

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

Data Communities - Datenmanagement jenseits von generischen und fachspezifischen Perspektiven

Erkenntnisse aus einem Workshop im Rahmen der FORGE 2021

Esther Marie Asefⁱ Elisabeth Huberⁱⁱ Sabine Imeriⁱⁱⁱ
Eva Ommert^{iv} Michaela Rizzolli^v Cosima Wagner^{vi}

2022

Zitiervorschlag

Asef, Esther Marie, Elisabeth Huber, Sabine Imeri, Eva Ommert, Michaela Rizzolli und Cosima Wagner. 2022. Data Communities - Datenmanagement jenseits von generischen und fachspezifischen Perspektiven. Erkenntnisse aus einem Workshop im Rahmen der FORGE 2021. *Bausteine Forschungsdatenmanagement. Empfehlungen und Erfahrungsberichte für die Praxis von Forschungsdatenmanagerinnen und -managern* Nr. 2/2022: S. 1-12. DOI: [10.17192/bfdm.2022.2.8434](https://doi.org/10.17192/bfdm.2022.2.8434).

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

ⁱFreie Universität Berlin. ORCID: [0000-0003-2411-4953](https://orcid.org/0000-0003-2411-4953)

ⁱⁱUniversität Bremen. ORCID: [0000-0003-4084-4315](https://orcid.org/0000-0003-4084-4315)

ⁱⁱⁱHumboldt-Universität zu Berlin. ORCID: [0000-0002-8844-4014](https://orcid.org/0000-0002-8844-4014)

^{iv}Universität Leipzig. ORCID: [0000-0001-9093-3477](https://orcid.org/0000-0001-9093-3477)

^vFreie Universität Berlin. ORCID: [0000-0001-8154-6563](https://orcid.org/0000-0001-8154-6563)

^{vi}Freie Universität Berlin. ORCID: [0000-0003-4957-3478](https://orcid.org/0000-0003-4957-3478)

Abstract

Die Frage, inwieweit Datenmanagement jenseits von generischen oder fachspezifischen Perspektiven denkbar ist, stand im Mittelpunkt eines Workshops im Rahmen der FORGE 2021. Im Workshop wurde das Konzept der „Data Communities“ (Cooper und Springer 2019) vorgestellt, seine Potenziale mit Blick auf die Sozial- und Geisteswissenschaften diskutiert und anschließend eruiert, welche strategischen wie operativen Kriterien sich daraus für adäquate forschungsunterstützende Datenmanagement-Services ableiten lassen. Der Beitrag fasst die wichtigsten Erkenntnisse aus dem Workshop zusammen und diskutiert, wie das bestehende Konzept um spezifisch sozial- und geisteswissenschaftliche Aspekte erweitert werden könnte.

1 Einleitung

Für adäquates Forschungsdatenmanagement (FDM) müssen – neben generischen Lösungen – vorwiegend solche Infrastrukturen und Werkzeuge angeboten sowie Praktiken etabliert werden, die den spezifischen Besonderheiten einzelner Forschungsbereiche möglichst umfassend gerecht werden – in diesem Punkt besteht weitgehend Einigkeit. Auch die (bereits bewilligten oder geplanten) NFDI-Konsortien gehen in diese Richtung: FDM ist primär fachspezifisch und richtet sich entlang von Disziplinen, Teildisziplinen oder Gruppen von Disziplinen aus.

Bei näherer Betrachtung ist diese Zuordnung aber keineswegs eindeutig. Vielmehr ist insbesondere in den Sozial- und Geisteswissenschaften die Ausdifferenzierung innerhalb einzelner Forschungsdisziplinen mitunter größer, als es die disziplinäre Ordnung nahelegt, wie im Bereich der „kleinen Fächer“¹, mit breit gefächerten Methoden und heterogenen Forschungspraktiken. Das gilt auch für interdisziplinäre Forschungszusammenhänge sowie interdisziplinär angelegte Fächer wie Game Studies, Environmental Humanities oder Regionalstudienfächer. Wie spezifisch fachspezifisches Datenmanagement hier sein kann oder mit Blick auf Infrastrukturen und andere Ressourcen so ausdifferenziert überhaupt sinnvoll realisierbar ist, scheint keineswegs klar. Vermutet werden kann aber, dass die Forschungsgegenstände oder die angewandten Methoden entscheidender für Notwendigkeiten und Verfahren im Umgang mit Forschungsdaten sind als die Zugehörigkeit zu einer Fachdisziplin.

Einen alternativen Ansatz der Skalierung und Strukturierung von Services und Infrastrukturen für FDM diskutieren Danielle Cooper und Rebecca Springer² in einem *Issue Brief* der US-amerikanischen Non-Profit-Organisation Ithaka S+R: Um Datenaustausch und gemeinsame Datennutzung zwischen Forschenden oder Forschungsgruppen zu

¹<https://www.kleinefaecher.de/>, zugegriffen 23. März 2022.

