



Die Nationale Forschungsdateninfrastruktur – eine Lösung infrastruktureller Bedarfe für die Inhaltsanalyse?

Annett Heft · Jakob Jünger · Julia Niemann-Lenz · Daniel Possler

Eingegangen: 19. Dezember 2022 / Angenommen: 14. Mai 2023 / Online publiziert: 5. Juli 2023
© The Author(s) 2023, korrigierte Publikation 2023

Zusammenfassung Obwohl die Inhaltsanalyse eine zentrale Stellung in der Kommunikations- und Medienforschung besitzt, existieren kaum Forschungsinfrastrukturen für diese Methode. Gleichzeitig werden in Deutschland seit 2018 große Dateninfrastrukturen in den 27 Konsortien der Nationalen Forschungsdateninfrastruktur (NFDI) aufgebaut. In diesem Beitrag gehen wir aus Perspektive der Forschenden der Frage nach, inwiefern die NFDI-Konsortien Lösungen für die infrastrukturellen Anforderungen in Bezug auf Inhaltsanalysen bieten. Zunächst beleuchten wir diese Anforderungen entlang des Forschungsdaten-Lebenszyklus und identifizieren Leerstellen. Dann explorieren wir, welche Bedarfe die NFDI-Konsortien decken können. Der Schwerpunkt liegt auf Konsortien, die sich auf die Sammlung und Aufbereitung von Text oder multimodalen Daten konzentrieren: *KonsortSWD*, *BERD@NFDI*, *Text+*, *NFDI4Memory*, *NFDI4Culture* und *NFDI4DataScience*. Unsere Untersuchung zeigt, dass die Konsortien bereits viele der Bedarfe abdecken.

✉ Dr. Annett Heft
Institut für Publizistik- und Kommunikationswissenschaft, Freie Universität Berlin,
Garystraße 55, 14195 Berlin, Deutschland
E-Mail: annett.heft@fu-berlin.de

Jun. Prof. Dr. Jakob Jünger
Institut für Kommunikationswissenschaft, Universität Münster, Bispinghof 9–14, 48143 Münster,
Deutschland
E-Mail: jakob.juenger@uni-muenster.de

Dr. Julia Niemann-Lenz
Fachbereich Sozialwissenschaften/Journalistik und Kommunikationswissenschaft, Universität
Hamburg, Sedanstraße 19, 20146 Hamburg, Deutschland
E-Mail: julia.niemann-lenz@uni-hamburg.de

Dr. Daniel Possler
Institut Mensch-Computer-Medien, Julius-Maximilians-Universität Würzburg,
Oswald-Külpe-Weg 82, 97074 Würzburg, Deutschland
E-Mail: daniel.possler@uni-wuerzburg.de

Allerdings gibt es weder ein Konsortium, in dem Kommunikationswissenschaftler:innen treibende Kräfte sind, noch wird die Inhaltsanalyse explizit berücksichtigt. Wir diskutieren, wie sich Forschungsinfrastrukturen für die Inhaltsanalyse durch die NFDI-Strukturen weiterentwickeln ließen.

Schlüsselwörter Forschungsinfrastrukturen · Inhaltsanalyse · Textdaten · Multimodale Daten · Nationale Forschungsdateninfrastruktur · Forschungsdaten · Forschungsdaten-Lebenszyklus

The national research data infrastructure—a solution for infrastructural needs of content analysis?

Abstract Content analysis has a central position in media and communication research, yet research infrastructures for the method are still scarce. At the same time, the 27 consortia of the National Research Data Infrastructure (NFDI) have started to establish large data infrastructures in Germany since 2018. In this paper, we explore from the perspective of researchers whether the NFDI consortia provide solutions to the infrastructural needs of content analysis. First, we illustrate these needs throughout the research data lifecycle and identify shortcomings. We then explore whether the NFDI consortia can meet these needs. The focus lies on consortia that concentrate on the collection and processing of text or multimodal data: *KonsortSWD*, *BERD@NFDI*, *Text+*, *NFDI4Memory*, *NFDI4Culture*, and *NFDI4DataScience*. Our exploration shows that many of the needs are already being addressed by the consortia. However, there is no consortium in which communication scholars are the driving force and content analysis does not receive explicit consideration. We discuss how research infrastructures for content analysis can be further developed through the NFDI structures.

Keywords Research infrastructures · Content analysis · Text data · Multimodal data · National Research Data Infrastructure · Research data · Research data life cycle

1 Einleitung

Forschungsinfrastrukturen sind „Instrumente und Institutionen, die allen Mitgliedern der jeweiligen wissenschaftlichen Gemeinschaft zur Unterstützung ihrer Forschung zur Verfügung stehen“ (Wissenschaftsrat 2011, S. 13). Mit der Nationalen Forschungsdateninfrastruktur (NFDI) existiert in Deutschland seit 2018 eine großangelegte Initiative zum Aufbau solcher Infrastrukturen. Dabei sollen „Forschungsdaten für das gesamte deutsche Wissenschaftssystem systematisch erschlossen, vernetzt und nachhaltig nutzbar gemacht“ (Hartl et al. 2021, S. 370) werden. Die Umsetzung erfolgt durch 27 Konsortien, die verschiedene Wissenschaftsbereiche und Datentypen abdecken (siehe Abschn. 3).

Die Kommunikationswissenschaft ist unseres Wissens an keiner dieser NFDI-Initiativen aktiv mit eigenen Infrastrukturangeboten beteiligt. Doch gerade für die

qualitative wie auch die quantitative Inhaltsanalyse wären kommunikationswissenschaftlich ausgerichtete Forschungsinfrastrukturen von besonderer Relevanz für Forschende, denn die Methoden nehmen eine zentrale Stellung in unserem Fach ein. Trotz vielversprechender Initiativen in den letzten Dekaden (bspw. Oehmer-Pedrazzi et al. 2022b) existieren bislang zu wenige Infrastrukturen für die Inhaltsanalyse: Es gibt kaum etablierte Unterstützungsangebote für die Untersuchungsplanung, Stichprobenziehung, Datenerhebung, -aufbereitung und -analyse sowie für die Archivierung und Vorbereitung der Nachnutzung von Datenkorpora qualitativer wie auch quantitativer Inhaltsanalysen. Strippel (2021) konstatiert daher mit Blick auf vorangehende Bemühungen im Fach: „Während Bedarfsanalysen und Lösungsvorschläge schnell erarbeitet und aufgeschrieben sind, ist deren Realisierung langwierig und ressourcenintensiv.“ (S. 142) Es scheint somit fraglich, ob die Kommunikations- und Medienforschung eine Forschungsinfrastruktur für die Inhaltsanalyse selbst aufbauen kann und sollte.

Unser Beitrag verfolgt vor diesem Hintergrund das Ziel zu explorieren, inwiefern die im Aufbau befindlichen NFDI-Strukturen die Bedarfe der Kommunikations- und Medienforschung an einer Forschungsinfrastruktur für die Inhaltsanalyse abdecken können.¹ Entlang des Forschungsdaten-Lebenszyklus arbeiten wir die Bedarfe heraus und erkunden auf dieser Grundlage, welche davon die NFDI-Konsortien im Hinblick auf die Inhaltsanalyse in ihrer gegenwärtigen Struktur decken können.

2 Forschungsinfrastrukturen für die Inhaltsanalyse: Angebote und Bedarfe

2.1 Die Inhaltsanalyse in der Kommunikations- und Medienforschung

Als empirische Sozialwissenschaft beschäftigt sich die Kommunikations- und Medienforschung mit den Bedingungen, Folgen und der Bedeutung medialer, öffentlicher sowie interpersonaler Kommunikation (vgl. DGPK 2008). Die Auseinandersetzung mit Medieninhalten ist dabei ein Bindeglied zwischen den Forschungsbereichen des Faches: Aus Perspektive der Kommunikatorforschung sind Medieninhalte Konsequenzen gesellschaftlicher, institutioneller, professioneller oder individueller Bedingungen der Kommunikator:innen. Aus Perspektive der Rezeptions- und Wirkungsforschung sind sie hingegen Faktoren, deren Folgen es zu ermitteln gilt. Die Beschäftigung mit Medieninhalten ist nötig, um die Wirkungskette von Kommunikator:innen zu Rezipient:innen nachzuzeichnen.

Die zentrale Methode der Medieninhaltsforschung ist die Inhaltsanalyse, nach Früh (2017) definiert als „empirische Methode zur systematischen, intersubjektiv nachvollziehbaren Beschreibung inhaltlicher und formaler Merkmale von Mitteilungen“ (S. 29). In der Kommunikationswissenschaft werden sowohl standardisierte

¹ Unsere Reise zur Klärung der Potenziale der NFDI für infrastrukturelle Bedarfe unseres Faches begann mit einem Workshop, den die Autor:innen gemeinsam mit Christian Strippel im Rahmen der Jahrestagung der DGPK 2022 ausgerichtet haben. Wir danken Christian Strippel, den Vertreter:innen der Konsortien und allen Workshopteilnehmer:innen für ihre Inputs und Anregungen.

(quantitative) als auch teilstandardisierte (qualitative) Verfahren eingesetzt. Analyziert wird vornehmlich Text. Es besteht aber die Forderung, Bildern und damit der Multimodalität der Medieninhalte Rechnung zu tragen (vgl. Geise und Rössler 2013, S. 308). In der Medienwissenschaft, in der die Analyse von Videomaterial traditionell stärker verankert ist, geschieht dies bereits, wobei hier vorrangig qualitative Verfahren und hermeneutische Zugänge Anwendung finden.

Insbesondere die standardisierte Medieninhaltsanalyse hat Wurzeln, die bis in die Anfänge der empirischen Kommunikationswissenschaft zurückreichen und eng mit der Entstehung der Disziplin verknüpft sind (vgl. Früh 2017, S. 11–12; Krippendorff 2018, S. 10–23; Merten 1995, S. 35–47). Das Fach begreift dementsprechend die Inhaltsanalyse als seine originäre Methode und hat darin einen hohen Professionalisierungsgrad erreicht. Beispielsweise stammen zentrale Methodenstandards sowie Überblicks- und Einführungswerke zur Methode in der Regel aus der Kommunikationswissenschaft selbst (z. B. Früh 2017; Krippendorff 2018; Neuendorf 2016; Oehmer-Pedrazzi et al. 2022a; Rössler 2017). Zunehmend werden Inhaltsanalysen aber auch in den angrenzenden Sozialwissenschaften und weiter entfernten Disziplinen eingesetzt. Insbesondere die Digitalisierung und die damit einhergehende hohe Verfügbarkeit von Daten sowie die Weiterentwicklung computergestützter Verfahren und Algorithmen befördern ein breiteres Interesse an der Methode, das über die Kommunikations- und Medienforschung hinausgeht.

