Media and Communication
2026 • Volume 14 • Article 11946
https://doi.org/10.17645/mac.11946

cogitatio

ARTICLE

Open Access Journal

A Community Data Trustee for Sensitive Data Sharing in Far-Right Online Research

Jan Rau [1], Nils Jungmann [2], Moritz Fürneisen [1], Gregor Wiedemann [1], Heidi Schulze [3], and Pascal Siegers [4]

[1] Media Research Methods Lab, Leibniz Institute for Media Research | Hans-Bredow-Institut, Germany
[2] Tern ZE GmbH, Germany
[3] Department of Media and Communication, Ludwig-Maximilians-Universität München, Germany
[4] Department of Survey Data Curation, GESIS – Leibniz Institute for the Social Sciences, Germany

Correspondence: Pascal Siegers (pascal.siegers@gesis.org)

Submitted: 15 January 2026 **Accepted:** 5 May 2026 **Published:** 16 July 2026

Issue: This article is part of the issue "Open Research Infrastructures and Resources for Communication and Media Studies" edited by Silke Fürst (University of Zurich), Johannes Breuer (Center for Advanced Internet Studies / University of Duisburg-Essen), Erik Koenen (University of Bremen), Dimitri Prandner (Johannes Kepler University of Linz), Christian Schwarzenegger (University of Bremen), and Christian Strippel (Weizenbaum Institute), fully open access at https://doi.org/10.17645/mac.i504

Abstract

Data sharing poses a major challenge for digital media and communication research, particularly in sensitive areas such as far-right online studies. This article introduces the innovative concept of a "community data trustee" (CDT), a research infrastructure aimed at fostering collaboration and the sharing of research data, such as digital account lists. Compiling these lists is a critical yet labor-intensive step in many research projects; sharing them could significantly reduce effort and improve data quality. However, especially in sensitive research areas, sharing remains rare due to legal uncertainties and limited incentives. This hinders the traceability and comparability of research findings and threatens overall research quality. To address these challenges, we introduce the CDT as a collective research infrastructure that promotes the shared use of extensive account lists. The CDT views actor directories as a communal asset, developed and utilized based on mutually agreed-upon guidelines. A principle of reciprocity underpins this model: Those who access the lists also contribute to their updates and expansions, returning them to the data pool. An online portal facilitates the exchange and collaborative maintenance of the data. The setup of the CDT includes appropriate technical measures to ensure compliance with data protection and security standards, along with a robust regulatory framework that creates a legally secure environment for sharing personal data. This approach aims to (a) incentivize data sharing, (b) foster trust and legal certainty among research projects, (c) enhance data quality through ongoing maintenance, and (d) enhance researcher safety.

Keywords
data privacy; data sharing; data trustee; far-right; social media

1. Introduction

The rise of the internet, along with the increased availability of observational digital trace data, has significantly transformed the social sciences in general and media and communication research in particular. New (sub-)disciplines and methods, such as computational social science, web science, internet studies, and digital anthropology, have emerged. Meanwhile, many established fields have undergone a "digital turn," integrating digital perspectives, data, and methodologies into their traditional frameworks, including media and communication research (Horst & Miller, 2020; Lazer et al., 2020; Theocharis & Jungherr, 2020). These developments have been driven both by new methodological opportunities and by a growing societal interest in the political and cultural consequences of digital media. Research attention has shifted from early hopes of online democratization (Castells, 2013) toward concerns over polarization, disinformation, and the increasing visibility and activity of extremist movements online, particularly the far-right (Lorenz-Spreen et al., 2023).

The far-right, as an umbrella term, encompasses illiberal radical, and anti-democratic extremist manifestations of right-wing politics whose common ideological core is motivated by *nativism* (i.e., the belief that the state should be inhabited exclusively by members of the native group, while non-native people are considered a threat) and *authoritarianism* (i.e., the vision of a strictly ordered society in which violations of authority are to be punished severely; Pirro, 2023). Far-right actors have maintained an online presence since as early as 1984 in the form of bulletin board systems. Since then, this presence has evolved, suggesting that far-right actors are particularly adept at leveraging the digital landscape to their advantage (Conway et al., 2019). They use digital media to disseminate propaganda, recruit members, coordinate activities, and attempt to influence electoral outcomes (Hoffmann & Rone, 2023; Miller-Idriss, 2020). This effectiveness has led to an increased demand for research focused on the far-right online and the societal and political risks it poses.