²Cooper, Danielle, und Rebecca Springer. „Data Communities: A New Model for Supporting STEM Data Sharing“. Ithaka S+R Issue Brief, 13. Mai 2019. <https://doi.org/10.18665/sr.311396>.

fördern, sollten Akteur*innen der unterstützenden Bereiche sowie Forschungsförderinstitutionen „Data Communities“ identifizieren und ihre Services gezielt an deren Bedarfen ausrichten. Data Communities werden dabei als fluide und häufig informelle Gruppen von Forschenden definiert, die in erster Linie durch die Arbeit mit und Praktiken des Teilens von „a certain type of data“³ miteinander verbunden sind – und nicht unbedingt durch disziplinäre Zuordnung. Besonderes Augenmerk legen die Autorinnen dabei auf „emergent Data Communities“⁴, also Gemeinschaften, die gerade erst im Entstehen begriffen sind und deshalb noch nicht über stabile Infrastrukturen für den Datenaustausch verfügen.

Data Communities entstehen in der Regel zunächst ohne formale Rahmung oder externen Impuls, bottom-up aus einem konkreten Bedarf und einer geteilten Datenpraxis einer Gruppe von Forschenden heraus. Entsprechend bestehe, wie Cooper und Springer beobachtet haben, unter den Forschenden auch insgesamt eine höhere Motivation, die Praxis des Datenteilens voranzutreiben und stärker zu institutionalisieren. Bereits etablierte Praktiken solcher Data Communities wie vorhandene Standards bei Dateiformaten, Metadaten oder der Dokumentation und Kontextualisierung von Forschungsdaten können dabei für die unterstützenden Bereiche und Infrastruktureinrichtungen Ausgangspunkte sein, stärker formalisierte Strukturen gezielt zu befördern. Das gilt auch für deren zumindest potenziell langfristigen Erhalt, auch wenn neue Data Communities nicht nur sehr schnell entstehen, sondern auch wieder verschwinden können, weil etwa die finanzielle Förderung eingestellt wird oder sich die Bedürfnisse einer Gemeinschaft mit der Zeit ändern.

Als Beispiel einer erfolgreich etablierten Data Community und ihrer infrastrukturellen Form führen Cooper und Springer eine Datenbank mit genetischen Sequenzen der Fruchtfliege an, die – angereichert mit weiteren spezifischen Informationen – nicht nur von Biolog_innen, sondern auch von Physiker_innen, Entomolog_innen und anderen genutzt und ergänzt wird. Als emergente Data Community wiederum werden Wissenschaftler_innen im Bereich der Forschung zur Luftverschmutzung beschrieben, die einerseits gewohnt sind, mit Daten zu arbeiten, die sie nicht selbst erhoben haben (z.B. von Behörden), und die andererseits ein großes Interesse an einer geteilten Datenbank mit Daten zur Luftqualität haben. Wie die Beispiele zeigen, ist das Konzept der Data Communities mit Blick auf die MINT-Fächer formuliert worden. Ließe sich das Konzept auch auf andere Fächergruppen wie die Sozial- und Geisteswissenschaften übertragen, und was wären dann potenzielle Data Communities? Was könnte hier als „a certain type of data“ gelten, der Data Communities wesentlich konstituiert? Was wären Ziele des Austauschs und wie nutzen Forschende in sozial- oder geisteswissenschaftlichen Data Communities geteilte Daten? Welche Ziele verfolgt eine bestimmte Data Community ggf. auch jenseits unmittelbarer Forschungsergebnisse mit der Publikation von Forschungsdaten? Und was für ein Verständnis von Gemeinschaft liegt der jeweiligen Community zugrunde?

³Ebd., 8.

⁴Ebd., 4.

2 Der Workshop „Data Communities“ auf der FORGE 2021

Diesen Fragen widmete sich ein Workshop⁵ am 08.09.2021 im Rahmen der *FORGE 2021*⁶ mit dem Titel „Über den Tellerrand geschaut: Data Communities jenseits von generischen/fachspezifischen Ansätzen“. Der Workshop richtete sich gleichermaßen an FDM-Praktizierende wie an Forschende und schloss neben den Geisteswissenschaften auch die Sozialwissenschaften mit ein. Wesentliches Ziel war, das Konzept der Data Communities in die Diskussion im deutschsprachigen Raum einzubringen und sein Potenzial für die Sozial- und Geisteswissenschaften zu eruieren.

Die Diskussion startete mit einem Input von **Danielle Cooper**⁷, einer der Autorinnen des Konzepts, zu dessen Anwendbarkeit für die Sozial- und Geisteswissenschaften. Cooper sieht insbesondere die Herausforderung, angesichts der großen Heterogenität der Datentypen in den Sozial- und Geisteswissenschaften geteilte Infrastrukturen zu entwickeln. Entsprechend würden z.B. umfangreiche Suchräume über unterschiedliche Datentypen hinweg und mit zentralen Recherchezugängen benötigt. Die Heterogenität geistes- und sozialwissenschaftlicher Daten würde sich etwa auch in der aktuellen Diskussion darüber spiegeln, was überhaupt als zu archivierende, nachnutzbare Forschungsdaten gelten könne. Cooper unterstrich die Notwendigkeit einer umfassenderen Definition von Daten. Hinzu kommt, dass Sozial- und Geisteswissenschaftler_innen spezifische Forschungspraktiken haben, die letztlich auch bedingen, ob oder welche Daten sie mit anderen Forschenden teilen können oder wollen. Bemühungen in Richtung gemeinsamer Datennutzung sollten nach Cooper durch bessere Belohnungssysteme gefördert werden. Wesentliche Erfolgsfaktoren beim nachhaltigen Aufbau von sozial- und geisteswissenschaftlichen Data Communities sieht sie in der engen Zusammenarbeit von Forschenden mit Bibliothekar_innen oder Informationswissenschaftler_innen, z.B. in den Bereichen Metadaten, Vokabulare oder Datenschutz sowie mit Blick auf eine langfristige Finanzierung der technischen Infrastrukturen.

