutzung braucht es die breite Verwendung von Normdaten, Thesauri, Ontologien etc. über Disziplinengrenzen hinweg. Dafür können vorhandene Vokabularsysteme genutzt und/oder erweitert werden, ein Mix verschiedener Systeme oder in einem koordinierten „Kraftakt“ ein gemeinsames kontrolliertes Vokabularsystem entwickelt werden, welches auf intelligenten Mappingverfahren zwischen bereits vorhandenen Vokabularsystemen basieren sollte.
- **Anreicherungen, Verweise, Kommentare, Versionierungen** etc.: Hierzu zählen Absprachen bezüglich generischer Mindestanforderungen bis hin zu fachwissenschaftlich ausgeprägten Mindeststandards für ein verzweigtes System der tieferen Anreicherung bzw. fachwissenschaftlichen Kommentierung.

Um das Beispiel der Klimaforschung (s. o.) oder der Biodiversität³⁶ noch einmal zu bemühen, die ebenfalls aus zahlreichen, quer über die großen Wissenschaftsgebiete wie Natur-, Geistes- und Sozialwissenschaften liegenden Teildisziplinen bestehen, besitzen auch die Geisteswissenschaften mit ihren zahlreichen Teildisziplinen und zum Teil sehr unterschiedlichen Praktiken und Kulturen des wissenschaftlichen Arbeitens das große Potential, geeignete, koordinierte Strukturen und Prozesse für eine gemeinsame Forschungsdateninfrastruktur aufzubauen. Diese würde als digitaler Wissensspeicher, als Bewahrinstantz des kulturellen Erbes, gegenüber der fragmentierten und fragilen Landschaft der derzeitigen digitalen Silos einen deutlichen Mehrwert für die Wissenschaft und die Gesellschaft bieten.

Dazu braucht es weniger technisch herausfordernde (Weiter-)Entwicklungen noch riesige Mengen an Speicherkapazität oder Rechnerleistung als vielmehr ein Paradigmenwechsel in der wissenschaftlichen Zusammenarbeit: Eine fachübergreifende Verständigung über Standards quer über alle Aspekte wie Semantik, Syntax, Technologien, Schnittstellen etc. hinweg. So könnte ein umfangreicher, gemeinsamer digitaler Wissensspeicher aufgebaut werden, der für die Bearbeitung von verschiedenen Fragestellungen aus unterschiedlichen (Teil)Disziplinen geeignet ist: Eine neue Kultur des wissenschaftlichen Austauschs, des neugierig Miteinander-Forschen-Wollens entsteht. Es darf bezweifelt werden, ob die Teilchenphysik oder Astrophysik ohne die Notwendigkeit, sich um Großgeräte „gruppieren“ zu müssen, heute miteinander weltweit so vernetzt wären und sich auf Standards, Good Practices etc. hätten einigen können. Allein durch die Notwendigkeit, den immensen Kosten dieser Großgeräte einen maximalen Nutzen für eine sehr große Anzahl an beteiligten Wissenschaftler*innen gegenüberzustellen, ist es gelungen, den Gedanken des Teilens von Forschungsdaten innerhalb weniger Jahrzehnte

³⁶ vgl. German Federation for Biological Data (gfbio). <https://www.gfbio.org/>.

weltweit so klar zu verbreitern. Dadurch sind auch erst weitere Vorteile der Zugänglichmachung von Forschungsdaten stärker sichtbar geworden wie z. B. Synergieeffekte durch die Vermeidung von Doppelfinanzierungen, Validierung von Forschungsergebnissen, Formulierung ganz neuer Forschungsfragen etc.

Dementsprechend formuliert gleich das zweite Kriterium, welches über die Auswahl der NFDI-Konsortien entscheidet, die Notwendigkeit, einen erwartbaren „Mehrwert für die Entwicklung disziplinübergreifender Metadatenstandards und die Schaffung eines verlässlichen und nachhaltigen Dienst-Angebots im Konsortium“³⁷ zu schaffen. Es steht also die Frage im Vordergrund, was jede(r) Einzelne dazu beitragen kann, eine geisteswissenschaftliche Forschungsdateninfrastruktur aufzubauen.

