Jungmann, Nils et al.

Working Paper

Eine Community-Datentreuhand für die gemeinsame Nutzung sensibler Daten in der Kommunikationswissenschaft

RatSWD Working Paper, No. 288

Provided in Cooperation with:
German Data Forum (RatSWD)

Suggested Citation: Jungmann, Nils et al. (2025) : Eine Community-Datentreuhand für die gemeinsame Nutzung sensibler Daten in der Kommunikationswissenschaft, RatSWD Working Paper, No. 288, Rat für Sozial- und Wirtschaftsdaten (RatSWD), Berlin, https://doi.org/10.17620/02671.97

This Version is available at:
https://hdl.handle.net/10419/322003

RatSWD Working Paper Series

RatSWD.
Rat für Sozial- und WirtschaftsDaten

288

Eine Community-Datentreuhand für die gemeinsame Nutzung sensibler Daten in der Kommunikationswissenschaft

Nils Jungmann, Pascal Siegers, Jan Philipp Rau, Moritz Fürneisen, Gregor Wiedemann, Heidi Schulze

Juli 2025

www.ratswd.de

RatSWD Working Papers

des Rates für Sozial- und Wirtschaftsdaten (RatSWD)

Die *RatSWD Working Papers*-Reihe startete Ende 2007. In dieser Online-Publikationsreihe werden konzeptionelle und historische Arbeiten, die sich mit der Gestaltung der statistischen Infrastruktur und der Forschungsinfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften beschäftigen, publiziert. Dies sind insbesondere Papiere zur Gestaltung der Amtlichen Statistik, der Ressortforschung und der akademisch getragenen Forschungsinfrastruktur sowie Beiträge, die Arbeit des RatSWD selbst betreffend. Auch Papiere, die sich auf die oben genannten Bereiche außerhalb Deutschlands und auf supranationale Aspekte beziehen, sind *besonders willkommen*.

RatSWD Working Papers sind nicht-exklusiv, d. h. einer Veröffentlichung an anderen Orten steht nichts im Wege. Alle Arbeiten können und sollen auch in fachlich, institutionell und örtlich spezialisierten Reihen erscheinen.

Die Inhalte der *RatSWD Working Papers* stellen ausdrücklich die Meinung der jeweiligen Autorinnen bzw. Autoren dar und nicht die des RatSWD. Die Zuwendungsgeber des RatSWD haben die Publikationen nicht beeinflusst.

Kontakt: Rat für Sozial- und Wirtschaftsdaten (RatSWD)
Geschäftsstelle | Am Friedrichshain 22 | 10407 Berlin | office@ratswd.de

Eine Community-Datentreuhand für die gemeinsame Nutzung sensibler Daten in der Kommunikationswissenschaft

Nils Jungmann[1*], Pascal Siegers[1], Jan Philipp Rau[2], Moritz Fürneisen[2], Gregor Wiedemann[2], Heidi Schulze[3]

Juli 2025

DOI: 10.17620/02671.97

1 GESIS - Leibniz-Institut für Sozialwissenschaften
2 Leibniz-Institut für Medienforschung | Hans-Bredow-Institut (HBI)
3 Ludwig-Maximilians-Universität München

***Kontakt:** Nils Jungmann | nils.jungmann@gesis.org

Abstract

Dieser Beitrag diskutiert das Konzept einer Community-Datentreuhand (CDT), das entwickelt wurde, um eine gemeinsame Erstellung und Nutzung sensitiver Daten in der Erforschung rechtsextremer Onlinekommunikation zu erleichtern. Die Besonderheit dieses Datentreuhandmodells ist, dass die Treuhand nicht zwischen einem Datengeber (als Monopolist) und Datennutzenden vermittelt, sondern die Nutzenden selbst in die Erstellung und Pflege des Datenbestandes eingebunden sind. Nützlich ist eine Community-Datentreuhand, wenn die Zusammenführung fragmentierter Datenbestände den wissenschaftlichen Wert der Daten erhöht. Wir zeigen am Beispiel der Forschung zur rechtsextremen Onlinekommunikation, wie eine Community-Datentreuhand aufgebaut sein kann, um die gemeinsame Datennutzung zu ermöglichen und die Datenqualität zu optimieren.

Zentraler Baustein akteursbasierter Ansätze in der politischen Onlinekommunikations-forschung sind Accountlisten einschlägiger Akteure und Netzwerke, die als Grundlage für automatisierte Datenerhebungen auf verschiedenen Online-Plattformen und Messenger-diensten genutzt werden. Die Erstellung dieser Listen ist sehr aufwändig. Durch die gemeinsame Nutzung dieser Listen können Forschende den Arbeitsaufwand erheblich verringern und die Qualität ihrer Daten verbessern. Allerdings ist das Teilen von Daten in sensiblen Bereichen aufgrund rechtlicher Unsicherheiten und begrenzter Anreize für einzelne Forschende nach wie vor die Ausnahme. Die CDT bietet Lösungen für diese Probleme an, indem sie eine Forschungsinfrastruktur einrichtet, die Accountlisten als gemeinschaftliche Ressource verwaltet, die von der Forschungsgemeinschaft nach selbst festgelegten Regeln genutzt werden können. Die Datentreuhand fußt auf dem Prinzip der Reziprozität, das heißt, Forschende, die für eigene Forschungsprojekte auf die Listen zugreifen, müssen Daten-prüfungen und ggf. Aktualisierungen vornehmen, wobei eine Online-Datenbank diesen Austausch und die gemeinsame Datenpflege technisch unterstützt. Dazu implementiert die CDT technische und organisatorische Maßnahmen für den Datenschutz und Datensicherheit und bietet eine Struktur, die (1) die gemeinsame Nutzung von Daten fördert, (2) Vertrauen und Rechtssicherheit schafft, (3) die Datenqualität verbessert und (4) die Sicherheit der Forschenden erhöht.

Keywords: Community-Datentreuhand, Forschungsinfrastruktur, Data Sharing, Datenqualität, Kommunikationswissenschaft

Inhalt

1. Einführung[1]

Die Digitalisierung als dritte industrielle Revolution hat alle Lebensbereiche durchdrungen. Große Online-Plattformen strukturieren Freizeit, Arbeit, Konsum, Mobilität und Kommunikation. Dabei entstehen riesige Datenmengen über das online-Verhalten der Menschen, die in digitalen Geschäftsmodellen monetarisiert werden, was Fragen nach Besitz und Kontrolle der individuellen Daten aufwirft. Die EU-Kommission hat deshalb Maßnahmen ergriffen, um die Kontrolle der Bürger über ihre Daten wiederzuerlangen. In der EU-Datenstrategie 2020 wurde dazu das Konzept des „Datentreuhänders" konturiert und im Data Governance Act von 2022 weiter entwickelt (European Commission, 2020). Datentreuhänder sollen als Intermediäre einen fairen Zugang zu persönlichen Daten ermöglichen und die Datensouveränität der Bürger stärken.

Auch die Bundesregierung hat in ihrer Datenstrategie von 2021 Datentreuhänder als geeignete Organisationsform für einen vertrauenswürdigen Austausch von Daten identifiziert. Das Ziel ist unter anderem, „datengetriebene Innovationen" zu fördern, was gleichermaßen der Entwicklung von Geschäftsmodellen wie der wissenschaftlichen Forschung dient. Das Problem liegt in erster Linie im Zugang zu Daten privater Akteure (insb. den großen Online-Plattformen) für den bislang kein Verfahren erfolgreich etabliert wurde. Der Rat für Informationsinfrastrukturen (RfII) unterstreicht die Ziele der Bundesregierung und fordert, Datentreuhänder auch und gerade für einen Zugang der Wissenschaft zu sensiblen Daten von öffentlichen und privaten Organisationen zu etablieren.

Die digitale Revolution hat auch die sozialwissenschaftliche Forschung erfasst und verändert die Praxis der Datenerhebung und -auswertung erheblich. Mit der Etablierung der Computational Social Sciences ist eine neue Disziplin entstanden, die sich den Methoden der Sammlung und Auswertung von digitalen Verhaltensdaten widmen. Mit einfachen Mitteln (Notebook, Python-distribution, Python-Auswertungspakete) lassen sich eigene Daten aus Online-Plattformen und Messengerdiensten extrahieren und für die eigene Forschung große Korpora aufbauen. KI-Unterstützt lassen sich auch komplexe Auswertungs- und Annotations-verfahren für unstrukturierte Daten von informierten Laien anwenden und große Datenmengen auswerten.

1 Dieser Beitrag ist im Verbundprojekt „Weiterentwicklung eines Community-Datentreuhandmodells für die Erforschung politischer Online-Kommunikation - Com-DTM" entstanden, das vom Bundesministerium für Forschung, Technologie und Raumfahrt gefördert wird (FKZ: 16DTM208A-B). Finanziert durch die Europäische Union – NextGenerationEU. Die geäußerten Ansichten und Meinungen sind ausschließlich die der Autor:innen und spiegeln nicht unbedingt die Ansichten der Europäischen Union oder der Europäischen Kommission wider. Weder die Europäische Union noch die Europäische Kommission können für sie verantwortlich gemacht werden.

Um den Zugang für die Wissenschaft auch zu sensiblen Daten zu organisieren, haben die meisten der etablierten Erhebungsprogramme (insbesondere die Umfrageprogramme) eigene Forschungsdatenzentren (FDZ) gegründet. Hier stehen die wissenschaftlichen Nutzenden der datengebenden Organisationen (als quasi-Monopolist) gegenüber. Insgesamt sind die existierenden FDZ in der Mehrzahl stark nach diesem Sender (Datengeber) – Empfänger (wissenschaftliche Nutzende) Schema (Blankertz et al., 2020) organisiert. Unbestreitbarer Vorteil dieses Modells ist, dass die Trägerorganisationen der FDZ die Kosten für die Datenbereitstellung ganz oder zum Großteil tragen.

Für eine fragmentierte Erhebungslandschaft, in der Forschende über die Instrumente verfügen, aus dem digitalen Datenuniversum passgenau eigene Daten zu sammeln, passt das für die großen Umfrageprogramme und amtlichen Statistiken etablierte Modell der FDZ nicht. Im Bereich der digitalen Daten, braucht nicht in erster Linie Strukturen, um große Datenbestände zentral zu sammeln und zur Verfügung zu stellen, sondern Angebote, die Forschende dabei unterstützen, die für die Beantwortung ihrer Forschungsfragen benötigten Daten aus dem digitalen „Datenuniversum" zu extrahieren (Breuer et al., 2023; Ohme et al., 2024).

Ein Beispiel für solche Ressourcen sind Accountlisten für Online-Plattformen und Messengerdienste. Akteursbasierte Zugänge der politischen Kommunikationsforschung beruhen darauf, für eine definierte Gruppe von Akteuren gezielt deren Onlineaktivitäten in sozialen Medien und Messenger Diensten zu erfassen. Ein Beispiel dafür ist die Datenbank öffentlicher Sprecher des Social Media Observatory am Hans-Bredow Institut (Schmidt et al., 2024) oder die Verzeichnisse von Kandidierenden zu Bundestags- oder Europawahlen (GLES, 2024; Sältzer et al., 2023). Anhand dieser Listen können Forschende mit den zur Verfügung stehenden Instrumenten gezielt Daten scrapen.

Im Fall von öffentlichen Sprechern bzw. Personen des öffentlichen Lebens gibt es eine Rechtsgrundlage für eine Veröffentlichung der Accountlisten. Sie werden damit zu einer Ressource, die allen Forschenden zur Verfügung steht und erleichtern so eigene Datenerhebungen. Problematisch ist das Teilen der Daten allerdings, wenn Accountlisten von natürlichen Personen erstellt werden, die unter die Schutzbestimmungen des Datenschutzrechtes fallen. Ein Beispiel dafür, ist die Erforschung rechtsextremer Onlinekommunikation (Rau et al., 2022). Um Daten über die Aktivitäten der Akteure untersuchen zu können, werden zum Teil plattformübergreifende Accountlisten erstellt. Diese Daten sind personenbezogen und – wegen des offensichtlichen Bezugs zu politischen Ideologien – auch als besondere Kategorien personenbezogener Daten zu werten. Solche Daten können nur geteilt werden, wenn eine Rechtsgrundlage dafür existiert und angemessene technisch-organisatorische Schutzmaßnahmen umgesetzt werden. Da dies für einzelne Forschungsprojekte einen hohen Aufwand bedeutet, führt dies im Ergebnis dazu, dass Accountlisten nicht geteilt werden. Können Forschungsprojekte also nicht auf existierende Accountverzeichnisse zurückgreifen, ist der größte Vorbereitungsaufwand für die Datenerhebung die Erstellung der Accountlisten

für die beforschte Zielpopulation. Zudem fallen in den Projekten unterschiedlicher Forschungsgruppen redundante Aufwände für die Datenerhebung an, obwohl die Listen in anderen Forschungskontexten bereits erstellt wurden.