Erzsébet Tóth-Czifra⁸, Expertin im Bereich FDM in den Geisteswissenschaften und Open Science Officer bei *DARIAH-EU*, knüpfte hier an und stellte die Aktivitäten von *DARIAH-EU* zur Förderung von Data Communities in den Geisteswissenschaften vor. Als größte Herausforderung beschrieb sie ein „cross sector continuum of data curation“: Akteur_innen wie Einrichtungen des kulturellen Erbes, Digital Humanities (DH)-Labs, Forschungs-Communities und Forschungsinfrastrukturen (inkl. Verlage) seien auf eine Kollaboration im „Datenkuratierungs-Workflow“ angewiesen, würden aber bis-

⁵Asef, Esther Marie, Elisabeth Huber, Sabine Imeri, Eva Ommert, Michaela Rizzolli, und Cosima Wagner. „Über den Tellerrand geschaut: ‚Data Communities‘ jenseits von generischen/fachspezifischen Ansätzen. Vorstellung und Diskussion eines Konzeptes.“ *FORGE 2021: Forschungsdaten in den Geisteswissenschaften - Mapping the Landscape - Geisteswissenschaftliches Forschungsdatenmanagement zwischen lokalen und globalen, generischen und spezifischen Lösungen*, 3. September 2021, Cologne. <https://doi.org/10.5281/zenodo.5379693>.

⁶<https://forge2021.uni-koeln.de/>, zugegriffen 23. März 2022.

⁷<https://orcid.org/0000-0001-7400-9735>.

⁸<https://orcid.org/0000-0002-5350-067X>.

lang in Open-Science-Agenden als getrennte „Silos“ betrachtet. Eine engere Zusammenarbeit zum übergreifenden Teilen von Kulturdaten sei vonnöten und würde daher in einer „cultural heritage data reuse charter“⁹ erstmals adressiert. Auf diese Weise sei der Aufbau nachhaltiger Data Communities möglich. Als zwei aktuelle Beispiele nannte sie zudem die Projekte *Social Sciences & Humanities Open Cloud (SSHOC)*¹⁰ und *Computational Literary Studies Infrastructure (CLS INFRA)*¹¹.

Anschließend stellten eingeladene Impulsgeber_innen Beispiele informeller Communities in ihren Forschungsfeldern vor, die auf unterschiedliche Weise mit Daten arbeiten und sich potenziell als (emergente) Data Communities begreifen lassen.

In seinem Bericht zur FDM-Praxis am Center for World Music an der Universität Hildesheim und dem European Centre for Jewish Music verwies **Samuel Mund**¹² u.a. auf spezifische Herausforderungen in internationalen Forschungsprojekten. Im Fall etwa der SDG Graduate School *Performing Sustainability. Cultures and Development in West-Africa*¹³ werden Infrastrukturen an Partneruniversitäten in Ghana und Nigeria aufgebaut, wobei rechtliche Vorgaben der jeweiligen Länder beachtet werden müssen. Eine Data Community entsteht in diesem Fall um die Graduiertenschule herum, ein verbindendes FDM-Querschnittsthema sei die Berücksichtigung der CARE-Prinzipien¹⁴ und die Schulung von Forschenden zu diesem Thema. Besonderes Augenmerk liege hier darauf, die Data Community (Forschende der Graduiertenschule) mit der „Source Community“ (den Forschungsteilnehmenden und Expert_innen, mit denen in der Forschung gemeinsam Daten produziert werden) zu vernetzen. Das Anliegen sei, gemäß der CARE-Prinzipien diese ansonsten meist getrennten Communities zusammenzubringen und idealerweise die Nutzung der Forschungsdaten nicht nur den Forschenden, sondern auch der „Source Community“ zu ermöglichen.

Dirk Goldhahn¹⁵ von der Sächsischen Akademie der Wissenschaften diskutierte das Data Community-Konzept am Beispiel von *CLARIN(-D)* bzw. des NFDI-Konsortiums *Text+*. Gemeinsamkeiten unter den Forschenden würden sich u.a. im Hinblick auf gängige Verfahren der Datenverarbeitung und -analyse sowie auf die Anforderungen an die Archivierung und Zugänglichmachung der Daten finden lassen. Die Services ließen trotzdem eine gewisse Flexibilität zu, bspw. durch Baukastentools oder Exportmöglichkeiten in verschiedene Dateiformate. Eine Data Community basiere nicht zuletzt auf der gegenseitigen Unterstützung der Mitglieder bei der Anwendung von Methoden und dem Ausloten von Analysemöglichkeiten.