A foundational aspect of far-right online studies, as well as other fields of digital media and communication research, involves compiling lists of accounts associated with relevant individuals or organizations for "account-based sampling." While this process serves as a crucial starting point for further data collection efforts (Wiedemann et al., 2023), it also presents significant challenges (Jost et al., 2023). The far-right online is vast and characterized by extreme dynamism, secrecy, and rapid evolution. Similar to other comparable movements, it may encompass thousands of relevant actors across various platforms—both mainstream and niche—within a single country, such as the UK, France, or Germany (Guhl et al., 2020; Miller-Idriss, 2020). The German far-right online ecosystem has evolved from earlier websites, forums, and other static digital communication spaces into a differentiated cross-platform infrastructure. Today, it comprises a decentralized but interconnected network of party-political actors, movement organizations, alternative media, influencers, and supporter milieus distributed across mainstream and alternative platforms (Guhl et al., 2020). This fragmentation, cross-platform connectivity, and ongoing migration between services make the identification and systematic sampling of relevant accounts especially challenging (Jost et al., 2023).

Compounding this complexity, the actors, platforms, and accounts involved are in a constant state of flux, driven by the inherent volatility of these movements and external factors like de-platforming efforts (Rogers, 2020). These dynamics raise significant concerns about potential data quality issues and the risks of measurement and representation errors in this type of research (Sen et al., 2021). Researchers often rely on curated account lists to guide data collection; however, these lists must be continuously updated to keep pace with the rapidly evolving ecosystem. Researchers monitoring even a subset of potentially harmful phenomena, such as the extremist core of the far-right online, often find themselves overwhelmed by the sheer volume of information and the effort needed to maintain up-to-date records.

These challenges can also be seen in other parts of the data collection process. The sheer number of potentially relevant platforms and their evolving data access policies place a substantial burden on individual researchers or research teams (Hase et al., 2024). Other social science fields have tackled similar challenges by establishing centralized data collection infrastructures and embracing open science principles, particularly data sharing. A prominent example is the European Social Survey, which evolved into a European Research Infrastructure Consortium, offering high-quality comparative survey data on various topics (Lindstrøm & Kropp, 2017). This collective approach allows researchers to leverage existing data and increase the visibility, usage, and citation of such datasets, ultimately benefiting the entire research field (Lindstrøm & Kropp, 2017). Furthermore, these mechanisms enhance reproducibility—a key component of scientific scrutiny—and can improve both the quality of data and research outcomes (Weller & Kinder-Kurlanda, 2016).

While authors like Lazer et al. (2020) have advocated for a similar infrastructure turn in digital and computational research, and recent coordination efforts (Hase et al., 2024; Sinnott et al., 2023) suggest that such a shift may be underway, these digital research fields face unique challenges that require targeted solutions.

Digital research, including media and communication research, often involves the use of personal data, and the sensitivity surrounding this digital data is heightened in research areas such as online extremism. The sensitive nature of this data brings significant ethical and legal constraints aimed at safeguarding individuals' privacy (Franzke et al., 2020; Lazer et al., 2020). Researchers must adhere closely to these restrictions, as any violations could compromise their ethical and legal responsibilities to protect research subjects and potentially result in serious legal consequences. Additionally, making extremist accounts publicly accessible risks unintended outcomes, such as inadvertently amplifying the reach or visibility of these actors (Rau et al., 2021). This makes it challenging to apply conventional scientific practices like open science and data sharing in these sensitive research fields. As a result of these challenges, research on radical and extremist actors online appears to be characterized by a lack of formalized data-sharing practices, which may leave projects dependent on the goodwill of colleagues and informal exchanges or lead to the inefficient recreation of similar lists and datasets. An alternative structure to promote data sharing among researchers, which remains largely unexplored in media and communication research or the social sciences in general (though it has been applied in fields like health-related research; see Beuthner et al., 2020), is the concept of data trustees. Data trustees are emerging entities in the realm of data governance, designed to manage and regulate access to sensitive data, ensuring its ethical and secure use while adhering to data protection regulations. Acting as intermediaries between data holders and data users, they strive to balance the interests of both sides while protecting individual privacy and ensuring regulatory compliance (Delacroix & Lawrence, 2019). Data holders can share their data with the data trustee, and data users can apply for access—within an institutional framework that upholds legal and ethical standards, particularly through technical and organizational data protection measures.