Das Modell der Community-Datentreuhand (CDT) soll hier eine Lösung anbieten. Es löst einerseits die rechtlichen Fragen für die beteiligten Forschenden auf einer zentralen Ebene und bietet andererseits eine gemeinsame technische Infrastruktur für die Bearbeitung der Daten. Die Kernidee ist, dass die beteiligten Forschenden bzw. Forschungsgruppen die Daten sowohl nutzen als auch an deren Erhebung aktiv beteiligt sind. So löst sich der Gegensatz zwischen Datengebenden und Nutzenden auf. Stattdessen werden die Daten als gemeinsames Gut kooperativ „bewirtschaftet".

Dieses Arbeitspapier soll am Beispiel der AVERA Community-Datentreuhand für die Rechtsextremismusforschung das Konzept der Community-Datentreuhand exemplarisch skizzieren. Die von uns vorgeschlagenen Lösungen lassen sich aber im Grundsatz auch auf andere Anwendungsfälle übertragen, weil sie drei Herausforderungen adressieren, die in der ein oder anderen Form alle Community-Datentreuhandmodelle betreffen: ein Vertrauens-problem, ein Anreizproblem und ein Datenqualitätsproblem.

Im Folgenden werden wir zunächst die Idee einer Community-Datentreuhand in die aktuelle Diskussion über Datentreuhandmodelle für die wissenschaftliche Forschung einordnen. Die darauffolgenden Kapitel adressieren je die Lösungsansätze für die drei Herausforderungen bei der Etablierung von Datentreuhändern: 1) Gewährleistung der Einhaltung rechtlicher und ethischer Verpflichtungen zum Schutz von Versuchspersonen (Vertrauensproblem); 2) Schaffung von Anreizen für Forschende zur Teilnahme an der CDT (Anreizproblem); 3) Aufrechterhaltung der Qualität der in der CDT verwalteten Daten (Datenqualitätsproblem). Das fünfte Kapitel widmet sich einer Herausforderung, die vor allem für Rechtsextremismus-forschung relevant ist: der Gewährleistung der Sicherheit der Forschenden vor rechts-extremen Übergriffen. Abschließend geben wir einen Einblick in die künftigen Schritte für die Weiterentwicklung der AVERA CDT und skizzieren das Potenzial des CDT-Konzepts für andere Anwendungsfälle insbesondere für die Computational Social Science (CSS) und die Kommunikationswissenschaft, wenn diese auf die Nutzung sensibler digitaler Mediendaten angewiesen sind.

2. Theorie der Datentreuhänder

Das Konzept der „Datentreuhänder" bekommt wachsende Aufmerksamkeit, seit die Europäische Kommission und die Bundesregierung es als Bausteine in ihre jeweiligen Datenstrategien aufgenommen haben. Die EU-Kommission führt mit dem Data Governance Act neutrale Datenintermediäre ein, die ohne eigene wirtschaftliche Interessen an den Daten zwischen Dateninhaber:innen und Datennutzenden vermitteln sollen. Sie sollen transparente Prozesse etablieren und so Vertrauen zwischen den Beteiligten Akteuren herstellen.

Insgesamt strebt die EU in der Datenstrategie einen verbesserten Schutz der Verbraucherrechte und verbesserte Formen der Einwilligungsverwaltung an, insbesondere durch eine Verstärkung der Datenteilhabe von Bürgern (EU Data Governance Act). Die Bundesregierung konzipiert Datentreuhänder deutlich breiter. Sie sollen in den Bereichen etabliert werden, in denen das Datenteilen (noch) nicht funktioniert, obwohl ein signifikanter wissenschaftlicher oder wirtschaftlicher Mehrwert davon zu erwarten wäre. In der Datenstrategie der Bundesregierung wird deutlich, dass der Begriff Organisationsformen des Datenteilens abbilden soll und flexibel an die Bedarfe der beteiligten Akteure angepasst werden kann, solange einige Grundprinzipien angewandt werden (Die Bundesregierung, 2021). Darunter fallen auch Modelle der gemeinschaftlichen Datennutzung, wie sie zum Beispiel in ersten Datenkooperativen umgesetzt werden (Pawelke et al., 2020).

Die Bundesregierung wählt also einen funktionalen Treuhandbegriff, der den Schwerpunkt auf die zu lösenden Aufgaben legt. So findet keine Verengung auf einzelne Organisation- oder Rechtsformen statt (z.B. staatliche, gemeinnützige oder genossenschaftliche Organisationen). Für die Bundesregierung sollen Datentreuhandmodelle die Kooperation von Akteuren mit mehr oder weniger unterschiedlichen Interessenlagen sicherstellen. Sie sollen Vertrauen zwischen den Akteuren schaffen und in die Qualität der von der Treuhand bereitgestellten Daten. Datentreuhänder tragen also nicht nur zur Lösung von Datenschutzproblematiken bei. Sie benötigen selbst die Kompetenz, um Daten bearbeiten und bewerten zu können und unterstützen die Nutzung über passende Beratungsangebote. Der Datenzugang soll dabei transparent und regelgeleitet erfolgen, um systematische Diskriminierungen zu verhindern.

Datentreuhänder treten folglich als Intermediär zwischen Akteuren auf, die Kontrolle über Daten haben, und Akteuren, die diese Daten für eigene Zwecke nutzen möchten. Bei sensiblen Daten stellt sich insbesondere die Frage, wie die Regeln des Datenschutzes eingehalten werden können, um die Vertraulichkeit personenbezogener und personenbeziehbarer Daten sicherzustellen, damit eine Übermittlung der Daten möglich ist (Kühling et al., 2020). Aus dieser Perspektive ist die Pseudonymisierung und Anonymisierung von Daten die Kernaufgabe von Datentreuhändern, wenn die personenbezogenen Daten für den wissenschaftlichen Nutzungszweck nicht unbedingt benötigt werden (was in den allermeisten Nutzungsszenarien vorausgesetzt werden kann). Diese Beschreibung entspricht in etwa der Rolle der FDZ in der Bundesrepublik, die für die Forschung Zugangswege zu sensiblen Daten aufgebaut und in der Breite etabliert haben.

In anderen Kontexten steht die Neutralität des Datentreuhänders im Mittelpunkt definitorischer Ansätze, weil nur so ein fairer Zugang für alle Interessenten gesichert werden kann, während die Interessen des Datengebers gewahrt bleiben (Werling et al. 2023). Allerdings steht die Neutralitätsverpflichtung in einem Spannungsverhältnis zur benötigten Datenkompetenz für die Anonymisierung und Qualitätssicherung. Die Erfahrung der Forschungsdatenzentren hat gezeigt, dass vor allem die eigene wissenschaftliche Forschung

die Datenkompetenz stärkt und zur Entwicklung innovativer Datenprodukte befähigt (z. B. indem Bestände verschiedener Datenzentren verknüpft werden).

In der Praxis sind Datentreuhandmodelle (wenn man von den existierenden FDZ absieht) noch nicht weit verbreitet. Erste Studien zeigen, dass vor allem rechtliche Unsicherheiten das Datenteilen erschweren. Ein zweiter wichtiger Grund für die Zurückhaltung beim Aufbau von Datentreuhändern sind die fehlenden Anreize für die Datengeber. Sie haben keinen oder nur mittelbaren Gewinn davon, anderen einen Zugang zu ihren Daten zu geben, tragen aber in jedem Fall einen Teil der rechtlichen Restrisiken (Kreutzer et al., 2024). Schließlich muss auch ein Weg gefunden werden, die Kosten für den Betrieb der Datentreuhand zu decken (Arlinghaus et al., 2021). In den Wissenschaften (und allen voran in den Geistes- und Gesellschaftswissenschaften) verfügen Forschende nur selten über größere Budgets, um Daten zu kaufen oder den Zugang zu Daten zu finanzieren.

Datentreuhänder wurden bislang vor allem in der Theorie als Modell für die Organisation des Datenteilens diskutiert. Sie können konzeptionell Lösungen für rechtliche Hürden darstellen, aber aufgrund der Asymmetrie in der Datenkontrolle fehlen Anreize für die Akteure, die über Daten tatsächlich verfügen. Das gilt nicht nur in Situationen, in denen das oben beschriebene Sender → Empfänger Prinzip greift, sondern auch, wenn fragmentierte Datenbestände durch deren Zusammenführung einen besonderen Wert für die Forschung erhalten.

Diese Situation charakterisiert verschiedene Anwendungsfälle in der Forschung mit Social-Media-Daten. Wir schlagen deshalb eine besondere Ausgestaltung der Datentreuhand vor, in der die Forschungscommunity einen Bestand an Accountdaten gemeinsam erstellt und bewirtschaftet, die wir als Community-Datentreuhand (CDT) bezeichnen.

3. Das Konzept der Community-Datentreuhand

3.1. Auswahl des Anwendungsfeldes

Die digitale Rechtsextremismusforschung stellt einen geeigneten Anwendungsfall aus den Computational Communication Sciences (CCS) dar, weil hier mit gleichermaßen daten-schutzrechtlich und politisch sensiblen Daten gearbeitet wird. Dieser Bereich ist nicht nur aufgrund seiner gesellschaftlichen Auswirkungen von Bedeutung, sondern ist auch von großem öffentlichem Interesse, was für die Bestimmung der Rechtsgrundlage für die Verarbeitung personenbezogener Daten durch den Datentreuhänder bedeutsam ist. Angesichts der Verbindungen zu Themen wie Extremismus, politische Gewalt und Terrorismus gelten die in diesem Zusammenhang gesammelten Daten als hochsensibel und erfordern entsprechend gestaltete Schutzmaßnahmen. Darüber hinaus gibt es in Deutschland eine aktive Community, die sich auf den digitalen Rechtsextremismus konzentriert und Bedarf an effizienteren Verfahren für den Datenaustausch hat.

Ein typischer Ausgangspunkt für die Erforschung der digitalen extremen Rechten wie auch für verschiedene andere Bereiche der Kommunikationswissenschaft, die automatisierte computergestützte Methoden für die Datenerhebung einsetzen, ist die Zusammenstellung von Accountlisten einschlägiger Personen, Netzwerke oder Organisationen (Wiedemann et al., 2023). Diese Listen sind häufig integraler Bestandteil des Datenerhebungsprozesses. Ihre Erstellung stellt jedoch eine große Herausforderung dar (Jost et al., 2023). Die mit dem digitalen Aufstieg der extremen Rechten verbundenen Phänomene haben ein enormes Ausmaß und weisen häufig eine extreme Dynamik, verdeckte Verhaltensweisen und eine schnelle Entwicklung auf. Ähnlich wie andere vergleichbare Bewegungen kann die extreme Rechte Zehntausende relevanter Akteure umfassen, die über verschiedene Plattformen, darunter sowohl Mainstream- als auch Nischenplattformen, in einem Land wie dem Vereinigten Königreich, Frankreich oder Deutschland aktiv sind (Guhl et al., 2020; Miller-Idriss, 2020). Darüber hinaus unterliegen diese Akteure und ihre Konten aufgrund der diesen Bewegungen inhärenten Volatilität und des externen Drucks, wie z. B. Deplatforming-Bemühungen, ständigen Veränderungen (Rogers, 2020). Diese Dynamik trägt zu den anhaltenden Bedenken hinsichtlich einer potenziell schlechten Datenqualität bei (Lazer et al., 2020), die zu Mess- und Darstellungsfehlern führen kann (Sen et al., 2021). Um diesen Problemen zu begegnen und eine hohe Datenqualität im Laufe der Zeit aufrechtzuerhalten, müssen Forschende Listen dieser Entitäten, einschließlich ihrer jeweiligen Social-Media-Konten und Webseiten, in einem zeitintensiven Prozess erstellen und aktualisieren. Dies ist sehr ressourcenintensiv, so dass es eine Herausforderung ist, mit der Schnelllebigkeit ihres Forschungsgegenstandes mitzuhalten. Folglich sind diejenigen, die versuchen, auch nur eine Teilmenge des digitalen Rechtsextremismus zu erforschen, mit dieser Aufgabe oft überfordert. Der Austausch solcher Accountverzeichnisse kann deshalb die Aufwände der Forschenden entscheidend reduzieren, wenn eine institutionelle Struktur die Transaktionskosten senkt.