⁹Tóth-Czifra, Erzsébet, und Laurent Romary. „The Heritage Data Reuse Charter: From Principles to Research Workflows“, 12. Februar 2020. <https://halshs.archives-ouvertes.fr/halshs-02475692>, zugegriffen 23. März 2022.

¹⁰<https://sshopencloud.eu/>, zugegriffen 23. März 2022.

¹¹<https://clsinfra.io/>, zugegriffen 23. März 2022.

¹²<https://orcid.org/0000-0003-0684-8945>.

¹³<https://www.uni-hildesheim.de/sustainability/>, zugegriffen 22. Mai 2022.

¹⁴<https://www.gida-global.org/care>, zugegriffen 23. März 2022.

¹⁵<https://d-nb.info/gnd/104558892X>.

Christoph Schmitt¹⁶ vom digitalen Wossidlo-Archiv *WossiDiA* an der Universität Rostock¹⁷ erläuterte die Herausforderungen bei der Übertragung einer umfangreichen historischen, kultur- und sprachwissenschaftlichen Zettelsammlung in eine digitale Datenbank. Das Ziel, das Material nicht nur digital lesbar, sondern auch verständlich und in seinem Entstehungsprozess nachvollziehbar zu präsentieren, sei neben der Digitalisierung selbst vor allem durch Transliteration und Übersetzung, der Erstellung von Hypergraphen sowie eines eigenen Thesaurus erreicht worden. Eine (internationale) Data Community würde insbesondere im Projekt *ISEBEL (Intelligent Search Engine for Belief Legends)*¹⁸ entstehen, das die Daten aus *WossiDiA* mit anderen europäischen Erzählarchiven verknüpfen soll. Auch die Einbeziehung von Lokalhistoriker_innen in die Erarbeitung des – wiederum mit *WossiDiA* verknüpften – *Ortschronikenportals Mecklenburg-Vorpommern*¹⁹ hätte dazu beigetragen, die *WossiDiA*-Community zu erweitern. Dass eine Data Community keine statische, in sich geschlossene Gemeinschaft ist, wird anhand der vielfältigen Nutzungsmöglichkeiten und dem weiteren Ausbau der Sammlung deutlich.

Maria Effinger²⁰ zeigte in ihrem Rückblick auf die Entstehung der FDM-Services der Universitätsbibliothek (UB) Heidelberg, wie wichtig die Zusammenarbeit mit konkreten Projekten aus den geisteswissenschaftlichen Disziplinen für das modulare System der UB aus Datenbanken für Bilder, digitale Editionen, digitalisierte Handschriften oder die Langzeitarchivierung sowie deren Weiterentwicklung ist. Die aktuelle Infrastruktur kann insofern als Ergebnis eines Bottom-up-Prozesses verstanden werden. In der UB Heidelberg stünden etwa „enhanced publications“ als „Ort“ der Verknüpfung und Veröffentlichung von nachhaltig archivierten Forschungsdaten mit Publikationen im Vordergrund, da beide besonders in den Geisteswissenschaften stärker als Gesamtheit gesehen werden. Die Langzeitarchivierung komplexer Datentypen, wie 3D-Daten, sei bislang noch nicht befriedigend gelöst, solle aber im Rahmen der aktiven Beteiligung an *NFDI4Culture*²¹ mitentwickelt werden. Auf diese Weise könnten Publikationen – und nicht Daten – Anlass und Motivation dafür sein, in Beziehungen mit anderen zu treten, um Publikationen als Ressource auszutauschen, wechselseitig zu nutzen und die gezielte Öffnung von Wissenschaft durch die Unterstützung von Publikationsprozessen in Archiven und Bibliotheken zu fördern.

¹⁶<https://d-nb.info/gnd/132174170> zugegriffen 23. März 2022.

¹⁷<https://www.wossidia.de/>, zugegriffen 23. März 2022.

¹⁸<https://isebel.eu/>, zugegriffen 23. März 2022.

¹⁹https://www.ortschroniken-mv.de/index.php/Ortschroniken_Mecklenburg-Vorpommern, zugegriffen 23. März 2022.

²⁰<https://orcid.org/0000-0001-6396-4876>.

²¹<https://nfdi4culture.de/>, zugegriffen 23. März 2022.

3 Workshop-Diskussion und Ergebnisse

In der Diskussion zeigte sich, dass sich die hier vorgestellten Forschungsnetzwerke und Infrastrukturen in einigen Eigenschaften und Dimensionen von der ursprünglichen Definition der Data Communities bei Cooper und Springer unterscheiden oder anders verstanden werden. Die wichtigsten und dringlichsten Punkte der Diskussion – Grenzen, Spannungsfelder und blinde Flecken – werden im Folgenden dargestellt.