We believe that adapting a data trustee model to align with the specific practices and needs of digital media and communication research could offer a viable solution to the challenges of sharing sensitive account lists.

To address these challenges, we introduce the concept of a "community data trustee" (CDT) and demonstrate its application in digital research on German-language far-right online ecosystems through AVERA (https://datenportal-rechtsextremismus.de/en/rex-forschungsdaten/digitale-rex-daten/avera), an empirical instantiation of the CDT model. We will first outline the development process, including our case selection and data collection approach. Next, we will present the general concept and core functionalities of the CDT. Following this, five key challenges identified during the CDT's development and implementation will be highlighted: (a) compliance with legal and ethical obligations to protect research subjects; (b) the creation of incentives for researcher participation in the CDT; (c) ensuring the quality of data managed within the CDT; (d) protecting the safety of the researchers involved; and (e) dealing with potential contractual constraints originating in platforms' terms of services (ToSs). We will elaborate on how the CDT addresses these challenges through its design and functionalities. Finally, the discussion will offer insights into future plans to extend the database and explore the potential for adapting the CDT concept to other digital research areas that involve sensitive data.

2. Methodology: Development of the Data Trustee

2.1. Case Selection

Germany's strong emphasis on privacy and data protection issues makes it an ideal location for developing a data trustee model that helps to prevent abuse and addresses various ethical and legal concerns. Moreover, Germany operates under the General Data Protection Regulation (GDPR) of the EU, one of the world's most stringent data protection frameworks, which governs the processing of personal data. While our specific data trustee will focus on the German language sphere, we anticipate that the general data trustee model we propose can be easily adapted and implemented in other European countries and beyond.

Far-right online research represents a compelling application for digital research involving sensitive data. This field is not only significant due to its societal implications but also garners considerable public interest, which is crucial for establishing a legal foundation for the data trustee model. Given its connections to issues such as extremism, political violence, and terrorism, the data collected within this context is considered highly sensitive and necessitates robust protective measures. Additionally, there is an active research community focused on the German far-right online, underscoring the demand for more efficient data-sharing practices within this area of study.

2.2. Data Collection

The development of the CDT has incorporated insights from two key external sources of expertise.

First, the community aspect of the data trustee is rooted in close collaboration with the German far-right online research community throughout the entire design process, employing a co-creation approach (Fdez-Arroyabe & Roye, 2017; Hohmann, 2021). Central to this co-creation methodology is the active involvement of future users, who contribute their knowledge and skills to innovate the product or service in

an "act of collective creativity" (Sanders & Stappers, 2008, p. 6; see also Piller et al., 2010). This collaborative methodology has become increasingly vital for developing sustainable research software in fields like media and communication studies, making it particularly well-suited for the CDT's development (Fecher et al., 2021). Over two years, we organized six co-creation workshops, engaging more than 50 researchers from relevant fields such as extremism research, political communication science, and sociology, as well as researchers representing research infrastructure projects. These workshops gathered input and requirements from the community to inform the organizational and technical development of the data trustee model and its governance. The primary goal of this process was to align the community's needs for efficient data sharing with legal and ethical obligations, particularly regarding the protection of individual privacy and safety.

Second, to develop an appropriate legal framework for the CDT, we collaborated with legal experts to conduct data protection and risk assessments. This process was crucial in defining a solid legal basis for the processing of personal data within the CDT, ensuring that they align with stringent data protection regulations, including the GDPR. As a result, a comprehensive set of organizational and technical measures was designed to safeguard sensitive personal information, minimize risks, and prevent any potential misuse of data.

3. Results: Challenges and Solutions for the Data Trustee

3.1. The Concept of the CDT

The CDT aims to address the challenges outlined in the introduction by establishing a new, collaborative approach to data management in digital research, facilitating cooperation among researchers. The CDT is designed to enable the efficient, ethical, and legally compliant sharing of sensitive digital research data—specifically, lists of entities belonging to the far-right online. These entities can represent different actor types such as individuals, parties, media, and others. As a collective research infrastructure, it reduces the burden of project-specific data collection and enhances the quality of data through peer review and critical evaluation among researchers. It incorporates a range of organizational and technical measures to ensure the protection of individuals' privacy and prevent the misuse of personal information.