Die konkrete Ausgestaltung der CDT ist im Austausch mit Vertretern der Community erfolgt, wobei ein Co-Creation Ansatz verwendet wurde (Fdez-Arroyabe & Roye, 2017; Hohmann, 2021). Im Mittelpunkt dieser Co-Creation steht die aktive Einbindung der zukünftigen Nutzenden, die ihr Wissen und ihre Fähigkeiten in einem "act of collective creativity" (Sanders & Stappers, 2008) aktiv in die Innovation eines Produkts oder einer Dienstleistung einbringen (Piller et al., 2010). Die Zusammenarbeit ist für die Entwicklung nachhaltiger Forschungs-software in Bereichen wie der Medien- und Kommunikationswissenschaft immer wichtiger geworden und eignet sich daher besonders gut für die Entwicklung der CDT (Fecher et al., 2021). Im Laufe von zwei Jahren haben wir sechs Co-Creation Workshops organisiert, an denen mehr als 50 Forschende aus relevanten Bereichen teilnahmen. In diesen Workshops wurden Beiträge und Anforderungen aus der Gemeinschaft gesammelt, um die organisatorische und technische Entwicklung des Datentreuhandmodells und seiner Verwaltung zu unterstützen. Das Hauptziel dieses Prozesses war es, die Bedürfnisse der Gemeinschaft nach effizientem Datenaustausch mit rechtlichen und ethischen

Verpflichtungen in Einklang zu bringen, insbesondere im Hinblick auf den Schutz der Privatsphäre und der Sicherheit des Einzelnen.

Darüber hinaus haben wir zur Entwicklung eines geeigneten Rechtsrahmens für die CDT mit Rechtsexperten und Datenschutzbeauftragten zusammengearbeitet, um Datenschutz- und Risikobewertungen durchzuführen (vgl. Abschnitt 4). Dieser Prozess war notwendig, um eine Rechtsgrundlage für die Datenverarbeitung durch die CDT zu bestimmen und sicherzustellen, dass sie die Vorgaben der DSGVO eingehalten werden. Infolgedessen wurden umfassende organisatorische und technische Maßnahmen zum Schutz sensibler personenbezogener Daten, zur Risikominimierung und zur Verhinderung eines möglichen Datenmissbrauchs ausgearbeitet.

3.2. Grundzüge der AVERA Datentreuhand

In diesem Prozess haben wir das Konzept einer CDT entwickelt, um Lösungen für die in der Einleitung beschriebenen Herausforderungen zu finden und dadurch einen neuen, kollaborativen Ansatz für eine gemeinsame Nutzung sensibler Daten in der digitalen Rechtsextremismusforschung (REX-Forschung) einzuführen. Die CDT soll die effiziente, ethische und rechtskonforme gemeinsame Nutzung sensibler digitaler Forschungsdaten ermöglichen - insbesondere Listen von Accounts, die online rechtsextreme Inhalte erzeugen oder weiterverbreiten. Als gemeinschaftliche Forschungsinfrastruktur verringert die Datentreuhand den Aufwand für die projektspezifische Datenerfassung und verbessert die Qualität der Daten durch Peer-Reviews und kritische Bewertung unter den Forschenden. Gleichzeitig implementiert die CDT angemessene technisch-organisatorische Maßnahmen, um den Schutz der Privatsphäre des Einzelnen zu gewährleisten und den Missbrauch persönlicher Daten zu verhindern.

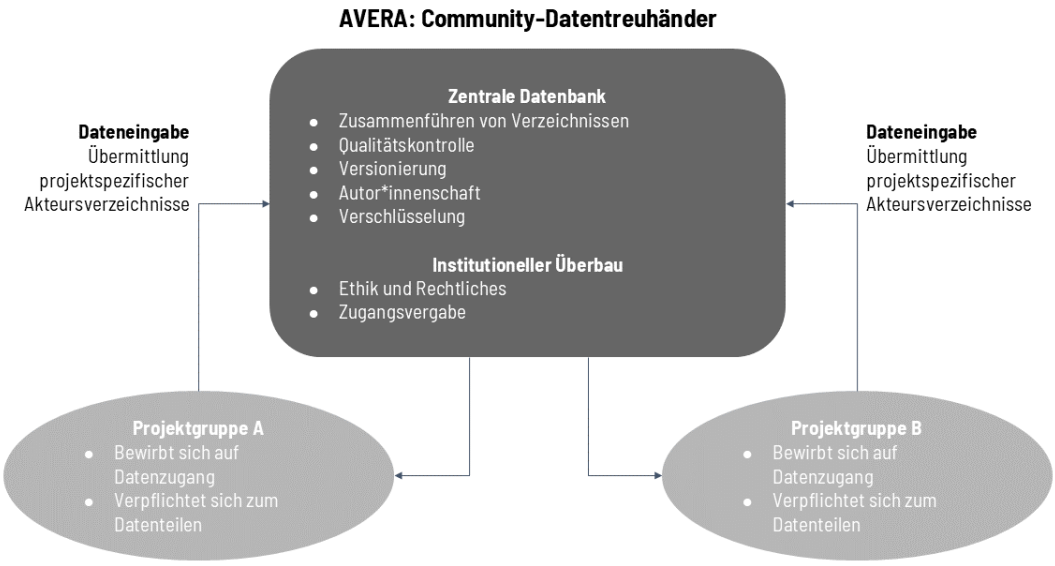

Abbildung 1: Das Konzept der Community-Datentreuhand

Wie in Abbildung 1 dargestellt, ist das Kernelement der Treuhand eine gemeinsam genutzte, dynamische Datenbank, die relevante Informationen über rechtsextrem kommunizierende Akteur*innen enthält, wie zum Beispiel Webseitendetails und Accountinformationen von Social-Media-Plattformen und Micro-blogging Diensten. Nicht gespeichert werden die von den Akteur*innen geposteten Inhalte. Diese Datenbank beruht auf einer eigens entwickelten Open-Source Software, die die gemeinsame Verwaltung, den Zugriff und die Aktualisierung über eine sichere Online-Plattform ermöglicht. Sie umfasst Funktionen wie die Verfolgung der Urheberschaft und Versionskontrolle für die Einträge in der Datenbank und implementiert einen Workflow für die kontinuierliche Qualitätssicherung der Einträge.

Über diese technische Ebene hinaus sorgt der institutionelle Rahmen der CDT dafür, dass die Datenverarbeitung ethischen und rechtlichen Standards gerecht wird, und setzt gleichzeitig Regeln für den Zugang der Forschenden durch. Die CDT wird von GESIS als Forschungsinstitut mit Infrastrukturauftrag mit Finanzierungen aus unterschiedlichen Drittmittelprojekten institutionell getragen. Als Träger der CDT stellt GESIS die Infrastruktur für die Datenbank und die Webapplikation bereit. Die personellen Ressourcen für den Betrieb werden aktuell aus Drittmittelprojekten mit einer Förderung des Bundesministeriums für Forschung, Technologie und Raumfahrt finanziert. GESIS ist zuständig für die Organisation (Abschluss der Nutzungsverträge, On-Boarding und Beratung der Nutzenden, Kuratierung des Datenbestandes, Bearbeitung von Auskunftsgesuchen). Die Regeln für die Nutzung wurden im Rahmen der Co-Creation Workshops mit der Community entwickelt. Das Regelwerk soll auch in Zukunft gemeinsam mit Vertreter*innen der Community weiterentwickelt werden. Die dafür notwendigen Strukturen müssen noch implementiert werden (vgl. Abschnitt 9). Als Träger der CDT trägt GESIS damit auch einen Teil der Risiken, die mit dem Betrieb der CDT verbunden sind, vor allen bezogen auf mögliche Beschwerden bezüglichen der Rechtsgrundlagen für die Verarbeitung personenbezogener Daten (siehe Abschnitt 4) oder Angriffe auf die Infrastruktur. Forschende können den Zugang zu den Daten bei GESIS beantragen und verpflichten sich im Gegenzug, ihre projektspezifischen Akteur*innenlisten und Aktualisierungen zur Datenbank beizusteuern (Reziprozitätsprinzip), um sicherzustellen, dass diese aktuell und umfassend bleibt. Diese Struktur ermöglicht es der CDT, die kontinuierliche Zusammenarbeit von Forschenden sowie die kontinuierliche Verbesserung und Pflege der Informationen in der Datenbank zu erleichtern.

Für die erfolgreiche Umsetzung der CDT müssen jedoch die genannten vier zentralen Herausforderungen bewältigt werden: 1) die Einhaltung der rechtlichen und ethischen Verpflichtungen zum Schutz der Untersuchungseinheiten; 2) die Schaffung von Anreizen für die Forschenden zur Nutzung des Angebotes der CDT; 3) die Sicherung einer hohen Datenqualität; 4) sowie der Schutz der Sicherheit der beteiligten Forschenden.

4. Einhaltung rechtlicher und ethischer Vorgaben zur Vertrauensbildung (Herausforderung 1)

4.1. Herausforderung Datenschutz

In Deutschland wird die Verarbeitung personenbezogener Daten durch die europäische Datenschutzgrundverordnung (DSGVO) und das Bundesdatenschutzgesetz (BDSG) geregelt. Gemäß Artikel 9 der DSGVO werden Informationen über politische Meinungen von Personen als besondere Kategorie personenbezogener Daten eingestuft, die besonders zu schützen sind. Die Verarbeitung solcher Daten erfordert eine passende Rechtsgrundlage und darauf abgestimmte technisch-organisatorische Schutzmaßnahmen.

Bei der Sammlung von Daten zu rechtsextremer Online-Kommunikation in den Sozialen Medien ist die Einholung einer informierten Einwilligung von den Akteuren in Kenntnis der Sachlage (Art. 6(1)(a) DSGVO) als Rechtsgrundlage für die Verarbeitung personenbezogener Accountinformationen unrealistisch. Extremistische Akteure werden einer solchen Datenverarbeitung zu Forschungszwecken nicht zustimmen. Daher ist eine alternative Rechtsgrundlage erforderlich, um die Datenverarbeitung zu ermöglichen.

Die sensible Natur der Daten in der Extremismusforschung erfordert eine strenge Rechtfertigung der Verarbeitung personenbezogener Daten der Untersuchungseinheiten und die Umsetzung geeigneter technischer und organisatorischer Maßnahmen zur Gewährleistung des Datenschutzes. Die Aufnahme in eine Datenbank potenziell extremistischer Akteure könnte für den Einzelnen erhebliche Risiken bergen - wie etwa Diffamierung und Stigmatisierung -, sollten Daten öffentlich werden. Dieses Risiko ist besonders problematisch bei Social-Media Konten von Privatpersonen, die kein öffentliches Profil und nur geringe Reichweite haben. Diese Personen sind sich möglicherweise nicht dessen bewusst, dass ihre Online-Kommunikation öffentlich ist, oder sie haben den spezifischen Kontext, in dem sie ursprünglich gepostet haben, verlassen.

Vor diesem Hintergrund sind Forschende oft unsicher, welche Rechtsgrundlage für die Übermittlung persönlicher Accountinformationen zu anderen Forschenden geeignet ist und wie sie in der Praxis umgesetzt wird. Wie Kreutzer et al. (2024) betonen, sind Rechts-unsicherheiten ein wesentlicher Grund für die Zurückhaltung von Forschenden bei der Weitergabe ihrer Daten (siehe auch Akdeniz et al., (2023)). Insbesondere besteht Unklarheit darüber, wann personenbezogene Daten ausreichend anonymisiert oder pseudonymisiert für eine Weitergabe sind. Im Falle von Accountlisten ist die Lage noch komplizierter, weil deren Anonymisierung die Nutzung für eigene Datensammlungen aus den Sozialen Medien unmöglich macht. Die Bestimmung einer Rechtsgrundlage und die Umsetzung geeigneter technischer und organisatorischer Maßnahmen zum Datenschutz ist für individuelle Forschende deshalb eine Hürde, deren Überwindung erheblichen Aufwand bedeutet (weil Verträge mit Datenschutzbeauftragten abgestimmt und zwischen den Kooperationspartnern

geschlossen werden müssen sowie die technisch-organisatorischen Schutzmaßnahmen umgesetzt werden müssen). Daher neigen Forschende bislang dazu, den "sicheren" Weg der Nichtweitergabe von Daten zu wählen, um Risiken zu vermeiden.