3.2 Disziplinübergreifende Handlungsfelder

Die Handlungsfelder für den Aufbau und Betrieb einer geisteswissenschaftlichen Forschungsdateninfrastruktur ergeben sich, neben der fachlich ausgewiesenen Expertise und der nationalen und internationalen Verankerung in den Fach-Communities, insbesondere in der Entwicklung eines stimmigen Konzepts „zu Datennutzung und -zugang sowie Auffindbarkeit und Nachnutzbarkeit der Daten, welches entlang der FAIR-Prinzipien ausgerichtet ist“³⁸. Einen guten Überblick über die FAIR-Prinzipien bietet die GO FAIR Initiative³⁹. International wird bereits über erste Ansätze zu FAIR-Metriken und -Evaluierungen diskutiert⁴⁰, mit deren Hilfe der Grad an „FAIRification“ bewertet werden kann. Zweifellos werden die FAIR-Prinzipien eine immer wichtigere Rolle spielen. Es sei an dieser Stelle betont, dass „FAIR“ nicht mit „Open“ gleichzusetzen ist.

Es lohnt sich, sich die FAIR-Prinzipien im Einzelnen anzuschauen und zu überlegen, was diese für den Diskurs rund um den NFDI-Prozess in den Geisteswissenschaften bedeuten könnten. Die folgende Beschreibung und Liste ist der GO FAIR-Webseite entnommen und führt jeweils zuerst in jedes der vier Haupt-Prinzipien ein, bevor einzelne Unterkriterien genannt werden. „Reusable“ stellt das höchste zu erreichende Ziel dar, da erst hier automatisiert „a la Google“ nach für die Fragestellung geeigneten Daten gesucht werden kann, eine relevante Trefferliste zu erwarten ist und diese Daten sofort unter Beachtung lizenzrechtlicher Bedingungen nachgenutzt werden können:

Findable: The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process.

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource

³⁷ GWK 2018, S. 2, Paragraph 5, Absatz b.

³⁸ GWK 2018, S. 2, Paragraph 5, Absatz g.

³⁹ vgl. GO FAIR: FAIR Principles. <https://www.go-fair.org/fair-principles/>.

⁴⁰ vgl. zum Beispiel die verschiedenen Arbeits- und Interessensgruppen der internationalen RDA-Initiative unter <https://www.rd-alliance.org/groups> wie die FAIR Data Maturity Model WG oder BoF - Measuring FAIRness of Digital Objects - RDA 13th Plenary meeting 2019.

Accessible: Once the user finds the required data, she/he needs to know how can they be accessed, possibly including authentication and authorisation.

- A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
 - A1.1 The protocol is open, free, and universally implementable
 - A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available

Interoperable: The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data

Reusable: The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

- R1. Meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (Meta)data are released with a clear and accessible data usage license
 - R1.2. (Meta)data are associated with detailed provenance
 - R1.3. (Meta)data meet domain-relevant community standards⁴¹

Mit den FAIR-Prinzipien liegen die Handlungsfelder auf der Hand und muss das Forschungsdatenmanagement für die Geisteswissenschaften neu formuliert und etabliert werden. Dabei kommt Metadaten und weiteren standardisierten Beschreibungsformaten (wieder) eine große Bedeutung zu, die sich von „Accessible“ bis zu „Reuse“ durchzieht. Ersichtlich wird auch, dass zur maximalen Unterstützung der FAIR-Prinzipien Absprachen und koordinierte *Good Practices* für eine einheitliche Beschreibung von Forschungsdaten ausschlaggebend sein werden, nicht technische Entwicklungen.

Für die geisteswissenschaftlichen Teildisziplinen und für die Geisteswissenschaften insgesamt müssen daher im Bereich des Forschungsdatenmanagements neue Kulturen und Praktiken gefunden und flächendeckend etabliert werden, die nachfolgende Fragestellungen berücksichtigen:

Im Bereich des **Datenmanagements** und der **Datenmanagementpläne** (DMPs):

- Welche Tools und Dienste unterstützen das Datenmanagement und die dynamische bzw. aktive Erstellung von DMPs (z. B. [RDMO](#))?
- Wie sehen die Anforderungen der jeweiligen Forschungsförderer, Fachgesellschaften, Zeitschriften, Fachdisziplinen etc. aus?
- Was sind gemeinsame Mindestanforderungen zur einheitlichen Beschreibung von Daten, Methoden, Verfahren, wissenschaftlicher Software etc.?
- Welche Mindestqualitätskriterien lassen sich für die einzelnen FAIR-Kriterien definieren? Wie könnten Entwicklungspfade Richtung „Reuse“ aussehen?