2.2 Die Inhaltsanalyse im Forschungsdatenlebenszyklus

Aus dem typischen Forschungsprozess der Inhaltsanalyse ergeben sich Anforderungen an eine Forschungsinfrastruktur, die wir im Folgenden entlang des Forschungsdaten-Lebenszyklus (vgl. Higgins 2008; Rümpel 2011) herausarbeiten. Er beschreibt den Prozess der Entstehung, Verarbeitung und Verwertung von Forschungsdaten. In Bezug auf die Inhaltsanalyse differenzieren wir zwischen (1) Untersuchungsplanung, (2) Stichprobenziehung, (3) Datenerhebung, -aufbereitung und -analyse sowie (4) der Archivierung, dem Teilen und der Nachnutzung inhaltsanalytischer Materialien. In diesen Bereichen betrachten wir jeweils die spezifischen Anforderungen der Methode, stellen beispielhaft bisherige Initiativen für Forschungsinfrastrukturen vor und zeigen Fehlstellen auf, um so eine Grundlage für die Evaluation der NFDI-Konsortien zu erhalten.

2.2.1 Untersuchungsplanung

Abhängig vom spezifischen Forschungsinteresse steht zu Beginn der Inhaltsanalyse die Aufarbeitung relevanter theoretischer Ansätze und des Forschungsstandes. Sie mündet in der dimensionalen Analyse und schließlich in der Operationalisierung, also der Bildung von konkreten Kategorien und Codes. Im Rahmen der standardisierten Inhaltsanalyse soll dabei ein Codebuch entstehen, das detaillierte Kategorienbeschreibungen und Codieranweisungen sowie Ankerbeispiele enthält. Im Rahmen qualitativ-hermeneutischer Forschung erfolgt zusätzlich eine fortlaufende, induktive Strukturierung und Kategorienbildung (vgl. Fürst et al. 2016, S. 220). In der Planungsphase sind alle Ressourcen hilfreich, die etwa aus vorangegangenen Projekten

als Vorlage dienen können. Hierzu müssen Erhebungen, insbesondere Langzeitanalysen, sichtbar und Erhebungsinstrumente zugänglich sein. Bislang liegen für die Inhaltsanalyse kaum Standardmessinstrumente vor. Kategorienhandbücher sind genauso wenig üblich wie die dezidierte Validierung von Erhebungsinstrumenten. Erfreulicherweise bietet die an der Universität Zürich ins Leben gerufene „Database of Variables for Content Analysis“ (DOCA; vgl. Oehmer-Pedrazzi et al. 2022b) seit Kurzem einen Ausgangspunkt für die Recherche von Operationalisierungen. In der Open-Access-Datenbank werden inhaltsanalytische Variablen gesammelt, systematisiert und bewertet. Wichtig ist dabei, dass Standards bei der Erhebung nicht nur definiert, sondern auch validiert werden.

Darüber hinaus ist der Austausch über Planungsaspekte der Inhaltsanalyse wichtig – wie er etwa auf Tagungen der DGPK-Fachgruppe Methoden 2012 und 2022 stattgefunden hat. Die Projektplanung umfasst außerdem die Organisation der Codierung: die Planung von Budget, zeitlichem Ablauf und Personal. Hier wäre durchaus – analog zu Befragungs-Panels oder kommerziellen Crowdcoding-Anbietern wie Amazon Mechanical Turk – infrastrukturelle Unterstützung bei der Rekrutierung von Codierer:innen denkbar. Eine Forschungsinfrastruktur könnte zudem praktische Hinweise zur Durchführung vorhalten, etwa in Form von Beratungsmöglichkeiten und Literatur.

2.2.2 Stichprobenziehung

In einem nächsten Schritt muss festgelegt werden, welches Material den Gegenstand der Inhaltsanalyse darstellt. Bei der Analyse publizistischer Medien erfolgt die Stichprobenziehung zumeist über Zeitungsarchive, Mediatheken oder kommerzielle Datenbankanbieter. Dabei ist eine einfache Zufallsauswahl meist nicht möglich, angewendet wird daher oft ein geschichtetes Verfahren: Zunächst werden zeitlicher und räumlicher Geltungsbereich bestimmt, dann Mediengattungen und konkrete Angebote ausgewählt, schließlich inhaltliche Aufgreifkriterien definiert. Um das Sample weiter zu reduzieren, folgt oft die Ziehung einer Zufallsstichprobe (vgl. Rössler 2017, S. 54; Wagner 2022). Bei Social-Media-Daten ist das Sampling noch herausfordernder. Zwar sind diese Daten meist halböffentlich zugänglich, sie lassen sich aufgrund ihrer Flüchtigkeit aber im Nachhinein meist nur lückenhaft erfassen (vgl. Bachl 2018). Zudem unterliegt der Zugriff oftmals den technischen Restriktionen der Plattform (vgl. Possler et al. 2019, S. 4). Für Social-Media-Daten werden daher stärker bewusste statt zufallsbasierter Auswahlstrategien (vgl. von Berg und Jünger 2022) oder eine „associative query snowballing technique“ (Rogers 2019, S. 124) angewendet.

Nötig sind in allen Fällen verlässliche und übersichtlich strukturierte Basisdaten zu Medieninhalten, die die Auswahl der Angebote anleiten können. Vergleichbare Angebots- und Nutzungsdaten fehlen vor allem für Onlineinhalte. Die Daten von Interessenverbänden wie der Informationsgesellschaft zur Feststellung der Verbreitung von Werbeträgern (IVW) oder dem Bundesverband Digitalpublisher und Zeitungsverleger (BDZV) sind auf teilnehmende Organisationen beschränkt. Darüber hinaus braucht es Archivzugänge zum gewünschten Datenmaterial oder Schnittstellen (APIs) zu digitalen Onlineinhalten. Zahlreiche Initiativen befassen sich bereits

mit dieser Herausforderung, seien es Digitalisierungsvorhaben, die Zugangssicherheit zu historischen Quellen schaffen wollen (bspw. „Kommunikationsgeschichte digitalisieren“; vgl. Koenen et al. 2018), Inventare publizistischer Quellen wie Meteor (vgl. Balluff et al. 2022) zu europäischen Nachrichtenquellen, Mediacloud (vgl. Roberts et al. 2021), das GDELT Project (2023) für internationale Quellen oder auf digitale Kommunikation ausgerichtete Projekte wie das Social Media Observatory (vgl. Hans-Bredow-Institut 2023). Publizistische Inhalte werden auch (teils mit erheblichen Kosten) über kommerzielle Anbieter wie LexisNexis oder WiSo verfügbar gemacht.

Trotz dieser Initiativen stellen Zugang, Vollständigkeit und Datenqualität von Medieninhaltsarchiven ein zentrales Problem dar. Das ist zunächst auf die Vielzahl an Archiven und Angeboten zurückzuführen, die mit verschiedenen Verfahren Daten aufbereiten und in ihren Prozessen unterschiedlich transparent sind. Darüber hinaus sind zahlreiche rechtliche, ethische und finanzielle Fragen immer wieder neu zu klären. Lokale Forschungsinfrastrukturen an Universitäten (etwa Ethikkommissionen) können Forschende zumeist nur in geringem Umfang beratend unterstützen.

2.2.3 Datenerhebung, -aufbereitung und -analyse

Der wesentliche Schritt quantitativer wie qualitativer Inhaltsanalysen ist die Codierung. Die Organisation der Datenerhebung ist zumeist aufwendig. Im Vorfeld erfolgen in der Regel Qualitätssicherungsmaßnahmen, wie Codierer:innenschulungen und Reliabilitätstests. Dabei ist der Prozess nicht nur bei qualitativen Verfahren iterativ ausgerichtet, sondern kann auch in standardisierten Inhaltsanalysen mehrere Runden mit Anpassungen am Erhebungsinstrument erfordern. Forschungssoftware kommt dabei eine zentrale Rolle zu, etwa für die Codierung oder Berechnung von Reliabilitätstests. Neben kommerziellen Angeboten (bspw. *MAXQDA* für qualitative Analysen) sind in den letzten Jahren einige einschlägige R- und Python-Pakete entwickelt worden – insbesondere für automatisierte Inhaltsanalysen (vgl. Welbers et al. 2017, S. 246). Für Letztere spielt auch die Zugänglichkeit von Server-Infrastruktur, Diktionären und vortrainierten Modellen eine wichtige Rolle. Hier existieren bislang vor allem kommerzielle Angebote (bspw. Server-Infrastruktur von Cloud Computing Diensten oder vortrainierte Modelle auf der Plattform Hugging Face [2023]).

Sind die Daten erhoben, folgen Datenbereinigung und -validierung. Gegebenenfalls muss nachcodiert oder automatisiert neu klassifiziert werden. Das Vorgehen ist zu dokumentieren. Diese Dokumentationen sind im Sinne einer transparenten und nachvollziehbaren Wissenschaft ebenfalls relevante Forschungsdaten, die es zu archivieren gilt.

Den nächsten Schritt markiert die Analyse der erhobenen Daten. Dabei entstehen erneut Forschungsmaterialien, die für die Dokumentation, Reproduzierbarkeit und Nachnutzung relevant und zu archivieren sind. Dies betrifft einerseits Produkte der Nutzung etablierter Software (bspw. Code-Maps oder Skripte und Outputs von Statistikprogrammen). Andererseits kann in diesem Schritt die Weiter- und Neuentwicklung von Forschungssoftware erfolgen (bspw. als R-Pakete). Im Positionspapier der DGPK-Arbeitsgruppe Forschungssoftware wurden hierzu bereits Herausforderungen und Perspektiven aufgezeigt (vgl. Hepp et al. 2021). Insbesondere wird

darauf verwiesen, dass öffentliche Repositorien (z. B. GitHub, GitLab) bereits ein Grundgerüst für die Dokumentation von Forschungssoftware bereitstellen, die einerseits den Vorteil eines verbreiteten Quasi-Standards bieten, durch ihre kommerzielle Natur aber rechtlichen Veränderungen unterliegen können, sodass zusätzliche Archivierungsmöglichkeiten empfohlen wurden.