3.1 Mit Blick auf den Begriff „Community“

Wie bereits in den Impulsvorträgen zeigten sich auch in der Diskussion zunächst unterschiedliche Verständnisse und Akzentuierungen von Data Communities. Insbesondere der Community-Begriff wurde – auch vor dem Hintergrund spezifischer Fachdebatten – unterschiedlich ausgelegt, so wie ohnehin der Aspekt „Community“ teils stärker in den Mittelpunkt gerückt wurde als der Aspekt „Data“. Konsens bestand darüber, dass eine potenzielle Data Community auch von Akteur_innen jenseits der unmittelbaren Forschung (mit) gestaltet werden kann bzw. dies in den Geisteswissenschaften häufig der Fall ist. Cooper und Springer zufolge bestehen Data Communities aus einem „network of researchers“²² (hier sind professionelle Forschende gemeint) – eine Definition, die sich mit Blick auf die Datenproduktion und Forschungsnetzwerke in den Sozial- und Geisteswissenschaften etwa hinsichtlich der Rolle von Kultur- und Gedächtnisinstitutionen wie Archiven, Museen oder Galerien nicht ohne weiteres halten lässt. Zwar wird in diesen Einrichtungen auch geforscht, gleichzeitig sind aber Bewahrung und Wissensvermittlung an eine breite Öffentlichkeit zentrale Aufgaben. Entsprechend sind es mit Blick auf eine nachhaltige Digitalisierung etwa von Objekten des kulturellen Erbes dann oft gerade diese Einrichtungen, die in den Austausch mit zukünftigen Nutzenden gehen, um die benötigten Formate, Metadaten und Schnittstellen, aber auch Nutzungsszenarien zu eruieren.

Eine Personengruppe, die häufig – auch in den Naturwissenschaften – unsichtbar bleibt²³ und auch im Konzept der Data Communities bisher nicht adressiert wurde, sind zudem sog. Bürgerwissenschaftler_innen bzw. Laienforschende. Unabhängig davon, ob Datensammlung und Datenauswertung durch „Laien“ selbstorganisiert oder in koordinierten Citizen-Science-Projekten erfolgt, tragen sie unbestreitbar in vielen Bereichen über alle Disziplinen hinweg erheblich zur Forschung bei.²⁴ Mit Citizen-Science-

²²Cooper und Springer, Data Communities, 8.

²³Vgl. Vohland, Katrin, Anne Land-Zandstra, Luigi Ceccaroni, Rob Lemmens, Josep Perelló, Marisa Ponti, Roeland Samson, und Katherin Wagenknecht. „Editorial: The Science of Citizen Science Evolves.“ In *The Science of Citizen Science*, dies. Hrsg., 1-12. Cham: Springer, 2021. https://doi.org/10.1007/978-3-030-58278-4_1. Siehe auch Cooper, Caren B., Jennifer Shirk, und Benjamin Zuckerberg. „The Invisible Prevalence of Citizen Science in Global Research: Migratory Birds and Climate Change“. *PLOS ONE* 9, Nr. 9 (2014): e106508. <https://doi.org/10.1371/journal.pone.0106508>.

²⁴Im Workshop wurden das Projekt Ortschroniken Mecklenburg-Vorpommern (https://www.ortschroniken-mv.de/index.php/Ortschroniken_Mecklenburg-Vorpommern) und

Projekten könnten demnach Data Communities „gefestigt“ werden oder entstehen. Zudem können auch beforschte Personen oder Gruppen als Teil einer Data Community gelten, etwa wenn sie als Co-Autor_innen bzw. Miteigentümer_innen der Daten verstanden werden – eine Diskussion, die z.B. in der ethnologischen Forschung insgesamt²⁵ und aktuell insbesondere mit Blick auf die Provenienzforschung in ethnologischen Museen und Prozesse der Repatriierung intensiv geführt wird.²⁶

In der Praxis können also an einem Forschungsprozess verschiedene heterogene Akteur_innen beteiligt sein, die sich aber nicht unbedingt auch zu einer Data Community zusammenfinden. In diesem Sinne müsste im Fortgang weiter ausgearbeitet und stärker systematisiert werden, welche Akteur_innen typischerweise an Prozessen sozial- und geisteswissenschaftlicher Datenproduktion beteiligt sind und welche Konsequenzen sich daraus ergeben. So ist etwa davon auszugehen, dass unterschiedliche Akteur_innen auch je eigene Interessen verfolgen, die etwa mit Blick auf die Infrastrukturentwicklung ausbalanciert werden müssen. Gleiches gilt für je spezifische Notwendigkeiten, die sich aus disparaten strukturellen Verankerungen ergeben.

3.2 Mit Blick auf die Struktur von Communities

Auf der Suche nach bestehenden und potenziellen Data Communities in den Sozial- und Geisteswissenschaften im deutschsprachigen Raum hat sich ebenso gezeigt, dass diese Praxisgemeinschaften nicht nur heterogene Akteur_innen umfassen können, sondern dass sie auch in verschiedensten Formen auftreten – groß, klein, strukturiert und unstrukturiert. Thematisiert wurde in diesem Zusammenhang insbesondere, welche Rolle der Umfang von Data Communities spielt. So stellte sich etwa die Frage, wie groß eine „informelle“ Gruppe von Personen werden kann, um weiterhin handlungsfähig zu sein. Wäre eine Community, die sich um relativ stabile Strukturen wie *DARIAH-EU* herum organisiert, bereits zu groß, zu wenig auf Daten konzentriert oder thematisch zu wenig fokussiert, als dass sie im Sinne des Konzepts (noch) als Data Community gelten könnte? Umgekehrt stellt sich die Frage, wie klein eine Community eigentlich sein kann, um als eigenständige Data Community anerkannt, entsprechend gefördert und unterstützt zu werden. Damit verbunden ist auch die Frage nach der Sprechfähigkeit von Communities, also der Fähigkeit, eigene Bedarfe und Interessen zu artikulieren, sich gegen andere Communities durchzusetzen oder sich in entsprechende Beteiligungsprozesse einzubringen – nicht zuletzt vor dem Hintergrund knapper personeller

das Projekt Hallische Heiratsgeschichten (<https://www.geschichte.uni-halle.de/struktur/hist-data/citizenscience/>), beide zugegriffen 23. März 2022) als Beispiele für die Sozial- und Geisteswissenschaften genannt.