⁴¹ GO FAIR: FAIR Principles. <https://www.go-fair.org/fair-principles/>.

- Welche Fachdisziplin hat in welchem Bereich bereits Vorbildcharakter?

Im Bereich der **Archivierung** und **Publikation**:

- Wie sieht eine gute datenwissenschaftliche Praxis aus? Welche allgemeinen und fachspezifischen Leitfäden lassen sich entwickeln?
- Was muss bzw. kann wie und ab wann zugänglich sein (z. B. *dark archive*, rein technische Langzeitarchivierung, Datenpublikation, Nachweis über Datensammlungen etc.)?
- Welche (zertifizierten, vertrauenswürdigen) Datenzentren und -repositorien gibt es bereits, welche Angebote lassen sich ausbauen und wo bestehen welche (weiteren) Bedarfe?

Im Bereich **Kompetenzen** und **Unterstützung**:

- Wie können (Nachwuchs)Wissenschaftler*innen im Bereich gute datenwissenschaftliche Praxis unterstützt und geschult werden?
- Wer kann sich wie um das Forschungsdatenmanagement kümmern? Welche neuen Karrierepfade und Berufsfelder entstehen in diesem Kontext? Wie kann dies durch geeignete Job-Beschreibungen und Aus- sowie Weiterbildungsangebote unterstützt werden?

Übergreifende Fragestellungen, die zumindest eine einheitliche Vorgehensweise verlangen:

- Sollen pro Disziplin Strategien, Policies, Leitlinien etc. entwickelt werden bzw. kann es z. B. als Referenzrahmen eine übergreifende geisteswissenschaftliche Policy geben, die auch Aspekte des digitalen kulturellen Erbes berücksichtigt?
- Sollten Strategien, Policies etc. eher vom Datentyp her entwickelt, also eine Ausdifferenzierung in Text, Bild, Ton und Video vorgenommen werden?
- Was ist der State-of-the-Art hinsichtlich Werkzeuge, digitale Dienste, Datenzentren, Zitierpraktiken, Datenpublikationen, Umgang mit wissenschaftlicher Software etc.?
- Wie könnte ein Referenzcurriculum für das Forschungsdatenmanagement und für den Umgang mit digitalen Daten im Allgemein aussehen? Welche Bausteine müssten grundlegend in die geisteswissenschaftliche Lehre (Bachelor und Master) integriert werden, um dem Anspruch „Investition in Köpfe“⁴² gerecht zu werden? Welche Grundkenntnisse in z. B. Statistik, Algorithmen, Technologien, Programmierung etc. sind für zukünftige Geisteswissenschaftler*innen unerlässlich?
- Wie können zukünftige Anreizsysteme, Anerkennungsverfahren und Karrierepfade entwickelt werden, um einer neuen Generation von Geisteswissenschaftler*innen das Arbeiten mit digitalen Daten inklusive der Publikation, der Entwicklung wissenschaftlicher Software, der Unterstützung des Forschungsdatenmanagements etc. als eigenständige wissenschaftliche Leistung zu vermitteln? Was kann auch hier von anderen Fachdisziplinen gelernt werden, die zum Teil Datenpublikationen in renommierten Fachzeitschriften mit über 100 Autor*innen veröffentlichen oder bei denen sogenannte *data manager* oder *data stewards* bereits etablierte Berufszweige sind?

⁴² RfII - Rat für Informationsinfrastrukturen: Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland. Göttingen 2016, S. 36.

4. Fazit & Ausblick

Insgesamt stehen die Geisteswissenschaften vor großen Herausforderungen. Im Rahmen des europäischen Prozesses um die [European Open Science Cloud](#) (EOSC) und des von Bund und Ländern gemeinsam initiierten NFDI-Programms bietet sich jedoch eine große Chance, sowohl die Geisteswissenschaften im digitalen Zeitalter neu zu positionieren als auch auf die Bedeutung digitaler Daten(sammlungen) auf eine andere Art und Weise aufmerksam zu machen.