As illustrated in Figure 1, the core of the CDT is a shared, dynamic database that contains relevant information about far-right actors, such as website details and account information across digital platforms, along with additional metadata like organizational affiliations. This database is built on custom-designed open-source software, enabling collaborative management, access, and updates through a secure online platform. It includes features such as authorship tracking and version control. This technical layer is live and running, even though the underlying software is continuously updated and improved (Fürneisen et al., 2026). Access can be requested through the Data Portal for Racism and Right-Wing Extremism Research (https://datenportal-rechtsextremismus.de/en/rex-forschungsdaten/digitale-rex-daten/avera). Beyond this technical layer, the CDT's institutional framework ensures that data processing complies with legal standards while also enforcing rules for researcher access. Researchers can apply for access to the data and, in return, commit to contributing their own project-specific actor lists and updates to the database, ensuring that it remains current and comprehensive.

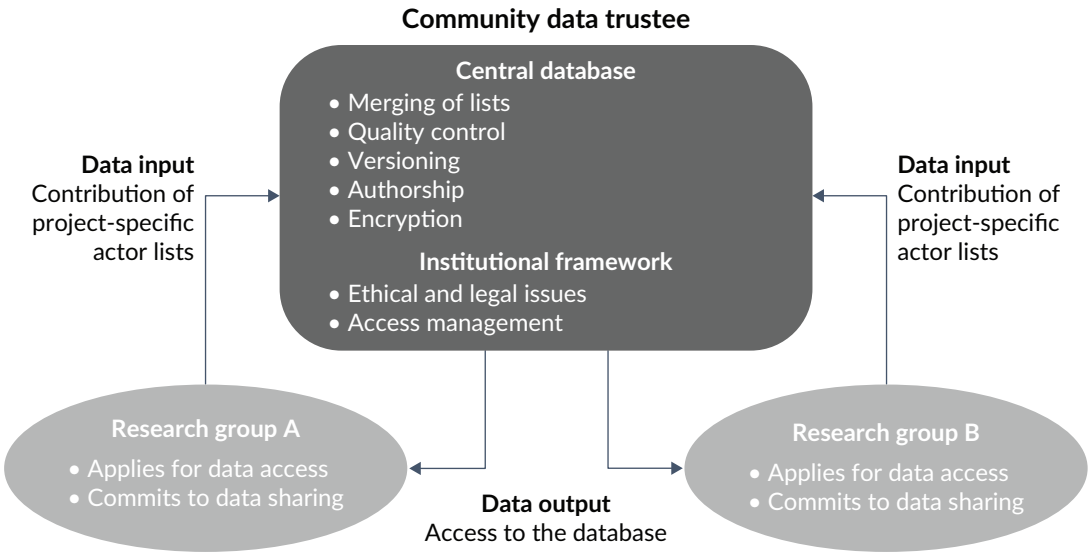

Figure 1. The CDT concept.

Data trustees have been established in several scientific domains. In health research, for example, trustees located at university hospitals primarily manage patients' informed consent and oversee the conditions under which patient data may be used for research. In most cases, data trustees are also responsible for preserving the anonymity of the data subjects when granting access to researchers, a role that closely resembles the function of trusted research environments (TREs), which provide controlled access to sensitive data while minimizing disclosure risks.

In contrast, the CDT is designed to enable access to non-anonymized data when this is necessary for research purposes. The CDT is therefore a collaborative model of data production and utilization. Its central premise is the transfer of authority and accountability to the research community itself. Within this framework, researchers collectively negotiate the rules governing the use of the data and the procedures for ensuring data quality, while the CDT provides the institutional structure that safeguards compliance with legal and ethical provisions.

The CDT addresses the five challenges introduced above—legal compliance, incentives for participation, data quality, researcher safety, and platforms' ToSs—which we now discuss in detail.

3.2. Challenge I: Compliance With Legal and Ethical Obligations to Protect Research Subjects

3.2.1. The Challenge

In Germany, the processing of personal data is governed by the European GDPR and its German counterpart, the Federal Data Protection Act (BDSG). According to Article 9 of the GDPR, information relating to individuals' political opinions is classified as a special category of personal data, which is considered particularly sensitive. Processing such data requires a robust legal basis and protective measures. In the context of the CDT, obtaining informed consent (Article 6(1)(a) GDPR) is not realistic. Therefore, an alternative legal basis must be used.