²⁵Vgl. Murillo, Luis F.R. "What Does 'Open Data' Mean for Ethnographic Research?" *American Anthropologist* 120, Nr. 3 (2018): 577-582. <https://doi.org/10.1111/aman.13088>.

²⁶Vgl. die Beiträge in Savoy, Bénédicte, Felicity Bodenstern, und Merten Lagatz (Hg.). *Translocations. Histories of Dislocated Cultural Assets*. Bielefeld: transcript, 2022 (im Erscheinen).

Ressourcen und struktureller Ungleichgewichte bzw. Nachteile für bestimmte Gruppen wie Promovierende in Einzelprojekten.²⁷

Aus der Perspektive von FDM-Unterstützenden kann die Frage nach der Größe relevant sein, wenn es das Ziel ist, emergente Data Communities möglichst früh zu identifizieren. Hat sich eine Community infrastrukturell nachhaltig etabliert, wie etwa im Fall der Datenbank zur Gensequenz-basierten Fruchtfliegen-Forschung, könnte Informatik eine zunehmend untergeordnete Rolle bei der Nutzung und Ergänzung der verfügbaren Daten spielen. Schaut man auf andere Beispiele wie etwa die *edition humboldt digital*²⁸, dann könnte das Projektteam womöglich bereits als Data Community verstanden werden, denn Forschende aus Disziplinen wie der Romanistik, den Geschichtswissenschaften oder der Geographie arbeiten hier mit Informatiker_innen gemeinsam an der Aufbereitung und Veröffentlichung von Alexander von Humboldts Nachlass. Während die technische Infrastruktur und die Einhaltung von Standards die Gruppe durchaus als Data Community ausweisen, bleibt offen, von wem die Daten auf welche Weise genutzt werden. Indem wissenschafts- und kulturhistorisch herausgehobene Quellen auch breiteren Öffentlichkeiten zur Verfügung gestellt werden, kommt womöglich auch hier eine Besonderheit zum Tragen, die schon mit Blick auf den Community-Begriff thematisiert wurde: Bibliotheken, Archive und Museen als Akteure in Data Communities haben mit ihren Aufgaben des Bewahrens und Vermittelns nicht nur Forschende im Blick. Damit in Zusammenhang steht, dass Daten und Informationen zum kulturellen Erbe nicht im strikten Sinn veralten und als überholt gelten, sondern fortlaufend ergänzt, angereichert und neu interpretiert werden. Am Beispiel des Projekts *Oral-History.Digital*²⁹, das lebensgeschichtliche Erzählungen aus verschiedenen Sammlungen bündeln will, lässt sich wiederum beobachten, wie Datenbestände in Zusammenarbeit mit mehreren Einrichtungen für eine breit definierte Gruppe von Nutzenden aufbereitet werden können. Hier wäre analog zur Frage, wie sich Data Communities zusammensetzen, aus der Perspektive FDM-unterstützender Einrichtungen auch die Handhabbarkeit heterogener Gruppen, etwa mit Blick auf Zugangs- und Bearbeitungsrechte, weiter zu diskutieren.

3.3 Mit Blick auf die Datentypen einer Community

Mehrfach angesprochen wurde, ob die Fokussierung auf bestimmte Datentypen als eine Art Kern von Data Communities für Disziplinen der Sozial- und Geisteswissenschaften sinnvoll sein kann oder ob sich Communities in diesen Fächergruppen nicht eher um Forschungsgegenstände herum bilden wie im Fall von *WossiDiA* oder um Methoden herum wie bei *CLARIN(-D)* – wobei das Teilen von Daten zunächst einen nachrangigen Stellenwert einnimmt – oder auch um Publikationsformen herum wie im Fall

²⁷Siehe auch <https://www.kleinefaecher.de/beitraege/blogbeitrag/dokumentation-des-informations-und-vernetzungshtml>, zugegriffen 23. März 2022.

²⁸<https://edition-humboldt.de/>, zugegriffen 23. März 2022.

²⁹<https://www.oral-history.digital/>, zugegriffen 23. März 2022.

der UB Heidelberg. Weitere Beispiele hierfür finden sich in institutionalisierter Form auch in Strukturen wie dem Forschungsdatenzentrum *DeZIM.fdz*³⁰, das Forschungsdaten aus der Integrations- und Migrationsforschung bündelt, oder der Plattform *Discuss Data*³¹, die das Teilen und gemeinsame Bearbeiten von Daten aus Forschungen zu und in postsowjetischen Staaten ermöglicht.