Durch die massive Förderung seitens der Europäischen Kommission und deutscher Förderorganisationen (z. B. DFG, BMBF, Stifterverband) wird die Bedeutung von digitalen Forschungsdaten und ihrem sachgemäßen (FAIRen) Umgang schnell und rasant wachsen. Auch der kommerzielle Dienstleister Google hat sich bereits der Thematik angenommen und eine Beta-Version seiner [Dataset Search](#) veröffentlicht, die es Nutzern ermöglichen soll, digitale Datensätze bzw. Datensammlungen mittels einer einfachen Stichwortsuche zu finden. Inwieweit lassen sich die Ziele von Google mit wissenschaftlichen Anforderungen wie den FAIR-Prinzipien vereinbaren? Müssen auch solche Initiativen in Bestrebungen für eine bessere Findbarkeit von Daten einbezogen werden?⁴³

Während viele Fachdisziplinen außerhalb der Geisteswissenschaften in den letzten Jahrzehnten umfangreiche, wertvolle Datensammlungen aufgebaut haben (z. B. in den Sozialwissenschaften das [Sozio-ökonomisches Panel](#)) bzw. gerade aufbauen (z. B. in der Teilchenphysik durch das FASER-Experiment⁴⁴ am CERN), stehen ähnliche Bestrebungen für den Aufbau umfangreicher geisteswissenschaftlicher Datensammlungen in Deutschland noch am Anfang. Hierdurch bietet sich die Gelegenheit des Aufbaus eines großen, digitalen Wissensspeichers, der einheitlich durchsucht und bei dem ein eigenes Forschungskorpus heruntergeladen und nachgenutzt werden kann. Dieser „Referenzkorpus“ könnte dazu beitragen, die Bedeutung der Geisteswissenschaften wieder fest in der Gesellschaft, bei den Bürger*innen, zu verankern.

Zur Realisierung des Korpus sind sicherlich mehrere iterative Schritte nötig. Einerseits sollte es von Anfang an ein niedrigschwelliges Angebot für alle Nutzer*innen einschließlich interessierten Laien sein, in den Datensammlungen zu stöbern, um z. B. literaturwissenschaftliche (oder völlig neue) Fragestellungen digital zu erkunden, andererseits könnten gezielt disziplinär interessierte Fachwissenschaftler*innen zu Kooperation und gemeinsamen, schrittweisen Entwicklung neuer Forschungsfragen und Nutzungsszenarien motiviert werden.

In den letzten Jahrzehnten ist es einigen Fachdisziplinen gelungen, zum Teil spektakuläre Fördersummen für ihre Forschungsanliegen und die Entwicklung und Bereitstellung der dafür notwendigen Forschungsinfrastrukturen einzuwerben. So gelang es der Klimaforschung als eine der jüngsten Fachdisziplinen, die auch immer wieder oder gerade wieder sehr heftig in die Kritik gerät und deren Forschungsergebnisse und -daten abgestritten oder ignoriert werden, sich durch eine gemeinsame Koordinierung und internationale Anschlussfähigkeit nicht nur innerhalb kürzester Zeit auf vorbildliche Art und Weise quer über verschiedene Fachdisziplinen⁴⁵ auf internationaler Ebene zu vernetzen, sondern auch durch die Interessengemeinschaft [Deutsches Klima Konsortium](#) (DKK) mit einer Stimme gegenüber Politik, Gesellschaft, Förderer

⁴³ vgl. Blumtritt, Jonathan / Rau, Felix: Metadaten im Zeitalter von Google Dataset Search. Zenodo 2019.

⁴⁴ vgl. z. B. Gast, Robert: CERN bewilligt neuen LHC-Detektor. Spektrum.de News 06.03.2019. <https://www.spektrum.de/news/cern-bewilligt-neuen-lhc-detektor/1627998>.

⁴⁵ Zu nennen sind hier neben Naturwissenschaften wie Meteorologie, Ozeanographie, Chemie und Physik auch Sozialwissenschaften wie Politologie, Ökonomie, Soziologie etc. sowie Technik- und Ingenieurwissenschaften.

etc. aufzutreten. Auch die rapide Entwicklung geeigneter (fachwissenschaftlicher) Standards, Leitfäden, Policies und der vorbildliche Umgang mit den eigenen Forschungsdaten zeigt in diesem Fall, wie eine Fachdisziplin eine Vorreiterrolle einnehmen kann. Die Geisteswissenschaften bilden diesbezüglich eher das Schlusslicht, nicht nur was die Einwerbung von Fördermitteln für digitale geisteswissenschaftliche Infrastrukturen betrifft. Die Investitionshöhe ist im Vergleich zu anderen Wissenschaften vergleichsweise gering. Dies liegt sicherlich auch daran, dass sie sich nicht so durchsetzen können, weil sie viel eher geneigt und gewohnt sind, die Unterschiede zwischen ihren Fächerkulturen zu betonen als mit „einer gemeinsamen Stimme zu sprechen“.