2.2.4 Archivierung, Teilen und Nachnutzung inhaltsanalytischer Materialien

Schließlich sind die Archivierung inhaltsanalytischer Materialien und die Ermöglichung des Teilens und der Nachnutzung dieser Forschungsdaten zentrale Forschungsschritte. Inhaltsanalytische Studien werden häufig fallbezogen entwickelt, was die Nachnutzung im Rahmen von Sekundär- und Meta-Analysen oder der universitären Lehre erheblich einschränken kann. Daher wäre eine institutionalisierte Forschungsinfrastruktur hilfreich, die einen Zugang zu den Forschungsdaten ermöglicht und eine Standardisierung vorgibt, insbesondere grundlegende Anforderungen an Metadaten, Datentypen und Datenformate. Möglichst offene, etablierte, standardisierte und anschlussfähige Formate sind wünschenswert (vgl. Hepp et al. 2021).

Im Sinne der intersubjektiven Nachvollziehbarkeit und Reproduzierbarkeit von Ergebnissen wäre es zudem wichtig, dass alle Bestände, das heißt Forschungs- und Erhebungsinstrumente, Medieninhalte als Rohdaten sowie die Datensätze der Codierung und Analysematerialien, gebündelt archiviert werden. Das ist besonders mit Blick auf die Rohdaten (d. h. Medieninhalte) schwierig, da rechtliche und ethische Fragen vor der Veröffentlichung zu klären sind. Diese reichen von Persönlichkeits- und Urheberrechten über die Anonymisierung bis zu spezifischen Vertrags- und Lizenzbedingungen kommerzieller Anbieter und Archive, die für die originäre Datenerhebung genutzt wurden. So decken Lizenzen häufig nur die projektbezogene Nutzung von Daten ab. Es braucht daher Datenzentren, welche die langfristige Ablage der gesamten Datenbestände ermöglichen und Forschende bei der Datenaufbereitung unterstützen sowie in rechtlichen, ethischen und finanziellen Fragen beraten. In Ermangelung einer solchen Infrastruktur erfolgt die Archivierung derzeit zumeist in Einzelprojekten in Repositorien wie dem Open Science Framework (vgl. Center for Open Science 2023).

Schließlich bedarf es gelungener fachinterner Kommunikation über Zugangswegen zu den Forschungsdaten, um deren Sichtbarkeit zu erhöhen und den Einbezug in zukünftige Projektplanungen zu ermöglichen. Bislang erfolgt diese Kommunikation primär im Rahmen einschlägiger Tagungen (etwa der DGPK-Fachgruppe Methoden).

2.3 Zwischenfazit: Bedarfe für die Inhaltsanalyse

Eine Forschungsinfrastruktur für die Inhaltsanalyse sollte demnach für die *Untersuchungsplanung* die Sichtbarkeit und Zugänglichkeit von Untersuchungsinstrumenten sicherstellen, Beratung bei der Studienkonzeption leisten und den fachinternen Austausch ermöglichen. Für den Schritt der *Stichprobenziehung* ist die Bereitstellung von Angebots- und Nutzungsdaten zu Medien und Plattformen nötig, die Verbesserung des Datenzugangs sowie Beratung zu rechtlichen und ethischen Fragen

Tab. 1 Bedarfe für die Inhaltsanalyse und Angebote zu deren Deckung innerhalb und außerhalb der NFDI

Bedarfe für die Inhaltsanalyse	Angebote außerhalb der NFDI (Beispiele)	Geplante Angebote aus der NFDI (Beispiele)
<i>Untersuchungsplanung</i>		
<p>(1) Zugänglichkeit von Untersuchungsinstrumenten</p> <p>(2) Validierung und Standardisierung inhaltsanalytischer Variablen</p> <p>(3) Austausch über Planungsaspekte</p> <p>(4) Infrastruktur für Rekrutierung und Beratung</p> <p><i>Stichprobenziehung und Datensammlung</i></p> <p>(1) Angebots- und Nutzungsdaten</p> <p>(2) Datenzugänge, u. a. zu Archiven und APIs</p> <p>(3) Beratung zu rechtlichen, ethischen, finanziellen Fragen</p>	<p>(1, 2) Database of Variables for Content Analysis</p> <p>(3) Panels auf Tagungen der Fachgesellschaften</p> <p>(4) Crowdsourcing Anbieter (bspw. MTurk)</p>	<p><i>Bislang keine</i></p>
<p>(1) Erhebungs- und Analysesoftware</p> <p>(2) Serverinfrastruktur/Rechenkapazität</p> <p>(3) Diktionäre, vortrainierte Modelle, Validierungsdaten</p> <p>(4) Aufbewahrung von Forschungssoftware (d. h. Übersicht und Sammlung von Angeboten)</p>	<p>(1) IVW, BDZV</p> <p>(1, 2) Meteor</p> <p>(2) Social Media Observatory, GDELT, Media Cloud; WISO, LexisNexis</p> <p>(3) Lokale Angebote der Universitäten (bspw. Ethikkommisstonen)</p>	<p>(2) Archivierung von Textdaten (<i>KonsortSWD</i>), Überblick über Zugänge zu unstrukturierten (Social-Media-)Daten (<i>BERD@NFDI</i>) bzw. Repositorien für Medientexten (<i>NFDI4DataScience</i>), Beratung zu Interoperabilität von Archiven/Formaten historischer Daten (<i>NFDI4Memory</i>)</p>
<i>Datenerhebung, -aufbereitung und -analyse</i>		
<p>(1) Erhebungs- und Analysesoftware</p> <p>(2) Serverinfrastruktur/Rechenkapazität</p> <p>(3) Diktionäre, vortrainierte Modelle, Validierungsdaten</p> <p>(4) Aufbewahrung von Forschungssoftware (d. h. Übersicht und Sammlung von Angeboten)</p>	<p>(1) Einschlägige R- und Python-Packages</p> <p>(2) Cloud-Computing-Angebote (bspw. AWS, Google Cloud, Microsoft Azure) oder von öffentlichen Einrichtungen (bspw. Gesellschaft für wissenschaftliche Datenverarbeitung)</p> <p>(3) Hugging Face</p> <p>(4) GitHub, GitLab, CRAN, PyPi</p>	<p>(1) Klassifikationssoftware für Textdaten bzw. unstrukturierte Daten (<i>KonsortSWD</i>, <i>Text+</i>, <i>BERD@NFDI</i>)</p> <p>Dienste zur Datenannotation, Kuratierung etc. (<i>NFDI4DataScience</i>)</p> <p>(2) Kapazitäten für High Performance Computing (<i>BERD@NFDI</i>)</p> <p>(3) Modelle für Textdaten bzw. unstrukturierte Daten (<i>BERD@NFDI</i>, <i>Text+</i>, <i>NFDI4DataScience</i>)</p> <p>(4) Tool Registry (<i>NFDI4Culture</i>)</p>

Tab. 1 (Fortsetzung)

Bedarfe für die Inhaltsanalyse	Angebote außerhalb der NFDI (Beispiele)	Geplante Angebote aus der NFDI (Beispiele)
<i>Archivierung und Nachnutzung</i>		
(1) Archivierung von Forschungsdaten	(1, 2) Repositorien der OSF	(1) Metadatenstandards für Textdaten bzw. unstrukturierte Daten (<i>BERD@NFDI</i> , <i>Text+</i> , <i>NFDI4Memory</i>), Metadatenformate für und Indizierung historischer Daten (<i>NFDI4Memory</i>)
(2) Archivierung von Erhebungsinstrumenten	(3) Lokale Angebote der Universitäten (bspw. Forschungsdatenmanagement-Abteilungen)	Archivierung von Textkorpora bzw. unstrukturierter Daten quantitativer und qualitativer Forschung (<i>KonsortSWD</i> , <i>BERD@NFDI</i>)
(3) Beratung zu rechtlichen, ethischen, finanziellen Fragen	(4) Tagungen der Fachgesellschaften	Entwicklung von Formaten für nicht-textliche Daten (<i>NFDI4Culture</i>)
(4) Kommunikationsräume		(2) Repositorien für Code, Software und Modelle (<i>NFDI4DataScience</i>)
		(3) Beratung zu Datenschutz, Anonymisierung von Textdaten bzw. unstrukturierten Daten (<i>KonsortSWD</i> , <i>Text+</i> , <i>BERD@NFDI</i>)
		Rechtsberatung zu nicht-textlichen, multimodalen Daten (<i>NFDI4Culture</i>), Rechtsberatung zu historischen Daten (<i>NFDI4Memory</i>)

Die Bezifferung der Angebote in den Spalten 2 und 3 gibt an, auf welche Bedarfe (Spalte 1) sie sich jeweils beziehen

und finanzielle Unterstützungsangebote. Für den Schritt der *Datenerhebung, -aufbereitung und -analyse* sollte die Infrastruktur Zugang zu Diktionären, Modellen und deren Validierungsdaten bieten und die Archivierung sowie Bereitstellung von Forschungssoftware ermöglichen. Für die *Archivierung und Nachnutzung* bedarf es schließlich Leistungen für die Standardisierung von Daten, Unterstützung und Beratung bei der Datenaufbereitung und Datenzentren, welche die Ablage der gesamten Forschungsmaterialien ermöglichen. Der Kreis schließt sich in der Anforderung, die Kommunikation über diese Infrastrukturen zu stärken, damit die vorhandenen Potenziale in die Planung künftiger Studien einfließen können.