Insgesamt sind Praktiken des Teilens von Daten auch im Forschungsprozess selbst in vielen dieser Felder (bislang) wenig etabliert. Zudem sind Datensätze in den Geisteswissenschaften und mindestens auch in den qualitativ arbeitenden Sozialwissenschaften gegenüber denen der Natur- und Lebenswissenschaften häufig weniger umfangreich und gleichzeitig ausgesprochen heterogen. In vielen Bereichen gibt es kaum etablierte Standards für Datenformate, Metadaten und Formen der Dokumentation und Kontextualisierung. Dies führt auch dazu, dass sich möglicherweise kein „einfaches“ Austauschformat herauskristallisiert und das Teilen der Daten nicht ohne weiteres bottom-up in einer selbstgebauten Infrastruktur münden kann. Auch für informelles Teilen müssten die meisten Communities gleich mehrere Formate und Standards bzw. neue Arten der Informationsstrukturierung berücksichtigen. Dies steht im Gegensatz zu den von Cooper und Springer identifizierten Merkmalen von etablierten Data Communities, für die es einfach ist, Daten zu teilen und die gemeinsam ausgehandelte Standards verwenden.³²

Gleichwohl gibt es längst Praktiken des Datenteilens bzw. der Verfügbarmachung von Daten und Materialien, auch hier wieder unter Beteiligung der Gedächtnisinstitutionen, die ihre Bestände großflächig digital bereitstellen. Die Anreicherung solcher bereits vorhandenen Daten und Materialien gehört zu den zentralen Praktiken in vielen geisteswissenschaftlichen Disziplinen, etwa wenn ein literarischer Text immer wieder von verschiedenen Personen unter variierenden Fragestellungen untersucht wird. Anreicherung bedeutet hier auch zunehmende Informationsverdichtung und Kontextualisierung und – sofern geeignete Infrastrukturen und Services das ermöglichen – auch Weitergabe von Forschungsdaten. Die datenproduzierende und datenarchivierende Data Community ist in diesem Sinne nicht unbedingt identisch mit künftigen nachnutzenden Data Communities. Und auch das, was zur Nachnutzung vorgesehen war, kann sich ändern, wenn z.B. zur Katalogisierung von dänischen Volksmärchen und Sagen verwendete Metadatenformulare selbst zu Informationsquellen – und damit Forschungsdaten – für die Genese von Erzählsammlungen werden.³³

³⁰<https://dezim-institut.de/forschungsdatenzentrum-dezimfdz/>, zugegriffen 23. März 2022.

³¹<https://www.discuss-data.net/>, zugegriffen 23. März 2022.

³²Vgl. Cooper und Springer, *Data Communities*, 14-16.

³³Holm Pedersen, Marianne, und Lars C. Kofoed Romer. "Field notes from a field of notes: Anthropology and the afterlife of notes in archives." In *Anthropology Inside Out: Fieldworkers Taking Notes*, Hrsg. Mikkel Rytter, Astrid Oberborbeck Andersen, Cecilie Rubow, Mette Lind Kusk, Anne Line Dalsgård, und Maria Nielsen, 33-53. Herefordshire: Sean Kingston Publishing, 2020.

4 Fazit

Angesichts heterogener Methoden, Forschungspraktiken und Daten in den Sozial- und Geisteswissenschaften halten wir Ansätze, die Datenmanagement auch jenseits von und ergänzend zu generischen und fachspezifischen Perspektiven denken, für eine gute Möglichkeit, FDM-Services und -Infrastrukturen noch stärker an der konkreten Praxis von Forschenden auszurichten. So würde die Zusammenarbeit zwischen Infrastrukturen und Forschung gefördert und Expertise bei unterstützenden Einrichtungen, aber auch bei den Forschenden für künftige Vorhaben aufgebaut werden. Eine Perspektive auf Data Communities kann zudem helfen, vor allem auch solche Forschungsgruppen systematischer zu unterstützen, die sich eben nicht ohne weiteres disziplinar zuordnen lassen.

Weil Data Communities Fachgrenzen überwinden, weist dieses Konzept aber auch darauf hin, wie Wissenschaft eigentlich praktiziert wird – nämlich interdisziplinär, vielseitig und in einem stetigen Änderungsprozess. Dieser Realität und den daraus resultierenden Bedarfen zu entsprechen, sollte eine wichtige Zielsetzung sein. In diesem Sinne könnte es für eine adäquate FDM-Unterstützung auch im Kontext der Sozial- und Geisteswissenschaften lohnend sein, nach Data Communities Ausschau zu halten. Im Workshop hat sich gezeigt, dass es hier vielleicht nicht so sehr der Blick auf „a certain type of data“ ist, sondern es eher methodisch definierte Kategorien, wie z.B. Sprachkorpora, digitale Quelleneditionen, Zeitzeugeninterviews, ethnografische Forschungsmaterialien oder auch thematische Linien sind, die Überlegungen zum Potenzial von Data Communities für diese Disziplinen anschlussfähig machen. Für die Zukunft würde sich daraus die Notwendigkeit ergeben, auch kleine Bottom-up-Communities zu fördern und vermehrt dort anzusetzen, wo die Motivation, Daten zu teilen, aus der konkreten Forschungspraxis erwächst. Wir schließen uns Cooper und Springer an, dass die Aufgabe, entstehende Data Communities zu identifizieren und zu unterstützen, solchen Akteur*innen zukommen sollte, die die Nutzung von Daten fördern wollen³⁴, also etwa Bibliotheken, Archiven, NFDI-Konsortien und weiteren Informationsinfrastrukturen und Kultureinrichtungen sowie Forschungsförderern. Offen bleibt, wie entstehende Data Communities (systematisch) gefunden werden können und nach welchen Kriterien entschieden werden kann oder sollte, welche Data Community unterstützt bzw. „stabilisiert“ werden kann und wie ein solches Unterstützungsmodell finanziert wird. Das im Workshop vorgestellte Beispiel der Services der UB Heidelberg zeigt, dass es möglich ist, notwendige Ressourcen dafür zu schaffen.