Ohne abzustreiten, dass die Herausforderungen in den geisteswissenschaftlichen Disziplinen aufgrund disziplinärer Unterschiede vielfältig sind und eine lange fragmentierte Praxis in den (Teil)Disziplinen die Konsensbildung verzögert bzw. behindert, wäre der erste wichtige Schritt eine detaillierte Kartierung der geisteswissenschaftlichen Fächer, um den Status Quo in den Bereichen geeigneter Vorbilder und (gemeinsamer) Bedarfe zu identifizieren, für die dann in vernetzter und abgestimmter Weise weitere gemeinsame Lösungen gefunden werden. Dabei sollte auch die Frage nach der Rolle der Gedächtniseinrichtungen (Bibliotheken, Museen, Archive etc.) als Informationsinfrastrukturanbieter im Zusammenspiel mit den Geisteswissenschaften einbezogen werden. Bei Fragen von Betriebsmodellen, Nachhaltigkeit, Vertrauenswürdigkeit etc. besitzen die geisteswissenschaftlichen Fächer bereits eine seit Jahrhunderten funktionierende Infrastruktur durch die Gedächtnisinstitutionen, die in den anderen, größtenteils natur- und sozialwissenschaftlichen Disziplinen erst teuer und mühsam aufgebaut werden muss(te). So stehen Bibliotheken, Museen und Archive als Sinnbild für neutrale, vertrauenswürdige und langlebige Einrichtungen, deren physische und mehr und mehr auch digitale Infrastruktur den Forschenden und Studierenden langfristig zur Verfügung steht.

Insgesamt kann die eingangs gestellte Frage nach der Utopie oder Dystopie einer gemeinsamen geisteswissenschaftlichen Forschungsdatenkultur zum jetzigen Zeitpunkt nicht abschließend beantwortet werden. Es bleibt abzuwarten, wie sich die geisteswissenschaftlichen Fächer im NFDI-Prozess positionieren werden und inwieweit es gelingen wird, über Traditionsgrenzen hinweg neue Formen der Kooperation aufzubauen und nachhaltig mit Leben zu füllen. Die Fachgesellschaften können dabei eine hilfreiche und unterstützende Rolle einnehmen, allerdings haben sie sich in der Vergangenheit beim Themenkomplex Forschungsdatenmanagement nicht besonders engagiert gezeigt. Vielleicht gelingt es den neueren Fachverbänden bzw. Interessensgemeinschaften wie dem [Verband Digital Humanities im deutschsprachigen Raum e.V.](#) (DHD) oder dem neuen *Geisteswissenschaftlichen Forum NFDI* eine neue Kultur der Kooperation zu initiieren.

Nicht umsonst heißt es in der Präambel der *Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI)*: „Dazu muss das Datenmanagement standardisiert sein: Nach den sogenannten FAIR-Prinzipien sollen Forschungsdaten auffindbar, zugänglich, interoperabel und nachnutzbar sei“. Es sollen also NFDI-Konsortien gefördert werden, die glaubhaft machen können, dass sie sich auf den Weg zu FAIRen Daten begeben und darlegen können, wie sie sich die Umsetzung in den nächsten zehn Jahren vorstellen. Dies ist eine gewaltige Aufgabe für die Geisteswissenschaften und ihre tradierten Verfahren, heterogenen Fächerkulturen und fragmentierten Disziplinen, aber wiederum eine enorme Chance für ein großes, geisteswissenschaftliches „Daten-Labor“ als Äquivalent zum Großgerät (z. B. in der Astrophysik).