3 Die Nationale Forschungsdateninfrastruktur

Einige Bedarfe werden durch Initiativen und Angebote auf nationaler und internationaler Ebene bereits gedeckt, das zeigen die oben angesprochenen Beispiele (siehe auch Tab. 1). Allerdings wurde auch deutlich, dass für etliche Bedarfe noch keine Infrastrukturen existieren. Eine mögliche Lösung stellt die im Aufbau befindliche NFDI dar. Die NFDI geht auf eine Empfehlung des Rats für Informationsinfrastrukturen (vgl. RfII 2016) zurück und wurde Ende 2018 von der Gemeinsamen Wissenschaftskonferenz (GWK) im Rahmen einer Bund-Länder-Vereinbarung beschlossen. Ziel der Vereinbarung ist nichts weniger als die „Steigerung der Effizienz des gesamten Wissenschaftssystems“ (GWK 2018, S. 1). Die vorerst bis Ende 2028 beschlossene Finanzierung erfolgt durch den Bund und die Länder. Angesichts des immensen Aufwands beim aktuellen Aufbau erscheint eine anschließende Förderung und damit eine dauerhafte Institutionalisierung wahrscheinlich. Die NFDI wird durch den Verein Nationale Forschungsdateninfrastruktur e.V. getragen. Neben den Gründungsmitgliedern (Bund und alle Bundesländer) sind hier 218 wissenschaftliche Einrichtungen vertreten (Stand Mai 2022), darunter alle größeren Universitäten, alle Akademien der Akademienunion, Einrichtungen wie die Deutsche Nationalbibliothek und die GESIS sowie politische Akteure wie das Bundesministerium für Bildung und Forschung. Auch einige Fachgesellschaften und -verbände sind Mitglieder des Vereins – die DGPK ist bislang kein direktes Mitglied.

Instrument der NFDI sind Konsortien (vgl. Kraft et al. 2022). Jedes Konsortium befasst sich mit der Forschungsdateninfrastruktur für eine bestimmte Fachcommunity und umfasst entsprechende wissenschaftliche Einrichtungen wie Universitäten, aber auch außeruniversitäre Institutionen. In drei Ausschreibungsrunden sind bislang 26 Konsortien und eine übergreifende Initiative eingerichtet worden. Die Konsortien sind das Ergebnis von umfangreichen, teilweise jahrelangen Koordinationsbemühungen in der Regel zwischen bestehenden Einrichtungen. Der Bereich Sozial- und Verhaltenswissenschaften wird von zwei Konsortien abgedeckt (*KonsortSWD* und *BERD@NFDI*), die Geisteswissenschaften von vier Konsortien (*NFDI4Culture*, *Text+*, *NFDI4Memory* und *NFDI4Objects*). Der überwiegende Teil der Konsortien richtet sich aber an Lebens- und Naturwissenschaftler:innen (siehe Tabelle A1 im ergänzenden Online-Anhang). Quer zur Konsortialstruktur liegen Sektionen, die sich mit übergreifenden Standards und Themen beschäftigen – insbesondere mit

ethischen und rechtlichen Aspekten, Metadaten und Ontologien, Ausbildungsfragen sowie mit der Verflechtung und Zusammenarbeit zwischen den Konsortien.

Gemeinsam ist den Konsortien und Sektionen, bei sehr unterschiedlichen fachlichen Traditionen, das Ziel: Es geht stets um die Frage, wie Datenbestände unter Berücksichtigung institutioneller, ethischer, rechtlicher und ganz besonders inhaltlicher Aspekte langfristig in hoher Qualität und nach den FAIR-Prinzipien² und CARE-Prinzipien³ verfügbar gemacht werden können. Dieses Ziel deckt sich wesentlich mit den Anforderungen an eine Infrastruktur für Inhaltsanalysen – es bleibt zu fragen, inwiefern die Konsortien der NFDI die konkreten Bedarfe bereits abdecken.

4 Exploration

Besonders vielversprechend für die Abdeckung der infrastrukturellen Bedarfe für die Inhaltsanalyse erscheinen Konsortien, die eine fachliche Nähe zur Kommunikations- und Medienforschung aufweisen und sich auf die Sammlung und Aufbereitung von einschlägigen Daten spezialisieren. Dazu zählen vornehmlich Textdaten und multimodale Inhalte. Sechs Konsortien entsprechen diesen Kriterien: Sie widmen sich implizit oder explizit Textdaten – als Ergebnis und Grundlage qualitativer und quantitativer Sozialforschung (*KonsortSWD* und *BERD@NFDI*) sowie hermeneutischer Analysen (*Text+*), als historische Artefakte (*NFDI4Memory*) oder als Teil von Kulturgütern (*NFDI4Culture*) sowie als Grundlage maschineller Lernverfahren (*NFDI4DataScience*). Zudem richten die ersten fünf sich explizit an Geistes- und Sozialwissenschaftler:innen und sind damit disziplinar einschlägig. Im Folgenden stellen wir diese sechs Konsortien näher vor und fragen, inwiefern sie die Erfordernisse für Inhaltsanalysen abdecken können. Da viele der Konsortien sich noch im Aufbau befinden, kann diese Exploration nicht dem Anspruch auf Vollständigkeit gerecht werden. Vielmehr präsentieren wir solche Angebote der Konsortien, die (a) besonders vielversprechend für die Inhaltsanalyse scheinen und (b) einen Schwerpunkt der Konsortien darstellen, sodass die konkrete Umsetzung bereits gut abschätzbar ist.⁴

² Diese Prinzipien sind auch für Inhaltsanalysen sinnvoll, wenngleich bislang im Fach oft vernachlässigt: *findable* (dauerhaft identifizierbar und durchsuchbar), *accessible* (durch standardisierte und offene Kommunikationsprotokolle abrufbar), *interoperable* (durch projektübergreifende Sprachen formalisiert) und *reusable* (entsprechend disziplinärer Standards erschlossen und unter freien Lizenzen nutzbar); siehe Wilkinson et al. (2016).

³ Die CARE-Prinzipien ergänzen die FAIR-Prinzipien um rechtlich-ethische Aspekte in Bezug auf die Souveränität der betroffenen Menschen. Sie gehen auf die „Global Indigenous Data Alliance“ zurück und setzen sich aus *collective benefit*, *authority to control*, *responsibility* und *ethics* zusammen (vgl. GIDA 2019).

⁴ Wir danken den Kolleg:innen aus den Konsortien für die Bereitstellung der Konsortialanträge und für hilfreiche Rückmeldungen zu den Zusammenfassungen.

4.1 KonsortSWD

KonsortSWD baut auf die langjährigen Bemühungen des Rates für Sozial- und Wirtschaftsdaten (RatSWD) zur Optimierung der Forschungsdateninfrastruktur in den Sozial-, Bildungs-, Verhaltens- und Wirtschaftswissenschaften auf. Der RatSWD hat in den vergangenen Jahren ein Netzwerk aus 39 Forschungsdatenzentren (FDZ) akkreditiert. Diese Datenzentren kuratieren und archivieren wissenschaftlich relevante Daten und ermöglichen Forscher:innen Zugang unter Einhaltung des Datenschutzes. Die Zentren sind etwa beim statistischen Bundesamt, aber auch der GESIS angesiedelt (bspw. „FDZ Allbus“, vgl. RatSWD 2022, S. 54). Ziel des KonsortSWD ist es, die bestehende Infrastruktur zu stärken und zu erweitern (vgl. Adena et al. 2019, S. 1).

Die DGPK unterstützte die Antragsstellung des Konsortiums mit einem „Letter of Support“, da es hochgradig einschlägig ist. Das Konsortium will explizit zur Verbesserung der Infrastruktur für die Analyse von Texten beitragen (vgl. Adena et al. 2019, S. 82) und deckt mehrere Bedarfe für die Inhaltsanalyse ab: Die *Stichprobenziehung* kann durch bessere Auffindbarkeit von Datensätzen profitieren. In der *Datenerhebung* können ggf. technische Dienste des Konsortiums genutzt werden. So ist etwa die Etablierung von Tools zur Erkennung von Entitäten in Texten (d.h. Personen, Orte, Objekte) geplant, um Textdaten mit anderen Datentypen zu verbinden (vgl. Blätte et al. 2022, S. 53). Die Stärken des Konsortiums liegen aber in der *Archivierung*. Die FDZ-Infrastruktur konzentriert sich bislang auf Umfragedaten und amtliche Statistiken, bietet aber bereits Möglichkeiten, Textkorpora zu archivieren (vgl. Blätte et al. 2022, S. 51). KonsortSWD will diese ausbauen und Text-Daten nach den FAIR-Prinzipien zugänglich machen (vgl. Adena et al. 2019, S. 82–83). Zukünftig sollen zudem qualitative Daten inkludiert werden (vgl. Adena et al. 2019, S. 65–66; Hollstein et al. 2021, S. 18). Qualitativ Forschende sollen im gesamten Forschungsprozess unterstützt werden, etwa durch Beratung zum Datenschutz oder zu Anonymisierungstools (vgl. Hollstein et al. 2021, S. 19). Schließlich schreibt das Konsortium regelmäßige Mittel zur Aufbereitung von Datensätzen für die Sekundärnutzung aus.

4.2 NFDI4Culture

NFDI4Culture bearbeitet unter anderem Bedarfe der Medienwissenschaft, aber auch der Kunstgeschichte oder der Architekturforschung. Es wird von einem Ensemble an außeruniversitären Einrichtungen (v. a. der Akademie Mainz) und Universitäten getragen und ist mit Gedächtnisinstitutionen (Bibliotheken und Museen) verknüpft. Im Fokus des Konsortiums steht die Bewahrung des materiellen und immateriellen Kulturerbes – dazu zählen Kulturgüter, Praktiken und auch künstlerische Performances (vgl. Kammerer et al. 2021, S. 24). Ziele sind die nachhaltige und langfristige Zugänglichkeit zu Daten entsprechend der FAIR-Prinzipien sowie die Klärung von komplexen Rechtsfragen unter Berücksichtigung der CARE-Prinzipien. Dieser Schwerpunkt erklärt sich unter anderem daraus, dass an der Genese von Kulturgütern eine Vielzahl an Akteuren beteiligt ist – die Rechte an Texten liegen etwa bei anderen Personen als die Rechte an Bildern, und fallweise sind postkoloniale

oder transkulturelle Aspekte zu berücksichtigen. Entsprechend sollen Empfehlungen und Datenformate entwickelt sowie Beratungsangebote aufgebaut werden (vgl. Altenhöner et al. 2020, S. 16; Rack et al. 2020, S. 255–56).