4.2. Die Rechtsgrundlage des CDT: Forschungsprivileg

Die CDT definiert eine Rechtsgrundlage für die Verarbeitung und Übermittlung der Accountinformationen in der Datenbank für alle beteiligten Forschenden. Daraus werden angemessene Schutzmaßnahmen für die Untersuchungseinheiten abgeleitet und die rechtlichen Unsicherheiten für die Forschenden verringert. Die rechtlichen Grundlagen werden also nicht mehr bilateral zwischen den Forschungsprojekten geklärt, sondern zentral für die Community (soweit sich Forschende an der Datentreuhand beteiligen).

Die DSGVO enthält verschiedene, gleichwertige Rechtsgrundlagen für die Verarbeitung personenbezogener Daten. Eine Alternative zur Einwilligung ist die Ausnahmeregelung für die wissenschaftliche Forschung, die in Art. 89 DSGVO vorgesehen ist. Diese Ausnahmeregelung erlaubt die Verarbeitung auch von besonderen Kategorien personenbezogener Daten, wenn sie zur Erreichung des Forschungsziels erforderlich ist und die Interessen der CDT als datenverarbeitende Instanz die Interessen der betroffenen Person erheblich überwiegen. Damit diese Ausnahmeregelung zur Anwendung kommt, muss die wissenschaftliche Forschung mehrere Kriterien erfüllen. Dazu gehören die Transparenz des Forschungsprozesses und der Forschungsergebnisse, die Unabhängigkeit der Forschung und ein Forschungszweck im öffentlichen Interesse (Buchner et al., 2021).

In Verbindung mit Art. 6(1)(e) DSGVO argumentieren wir, dass die Erforschung der Online-Aktivitäten rechtsextremer Akteur*innen im öffentlichen Interesse liegt, da sie zum Schutz der freiheitlich-demokratischen Grundordnung beiträgt. Gemäß Art. 27(1) BDSG ist dann eine Interessenabwägung zwischen den Forschungszielen und den Rechten der betroffenen Personen vorzunehmen.

Die Kriterien für ein öffentliches Interesse sind erfüllt, weil das Wachstum der digitalen rechtsextremen Onlineaktivitäten mit erheblichen politischen und gesellschaftlichen Risiken verbunden, darunter rechte Hassrede, Terrorismus, politische Gewalt und Bedrohungen für die verfassungsmäßige Ordnung sowie für ethnische und religiöse Minderheiten. Diese Risiken unterstreichen das öffentliche Interesse an der Erforschung rechtsextremer Onlinekommunikation und damit der Datenverarbeitung durch die CDT. Die Datentreuhand stellt für die Forschung einen entscheidenden Baustein für die Datenerhebung bereit, und stärkt so das Wissen über Erfolgsfaktoren der digitalen Rechten, erleichtert die Erforschung der ihm zugrundeliegenden Ursachen und unterstützt die Entwicklung von evidenzbasierten Strategien zur Prävention und Intervention.

4.3. Einschränkungen und Grenzen

Obwohl die Verarbeitung personenbezogener Daten rechtsextremer Akteur*innen grundsätzlich gerechtfertigt ist, bedeutet dies keine uneingeschränkte Billigung solcher Praktiken. Die von der DSGVO vorgeschriebenen Schutzmaßnahmen müssen strikt eingehalten werden, und die Ausnahmeregelung für die Forschung enthält - wie die DSGVO insgesamt - eine Reihe von Grundsätzen und Garantien, die für eine rechtmäßige Datenverarbeitung umgesetzt werden müssen.

Einer der schwierigsten Aspekte ist die Verarbeitung von Daten über Personen, die möglicherweise mit der digitalen extremen Rechten in Verbindung stehen, aber nicht den extremistischen Kern darstellen. Während die unmittelbare Bedrohung durch diese Gruppe geringer ist als durch eindeutig extremistische Personen, sind die potenziellen stigmatisierenden Auswirkungen der Aufnahme in eine solche Datenbank für diese Personen erheblich.

Dennoch ist die Erfassung der digitalen öffentlichen Kommunikation dieser Personen ein wichtiges Ziel der digitalen Rechtsextremismusforschung. Diese Notwendigkeit wird unter anderem durch die Herausforderung unterstrichen, Akteure in digitalen Räumen zu unterscheiden und entsprechend ihrer ideologischen Position zu klassifizieren. Viele dieser Räume zeichnen sich durch komplexe und vielschichtige ideologische Strukturen aus, in denen extremistische Ideologien mit anderen politischen Ausdrucksformen koexistieren. Diese Kommunikation ist anfällig für die Beeinflussung durch antidemokratische Ideologien, und ein erheblicher Teil ihrer Inhalte kann als extremistisch eingestuft werden (Guhl et al., 2020). Extremistische Akteure zielen häufig strategisch auf nicht-extremistische Räume und Personen ab, da diese als kritische Einfallstore für die digitale Verbreitung extremistischer Ideologien und Narrative dienen (Rothut et al., 2024). Um die Verbreitung extremistischer Ideologien und die potenzielle Radikalisierung von Akteuren effektiv erforschen und überwachen zu können, ist es unerlässlich, auch diejenigen einzubeziehen, die sich außerhalb, aber in unmittelbarer Nähe des extremistischen Kerns befinden.

Jeder Eingriff in die Privatsphäre des Einzelnen muss jedoch sorgfältig begründet werden, wobei die Schwere des Eingriffs in einem angemessenen Verhältnis zum möglichen Nutzen für die Forschenden und die Öffentlichkeit stehen muss. Deshalb werden zusätzliche Maßnahmen ergriffen, um die Schwere eines solchen Eingriffs zu begründen oder die Aufnahme in die Datenbank zu verhindern.

4.4. Technisch-organisatorische Schutzmaßnahmen

4.4.1. Organisatorische Schutzmaßnahmen

Die organisatorischen Maßnahmen zum Schutz der von der Datenverarbeitung im Rahmen des CDT betroffenen Personen lassen sich in zwei Hauptbereiche unterteilen: erstens die

Arten von Daten, die im CDT verarbeitet werden können, und zweitens die Bedingungen, unter denen der Zugang zu diesen Daten gewährt wird.

Einschränkungen bei den Merkmalen in der Datentreuhand

Zu den Arten von Daten, die im Rahmen des CDT verarbeitet werden dürfen, gehören: 1) relevante Webseiten- oder Plattform-Metadaten (z. B. URL, Accountname, Account-ID, Profil-ID); und 2) eine Rechtfertigung für die Aufnahme der Daten der Person in die Datenbank (z. B. die Einstufung der Person als rechtsextremistisch auf der Grundlage von Web- und Social Media Inhalten). Diese Rechtfertigung ergibt sich aus einer Abwägung der öffentlichen und individuellen Interessen. Die dritte Kategorie umfasst zusätzliche Informationen über die betreffende Person, z.B. ob sie von Forschenden als extremistisch oder populistisch eingestuft wird. Diese Informationen können sich mit der zweiten Kategorie überschneiden, die Informationen enthält, die die Aufnahme der Person in die Datenbank rechtfertigen. Die Begründung für die Aufnahme in die Datenbank ist jedoch eine rechtliche Verpflichtung, während die Einstufung in eine bestimmte Kategorie Gegenstand wissenschaftlicher Debatten und Ausarbeitungen sein kann.

Eine notwendige Voraussetzung für die Aufnahme in die Datenbank ist die Zugehörigkeit aller Personen zu einer bestimmten Gruppe relevanter ideologischer Akteur:innenkategorien. Dazu können Personen gehören, die dem rechtsextremen Spektrum zuzuordnen sind, Personen, die in ideologisch gemischte Szenen eingebunden sind, in denen rechtsextremes und extremistisches Gedankengut verbreitet wird, oder Personen, die dem Rechts-konservatismus zuzuordnen sind (weil sie ein Ziel extremistischer Agitation sind). Personen, die keinem dieser Spektren zugeordnet werden können, dürfen nicht in die Datenbank aufgenommen werden.

Als zusätzliche Schutzmaßnahme müssen die Accounts eine Mindestschwelle an diskursiver Relevanz erfüllen, um in die Datenbank aufgenommen zu werden. Konten, die als zum extremistischen Kern gehörend eingestuft werden, können aufgenommen werden, wenn sie mindestens 100 Follower auf mindestens einer Social-Media-Plattform haben. Für andere Personen ist ein Minimum von 500 Followern auf mindestens einer Plattform erforderlich.

Für die Verarbeitung der vom CDT erhaltenen Daten wurde eine Reihe von Regeln aufgestellt. Diese Daten dürfen nicht veröffentlicht oder an Dritte weitergegeben werden, und sie werden gelöscht, sobald die Grundsätze der guten wissenschaftlichen Praxis dies zulassen.

Einschränkungen im Datenzugang und den Nutzungszwecken

Der Zugang zum CDT unterliegt strengen Beschränkungen. Um Zugang zur Datenbank zu erhalten, müssen Forschende gegenüber der CDT nachweisen, dass ihr Forschungszweck in den Bereich der digitalen Rechtsextremismusforschung fällt und dass sie über die erforderlichen Qualifikationen zur Durchführung wissenschaftlicher Untersuchungen

verfügen. Forschende werden daher auf der Grundlage der folgenden drei Kriterien zur Nutzung der CDT zugelassen:

1) Forschungszweck: Die Nutzung von Kontenlisten zu nicht-wissenschaftlichen Zwecken ist ausdrücklich untersagt. Das Ziel ist ausschließlich die Gewinnung von Erkenntnissen über das gesellschaftliche Phänomen des Rechtsextremismus, die dem allgemeinen öffentlichen Interesse dienen. Um Zugang zum CDT zu erhalten, müssen Forschende eine Begründung für ihr Forschungsprojekt vorlegen und darlegen, wie es mit dem Zweck der CDT in Einklang steht.

2) Institutionelle Zugehörigkeit: Um Zugang zu den Daten zu erhalten, müssen die Forschenden als wissenschaftliche Mitarbeitende an einer öffentlich oder privat finanzierten Forschungseinrichtung wie einer Universität, einer Fachhochschule oder einem außer-universitären Forschungsinstitut beschäftigt sein. Diese Anforderung belegt das Vorliegen eines Forschungszwecks im öffentlichen Interesse, bestätigt ihre Qualifikation für die wissenschaftliche Arbeit und gewährleistet die Fähigkeit zur Einhaltung der grundlegenden Datenschutzstandards der DSGVO und der IT-Sicherheit.

3) Einschlägiger Hintergrund in der Rechtsextremismusforschung: Zugang zur Datenbank erhalten ausschließlich Personen, die einen einschlägigen Hintergrund in der digitalen Rechtsextremismusforschung nachweisen können. Auch dieses Kriterium stellt sicher, dass die Forschung im öffentlichen Interesse durchgeführt wird und, dass mit sensiblen Daten verantwortungsvoll umgegangen wird. Als Qualifikationsnachweis können einschlägige Veröffentlichungen, Promotionsprojekte, geförderte Forschungsinitiativen oder die Zugehörigkeit zu einem Forschungszentrum oder -institut mit Schwerpunkt auf dem betreffenden Thema dienen.

Wenn die Kriterien für den Datenzugang erfüllt und von der CDT formell bestätigt sind, wird ein Nutzungsvertrag geschlossen und ein Onboarding für die Nutzenden durchgeführt. Das Schulungsprogramm umfasst einen Überblick über die Regeln für den Umgang mit den im Rahmen des Projekts erhobenen Daten sowie eine praktische Einführung in die Verwendung des Dateneditors. Diese Einführung gewährleistet den verantwortungsvollen Umgang mit den Daten.