Falls nur einige der nachfolgenden Fragen mit „Ja“ beantwortet werden können, könnten sich die Geisteswissenschaften glaubhaft und zukunftsorientiert eine eigene Forschungsdateninfrastruktur aufbauen: Besteht ein gemeinsamer Nenner, um geisteswissenschaftliche Sammlungen zu „atomisieren“ (Korpora, Beschreibungen, Metadaten) damit diese als „Großgeräte“ und „Labore“ und Ausgangspunkt für verschiedene Fragestellungen dienen können? Macht es Sinn, für jede einzelne geisteswissenschaftliche Disziplin ein eigenes Daten-Repository aufzubauen?⁴⁶ Oder braucht es nur wenige, international vernetzte, Repositorien in Deutschland (vgl. HathiTrust Digital Library)? Und natürlich muss auch geklärt werden, wo und ab wann disziplinäre Grenzen eine Rolle spielen oder ob sich die Daten einer objektiven Atomisierung verweigern, da sie auf subjektiven Interpretationen beruhen.

Literaturverzeichnis

Blumtritt, Jonathan / Rau, Felix: Metadaten im Zeitalter von Google Dataset Search. Zenodo 2019. DOI: <http://doi.org/10.5281/zenodo.2613444> (Zugriffsdatum: 24.04.2019).

BMBF - Bundesministerium für Bildung und Forschung: Forschung und Innovation für die Menschen: Die Hightech-Strategie 2025. Berlin 2018. https://www.bmbf.de/upload_files-tore/pub/Forschung_und_Innovation_fuer_die_Menschen.pdf (Zugriffsdatum: 24.04.2019).

Borgman, Christine L.: Big Data, Little Data, No Data. Scholarship in the Networked World. Cambridge (Mass.), London: MIT Press 2015.

Buddenbohm, Stefan / Engelhardt, Claudia / Wuttke, Ulrike: Angebotsgenese für ein geisteswissenschaftliches Forschungsdatenzentrum. In: Zeitschrift für digitale Geisteswissenschaften 2016,1. DOI: http://dx.doi.org/10.17175/2016_003 (Zugriffsdatum: 24.04.2019).

Cremer, Fabian / Klaffki, Lisa / Steyer, Timo: Der Chimäre auf der Spur: Forschungsdaten in den Geisteswissenschaften. In: O-Bib 5,2 (2018), S. 142-162. DOI: <https://doi.org/10.5282/o-bib/2018H2S142-162> (Zugriffsdatum 24.04.2019).

DARIAH-DE. <https://de.dariah.eu/> (Zugriffsdatum: 24.04.2019).

DARIAH-DE: Weiterführende Informationen. <https://de.dariah.eu/weiterfuehrende-informationen> (Zugriffsdatum: 24.04.2019).

DataCite Metadata Schema. <https://schema.datacite.org/> (Zugriffsdatum: 24.04.2019).

Deutsches Klimarechenzentrum (DKRZ). <https://www.dkrz.de/> (Zugriffsdatum: 24.04.2019).

DFG - Deutsche Forschungsgemeinschaft: Vorschläge zur Sicherung guter wissenschaftlicher Praxis / Proposals for Safeguarding Good Scientific Practice, Denkschrift / Memorandum, Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“ / Recommendations of the Commission on Professional Self Regulation in Science. Bonn: Wiley-VCH 2013. https://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf (Zugriffsdatum: 24.04.2019).

⁴⁶ vgl. z. B. DHd-AG Datenzentren 2017.

DFG – Deutsche Forschungsgemeinschaft: Leitlinien zum Umgang mit Forschungsdaten. 2015. https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf (Zugriffsdatum: 24.04.2019).

DFG - Deutsche Forschungsgemeinschaft: Umgang mit Forschungsdaten. https://www.dfg.de/foerderung/antrag_gutachter_gremien/antragstellende/nachnutzung_forschungsdaten/ (Zugriffsdatum: 24.04.2019).

DHd - Digital Humanities im deutschsprachigen Raum. <https://dig-hum.de/> (Zugriffsdatum: 24.04.2019).

DHd AG Datenzentren: Geisteswissenschaftliche Datenzentren im deutschsprachigen Raum - Grundsatzpapier zur Sicherung der langfristigen Verfügbarkeit von Forschungsdaten (Version 1.0). DHd AG Datenzentren, 2018. DOI: <http://doi.org/10.5281/zenodo.1134760> (Zugriffsdatum: 24.04.2019).

DKK - Deutsches Klima-Konsortium. <https://www.deutsches-klima-konsortium.de/> (Zugriffsdatum: 24.04.2019).