Für kommunikationswissenschaftliche Inhaltsanalysen wären eine Rechtsberatung und die Informationsangebote des Konsortiums insbesondere dann einschlägig, wenn es um Bildanalysen und die Archivierung komplexer multimodaler Daten geht: „NFDI4Culture aims to create a user-centered, research-driven infrastructure for research data on non-textual material and immaterial cultural heritage.“ (Altenhöner et al. 2020, S. 2) Zu den erwartbaren Angeboten gehört wie in anderen Konsortien auch die Einrichtung einer Tool Registry (vgl. Altenhöner et al. 2020, S. 46). Hierüber könnte potenziell auch Forschungssoftware der Kommunikationswissenschaft veröffentlicht und recherchiert werden. Insofern stellen die Einrichtungen des Konsortiums perspektivisch in der *Planungsphase* von Inhaltsanalysen und bei der *Archivierung* nicht-textlicher Daten (und Tools) Ansprechpartner:innen und Informationsangebote bereit.

4.3 BERD@NFDI

Das Konsortium für „Business, Economic and Related Data“ wird maßgeblich von Universitäten und Einrichtungen im Bereich der Wirtschaftswissenschaften getragen. Mitglieder sind darüber hinaus sozialwissenschaftliche Einrichtungen wie die GESIS und der RatSWD – die Kommunikationswissenschaft wird bei den für BERD@NFDI relevanten sozialwissenschaftlichen Disziplinen explizit genannt (vgl. BERD@NFDI 2020, S. 5).

Ausgangspunkt des Konsortiums sind unstrukturierte Daten, die in informationstechnischen Umgebungen durch das Verhalten von Akteuren anfallen. Angesichts des Umfangs, der Vielfalt und der Unübersichtlichkeit unstrukturierter Datenquellen zielt das Konsortium darauf ab, mit einer kontinuierlichen Literaturanalyse Transparenz über Datenzugänge und Ansätze zur Datensammlung herzustellen. Zudem sollen Standards zur Datenaufbereitung vorangetrieben werden: Neben Verfahren zur Anonymisierung, Klassifizierung, Normalisierung und Qualitätsbeurteilung soll ein Metadaten-Standard für unstrukturierte Daten entwickelt werden (vgl. Gehrlein et al. 2022, S. 64). Das Konsortium plant außerdem eine Plattform für das Teilen von Daten, die Möglichkeit zur Speicherung, Evaluation und Veröffentlichung von Algorithmen und verfügt über Kapazitäten für High Performance Computing (vgl. BERD@NFDI 2020, o.S.).

Wie alle Konsortien verfolgt BERD@NFDI die Einbindung in die Fachgemeinschaften, nimmt dazu Bedarfsanalysen vor und plant Trainings- und Weiterbildungsprogramme. Diese widmen sich rechtlichen, ethischen und methodischen Aspekten sowie der Etablierung einer Wissenschaftskultur, die das Teilen von Daten wertschätzt. Mit Blick auf die Inhaltsanalyse sind die Themen des Konsortiums vor allem für die automatisierte *Erhebung* von Social-Media-Daten und die *Datenaufbereitung und -analyse* durch automatisierte Inhaltsanalysen relevant. Zudem scheint eine Nutzung von Metadatenschemata und Repositorien zur *Archivierung* inhaltsanalytischer Daten vielversprechend.

4.4 NFDI4DataScience

Datenwissenschaft ist ein interdisziplinäres Feld, das Konsortium benennt unter anderem die Anwendungsbereiche Sprachtechnologie sowie Lebens-, Informations- und Sozialwissenschaften. NFDI4DataScience strebt nicht nur eine Infrastruktur für Daten im engeren Sinne, sondern darüber hinaus für Code, Software, Modelle sowie (wissenschaftliche) Veröffentlichungen an. Wesentliches Ziel ist es, die Transparenz und Reproduzierbarkeit dieser Objekte (durch Research Knowledge Graphs, Registries und Repositorien) zu verbessern. Angedacht sind außerdem Dienste für die Erzeugung, Verarbeitung, Annotation und Kuratierung von Daten. Die Interoperabilität der digitalen Artefakte soll sichergestellt werden, indem die Ressourcen des Konsortiums in eigenen und bestehenden Infrastrukturen verfügbar gemacht werden (vgl. Schimmler et al. 2020, S. 6–7).

In Bezug auf die Inhaltsanalyse könnten diese Repositorien einen Ort darstellen, an dem Mediendaten, Codebücher, Datensätze und aber auch Modelle automatisierter Inhaltsanalysen abgelegt werden und für die *Untersuchungsplanung* und *Datenerhebung und -analyse* zugänglich sind. Die Ergebnisse manueller standardisierter Inhaltsanalysen stellen potenzielle Trainingsdaten für Verfahren des maschinellen Lernens dar. Umgekehrt werden in der Kommunikations- und Medienforschung maschinelle Verfahren, die Kernthema des Konsortiums sind, für die automatisierte Codierung eingesetzt. Das Fach könnte zudem von Weiterbildungsprogrammen profitieren.

4.5 Text+

Text+ versammelt hauptsächlich geisteswissenschaftliche Disziplinen, insbesondere die Linguistik, die Philologien und die Literaturwissenschaft. Das Konsortium ist stark mit außeruniversitären Einrichtungen wie der Deutschen Nationalbibliothek und den wissenschaftlichen Akademien (Berlin-Brandenburg, Nordrhein-Westfalen) vernetzt und stellt Verbindungen zu europäischen Initiativen (CLARIN ERIC, DARIAH ERIC) her (vgl. Hinrichs et al. 2022, S. 15).

Inhaltlich gruppiert sich das Konsortium Text+ um drei Datendomänen (siehe auch Trippel 2022, S. 4). Editionen umfassen zum Beispiel mittelalterliche Texte oder kritische Editionen der Werke u. a. von Leibniz, Marx, Humboldt oder Kant. Kollektionen umfassen Zusammenstellungen von literarischen Werken, Dissertationen und Zeitungen. Lexikalische Ressourcen sind beispielsweise Diktionäre und annotierte Korpora (z. B. GermaNet). Für diese Domänen sollen viele bestehende Datenzentren ausgebaut, weitere Bestände also integriert werden. Die Aufnahmekriterien umfassen unter anderem die Relevanz der Forschung für die Community, die Erweiterung von Diversität auch unter Berücksichtigung kleiner Projekte sowie das Befolgen von FAIR- und CARE-Prinzipien. Projekte können eine Mittelförderung beantragen, um ihre Daten für die bestehende Infrastruktur aufzubereiten.

Das Konsortium besticht durch eine hohe Expertise bei Metadaten und Standardisierung, etwa zur Erfassung von Textdaten im TEI-Format⁵. Ein Schwerpunkt liegt über alle Datendomänen hinweg auf der Etablierung von Interoperabilität und Normdaten über Einrichtungen wie die Deutsche Nationalbibliothek. Auch entsprechende rechtlich-ethische Beratungsangebote sind angedacht (vgl. Hinrichs et al. 2022, S. 31). Die *Archivierung* von Inhaltsanalysen könnte von Standards zum Metadaten-Crawling profitieren. Zu erwarten sind zudem Informationsangebote sowie Austauschformate zu Tools der (automatisierten) Textaufbereitung und -extraktion, die die *Untersuchungsplanung* und *Datenerhebung und -analyse* unterstützen können. Da Grundlagen von Inhaltsanalysen häufig Texte sind, könnte eine Ausrichtung an in den Geisteswissenschaften etablierten Standardformaten und die Nutzung persistenter Identifikatoren die *Nachnutzbarkeit* von Daten verbessern. Die Datenzentren scheinen bislang allerdings weniger auf urheberrechtlich oder datenschutzrechtlich herausfordernde Daten eingerichtet zu sein – für Anonymisierungstechniken wird etwa an das KonsortSWD verwiesen (vgl. Hinrichs et al. 2022, S. 29).

4.6 NFDI4Memory

NFDI4Memory beschäftigt sich vorrangig aus geschichtswissenschaftlicher Perspektive mit historischen Daten. Dabei werden aber zahlreiche Aspekte bearbeitet, die auch für die Kommunikationswissenschaft relevant sind. Dazu gehören Austauschformate und Beratungsangebote, um Kompetenzen und die *Nachnutzung* von Forschungsdaten zu stärken sowie die Interoperabilität von Archiven und Daten zu erhöhen. Der Aufbau einer nachhaltigen *Archivierungsstruktur* erfordert zudem die Auseinandersetzung mit der Datenqualität – das Konsortium beschäftigt sich mit entsprechenden Richtlinien. Der Schwerpunkt des Konsortiums liegt auf der besseren Vernetzung bestehender Archive. Ein konzertiertes Datenzentrum, das für die Kommunikationswissenschaft unmittelbar einschlägig wäre, ist (soweit ersichtlich) nicht Gegenstand des Arbeitsprogramms.

Neben dem *Archivieren und Teilen von Daten* spielt die Frage eine Rolle, wie Datenbestände mit Meta-Daten angereichert, indiziert und durchsuchbar gemacht werden können. In diesem Zusammenhang spricht das Konsortium automatische Textanalysen an (vgl. NFDI4Memory 2022, S. 71). Auch rechtliche und ethische Aspekte werden in einem Arbeitsschwerpunkt aufgearbeitet (Data Culture). Dabei werden urheberrechtliche und ethische Probleme historischer Daten in den Blick genommen, die in der Kommunikationswissenschaft für das breitere Set an fachspezifischen Forschungsdaten von Bedeutung sind. All diese Angebote beziehen sich auf historische Daten allgemein, könnten aber eine „Blaupause“ für Inhaltsanalydaten und deren *Archivierung und Nachnutzung* bieten.

⁵ Der TEI-Standard der Text Encoding Initiative (<https://tei-c.org/>) ist in den digitalen Geisteswissenschaften weit verbreitet. In der Kommunikationswissenschaft ist TEI möglicherweise deshalb kaum bekannt, weil Texte selten langfristig für die maschinelle Nachnutzung bereitgestellt werden.