4.4.2. Technische Schutzmaßnahmen

Um die Datensicherheit zu gewährleisten und einen reibungslosen Betrieb des Treuhänders zu ermöglichen, werden verschiedene technische Maßnahmen ergriffen. Diese Maßnahmen lassen sich grob in zwei Gruppen einteilen: 1) Lösungen für die Bereitstellungsinfrastruktur und den Netzzugang und 2) Entscheidungen zum Softwaredesign.

Die meisten Maßnahmen der ersten Kategorie sind Standardmaßnahmen für Weban-wendungen; sie sind jedoch aufgrund der Sensibilität der von der CDT verwalteten Daten besonders wichtig. Zu diesen Maßnahmen gehören unter anderem die folgenden: (1) der

Treuhänder ist nur über verschlüsselte Kommunikation erreichbar; (2) die Bereitstellung ist durch Firewall-Regeln von anderen Diensten, die im selben Rechenzentrum gehostet werden, isoliert; (3) die Bereitstellungsinfrastruktur und Software-Abhängigkeiten werden auf dem neuesten Stand gehalten, um Schwachstellen zu minimieren; (4) jeder unregelmäßige Zugriff, d. h. nicht über die bereitgestellte Nutzendenoberfläche, wird überwacht; (5) es wird eine mehrstufige Backup-Strategie eingesetzt, um Datenverluste zu verhindern, und (6) der Zugriff auf die Datenbank wird nur Nutzenden gewährt, deren Identität durch eine Zwei-Faktor-Authentifizierung überprüft worden ist.

Die Entscheidungen zum Softwaredesign sind auf die besonderen Anforderungen der gemeinsamen Arbeit an sensiblen Daten zugeschnitten. Die Daten sind je nach Forschungszweck in verschiedene Berechtigungsgruppen mit unterschiedlichem Lese- und Schreibzugriff unterteilt.

Um eine neue Entität in die Datenbank aufzunehmen, müssen die Forschenden eine Begründung liefern. Diese Begründung muss die Kategorie der rechtsextremen Akteur*innen, zu der die Entität gehört, sowie die wissenschaftlichen Definitionen, die ihrer Klassifizierung zugrunde liegen, enthalten. Zusätzliche Begründungen können später ergänzt werden, um die Aufnahme der Entität zu untermauern.

Dem Grundsatz der Datenminimierung entsprechend werden Entitäten, für die über einen längeren Zeitraum keine Aktivität in den Sozialen Medien inaktiv waren oder veraltete Begründungen haben, gelöscht.

Auch die Konten der wissenschaftlichen Nutzenden müssen regelmäßig erneuert werden, um die Berechtigung zum Zugriff auf die Datenbank zu erhalten (siehe Abschnitt 4.4.1).

5. Anreize zur Überwindung des Kollektivgutproblems (Herausforderung 2)

5.1. Herausforderung Community Beteiligung

Trotz der potenziellen Vorteile der gemeinsamen Nutzung von Daten für die Forschungsgemeinschaft, wie zum Beispiel eine effizientere Ressourcennutzung und eine verbesserte Datenqualität, stellen nicht nur rechtliche Unsicherheiten eine Hürde für das Teilen von Daten dar. Für einzelne Forschende ist der Anreiz, Kolleg*innen Daten zur Verfügung zu stellen, begrenzt (Kreutzer et al., 2024).

Der Grund ist, dass Forschende zunächst keinen direkten Nutzen vom Teilen ihrer Daten haben. Stattdessen verschaffen sie ihren Kolleg*innen, mit denen sie um Karriereförderung und Forschungsfinanzierung konkurrieren, mitunter einen Wettbewerbsvorteil, weil diese von den Vorarbeiten profitieren. Dieses Szenario führt zu einem Kollektivgutproblem innerhalb der Forschungsgemeinschaft: Sobald jemand dem Datentreuhänder eigene Daten zur

Verfügung stellt, erhalten alle Teilnehmenden für ihre Zwecke Zugang zu diesen Daten, wodurch die Wettbewerbsposition des Datengebenden potenziell geschwächt wird.

Ein wirksames Modell für die gemeinsame Nutzung von Daten muss daher eine Lösung für das Kollektivgutproblem bieten, indem es die Forschenden davon überzeugt, dass der Datenaustausch fair ist. Es sollte sicherstellen, dass der individuelle Nutzen eines Beitrags zum CDT die Kosten bzw. Aufwände überwiegt, die mit der Bereitstellung ihrer persönlichen Accountlisten und der aktiven Beteiligung an der Pflege und Aktualisierung der Datenbank verbunden sind.

5.2. Lösungen des CDT

In Gesprächen mit der Forschungsgemeinschaft wurde deutlich, dass zwei Hauptaspekte adressiert werden müssen, um Anreize für die Teilnahme an der CDT zu schaffen. Erstens wurde betont, dass der größte Anreiz der CDT der Zugang zu seiner Datenbank ist, wodurch die Forschenden, die die kollektiv gesammelten Daten für ihre Studien nutzen können, einen erheblichen Forschungsvorteil erhalten. Dieser Vorteil kann genutzt werden, um weitere Beiträge durch ein *Reziprozitätssystem* zu fördern: Forschende erhalten nur dann Zugang zur CDT, wenn sie sich bereit erklären, die Ergänzungen und Korrekturen an den Daten in die Datenbank zur kollektiven Nutzung zu übermitteln. Ein solcher reziproker Ansatz trägt dazu bei, das Risiko eines Wettbewerbsverlusts für die teilnehmenden Forschenden zu verringern, da sie nicht nur ihre eigenen Daten weitergeben, sondern auch Zugang zu zusätzlichen Informationen von anderen Forschungsgruppen erhalten können. Eine potenzielle Herausforderung für dieses System könnte sich jedoch aus dem Ungleichgewicht zwischen sehr aktiven und weniger aktiven Teilnehmenden ergeben. Es steht zu befürchten, dass die aktiveren Teilnehmenden sich scheuen, ihren Wettbewerbsvorteil zu opfern.

Dies wird (teilweise) durch den zweiten Mechanismus der Sichtbarkeit aufgefangen: Zusätzlich zu der Diskussion über den Forschungsvorteil und die Reziprozität wurde vorgeschlagen, dass die individuellen Beiträge zum Datenbestand als wissenschaftliche Leitung sichtbar gemacht werden sollen, damit die Mitwirkenden für ihren Beitrag zum Aufbau der Datenbank akademische Anerkennung erhalten können (vgl. Abschnitt 5.2.1). Es wurden zwei verschiedene Dimensionen der Sichtbarkeit identifiziert: Erstens ist es notwendig, die Beiträge jedes Einzelnen zu würdigen, da dies der guten wissenschaftlichen Praxis für Veröffentlichungen und Datenerhebungen entspricht. Zweitens: Je größer der Bekanntheitsgrad und die Relevanz der CDT innerhalb der Forschungsgemeinschaft sind, desto wertvoller werden die einzelnen Beiträge. Die Annahme dahinter ist, dass ein Reputationstransfer von einem angesehenen Projekt auf die beteiligten Forschenden gibt, der künftige Karriereaussichten der Mitwirkenden verbessert.

5.2.1. Organisatorische Lösungen

Reziprozität als Bedingung für die Datennutzung

Alle Forschenden, die die Datenbank nutzen, verpflichten sich, zur Pflege der Datenbank beizutragen, indem sie neue Entitäten hinzufügen und/oder bestehende Entitäten aktualisieren, validieren oder löschen. Zu den Mindestverpflichtungen gehören die Validierung von zur Überprüfung gekennzeichneten Einträgen (vgl. den Abschnitt technische Lösungen weiter unten) und das Rückspielen von Aktualisierungen, die im Zuge eigener Datenerhebungsprozesse vorgenommen wurden. Der Umfang des Beitrags kann je nach dem Umfang der zu Forschungszwecken aus der Datenbank abgerufenen Daten variieren. Diese Verpflichtung zur Reziprozität ist im Code-of-Conduct festgeschrieben, den Forschende unterzeichnen müssen, bevor sie Zugang zu den Daten der CDT erhalten.

Eine unabdingbare Voraussetzung für die Anwendung des Reziprozitätsprinzips ist, dass eine Gruppe altruistischer Akteure eine signifikante Mindestanzahl von Konten zur Datenbank beisteuert, sodass andere Nutzende einen Mehrwert darin sehen, dem CDT beizutreten. Für die aktuelle Forschung wird der anfängliche Datenpool von der Forschungsgruppe, die die CDT entwickelt hat, in Zusammenarbeit mit Forschungsprojekten, die in den frühen Phasen des Entwicklungsprozesses beteiligt waren, bereitgestellt. Ihre Beiträge werden im Hinblick auf ihre Sichtbarkeit anerkannt (siehe unten).

Sichtbarkeit durch wissenschaftliche Reputation

Mit Blick auf die erste Dimension der Sichtbarkeit - die Hervorhebung individueller Beiträge zur Datenbank - wird die Datenbank so konzipiert, dass sie zitierfähig ist (vgl. technische Lösungen weiter unten). Um dies zu erreichen, werden Inhalt und Struktur der Datenbank in einer Datenpublikation dokumentiert, die in regelmäßigen Abständen aktualisiert und veröffentlicht wird. Die Beiträge der Forschenden können dann (kollektiv) zitiert werden, wenn die Datenbank für die Datenerhebung verwendet wird. Dies stellt einen erheblichen Anreiz dar, da Zitate in der Wissenschaft die wichtigste Währung sind, die die Relevanz und den Einfluss der Forschung widerspiegeln. Darüber hinaus kann die Namensnennung als Indikator für die Qualität und Integrität der Datenbank dienen, insbesondere wenn prominente Wissenschaftler*innen zur CDT beitragen. Dies deckt sich mit dem Argument bezüglich der Reputation der CDT. Die Möglichkeit, eine Person als Mitwirkenden namentlich zu nennen, kann als Nachweis der Glaubwürdigkeit dienen, was für eigene Forschungsvorschläge oder Bewerbungen von Vorteil sein kann.

Wenn alle Beitragenden als Autor*innen auf dem Datenpapier aufgeführt sind, stellt sich zwangsläufig die Frage der Autor*innenreihenfolge. Die weiter unten beschriebenen technischen Tracking- und Quantifizierungssysteme ermöglichen die Umsetzung eines meritokratischen Systems, bei dem die Autor*innen in absteigender Reihenfolge nach der Anzahl ihrer Beiträge aufgeführt werden. Obwohl auch ein egalitärer Ansatz - z. B. eine alphabetische oder zufällige Reihenfolge - in Betracht gezogen wurde, hat die Community

ein meritokratisches Prinzip für die Autorenreihenfolge bevorzugt, um die Wirksamkeit des Anreizes zu erhöhen und potenzielle Ungleichgewichte im Zusammenhang mit Forschungsvorteilen abzumildern (siehe oben). Dieses Prinzip würdigt die Leistung besonders bedeutender Beitragender, die andernfalls befürchten müssten, aufgrund von Unterschieden bei den Datenbeiträgen einen Wettbewerbsvorteil zu verlieren. Um die Sicherheit der Forschenden zu gewährleisten, haben die Mitwirkenden aber auch die Möglichkeit, sich gegen eine Nennung als Autor*in zu entscheiden (siehe auch Abschnitt 7.2).

Die zweite Dimension der Sichtbarkeit bezieht sich auf die akademische Reputation der CDT als wissenschaftlicher Dateninfrastruktur. In unserem Projekt wollen wir diese Reputation durch zwei Schlüsselstrategien aufbauen: 1) die Einbeziehung prominenter Mitglieder der Forschungscommunity als Nutzende (als Markenbotschafter sozusagen), ergänzt durch die Einrichtung eines Gremiums von Nutzenden in der Governance der Datentreuhand und 2) die Dokumentation des wissenschaftlichen Impacts der CDT durch das Tracking und die Sichtbarmachung von wissenschaftlichen Arbeiten, in denen die CDT zitiert wird. Da dieser Aspekt jedoch mit anderen Herausforderungen, einschließlich der Sicherheit der Forschenden (siehe Herausforderung 4), kollidieren kann, konzentriert sich das Projekt in erster Linie darauf, Sichtbarkeit innerhalb der akademischen Gemeinschaft und nicht in der breiten Öffentlichkeit aufzubauen. Zu den Ausnahmen gehören Maßnahmen zur Gewährleistung der allgemeinen Projekttransparenz, wie die Bereitstellung öffentlich zugänglicher Informationen auf Webseiten und anderen Plattformen, die eine öffentliche und rechtliche Prüfung des Projekts erleichtern.