DTA - Deutsches Textarchiv. <http://www.deutschestextarchiv.de> (Zugriffsdatum: 24.04.2019).

DTA - Deutsches Textarchiv. DTA - Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache. <http://www.deutschestextarchiv.de/doku/referenzkorpus> (Zugriffsdatum: 24.04.2019).

Dublin Core Metadata Element Set, Version 1.1: Reference Description. <http://www.dublin-core.org/specifications/dublin-core/dces/> (Zugriffsdatum: 24.04.2019).

EOSC - European Open Science Cloud. <https://www.eosc-portal.eu/> (Zugriffsdatum: 24.04.2019).

European Commission Expert Group on FAIR Data: Turning FAIR into Reality. Final Report and Action Plan from the European Commission Expert Group on FAIR Data. European Commission: Brüssel 2018. DOI: <https://doi.org/10.2777/1524> (Zugriffsdatum: 24.04.2019).

Europeana: Europeana Data Model. <https://pro.europeana.eu/resources/standardization-tools/edm-documentation> (Zugriffsdatum: 24.04.2019).

FORCE11: The FAIR Data Principles. <https://www.force11.org/group/fairgroup/fairprinciples> (Zugriffsdatum: 24.04.2019).

Gast, Robert: CERN bewilligt neuen LHC-Detektor. Spektrum.de News 06.03.2019. <https://www.spektrum.de/news/cern-bewilligt-neuen-lhc-detektor/1627998> (Zugriffsdatum: 24.04.2019).

Geisteswissenschaftliches Forum NFDI. <https://nfdi.hypotheses.org/> (Zugriffsdatum: 24.04.2019).

German Federation for Biological Data (gfbio). <https://www.gfbio.org/> (Zugriffsdatum 24.04.2019).

GO FAIR: FAIR Principles. <https://www.go-fair.org/fair-principles/> (Zugriffsdatum: 24.04.2019).

Google: Google Data Set Search Beta, Webseite. <https://toolbox.google.com/datasetsearch> (Zugriffsdatum: 24.04.2019).

GWK - Gemeinsame Wissenschaftskonferenz: Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI) vom 26. November 2018 (BAnz AT 21.12.2018 B10). <https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Papers/NFDI.pdf> (Zugriffsdatum: 24.04.2019).

GWK - Gemeinsame Wissenschaftskonferenz: Informationsinfrastrukturen / NFDI. <https://www.gwk-bonn.de/themen/weitere-arbeitsgebiete/informationsinfrastrukturen-nfdi/> (Zugriffsdatum: 24.04.2019).

HathiTrust: Our Digital Library. https://www.hathitrust.org/digital_library (Zugriffsdatum: 24.04.2019).

HathiTrust: Our Research Center. <https://www.hathitrust.org/htrc> (Zugriffsdatum: 24.04.2019).

HathiTrust: Zephir. <https://www.hathitrust.org/zephir> (Zugriffsdatum: 24.04.2019).

Klaffki, Lisa / Schmunk, Stefan / Stäcker, Thomas: Stand der Kulturgutdigitalisierung in Deutschland: Eine Analyse und Handlungsvorschläge des DARIAH-DE Stakeholdergremiums „Wissenschaftliche Sammlungen“, Göttingen 2018 (DARIAH-DE Working Papers, 26). URN: <http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2018-1-3> (Zugriffsdatum: 24.04.2019).

Institute of Museum and Library Services (IMLS). <https://www.ims.gov/> (Zugriffsdatum: 24.04.2019).

Knoche, Michael: Die Idee der Bibliothek und ihre Zukunft. Göttingen: Wallstein 2018.

Lingenhöhl, Daniel: Der tiefe Fall eines Ernährungspapstes. Spektrum.de News 06.03.2019. Webseite. <https://www.spektrum.de/news/der-tiefe-fall-eines-ernaehrungspapstes/1593004> (Zugriffsdatum: 24.04.2019).

Max-Planck-Institut für Astronomie: Startschuss für modernste Kartierung des Himmels. Webseite. <http://www.mpia.de/news/wissenschaft/2000-03-sdss> (Zugriffsdatum: 24.04.2019).

NASA Science: Astrophysics Data Centers. <https://science.nasa.gov/astrophysics/astrophysics-data-centers> (Zugriffsdatum: 24.04.2019).