4.7 Zwischenfazit: Angebote für die Inhaltsanalyse

Zusammenfassend lässt sich festhalten: Viele der Bedarfe werden in den evaluierten NFDI-Konsortien bereits abgedeckt. Jedoch gibt es kein Konsortium, bei dem Kommunikationswissenschaftler:innen zentrale, treibende Kräfte sind. Zudem erfährt die Inhaltsanalyse (abgesehen vom Stichwort „Text Mining“ und der Berücksichtigung von Texten oder multimodalen Inhalten als Datensorte) in den meisten Konsortien keine besondere Würdigung oder explizite Berücksichtigung. Es gibt zudem keinen Verbund, der alle beschriebenen Herausforderungen und Fehlstellen gleichzeitig abdeckt. Dies anzunehmen wäre aufgrund der Vielfalt der Methode wohl auch vermessen.

Die disziplinäre Verortung der Konsortien legt die Zuordnung zu den verschiedenen Verfahren der Inhaltsanalyse nahe. Naturgemäß werden bei *Text+*, *NFDI4Culture* und *NFDI4Memory* eher Daten und Prozesse abgedeckt, die weniger standardisierten Forschungszugängen entgegenkommen, während *KonsortSWD* einerseits die qualitative Sozialforschung in den Blick nimmt, andererseits auch Angebote für standardisierte Methoden bereithält und zunehmend auch textuelle Daten berücksichtigt. Automatisierte Verfahren lassen sich sowohl bei *BERD@NFDI* als auch bei *NFDI4DataScience* verorten. Die soeben aufgezeigte Zuordnung entlang des Standardisierungsgrades und des methodischen Zugangs wird der komplexen Problemstellung von Inhaltsanalysen aber nur teilweise gerecht. Einzelne Stärken mit Bezug auf die Herausforderungen im Forschungsdaten-Lebenszyklus der Inhaltsanalyse lassen sich in allen Konsortien identifizieren (siehe Tab. 1):

Mit Blick auf die *Archivierung und Nachnutzung* von Daten bieten oder planen *NFDI4Culture*, *BERD@NFDI*, *Text+*, oder *NFDI4Memory* Beratungsangebote zu rechtlichen und ethischen Fragen. Auch die von *KonsortSWD* geplante Beratung in Urheberrechts- und Datenschutzfragen ist für die Publikation und Datenveröffentlichung von Inhaltsanalysen nahezu immer relevant. Hinsichtlich Standardisierung und Interoperabilität von Daten sind u. a. *Text+*, *NFDI4Memory*, *KonsortSWD* und *NFDI4DataScience* tätig. Gerade Interoperabilität ist jederzeit wünschenswert, denn Forschungsergebnisse sollten gemeinsam mit Medienstichproben sowie Erhebungsinstrumenten und Codierungen auffindbar sein, um die Nutzbarkeit zu erhöhen. Anonymisierungsstrategien, wie sie von *KonsortSWD* und *BERD@NFDI* thematisiert werden, sind unabhängig vom methodischen Zugang wichtig. Unstrukturierte (Social-Media-)Daten werden nicht nur im Rahmen der von *BERD@NFDI* protegierten maschinellen Auswertungsverfahren analysiert, sondern auch im Rahmen manueller sowie qualitativer Analysen. Die *Datenerhebung, -aufbereitung und -analyse* profitiert von Diktionären und Modellen, wie sie von *Text+* oder *BERD@NFDI* angedacht sind, ebenso wie von Übersichten über Tools und Forschungssoftware im Rahmen von *NFDI4Culture* oder *NFDI4DataScience*. Für die *Stichprobenziehung* sind sowohl sich daraus entwickelnde bessere Möglichkeiten der Datennachnutzung als auch originäre Archivzugänge relevant, wie sie *KonsortSWD*, *NFDI4Culture* oder *NFDI4Memory* anstreben.

Gleichzeitig zeigen sich Fehlstellen in Bezug auf die besonderen Anforderungen der Kommunikations- und Medienforschung im Kontext inhaltsanalytischer Methoden: Im Schritt der *Untersuchungsplanung* sind dies die Sichtbarkeit und Zugäng-

lichkeit explizit inhaltsanalytischer Instrumente und Studien, die darauf bezogene Qualitätssicherung und der methodenspezifische Austausch, die Weiterentwicklung der Methode und die Weiterbildung. Gerade weil die Inhaltsanalyse eine auch für andere Disziplinen zunehmend wichtige Methode ist, könnte die Forschungsgemeinschaft von der spezifischen Expertise der Kommunikationswissenschaft profitieren. Für die *Stichprobenziehung* bleiben die Verfügbarkeit vergleichbarer Basisdaten zu Medien und deren Anbieter- und Nutzungscharakteristika ebenso eine Leerstelle wie die nach wie vor große Herausforderung, Archivzugänge zu klassischen Medien und insbesondere zu den flüchtigeren digitalen Kommunikationsinhalten verschiedener Plattformen und im Netz allgemein in einer Form zu realisieren, die rechtlichen und ethischen Anforderungen entspricht und finanziell tragfähig ist. Auch für die *Datenarchivierung und -nachnutzung* ist zu konstatieren, dass inhaltsanalytische Daten bei der Entwicklung von Metadatenstandards besser berücksichtigt werden könnten und umgekehrt Inhaltsanalysen sich an solchen Standards orientieren sollten. Auch die Frage geeigneter Datenzentren für eine zentrale oder dezentrale, gleichwohl einfach zugängliche und langfristige Archivierung inhaltsanalytischer Daten ist eine wichtige Aufgabe für das Fach.

5 Fazit und Ausblick

Die Inhaltsanalyse ist eine für die Kommunikationswissenschaft zentrale, identitätsstiftende Methode, die jedoch mit bestehenden Forschungsinfrastrukturen nur unzureichend abgedeckt wird. Gerade im Kontext der Forderung nach Open Science (z. B. Dienlin et al. 2021) besteht Handlungsbedarf, damit Transparenz, Reproduzierbarkeit und Nachnutzbarkeit von Inhaltsanalysen gesteigert werden können. Mit der neu entstehenden NFDI besteht nun ein Momentum, dies grundlegend zu ändern. Im vorliegenden Artikel haben wir deshalb zunächst die Schritte der Inhaltsanalyse entlang des Forschungsdaten-Lebenszyklus beschrieben und daraus resultierende Anforderungen für Forschungsinfrastrukturen abgeleitet. Im Anschluss haben wir diejenigen Konsortien der NFDI vorgestellt, die ein hohes Potenzial aufweisen, den Herausforderungen der Inhaltsanalyse zu begegnen.

Tab. 1 bilanziert abschließend die Angebote der NFDI-Konsortien und kontextualisiert diese mit bestehenden Forschungsinfrastrukturen für die Inhaltsanalyse (siehe hierzu auch Abschn. 2.2). Dabei zeigt sich, dass die Konsortien bisherige infrastrukturelle Leerstellen bei der Codierung und Aufbereitung sowie der Archivierung und Nachnutzung von inhaltsanalytischen Daten besonders gut decken können. Doch gerade die Punkte, die hochgradig spezifisch für die kommunikationswissenschaftliche Inhaltsanalyse sind, müssten erst noch in NFDI-Konsortien eingebracht werden. Dazu zählen die Unterstützung bei der Untersuchungsplanung, die Zugänglichkeit von vergleichbaren Angebots- und Nutzungsdaten sowie vollständige Archivzugänge für das Ziehen von Medieninhaltsstichproben und die Entwicklung von Standards für inhaltsanalytische Daten. Damit die NFDI-Angebote tatsächlich eine Lösung für die infrastrukturellen Herausforderungen der Inhaltsanalyse darstellen können, scheinen das Engagement des Faches und die gemeinsame Weiterentwicklung der Angebote unabdingbar. Die unseres Wissens bislang ausstehende operative Einbindung

der Kommunikationswissenschaft in die NFDI birgt die Gefahr, dass Wissenschaftler:innen aus unserem Fach nicht oder zu spät an der NFDI partizipieren und somit die disziplinäre Perspektive auf die Inhaltsanalyse keine Berücksichtigung findet.

Möglichkeiten zur Beteiligung an den Konsortien scheinen aufgrund ihrer Bedarfsorientierung und Dialogbereitschaft durchaus zu bestehen. Einige Konsortien wie *BERD@NFDI* erwähnen die Kommunikationswissenschaft sogar explizit als Zielgruppe, und mit anderen wie *KonsortSWD* bestehen bereits erste Austauschbeziehungen auf institutioneller und individueller Ebene. So wurde das Konsortium von der DGPK mit einem „Letter of Support“ bedacht; wie sich diese Unterstützung aber inhaltlich, in der Arbeit des Konsortiums, optimal niederschlagen sollte, muss diskutiert werden. Wichtig erscheint bei der Etablierung von Infrastrukturen, dass sie die spezifischen Anforderungen einer Community abdecken. Die Medienwissenschaften bringen sich bereits in *NFDI4Culture* ein; hier bestehen möglicherweise auch Anknüpfungspunkte für klassische Kommunikationswissenschaftler:innen.