Durch eine gezielte Fokussierung auf informelle Netzwerke, in denen Forschungsdaten geteilt und ausgetauscht werden, können konkrete Bedarfe zudem auch ohne wiederholte, aufwändige Bedarfserhebungen identifiziert und bearbeitet werden. Vorhandene Praktiken des Teilens können aufgegriffen und weiterentwickelt, ggf. auch mit nachhaltigen Services und Infrastrukturen gefestigt werden. Dies soll Bedarfserhebungen nicht ersetzen, könnte aber ein komplementäres Instrument sein. Eine zentrale

³⁴Vgl. Cooper und Springer, Data Communities, 4.

Frage hierbei ist, inwieweit es gelingen kann, flexibel auf die sich ständig ändernden Bedarfe und Anforderungen aus der Forschungspraxis zu reagieren und auch kurzfristig ergebnisorientierte Lösungen für sich entwickelnde Data Communities anzubieten. Denn während Nachhaltigkeit im Kontext von FDM stets im Fokus steht, gerät die Frage, wie sich flexibilisierende Elemente in eine bestehende Infrastruktur „einbauen“ bzw. in den Prozess der Infrastrukturentwicklung integrieren lassen, zumeist aus dem Blick. Auch förderpolitische Rahmenbedingungen stehen womöglich der kurzfristigen Entwicklung dynamischer und zunächst (noch) nicht oder nicht unbedingt auf Dauerhaftigkeit und Nachhaltigkeit bedachter Data Communities entgegen. Freilich ist Nachhaltigkeit wichtig, gleichwohl stellen Aspekte wie Flexibilität, Dynamik und Fluidität – die bei der (Weiter-)Entwicklung von Informationsinfrastrukturen wenig adressiert werden – wesentliche Charakteristika erfolgreicher Data Communities dar. Im Hinblick auf den Aufbau nachhaltiger Forschungsdateninfrastrukturen etwa im Rahmen der NFDI-Konsortien, die Forschungsdaten systematisch erschließen, nachhaltig sichern und zugänglich machen sollen, wird sich zeigen, inwieweit sich beides – die Dynamik von Data Communities einerseits und die Stabilität von Infrastrukturen andererseits – zusammenzudenken und in eine angemessene Balance bringen lässt.

Wenn es ein Ziel ist, Datensilos zu vermeiden, dann könnten gerade emergente Data Communities Impulse für die Weiterentwicklung oder auch Verknüpfung bestehender Infrastrukturen setzen. Mit Blick auf die potenziell stärkere Heterogenität von Datengebernden und Datennutzenden und damit auch von möglichen Data Communities in den Sozial- und Geisteswissenschaften wäre weiter zu diskutieren, welche Herausforderungen hinsichtlich zukünftiger vielfältiger Nutzungsgruppen angegangen werden müssen. Die Erfahrungen von Museen und anderen Gedächtniseinrichtungen mit der digitalen Aufbereitung und Präsentation von Inhalten für verschiedene Öffentlichkeiten können hier sicher fruchtbar gemacht werden. Ebenso ist – insbesondere mit den Methoden der User-Experience-Forschung – genauer zu untersuchen, wie Forschungsdaten von denjenigen, die sie (nach)nutzen wollen, gesucht und gefunden werden können.³⁵

Die Diskussion im Workshop hat gezeigt, dass die Übertragung des für den MINT-Bereich definierten Data Community-Konzepts auf die Geistes- und Sozialwissenschaften eine Reihe von Fragen aufwirft und womöglich Erweiterungen benötigt. Gleichwohl ermöglicht das Konzept, Forschungsunterstützung aus einer weiten Perspektive zu denken und vor allem den Aspekt der Fluidität der Forschung in nachhaltigen Infrastrukturen zu reflektieren. Es bietet gleichzeitig Impulse, jenseits von allgemeinen allumfassenden Services konkrete Bereitschaft zum Teilen von Daten in Forschungsgruppen aufzugreifen und zu fördern. Den Bibliotheken, Archiven und Museen kommt hierbei eine zentrale Rolle zu: durch die Bereitstellung von Werkzeugen, durch die Bereitschaft zum Dialog mit unterschiedlichen Nutzergruppen und einer gewissen Neugier für neue potentielle Partner*innen.

³⁵Siehe Gregory, Kathleen. "Findable and reusable? Data discovery practices in research". Dissertation, Universität Maastricht, 2021. <https://doi.org/10.26481/dis.20210302kg>.