Neuroth, Heike / Engelhardt, Claudia / Klar, Jochen / Enke, Harry: Aktives Forschungsdatenmanagement - Research Data Management Organiser. In: ABI Technik 38,1 (2018), S. 55-64. DOI: <http://doi.org/10.1515/abitech-2018-0008> (Zugriffsdatum: 24.04.2019).

NOAA - National Oceanic and Atmospheric Administration - Earth System Research Laboratory (ESRL) - Physical Sciences Division (PSD): Data and Imagery. <https://www.esrl.noaa.gov/psd/data/> (Zugriffsdatum: 24.04.2019).

Open Research Data (ORD). <https://data.europa.eu/euodp/de/data/dataset/open-research-data-the-uptake-of-the-pilot-in-the-first-calls-of-horizon-2020> (Zugriffsdatum: 24.04.2019).

Open Research Data Task Force: Realising the Potential. Final Report of the Open Research Data Task Force“. Open Research Data Task Force, 2018. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/775006/Realising-the-potential-ORDTF-July-2018.pdf (Zugriffsdatum: 24.04.2019).

Pronk, Tessa E.: The Time Efficiency Gain in Sharing and Reuse of Research Data, In: Data Science Journal 18,1 (2019), S. 10. DOI: <http://doi.org/10.5334/dsj-2019-010> (Zugriffsdatum: 24.04.2019).

Raspe, Martin / Schelbert, Georg: Genau, wahrscheinlich, eher nicht. Beziehungsprobleme in einem kunsthistorischen Wissensgraph. In: Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. Hg. von Andreas Kuczera / Thorsten Wübbena / Thomas Kollatz. 2019 (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 4). DOI: http://dx.doi.org/10.17175/sb004_012 (Zugriffsdatum: 24.04.2019).

RDA DE: RDA Deutschland Tagung 2019. <https://www.rda-deutschland.de/events/tagung-2019> (Zugriffsdatum: 24.04.2019).

RDMO - Research Data Management Organiser. <https://rdmorganiser.github.io/> (Zugriffsdatum: 24.04.2019).

RfII - Rat für Informationsinfrastrukturen: Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland. Göttingen 2016. <http://www.rfii.de/?p=1998> (Zugriffsdatum: 24.04.2019).

Sahle, Patrick / Kronenwett, Simone: Jenseits der Daten. Überlegungen zu Datenzentren für die Geisteswissenschaften am Beispiel des Kölner ‚Data Center for the Humanities‘. In: LIBREAS. Library Ideas, 23 (2013), S. 76–96. <http://dx.doi.org/10.18452/9043> (Zugriffsdatum: 24.04.2019).

Schöch, Christof: Aufbau von Datensammlungen. In: Jannidis, Fotis / Hubertus Kohle / Malte Rehbein (Hg). Digital Humanities. Eine Einführung. Stuttgart: J.B. Metzler, 2017, S. 223-233.

Schwerpunktinitiative „Digitale Informationen“: Forschungsdaten. <https://www.allianzinitiative.de/de/archiv/forschungsdaten/> (Zugriffsdatum: 24.04.2019).

SDSS - Sloan Digital Sky Survey. <https://www.sdss.org/> (Zugriffsdatum: 24.04.2019).

SDSS - Sloan Digital Sky Survey. How to Cite SDSS. <https://www.sdss.org/collaboration/citing-sdss/> (Zugriffsdatum: 24.04.2019).

SOEP - Sozio-oekonomische Panel. <https://www.diw.de/soep> (Zugriffsdatum: 24.04.2019).

TextGrid: Die Digitale Bibliothek bei TextGrid. Webseite. <https://textgrid.de/digitale-bibliothek> (Zugriffsdatum: 24.04.2019).

The Santa Barbara Statement on Collections as Data. <https://collectionsas-data.github.io/statement/> (Zugriffsdatum: 24.04.2019).

ZVDD - Zentrales Verzeichnis Digitalisierter Drucke. <http://www.zvdd.de/startseite/> (Zugriffsdatum: 24.04.2019).

Wilkinson, Mark D. et al.: The FAIR Guiding Principles for Scientific Data Management and Stewardship. in: Scientific Data 3 (2016). <https://doi.org/10.1038/sdata.2016.18> (Zugriffsdatum: 24.04.2019).