Wir hoffen, dass die hier präsentierte Übersicht über relevante Angebote der NFDI-Konsortien eine fachpolitische Diskussion anregt, wie wir die infrastrukturellen Bedarfe für die Inhaltsanalyse befriedigen können. Konkret regen wir auf Basis unserer Darstellungen an, dass wir uns als Kommunikationswissenschaftler:innen stärker in die NFDI einbringen. Im Folgenden schlagen wir einige kurz- und mittelfristige Maßnahmen als erste Impulse für den fachpolitischen Austausch vor:

- Forschungsprojekte sollten prüfen, inwiefern die *Integration von Projektdaten* in die Infrastruktur eines Konsortiums möglich ist. Da die Konsortien weitgehend auf bekannten Datenzentren (bspw. GESIS) oder auf im Aufbau begriffenen Angeboten (bspw. Qualiservice) aufsetzen, können sie als Unterstützungsstruktur verstanden werden. Mittel für die Aufbereitung und Archivierung von Forschungsdaten können bei einigen Konsortien beantragt werden (bspw. KonsortSWD). Zudem können solche Mittel Teil von Drittmittelanträgen (bspw. bei der DFG) sein. Sollten Projekte bereits eigenständige Repositorien aufgebaut haben, schlagen wir vor zu prüfen, inwiefern diese als eigenes Datenzentrum langfristig in die NFDI integriert werden können.
- Für den Aufbau einer nachhaltigen Infrastruktur sind *standardisierte Austauschformate* für Daten unabdingbar. In einigen Disziplinen haben sich bereits Standards wie TEI für Textdaten oder RDF für die interoperable, systematische Beschreibung von Objekten herausgebildet. Wir können von der Erfahrung dieser Disziplinen profitieren. Um Wissen aufzubauen, sollten Beratungsangebote der Konsortien wahrgenommen werden. Die Erarbeitung von Richtlinien für unser Fach könnte in einer Arbeitsgruppe der Fachgesellschaften geleistet werden – etwa in der AG Forschungsinfrastrukturen. Die langfristige Etablierung dieser Richtlinien kann durch den Einbezug in die Lehre gelingen.
- Mittelfristig wäre zudem der Aufbau einer *Dateninfrastruktur für Medieninhalte* wünschenswert, die systematisch und möglichst vollständig das Material bereitstellt, das typischerweise in Inhaltsanalysen verwendet wird. Dazu würden Datenbanken mit journalistischer Berichterstattung, aber auch Archive von Social-Media-Daten gehören. Die NFDI stellt dafür eine föderierte Infrastruktur zur Verfügung. Denkbar wäre eine durch die Fachgesellschaften getragene, auf Dauer ge-

stellte Arbeitsgruppe, die ausgehend von einer Bestandsanalyse langfristige Strategien und Finanzierungsmodelle entwickelt.

- Um eine langfristige Berücksichtigung der Inhaltsanalyse in der NFDI zu erzielen, reicht die Wahrnehmung der bestehenden Angebote nicht aus. Mit Blick auf eine *institutionelle Verankerung* sind die Fachgesellschaften gefordert: Sie sollten darauf hinwirken, in den Gremien und auf Veranstaltungen der NFDI vertreten zu sein, das heißt Vertreter:innen benennen und die Kommunikation mit den Konsortien suchen. Eine Beteiligung wäre etwa auf Tagungen oder in Gremien und Arbeitsgruppen der Konsortien denkbar. Um die Bedürfnisse unseres Faches optimal einzubringen, sollte auch eine ergänzende empirische Erfassung der infrastrukturellen Bedarfe in Bezug auf die Inhaltsanalyse erfolgen.

Auch wenn die NFDI in der aktuellen Form nicht spezifisch auf die Inhaltsanalyse ausgerichtet ist, so bestehen doch viele Anknüpfungspunkte. Insbesondere die verteilte Infrastruktur ermöglicht es, mit kleineren Projekten erste Erfahrungen zu sammeln, praktikable Lösungen zu erarbeiten und durch fachpolitische Beteiligung langfristig an der NFDI mitzuwirken. Es liegt nun in der Hand der Kommunikations- und Medienforscher:innen die Initiative zu ergreifen, sich mit ihrer Perspektive in die NFDI einzubringen und somit für eine stärkere Verankerung der Inhaltsanalyse zu sorgen.

Zusatzmaterial online Zusätzliche Informationen sind in der Online-Version dieses Artikels (<https://doi.org/10.1007/s11616-023-00789-5>) enthalten.

Förderung Zur Erstellung dieses Artikels erhielten die Autor:innen keine Zuschüsse von öffentlichen, kommerziellen oder gemeinnützigen Einrichtungen.

- **Name der Ressource:** Nationale Forschungsdateninfrastruktur (NFDI)
- **Namen der Autor:innen:** Annett Heft, Jakob Jünger, Julia Niemann-Lenz, Daniel Possler
- **Link:** <https://www.nfdi.de/>
- **Kurze Beschreibung:** Die NFDI verfolgt das Ziel, wissenschaftliche Daten zu vernetzen und langfristig verfügbar zu machen. Das großangelegte Infrastrukturprojekt wurde Ende 2018 von der Gemeinsamen Wissenschaftskonferenz beschlossen und vertritt über 200 wissenschaftliche Einrichtungen in Deutschland. Die wissenschaftlichen Disziplinen, von den Naturwissenschaften über die Lebenswissenschaften bis zu den Geistes- und Sozialwissenschaften, werden mit 26 fachlich ausgerichteten Konsortien abgedeckt. Die Konsortien erarbeiten jeweils spezifische Ressourcen, Standards und Beratungsangebote.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Adena, M., Aßmann, C., Bambej, D., Blask, K., Blätte, A., Bosnjak, M., Buck, D., Bug, M., Busch, A., Fräßdorf, A., Fuß, D., Goebel, J., Hollstein, B., Jungbauer-Gans, M., Kern, D., Klas, C., Kretzer, S., Liebig, S., Mayer-Ahuja, N., Müller, B., Mutschke, P., Perry, A., Riphahn, R. T., Rittberger, M., Schiffner, D., Schneider, S., Schupp, J., Siegers, P., Singh, R. K., Solga, H., Tochtermann, K., Toepfer, R., Wolf, C., & Zindler, S. (2019). National Research Data Infrastructure (NFDI) consortium proposal. Consortium for the Social, Behavioural, Educational, and Economic Sciences (KonsortSWD) (Projektnr. 442494171). <https://zenodo.org/record/3968372/files/KonsortSWD.pdf>. Zugegriffen: 26. Apr. 2023.
- Altenhöner, R., Blümel, I., Boehm, F., Bove, J., Bicher, K., Bracht, C., Brand, O., Dieckmann, L., Effinger, M., Hagener, M., Hammes, A., Heller, L., Kailus, A., Kohle, H., Ludwig, J., Münzmay, A., Pittroff, S., Razum, M., Röwenstrunk, D., Sack, H., Simon, H., Schmidt, D., Schrade, T., Walzel, A.-V., & Wiermann, B. (2020). NFDI4Culture—Consortium for research data on material and immaterial cultural heritage. *Research Ideas and Outcomes*, 6, e57036. <https://doi.org/10.3897/rio.6.e57036>.
- Bachl, M. (2018). *An evaluation of retrospective Facebook content collection*. OSF preprints. <https://doi.org/10.31219/osf.io/6txge>.
- Balluff, P., Lind, F., Boomgaarden, H. G., & Waldherr, A. (2022). Mapping the European media landscape—Meteor, a curated and community-coded inventory of news sources. *European Journal of Communication*, 38(2), 181–194. <https://doi.org/10.1177/02673231221112006>.
- BERD@NFDI (2020). Storage & computing resources. <https://www.berd-nfdi.de/services/tools/speicher-und-rechenressourcen/>. Zugegriffen: 26. Apr. 2023.
- von Berg, M.-L., & Jünger, J. (2022). Stichprobenverfahren in der Online-Welt: Kriterien für die plattformübergreifende Auswahl von Akteuren und Inhalten. In J. Jünger, U. Gochermann, C. Peter & M. Bachl (Hrsg.), *Grenzen, Probleme und Lösungen bei der Stichprobenziehung* (S. 230–271). Köln: Halem.
- Blätte, A., Fräßdorf, A., Heuer, J.-O., Hoffstätter, U., Leonhardt, C., Menze, L., Müller, B., Mozygemba, K., Pattloch, D., Rakers, J., Reineke, S., Runge, T., Schlücker, F., Schmidt, T., Wenzig, K., & Wolf, C. (2022). Eine Dateninfrastruktur für die Gesellschaftswissenschaften. Unterstützung in der Arbeit mit Forschungsdaten durch KonsortSWD. *Zeitschrift für Bibliothekswesen und Bibliographie*, 69(1), 48–58. <https://doi.org/10.3196/1864295020691276>.
- Center for Open Science (2023). Open science framework. <https://osf.io/>. Zugegriffen: 26. Apr. 2023.
- DGPuK (2008). Kommunikation und Medien in der Gesellschaft: Leistungen und Perspektiven der Kommunikations- und Medienwissenschaft. Eckpunkte für das Selbstverständnis der Kommunikations- und Medienwissenschaft. <https://www.dgpuk.de/de/selbstverst%C3%A4ndnis-der-dgpuk.html>. Zugegriffen: 26. Apr. 2023.
- Dienlin, T., Johannes, N., Bowman, N. D., Masur, P. K., Engesser, S., Kümpel, A. S., Lukito, J., Bier, L. M., Zhang, R., Johnson, B. K., Huskey, R., Schneider, F. M., Breuer, J., Parry, D. A., Vermeulen, I., Fisher, J. T., Banks, J., Weber, R., Ellis, D. A., Smits, T., Ivory, J. D., Trepte, S., McEwan, B., Rinke, E. M., Neubaum, G., Winter, S., Carpenter, C. J., Krämer, N., Utz, S., Unkel, J., Wang, X., Davidson, B. I., Kim, N., Stevenson Won, A., Domahidi, E., Lewis, N. A., & de Vreese, C. (2021). An agenda for open science in communication. *Journal of Communication*, 71(1), 1–26. <https://doi.org/10.1093/joc/jqz052>.
- Früh, W. (2017). *Inhaltsanalyse* (9. Aufl.). Konstanz: UVK. <https://doi.org/10.36198/9783838547350>.
- Fürst, S., Jecker, C., & Schönhagen, P. (2016). Die qualitative Inhaltsanalyse in der Kommunikationswissenschaft. In S. Auerbeck-Lietz & M. Meyen (Hrsg.), *Handbuch nicht standardisierte Methoden in der Kommunikationswissenschaft* (S. 209–225). Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-01656-2_13.