5.2.2. Technische Lösungen für Reziprozität und Sichtbarkeit

Technische Implementation von Reziprozität

In den Co-Creation Workshops war es Konsens, dass Beiträge zur Datentreuhand auch schon beim Zugang zu den Daten geleistet werden sollen. Wenn Forschende Daten aus der Datenbank herunterladen und für ihre Forschung nutzen möchten, werden 5 Prozent der Einträge automatisch zur Überprüfung markiert. Dieser Auswahlprozess basiert auf der Identifizierung von Datenpunkten, die am längsten nicht mehr aktualisiert oder validiert wurden (vgl. Abschnitt 6: Herausforderung III: Sicherstellung der Datenqualität). Diese Verpflichtung sichert, dass alle Forschenden, die Zugang zu den Datenerhalten, an der Pflege des Bestandes mitwirken, auch wenn anschließend keine weiteren Aktualisierungen vorgenommen werden (z.B., weil ein Projekt scheitert).

Technische Implementation der Sichtbarkeit

Um sicherzustellen, dass die Beiträge der Forschenden sichtbar sind, ist ein technisches System erforderlich, um diese Beiträge nachzuverfolgen und zu quantifizieren. Daher ist die Datenbank so konzipiert, dass alle von den Forschenden vorgenommenen Ergänzungen protokolliert werden, einschließlich des Hinzufügens neuer Entitäten und der Aktualisierung, Validierung und Löschung bestehender Entitäten. Die Forschenden können eine "Update

Session" einleiten, in die weitere Forschende aufgenommen werden können. Die Beiträge, die während dieser Sitzung geleistet werden, werden nicht nur dem Hauptverfassenden der Aktualisierung zugeschrieben, sondern auch den anderen teilnehmenden Forschenden, wodurch die gemeinsamen Bemühungen, die der Aktualisierung zugrunde liegen, anerkannt werden (damit die Leistung ganzer Projektgruppen berücksichtigt wird). Dieses System erleichtert die Quantifizierung der von Einzelpersonen oder Gruppen geleisteten Arbeit.

6. Datenqualität gewährleisten (Herausforderung 3)

6.1. Herausforderung Datenqualität

Wie in der Einleitung dargelegt, stellt die Erstellung und Pflege von Accountlisten für Online-Phänomene wie die digitale extreme Rechte eine große Herausforderung dar. Diese Bewegungen zeichnen sich durch ihre hohe Volatilität aus, mit häufig wechselnden Akteur*innen und Plattformen. Das Teilen von Daten ist aufgrund ihrer Sensibilität begrenzt, was dazu führt, dass bestehende Listen oft veraltet sind und eine kritische Überprüfung durch andere Forschende fehlt. Diese Situation akzentuiert Probleme wie den fehlenden Konsens über die ideologische Kategorisierung von Akteur*innen und zentrale Definitionen im Forschungsfeld und der Schwierigkeit, Fehler oder Lücken in den Datenpools zu identifizieren.

Die CDT kann auch zur Verbesserung der Datenqualität beitragen (was aus Sicht der Forschenden eine der zentralen Motive für die Nutzung der Daten darstellen wird) indem sie zwei Anforderungen umsetzt: 1) Sie muss einen funktionalen und organisierten kollaborativen Prozess für die Erstellung und Pflege dieser Accountlisten erleichtern und sicherstellen; 2) Sie benötigt eine Datenbank, die in der Lage ist, alle notwendigen Informationen im Zusammenhang mit der Erstellung und Pflege dieser Accountlisten zu verarbeiten, zu speichern und bereitzustellen.

6.2. Lösungen des CDT

6.2.1. Organisatorische Lösungen

Die Trägerorganisation der CDT (d.h. GESIS) übernimmt Aufgaben der Datenkuratierung. Für die Verwaltung von Entitäten und zentralen Datenpunkten, wie z. B. Plattformdaten (ID, Accountname, Link usw.), bewertet die CDT, ob die von den Nutzenden (Forschende) eingereichten Änderungen bestätigt werden können. Wenn dies der Fall ist, werden die Änderungen nach Prüfung genehmigt (oder verworfen), und die Datenbank wird entsprechend aktualisiert. Wenn neue Forschende ihre Daten beisteuern und Überschneidungen zu den existierenden Daten bestehen, werden die beschriebenen Prozesse eingeleitet, die Gegenkontrollen ermöglichen.

Die Verpflichtung zur Überprüfung von 5 Prozent der heruntergeladenen Daten schafft deshalb nicht nur faire Bedingungen zwischen den Forschenden, die einen Anreiz zur Teilnahme bieten (siehe Herausforderung 2), sondern trägt auch zur laufenden Überprüfung der gesamten Datenbank bei. Dieser Überprüfungsmechanismus gewährleistet eine regelmäßige Validierung der Genauigkeit der Datenbank und trägt somit der hohen Volatilität des Forschungsbereichs Rechnung und verbessert die Gesamtqualität der Datenbank.

6.2.2. Technische Lösungen

Die Datenbank ist so konzipiert, dass sie Entitäten – Einzelpersonen, Kanäle oder Gruppen - speichert und verschiedene Arten von Informationen zu jeder Entität verknüpft (die von Entität zu Entität unterschiedlich sein können, z. B. die Plattformen, auf denen eine bestimmte Entität aktiv ist). Die Flexibilität der Datenbank ermöglicht das einfache Hinzufügen neuer Entitäten (relevante Personen oder Organisationen), Plattformen, Konten oder zusätzlicher Informationen. Das implementierte Versionskontrollsystem verfolgt diese Änderungen über die Zeit und stellt sicher, dass frühere Versionen der Datenbank zugänglich bleiben, um die Reproduzierbarkeit der Forschung zu ermöglichen. Die Datenbank berücksichtigt Mehrdeutigkeiten, wo dies notwendig ist (z. B. kann eine Entität mehr als eine ideologische Kategorisierung haben, was die Pluralität in der akademische Debatte widerspiegelt), und sorgt gleichzeitig für Eindeutigkeit, wo dies für die Nutzung der Daten in Webscraping Verfahren notwendig ist (z. B. nur eine Plattform-ID pro Plattformkonto; dies schließt jedoch nicht aus, dass es mehrere Plattformkonten mit mehreren IDs gibt).

Die Datenbank unterstützt den gleichzeitigen Zugriff und die Aktualisierung, um die Zusammenarbeit zu erleichtern. Datenaktualisierungen können manuell über die Software-Schnittstelle oder durch Massen-Uploads erfolgen. Eine Technologie zur Erkennung von Ähnlichkeiten ist integriert, um Duplikate zu erkennen, Aktualisierungen zu unterstützen und ein Aufblähen der Datenbank zu verhindern. Wenn Duplikate identifiziert werden, können sie zusammengeführt werden aber die CDT behält die Befugnis, diese Zusammenführungsanträge zu genehmigen oder abzulehnen, um fehlerhafte Aktualisierungen zu vermeiden. Jede Datenänderung wird protokolliert und dem jeweiligen Beitragenden zugeordnet, um die Verantwortlichkeit im Datenmanagementprozess zu gewährleisten.

Wie bereits unter Herausforderung 2 erwähnt, wurde ein Datenüberprüfungsmechanismus eingeführt: Wenn Forschende Daten aus der Datenbank herunterladen und für ihre Forschung verwenden, werden 5 Prozent der Daten automatisch zur Überprüfung markiert. Dafür werden die Datenpunkte identifiziert, die am längsten nicht mehr aktualisiert oder validiert wurden.

7. Schutz der Forschenden (Herausforderung 4)

7.1. Forschung im Spannungsfeld von Sichtbarkeit und Sicherheit

Eine vierte Herausforderung ist nicht direkt an das Konzept der Community-Datentreuhand gebunden, sondern ergab sich aus den Diskussionen in den von uns durchgeführten Co-Creation-Workshops. Sensible digitale Forschungsthemen wie der digitale Rechtsextremismus sind oft hochgradig politisiert und können Themen wie gewalttätigen Extremismus und Terrorismus umfassen, die einen direkten Bezug zu physischer Gewalt aufweisen. Forschende berichteten, dass sie sich aufgrund ihrer wissenschaftlichen Arbeit latent von rechtsextremen Netzwerken bedroht fühlen. In der Vergangenheit waren sowohl Forschende selbst als auch ihre Kolleg*innen Ziel von Hasskampagnen, insbesondere wenn ihre Arbeit in den traditionellen Medien oder auf Social-Media-Plattformen Aufmerksamkeit erhalten hat. Ähnliche Risiken sind auch in anderen politisierten Forschungsbereichen zu beobachten, z. B. in der Klimaforschung oder im öffentlichen Gesundheitswesen während der COVID-19-Pandemie, wo Forschende schwerwiegenden Beschimpfungen, einschließlich Online-Belästigungen und Todesdrohungen, ausgesetzt waren (Global Witness, 2023; Nogrady, 2021). In der Kommunikationswissenschaft sind Drohungen in Online-Umgebungen besonders häufig (Obermaier et al., 2024; Seeger et al., 2024).

Auf persönlicher Ebene können betroffene Forschende Angstzustände, Schlafprobleme und Angst um ihre Sicherheit erleben (Global Witness, 2023). Auf gesellschaftlicher Ebene lassen sich die Folgen eines Rückzugs aus dem öffentlichen Diskurs oder des sogenannten "Silencing" beobachten (Seeger et al., 2024). Dies ist besonders besorgniserregend, da unterrepräsentierte Gruppen, wie z. B. Wissenschaftlerinnen, überproportional von Online-Missbrauch betroffen zu sein scheinen (Hodson et al., 2018; Obermaier et al., 2024; Veletsianos et al., 2018). Aufgrund dieser Erfahrungen sind Forschende besorgt, dass ihr Beitrag zur CDT sie virtueller Belästigung aussetzen könnte.

In den Co-Creation-Workshops zeigte sich daher ein klares Dilemma: Einerseits wünschen Forschende eine hohe Sichtbarkeit innerhalb der Forschungsgemeinschaft, bei politischen Entscheidungsträgern und möglicherweise sogar in der breiten Öffentlichkeit (siehe den Abschnitt über Anreize). Wissenschaftskommunikation ist eine Kernaufgabe von Forschenden, insbesondere in Bereichen von großer gesellschaftlicher Bedeutung. Auf persönlicher Ebene können eine erfolgreiche Wissenschaftskommunikation und das Erreichen einer hohen öffentlichen Sichtbarkeit entscheidend sein, um wissenschaftliche Karrieren voranzutreiben.

Andererseits möchten Forschende möglicherweise die Exposition ihrer Arbeit in rechtsextremen Milieus minimieren, um die oben beschriebenen Risiken von digitalen oder körperlichen Angriffen zu vermeiden. In Anbetracht der Tatsache, dass die CDT Daten über rechtsextreme Akteur*innen verwaltet, besteht die Sorge, dass Forschende zur Zielscheibe

solcher Angriffe werden könnten, nur weil sie an der Erhebung und Analyse solcher Daten beteiligt sind.

Universitäten und Forschungsinstitute wurden unisono kritisiert, weil es ihnen an Strukturen zur Unterstützung von Forschenden fehlt, die online angegriffen werden. In den schlimmsten Fällen sind die Forschenden auf sich allein gestellt, was zu einer erheblichen psychischen Belastung werden kann. Darüber hinaus erfordert die Verfolgung rechtlicher Schritte gegen Beleidigungen und Drohungen erhebliche materielle Ressourcen. Dadurch besteht das Risiko, dass Forschende sich von ihrer Arbeit zurückziehen, was sowohl für sie persönlich als auch für die Gesellschaft insgesamt negative Folgen haben kann.