- Gehrlein, S., Klein, A., Schumm, I., & Tochtermann, K. (2022). BERD@NFDI für unstrukturierte Daten in den Wirtschafts- und Sozialwissenschaften. *Zeitschrift für Bibliothekswesen und Bibliographie*, 69(1–2), 59–67. <https://doi.org/10.3196/1864295020691280>.
- Geise, S., & Rössler, P. (2013). Standardisierte Bildinhaltsanalyse. In W. Möhring & D. Schlütz (Hrsg.), *Handbuch standardisierte Erhebungsverfahren in der Kommunikationswissenschaft* (S. 307–326). Wiesbaden: Springer. <https://doi.org/10.1007/978-3-531-18776-1>.
- GIDA (The Global Indigenous Data Alliance) (2019). CARE principles for indigenous data governance. <https://www.gida-global.org/care>. Zugegriffen: 26. Apr. 2023.
- GWK (Gemeinsame Wissenschaftskonferenz) (2018). Bekanntmachung der Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI). Amtliche Veröffentlichung. *Bundesanzeiger*. <https://www.bundesanzeiger.de/pub/de/amtliche-veroeffentlichung/1>. Zugegriffen: 14. Dez. 2022.
- Hans-Bredow-Institut (2023). Social media observatory. <https://smo.leibniz-hbi.de>. Zugegriffen: 26. Apr. 2023.
- Hartl, N., Wössner, E., & Sure-Vetter, Y. (2021). Nationale Forschungsdateninfrastruktur (NFDI). *Informatik Spektrum*, 44(5), 370–373. <https://doi.org/10.1007/s00287-021-01392-6>.
- Hepp, A., Hohmann, F., Belli, A., Boczek, K., Haim, M., Heft, A., Jünger, J., Jürgens, P., Koenen, E., von Nordheim, G., Rinsdorf, L., Rothenberger, L., Schatto-Eckrodt, T., & Unkel, J. (2021). Forschungssoftware in der Kommunikations- und Medienwissenschaft: Stand, Herausforderungen und Perspektiven. <https://www.dgpuk.de/de/forschungssoftware.html>. Zugegriffen: 26. Apr. 2023.
- Higgins, S. (2008). The DCC curation lifecycle model. *International Journal of Digital Curation*, 3(1), 134–140. <https://doi.org/10.2218/ijdc.v3i1.48>.
- Hinrichs, E., Leinen, P., Geyken, A., Speer, A., & Stein, R. (2022). *Text+. Language- and text-based research data infrastructure*. Zenodo. <https://doi.org/10.5281/zenodo.6452002>.
- Hollstein, B., Müller, B., Siegers, P., & Wolf, C. (2021). KonsortSWD: Vom Netzwerk zur integrierten Dateninfrastruktur der Gesellschaftsforschung. *Bausteine Forschungsdatenmanagement*, 2, 10–22. <https://doi.org/10.17192/bfdm.2021.2.8330>.
- Hugging Face (2023). Hugging face. <https://huggingface.co>. Zugegriffen: 26. Apr. 2023.
- Kammerer, D., Schrade, T., & Stellmacher, M. (2021). NFDI4Culture. Konsortium für Forschungsdaten zu materiellen und immateriellen Kulturgütern. *Bausteine Forschungsdatenmanagement*, 2, 23–33. <https://doi.org/10.17192/bfdm.2021.2.8341>.
- Koenen, E., Schwarzenegger, C., Bolz, L., Gentzel, P., Kramp, L., Pentzold, C., & Sanko, C. (2018). Historische Kommunikations- und Medienforschung im digitalen Zeitalter. Ein Kollektivbeitrag der Initiative „Kommunikationsgeschichte digitalisieren“ zu Konturen, Problemen und Potentialen kommunikations- und medienhistorischer Forschung in digitalen Kontexten. *Medien & Zeit*, 33(2), 4–19.
- Kraft, S., Seitz-Moskaliuk, H., Sure-Vetter, Y., Wössner, E., Bohmer, N., Eufinger, J., Fluck, J., Koepler, O., Korbel, J., Müller, B., Pittroff, S., Martins Rodrigues, C., Schwetje, T., Von Suchodoletz, D., & Weber, J.S. (2022). Nationale Forschungsdateninfrastruktur (NFDI). In V. Heuveline & N. Bisheh (Hrsg.), *E-Science-Tage 2021. Share Your Research Data* (S. 374–383). Heidelberg: heiBOOKS. <https://doi.org/10.11588/heibooks.979.c13752>.
- Krippendorff, K. (2018). *Content analysis. An introduction to its methodology* (4. Aufl.). Thousand Oaks: SAGE.
- Merten, K. (1995). *Inhaltsanalyse. Einführung in Theorie, Methode und Praxis* (2. Aufl.). Wiesbaden: Springer.
- Neuendorf, K. A. (2016). *The content analysis guidebook* (2. Aufl.). Thousand Oaks: SAGE.
- NFDI4Memory (2022). Consortium for the historically oriented humanities. Consortium proposal for the National Research Data Infrastructure (NFDI). <https://opus4.kobv.de/opus4-h-da/frontdoor/index/index/docId/352>. Zugegriffen: 26. Apr. 2023.
- Oehmer-Pedrazzi, F., Kessler, S.H., Humprecht, E., Sommer, K., & Castro, L. (Hrsg.). (2022a). *Standardisierte Inhaltsanalyse in der Kommunikationswissenschaft – Standardized Content Analysis in Communication Research. Ein Handbuch – A Handbook*. Wiesbaden: Springer. <https://doi.org/10.1007/978-3-658-36179-2>.
- Oehmer-Pedrazzi, F., Kessler, S.H., Humprecht, E., Sommer, K., & Castro Herrero, L. (Hrsg.). (2022b). DOCA—Database of categories for content analysis. <https://www.hope.uzh.ch/doca>. Zugegriffen: 26. Apr. 2023.
- Possler, D., Bruns, S., & Niemann-Lenz, J. (2019). Data is the new oil—But how do we drill it? Pathways to access and acquire large data sets in communication science. *International Journal of Communication*, 13, 3895–3911.
- The GDELT Project (2023). The GDELT Project. <https://www.gdeltproject.org>. Zugegriffen: 26. Apr. 2023.

- Rack, F., Boehm, F., Pasdzierny, M., & Schmidt, D. (2020). NFDI4Culture: Forschungsdaten in den Kulturwissenschaften. In S. Schrör, G. Fischer, S. Beaucamp & K. Hondros (Hrsg.), *Tipping Points. Interdisziplinäre Zugänge zu neuen Fragen des Urheberrechts* (S. 253–274). Baden-Baden: Nomos. <https://doi.org/10.5771/9783748910664>.
- RatSWD (Rat für Sozial- und Wirtschaftsdaten) (2022). Tätigkeitsbericht 2020 der vom RatSWD akkreditierten Forschungsdatenzentren (FDZ). https://www.konsortswd.de/wp-content/uploads/RatSWD_Taetigkeitsbericht-FDZ-2020.pdf. Zugegriffen: 26. Apr. 2023.
- RfII (Rat für Informationsinfrastrukturen) (2016). Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland. <https://rfii.de/?p=1998>. Zugegriffen: 26. Apr. 2023.
- Roberts, H., Bhargava, R., Valiukas, L., Jen, D., Malik, M. M., Bishop, C. S., Ndulue, E. B., Dave, A., Clark, J., Eiting, B., Faris, R., Shah, A., Rubinovitz, J., Hope, A., D'Ignazio, C., Bermejo, F., Benkler, Y., & Zuckerman, E. (2021). Media cloud: massive open source collection of global news on the open web. *Proceedings of the International AAAI Conference on Web and Social Media*, 15(1), 1034–1045. <https://doi.org/10.1609/icwsm.v15i1.18127>.
- Rogers, R. (2019). *Doing digital methods*. London: SAGE.
- Rössler, P. (2017). *Inhaltsanalyse* (3. Aufl.). Konstanz: UVK. <https://doi.org/10.36198/9783838547060>.
- Rümpel, S. (2011). Der Lebenszyklus von Forschungsdaten. In S. Büttner, H.-C. Hobohm & L. Müller (Hrsg.), *Handbuch Forschungsdatenmanagement* (S. 25–34). Bad Honnef: Bock + Herchen.
- Schimmler, S., Möller, S., Rehm, G., Sack, H., Beyan, O., Lange-Bever, C., Dietze, S., Wagner, C., Hotz, L., Abedjan, Z., Seidel, R., Ackermann, M. R., Wagner, M., Nagel, W. E., Auer, S., Stocker, M., Hauswirth, M., Markl, V., Möller, S., Neumuth, T., Rebholz-Schuhmann, D., & Tochtermann, K. (2020). NFDI4DataScience—Letter of intent. https://www.dfg.de/download/pdf/foerderung/programme/nfdi/absichtserklaerungen_2020/2020_nfdi_4datascience.pdf. Zugegriffen: 26. Apr. 2023.
- Strippel, C. (2021). Forschungsinfrastrukturen für die Kommunikations- und Medienforschung im deutschsprachigen Raum. Initiativen, Bedarfe und Perspektiven. *Medien & Kommunikationswissenschaft*, 69(1), 136–157. <https://doi.org/10.5771/1615-634x-2021-1-136>.
- Trippel, T. (2022). Mit Text+ Forschungsdaten digital vernetzen – ein Fall für die Sprachwissenschaft? *Sprachreport*, 38(1), 1–7. https://doi.org/10.14618/sr-1-2022_trip.
- Wagner, A.-L. (2022). Die Publizistische Stichprobe als Auswahlverfahren lokaljournalistischer Medienangebote: Potenziale, Grenzen und Erweiterungen am Beispiel einer inhaltsanalytischen Studie. In J. Jünger, U. Gochermann, C. Peter & M. Bachl (Hrsg.), *Grenzen, Probleme und Lösungen bei der Stichprobenziehung* (S. 272–300). Köln: Halem.
- Welbers, K., Van Atteveldt, W., & Benoit, K. (2017). Text analysis in R. *Communication Methods and Measures*, 11(4), 245–265. <https://doi.org/10.1080/19312458.2017.1387238>.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, J. S., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Saik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mullingen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., & Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3, 1–9. <https://doi.org/10.1038/sdata.2016.18>.
- Wissenschaftsrat (2011). Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften. <https://www.wissenschaftsrat.de/download/archiv/10465-11.html>. Zugegriffen: 26. Apr. 2023.

Dr. Annett Heft ist Leiterin der Forschungsgruppe „Dynamiken der digitalen Mobilisierung“ am Weizenbaum-Institut für die vernetzte Gesellschaft und der Freien Universität Berlin.

Dr. Jakob Jünger ist Juniorprofessor für „Digital Media & Computational Methods“ am Institut für Kommunikationswissenschaft der Universität Münster.

Dr. Julia Niemann-Lenz ist Postdoctoral Researcher & Scientific Coordinator im Data Literacy Education DDLitLab der Universität Hamburg.

Dr. Daniel Possler ist Postdoctoral Researcher am Institut Mensch-Computer-Medien der Julius-Maximilians-Universität Würzburg.