7.2. Unterstützung durch die CDT

Ein Datentreuhänder kann das Spannungsfeld zwischen Sichtbarkeit und Sicherheit nicht auflösen. Angesichts der fehlenden Unterstützung vor allem an den Universitäten wurde jedoch die Frage an die CDT formuliert, ob Forschungsinfrastrukturen nicht auch einen Beitrag zur Sicherheit der Forschenden leisten können. Eine einfache Möglichkeit ist, die Sichtbarkeit der Beitragenden am Datenbestand der CDT zu minimieren. Deshalb haben Forschende die Möglichkeit, nicht als Autor*innen der CDT geführt zu werden (siehe oben). Diese Entscheidung bedeutet zwar, dass ihre Beiträge nicht zitierfähig sind oder Sichtbarkeit erlangen, sie können jedoch weiterhin auf die Ressource zugreifen und zu ihrer Forschung beitragen. Wenn eine in der Datenbank aufgeführte Person einen Antrag auf Informations-zugang zu ihren Daten stellt, muss der CDT dem nachkommen und die Informationen freigeben, wobei jedoch alle Einzelheiten, die die Identität der an der Bearbeitung des Eintrags beteiligten Forschenden preisgeben könnten, unkenntlich gemacht werden. Schließlich umfasst das Training im Rahmen des Onboardings (siehe Herausforderung I) ein Modul zu Fragen der persönlichen Sicherheit, das Hinweise und Ressourcen zum Schutz vor potenziellen Bedrohungen bietet.

Dieses Modul basiert auf Best Practices für den Umgang mit gefährlichen Situationen, die gerade innerhalb der Forschungsgemeinschaft entstehen (Tollefson, 2024). Die erste Regel besteht darin, die Kommunikation von Forschungsergebnissen auf Organisationen und Projekte und nicht auf die (privaten) Konten einzelner Forschenden zu konzentrieren. Dieser Ansatz verlagert den Fokus rechtsextremer Akteur*innen weg von den Forschenden selbst und hin zu den Forschungsorganisationen. Eine zweite Maßnahme ist die Erarbeitung von Kommunikationsstrategien für den Umgang mit solchen Kampagnen, die festlegen, welche Inhalte von wem und über welche Kanäle verbreitet werden. Dies ermöglicht ein koordiniertes Vorgehen aller beteiligten Parteien.

Weitere Maßnahmen können nicht durch die CDT unterstützt werden. So betonten die Forschenden in den Co-Creation-Workshops die Notwendigkeit der Unterstützung durch die Arbeitgebenden bei der Sicherstellung von Rechtshilfe, wenn sie sich verteidigen oder Missbrauch melden müssen. Forschungseinrichtungen sollten proaktiv Forschende mit

spezialisierten Anwaltskanzleien zusammenbringen und eine vollständige oder teilweise Übernahme der Rechtskosten anbieten. Wirksame rechtliche Reaktionen auf Beleidigungen und Drohungen können Online-Kampagnen abschrecken, indem sie die Schwelle für ein solches Verhalten erhöhen. Diese Maßnahmen würden auch dem CDT zugutekommen, wenn sie von der Host-Organisation umgesetzt werden. Weiterhin können für besonders exponierte Mitarbeitende Sicherheitschecks durchgeführt werden, um zu prüfen, ob Adressen im Netz verfügbar sind und im Fall des Falles Gegenmaßnahmen einzuleiten. Eine letzte Maßnahme, die von Arbeitgebern unterstützt werden kann, ist die Beantragung von Auskunftssperren, um personenbezogene Daten der Forschenden zu schützen.

Die Möglichkeiten einer Forschungsdateninfrastruktur zur persönlichen Sicherheit der Forschenden beizutragen sind folglich begrenzt. Vor allem Arbeitgeber in der Wissenschaft müssen noch stärker für Sicherheitsfragen sensibilisiert werden.

8. Anwendungsfälle außerhalb der digitalen Rechtsextremismusforschung

Die wachsende Bedeutung neuer Datentypen und Analysemethoden in den Sozialwissenschaften verändert die Infrastrukturbedarfe der Forschenden. Für die Computational Social Sciences fordern (Lazer et al., 2020) eine infrastrukturelle Wende um mit den kritischen Herausforderungen digitaler Daten umgehen zu können, darunter unzureichender Datenaustausch, unzureichende ethische Regeln für die Verarbeitung persönlicher Daten und schlechte Datenqualität. Das Modell einer Community-Datentreuhand kann Lösungen für die Probleme anbieten, wenn eine kooperative Erstellung und Nutzung von Daten deren Nutzungswert erhöhen. Am Beispiel der Accountlisten zur Erforschung rechtsextremistischer Onlinekommunikation lassen sich die Vorteile dieser Form kollektiven Datenmanagements gut illustrieren und Vorschläge für organisatorische und technische Lösungen zentraler Herausforderungen bei der kollektiven Zusammenarbeit skizzieren, die im Rahmen eines Co-Creation Ansatzes mit Vertretern der Community entwickelt wurden. Der Träger der Datentreuhand hat in diesem Modell in erster Linie die Rolle, einen institutionellen Rahmen für die Kooperation bereitzustellen, die technische Infrastruktur zu betreiben und über die Einhaltung der Regeln zu wachen (z.B. bei der Genehmigung der Nutzungsverträge). Im Zuge der Datenkuratierung trägt er aber auch Mitverantwortung für die Datenqualität, sodass keine vollständige Neutralität gegeben ist.

Die Frage ist naheliegend, ob dieses Modell auch auf andere Anwendungsfälle in den Computational Social Science oder Sozialwissenschaften übertragbar ist. Zu erwarten ist, dass der Nutzen einer CDT umso größer ist, wenn eine zentrale Erhebung von Daten nicht möglich oder nicht sinnvoll ist oder die Daten sehr volatil sind, sodass die Aktualität der Daten mit großem Aufwand gesichert werden muss.

Dabei müssen nicht unbedingt sensible Daten im Vordergrund stehen. Community-Datentreuhänder sind gleichermaßen für Anwendungsfälle geeignet, in denen aus zahlreichen Beiträgen ein gemeinsamer Datenkorpus erzeugt wird, auch wenn dieser keine personenbezogenen Daten enthält. Dann liegt der Schwerpunkt auf dem fairen Interessenausgleich zwischen den beteiligten Forschenden.

Seitdem die großen Online-Plattformen den Abruf von Daten für die wissenschaftliche Forschung über ihre Schnittstellen sukzessive eingeschränkt haben, ist die Verfügbarkeit von Daten gerade für die reichweitenstarken Plattformen begrenzt. In der Forschung werden deshalb aktuelle Verfahren der Datenspende erprobt. Dabei werden Nutzende der Plattformen gebeten, ihre eigenen Daten herunterzuladen und für die Forschung zur Verfügung zu stellen. Da die Rekrutierung von Datenspendern sehr aufwändig ist, kann auch für diesen Anwendungsfall eine CDT über den kumulativen Aufbau eines Datenbestandes, den wissenschaftlichen Wert einzelner Datenspenden deutlich steigern. Datenspenden sind auch im Bereich der Mobilitätsdaten eine Lösung, wenn z. B. Ortungsdaten aus mobilen Endgeräten für die Forschung benötigt werden.

9. Schlussbemerkungen

Auf der Grundlage des Konzepts der Datentreuhänder führt die CDT eine neuartige Forschungsinfrastruktur ein, die die Zusammenarbeit zwischen Forschenden fördert, insbesondere bei der Erstellung und Pflege von Accountlisten. Diese Listen sind für viele automatisierte Datenerhebungsprozesse in den digitalen Medien entscheidend, besonders in der Kommunikationswissenschaft.

Die CDT bietet einen Rahmen, der die ethische und rechtskonforme Verwaltung dieser sensiblen Forschungsdaten während des gesamten Prozesses der Datenweitergabe gewährleistet. Er nutzt das Forschungsprivileg der EU-DSGVO und schafft eine kontrollierte Umgebung mit organisatorischen und technischen Sicherheitsvorkehrungen für kollaboratives Datenmanagement. Auf diese Weise minimiert die CDT die rechtlichen Risiken für die einzelnen Forschenden und trägt dazu bei, Kooperationsbarrieren zu überwinden.

Kernelement der CDT ist das Prinzip der Reziprozität, das Forschenden Vertrauen in einen fairen Datenaustausch geben soll. Darüber hinaus werden individuelle Beiträge durch ein zitierbares Datenpapier anerkannt. Die Datenqualität innerhalb der CDT wird durch ständige Aktualisierungen und Peer-Reviews aufrechterhalten und verbessert, um die Genauigkeit und Anpassungsfähigkeit an die sich weiterentwickelnde Natur von Online-Phänomenen zu gewährleisten.

Obwohl der Schutz der Sicherheit von Forschenden - insbesondere in politisch sensiblen Bereichen - weiterhin eine Herausforderung darstellt, bietet die CDT Mechanismen für den Ausstieg aus der öffentlichen Anerkennung und stellt Ressourcen zur Risikominderung bereit.

Es bedarf jedoch weiterer institutioneller Unterstützung durch Forschungsorganisationen, um Forschende besser vor möglicher Belästigung oder Missbrauch zu schützen.

Die nächste Herausforderung für die Implementierung der CDT im Kontext der deutschen digitalen Rechtsextremismusforschung ist die Etablierung eines nachhaltigen Modells für ihren kontinuierlichen Betrieb bei gleichzeitiger Erweiterung der Nutzendenbasis. Derzeit befindet sich die CDT in einer ersten Betriebsphase, in der vertrauenswürdige Forschende aus unseren Co-Creation-Workshops aktiv ihre Daten einspeisen und teilen. Um einen langfristigen Erfolg zu gewährleisten, ist es entscheidend, das Engagement dieser etablierten Gruppe aufrechtzuerhalten und gleichzeitig neue Nutzende zu integrieren und das Projekt über die bestehenden Netzwerke hinaus zu erweitern.

Die Aktivitäten der Nutzenden im Rahmen der CDT gehen über die einfache Nutzung und Pflege der Datenbank hinaus; sie umfassen auch einen Beitrag zur Verwaltung der CDT. In einem sich rasch entwickelnden Bereich wie der CCS und den CSS wird das Datentreuhandmodell unweigerlich auf Herausforderungen stoßen, die in der Anfangsphase des Projekts noch nicht absehbar waren. Daher müssen sich die Governancestrukturen der CDT kontinuierlich weiterentwickeln, wobei die Forschungsgemeinschaft in diesen Prozess nachhaltig eingebunden werden muss.

Darüber hinaus sollten Fragen der Verwaltung - wie etwa mögliche Verstöße gegen das Reziprozitätsprinzip unter den teilnehmenden Forschenden - nicht immer von oben nach unten durch die Trägerorganisation der CDT geregelt werden. Stattdessen sollten die Lösungen gemeinsam entwickelt werden. Dieser kollaborative Ansatz für die Verwaltung erfordert eine kontinuierliche Beteiligung der Nutzenden. Eine mögliche Strategie zur Sicherstellung dieser Beteiligung ist die Einrichtung eines wissenschaftlichen Beirats durch die Trägerorganisation, der die CDT bei ihren künftigen Entwicklungs- und Verwaltungsprozessen berät.

Eine zweite Herausforderung neben der kontinuierlichen Beteiligung der Forschungsgemeinschaft ist die langfristige Finanzierung der CDT als Forschungsinfrastruktur. Forschungsinfrastrukturprojekte leiden oft unter kurzfristigen Finanzierungszyklen, die ausreichen können, um neue Projekte zu initiieren, erstes Fachwissen aufzubauen, Institutionen zu etablieren und Software zu entwickeln. Allerdings stehen solche Projekte häufig vor Herausforderungen, wenn die Finanzierung ausläuft, da es oft keine Strategie gibt, um ihre Leistungen aufrechtzuerhalten. Für nachhaltige Forschungsinfrastrukturen wie die CDT sind mittel- und langfristige Verpflichtungen seitens staatlicher und anderer Finanzierungseinrichtungen von entscheidender Bedeutung. Diese kontinuierliche Unterstützung ist notwendig, um Fortschritt sicherzustellen und den langfristigen Nutzen der Infrastruktur für die Forschungsgemeinschaft zu maximieren (Rat für Informationsinfrastrukturen, 2016).

Über den speziellen Fall der digitalen Rechtsextremismusstudien hinaus glauben wir, dass die CDT eine transformative Forschungsinfrastruktur darstellt, die sich an ein breites Spektrum von (sensiblen) kommunikationswissenschaftlichen und CSS-Forschungsbereichen anpassen lässt, die auf digitale Accountlisten angewiesen sind. Diese Listen sind integraler Bestandteil vieler digitaler Datenerhebungsprozesse und werden in Bereichen wie Desinformation, Hassrede und Propagandastudien häufig verwendet. Durch die Schaffung eines sicheren, ethischen und kollaborativen Rahmens für die gemeinsame Nutzung und Verwaltung von Daten geht die CDT eine der dringendsten Herausforderungen in der digitalen Forschung an: den Zugang zu sensiblen Daten. Dank ihrer Flexibilität kann sie unterschiedliche Forschungsziele berücksichtigen, während ihre dynamische und rechtskonforme Infrastruktur gewährleistet, dass er mit der sich ständig verändernden digitalen Landschaft Schritt halten kann. Damit ist die CDT ein vielversprechendes Modell für die kollektive Datenverwaltung digitaler Accountlisten.

Darüber hinaus könnte das CDT-Framework auch auf andere Datentypen, wie z. B. Datenspenden, ausgeweitet werden. Eine solche Erweiterung würde zwar eine Reihe von Anpassungen und Weiterentwicklungen des bestehenden Modells erfordern, aber die Grundprinzipien der Datentreuhänderschaft und die aus der Entwicklung der CDT gewonnenen Erkenntnisse bieten eine solide Grundlage für die Anpassung der Infrastruktur an verschiedene Arten von digitalen Daten.

Literaturverzeichnis

Akdeniz, E., Borschewski, K. E., Breuer, J. & Voronin, Y. (2023). Sharing social media data: The role of past experiences, attitudes, norms, and perceived behavioral control. *Frontiers in big data*, *5*. https://doi.org/10.3389/fdata.2022.971974

Arlinghaus, T., Kus, K., Kajüter, P. & Teuteberg, F. (2021). Datentreuhandstellen gestalten: Status quo und Perspektiven für Geschäftsmodelle. *HMD Praxis der Wirtschaftsinformatik*, *58*(3), 565–579. https://doi.org/10.1365/s40702-021-00727-x

Blankertz, A., Kuzev, P., Richter, F., Richter, H., Schallbruch, M. & Braunmühl, P. von. (2020). *Datentreuhandmodelle: Themenpapier*. Berlin. Stiftung Neue Verantwortung. https://doi.org/10.13140/RG.2.2.24915.30246

Breuer, J., Kmetty, Z., Haim, M. & Stier, S. (2023). User-centric approaches for collecting Facebook data in the 'post-API age': experiences from two studies and recommendations for future research. *Information, Communication & Society*, *26*(14), 2649–2668. https://doi.org/10.1080/1369118X.2022.2097015

Buchner, B., Haber, A. C., Hahn, H. K., Prasser, F., Kusch, H., Sax, U. & Schmidt, C. O. (2021). Das Modell der Datentreuhand in der medizinischen Forschung. *Datenschutz und Datensicherheit - DuD*, *45*(12), 806–810. https://doi.org/10.1007/s11623-021-1534-y

Die Bundesregierung. (2021). *Datenstrategie der Bundesregierung: Eine Innovationsstrategie für gesellschaftlichen Fortschritt und nachhaltiges Wachstum*. https://www.publikationen-bundesregierung.de/resource/blob/2277952/1845634/1a4f7ea800bb838562e16fdfe4ffb354/datenstrategie-der-bundesregierung-download-bpa-data.pdf?download=1

European Commission. (2020). *A European strategy for data*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52020DC0066

Europäische Kommission (2022): Data Governance Act, Verordnung (EU) 2022/868, ABl. L 152, 3.6.2022, S. 1–44.

Fdez-Arroyabe, P. & Roye, D. (2017). Co-creation and Participatory Design of Big Data Infrastructures on the Field of Human Health Related Climate Services. In C. Bhatt, N. Dey & A. S. Ashour (Hrsg.), *Studies in Big Data. Internet of Things and Big Data Technologies for Next Generation Healthcare* (Bd. 23, S. 199–226). Springer International Publishing. https://doi.org/10.1007/978-3-319-49736-5_9

Fecher, B., Kahn, R., Sokolovska, N., Völker, T. & Nebe, P. (2021). Making a Research Infrastructure: Conditions and Strategies to Transform a Service into an Infrastructure. *Science and Public Policy*, *48*(4), 499–507. https://doi.org/10.1093/scipol/scab026

GLES. (2024). *GLES Kandidierendenstudie Europawahl 2024, Social Media Accounts*. https://doi.org/10.4232/1.14384

Global Witness. (2023). *Global Hating: How online abuse of climate scientists harms climate action*. Global Witness. https://gw.cdn.ngo/media/documents/Global_Hating_-_April_2023_8azJDgp.pdf

Guhl, J., Ebner, J. & Rau, J. (2020). *Das Online-Ökosystem rechtsextremer Akteure*. https://www.bosch-stiftung.de/de/publikation/das-online-oekosystem-rechtsextremer-akteure

Hodson, J., Gosse, C., Veletsianos, G. & Houlden, S. (2018). I get by with a little help from my friends: The ecological model and support for women scholars experiencing online harassment. *First Monday*. Vorab-Onlinepublikation. https://doi.org/10.5210/fm.v23i8.9136

Hohmann, F. (2021). Co-Creation als Entwicklungsmethode. Zu Möglichkeiten und Grenzen partizipativer Forschungssoftwareentwicklung am Beispiel der Sortiersoftware MeSort und Tagebuchsoftware MeTag. *Medien & Kommunikationswissenschaft*, *69*(1), 97–116. https://doi.org/10.5771/1615-634X-2021-1-97

Jost, P., Heft, A., Buehling, K., Zehring, M., Schulze, H., Bitzmann, H. & Domahidi, E. (2023). Mapping a Dark Space: Challenges in Sampling and Classifying Non- Institutionalized Actors on Telegram. *Medien & Kommunikationswissenschaft*, *71*(3-4), 212–229. https://doi.org/10.5771/1615-634X-2023-3-4-212

Kreutzer, S., Heimer, T., Nachtigall, H., Pschorn, L., Bauer, F., Blind, K., Martin, N., Grafenstein, M. von, Streblow, R., Du, J. & Schölzel, J. D. (2024). *Wissenschaftliche Begleitung und Vernetzung der Projekte zur Entwicklung und praktischen Erprobung von Datentreuhandmodellen in den Bereichen Forschung und Wirtschaft : Bericht zu Arbeitspaket 1.2 : Anforderungen und Umsetzungshemmnisse für Datentreuhandmodelle ; Die Studie wird im Auftrag des Bundesministeriums für Bildung und Forschung (kofinanziert durch das Programm „NextGenerationEU" der Europäischen Union) durchgeführt*. https://doi.org/10.18154/RWTH-2024-04375

Kühling, J., Sackmann, F. & Schneider, H. (2020). *Datenschutzrechtliche Dimensionen Datentreuhänder: Kurzexpertise*. *Forschungsbericht / Bundesministerium für Arbeit und Soziales: Bd. FB550*. SSOAR, GESIS – Leibniz-Institut für Sozialwissenschaften e.V; Bundesministerium für Arbeit und Soziales.

Lazer, D. M. J., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., Nelson, A., Salganik, M. J., Strohmaier, M., Vespignani, A. & Wagner, C. (2020). Computational social science: Obstacles and opportunities. *Science (New York, N.Y.)*, *369*(6507), 1060–1062. https://doi.org/10.1126/science.aaz8170

Miller-Idriss, C. (2020). *Hate in the homeland: The new global far right*. Princeton University Press.

Nogrady, B. (2021). 'I hope you die': How the COVID pandemic unleashed attacks on scientists. *Nature*, *598*(7880), 250–253. https://doi.org/10.1038/d41586-021-02741-x

Obermaier, M., Seeger, C., Frischlich, L., Schmid, U. K. & Riesmeyer, C. (2024). An easy target? Factors behind uncivil attacks against communication scientists. *Charms Reports*, *No. 2*. https://osf.io/jswav

Ohme, J., Araujo, T., Boeschoten, L., Freelon, D., Ram, N., Reeves, B. B. & Robinson, T. N. (2024). Digital Trace Data Collection for Social Media Effects Research: APIs, Data Donation, and (Screen) Tracking. *Communication Methods and Measures*, *18*(2), 124–141. https://doi.org/10.1080/19312458.2023.2181319

Pawelke, A., Fetic, L. & Bertelsmann Stiftung. (2020). *Daten teilen, aber wie?* https://doi.org/10.11586/2020079

Piller, F. T., Ihl, C. & Vossen, A. (2010). A Typology of Customer Co-Creation in the Innovation Process. *SSRN Electronic Journal*. Vorab-Onlinepublikation. https://doi.org/10.2139/ssrn.1732127

Rat für Informationsinfrastrukturen. (2016). *Leistung aus Vielfalt: Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland*. Göttingen, 160 S. https://rfii.de/?p=1998

Rau, J., Kero, S., Hofmann, V., Dinar, C. & Heldt, A. P. (2022). *Rechtsextreme Online-Kommunikation in Krisenzeiten: Herausforderungen und Interventionsmöglichkeiten aus Sicht der Rechtsextremismus- und Platform-Governance-Forschung*. https://doi.org/10.21241/ssoar.78072

Rogers, R. (2020). Deplatforming: Following extreme Internet celebrities to Telegram and alternative social media. *European Journal of Communication*, *35*(3), 213–229. https://doi.org/10.1177/0267323120922066

Rothut, S., Schulze, H., Rieger, D. & Naderer, B. (2024). Mainstreaming as a meta-process: A systematic review and conceptual model of factors contributing to the mainstreaming of radical and extremist positions. *Communication Theory*, *34*(2), 49–59. https://doi.org/10.1093/ct/qtae001

Sältzer, M., Stier, S., Bäuerle, J., Blumenberg, M., Mechkova, V., Pemstein, D., Seim, B. & Wilson, S. (2023). *Twitter-Accounts der Kandidierenden zur Bundestagswahl 2021 (GLES)*. https://doi.org/10.4232/1.14233

Sanders, E. B.-N. & Stappers, P. J. (2008). Co-creation and the new landscapes of design. *CoDesign*, *4*(1), 5–18. https://doi.org/10.1080/15710880701875068

Schmidt, J.-H., Merten, L. & Münch, F. V. (2024). *Die „Datenbank Öffentlicher Sprecher" (DBÖS)*. v2. Dezember 2024. https://doi.org/10.17605/OSF.IO/SK6T5

Seeger, C., Frischlich, L., Obermaier, M., Schmid, U. K., Riesmeyer, C. & Menke, M. (2024). The dark side of science communication — communication scientists' experiences with uncivil attacks. *Charms Reports*, *No. 1*. https://osf.io/xcrkp/

Sen, I., Flöck, F., Weller, K., Weiß, B. & Wagner, C. (2021). A Total Error Framework for Digital Traces of Human Behavior on Online Platforms. *Public Opinion Quarterly*, *85*(S1), 399–422. https://doi.org/10.1093/poq/nfab018

Tollefson, J. (2024). Harassed? Intimidated? Guidebook offers help to scientists under attack. *Nature*. Vorab-Onlinepublikation. https://doi.org/10.1038/d41586-024-03104-y

Veletsianos, G., Houlden, S., Hodson, J. & Gosse, C. (2018). Women scholars' experiences with online harassment and abuse: Self-protection, resistance, acceptance, and self-blame. *New Media & Society*, *20*(12), 4689–4708. https://doi.org/10.1177/1461444818781324

Wiedemann, G., Münch, F. V., Rau, J. P., Kessling, P. & Schmidt, J.-H. (2023). Concept and challenges of a social media observatory as a DIY research infrastructure. *Publizistik*, *68*(2-3), 201–223. https://doi.org/10.1007/s11616-023-00807-6

Impressum

Kontakt:
Rat für Sozial- und Wirtschaftsdaten (RatSWD)
Geschäftsstelle
Am Friedrichshain 22
10407 Berlin
office@ratswd.de
https://www.ratswd.de

Die Geschäftsstelle des RatSWD wird als Teil von KonsortSWD im Rahmen der NFDI durch die Deutsche Forschungsgemeinschaft (DFG) gefördert - Projektnummer: 442494171.

Berlin, Juli 2025

DOI: 10.17620/02671.97

Zitationsvorschlag:
Jungmann, N., Siegers, P., J. P., Fürneisen, M., Wiedemann, G. & Schulze, H. (2025). *Eine Community-Datentreuhand für die gemeinsame Nutzung sensibler Daten in der Kommunikationswissenschaft*. (RatSWD Working Paper 288/2025). Berlin. Rat für Sozial- und Wirtschaftsdaten (RatSWD). https://doi.org/10.17620/02671.97