This is important given that, due to the sensitive nature of extremism-related data, rigorous justification is required for the processing of personal data relating to research subjects, as well as the implementation of suitable technical and organizational measures to ensure data protection. Inclusion in a database of right-wing actors could pose significant risks to individuals, such as defamation and stigmatization, especially if the data were to be leaked. This risk is particularly acute in cases where the collected social media accounts belong to private individuals who lack a public profile or a large group of followers in the network. These individuals might not realize that their online communications are public or may spread beyond the specific context in which they originally posted.

Consequently, researchers are often uncertain about what the legal basis for sharing account information with peers should entail and how this could be implemented in practice. As Kreutzer et al. (2024) point out, legal uncertainties are a significant reason for researchers' reluctance to share data (see also Akdeniz et al., 2023). Defining a legal basis and implementing appropriate technical and organizational measures for data protection requires substantial effort when anonymizing the data would make meaningful reuse impossible. For individual researchers or research projects, the cost of establishing this framework is too high. As a result, researchers tend to opt for the "safe" route of not sharing data, avoiding potential legal risks.

3.2.2. The Legal Basis of the CDT: Research Exemption

A viable option for the CDT is the scientific research exemption provided by Article 89 of the GDPR. This exemption permits the processing of even special category personal data if they are necessary to achieve the research objective and if the interests of the CDT as a data processor significantly outweigh those of the data subject. However, for this exemption to apply, several criteria related to scientific research must be met. These criteria include transparency of the research process and results, the independence of the research, and a clear research purpose aimed at advancing knowledge for the public interest (Buchner et al., 2021).

Combined with Article 6(1)(e), we argue that research into the online activities of far-right—especially extremist—actors is in the public interest because it contributes to the protection of democratic governance. In the case of the Federal Republic of Germany, these rules are set out in Section 27 of the BDSG. In particular, a balancing of interests must be carried out between the research objectives and the rights of the individuals involved.

The criteria for the research exemption are met in the case of the CDT. The growth of the far-right online is associated with significant political and societal risks, including right-wing terrorism, political violence, and threats to the constitutional order, as well as to ethnic and religious minorities. These risks amplify the importance of the CDT for researchers and the public, as it provides a crucial foundation for understanding the far-right's online success, investigating its underlying causes, and facilitating the development of evidence-based strategies for prevention and intervention.

3.2.2.1. Constraints and Limitations

Despite the fundamental justification for processing personal data belonging to far-right actors, this does not imply an unconditional endorsement of such practices. The protective measures mandated by the GDPR must be strictly adhered to, and the research exemption—along with the GDPR as a whole—comes with a range of principles and safeguards that need to be implemented for lawful data processing.

One of the most challenging aspects involves processing data on individuals who may be relevant for far-right online research but do not represent the extremist core. While the immediate threat connected with this group is comparatively lower than that posed by clearly extremist individuals, the potentially stigmatizing effects of being included in such a database are significant for these individuals.

Nonetheless, capturing the digital public communication of such individuals is a legitimate and crucial objective for research. These spaces are susceptible to the extremist influence of right-wing ideologies, and a substantial proportion of their content can be categorized as extreme (Guhl et al., 2020). To circumvent the content moderation practices of social media platforms, extremist actors often obfuscate their true intention by purposefully presenting themselves as less extreme or using stylistic techniques that reduce reactance, such as humor (Schmid et al., 2024). This increases the difficulty for researchers to effectively differentiate between extremist and less or non-extremist actors connected to extremist circles. At the same time, extremist actors often strategically target non-extremist spaces and individuals, as they serve as critical gateways for the spread of extremist ideologies and narratives (Rothut et al., 2024). To effectively research and monitor the dissemination of extremist ideologies and the potential radicalization of those involved, it is essential to include actors and discourse spaces seemingly situated outside but in close proximity to the extremist core.

Any intrusion into individuals' privacy must be carefully justified, with the severity of that intrusion being proportionate to the potential benefits for researchers and the public. Additional measures are to be added to mitigate the severity of such intrusions (see below).

3.2.3. Solutions Provided by the CDT

3.2.3.1. Organizational Solutions

The organizational measures designed to protect research subjects affected by data processing in the context of the CDT can be divided into three main areas: first, the types of data that can be processed via the CDT; second, the conditions under which access to that data is granted; and third, data subject rights in regard to data deletion.

3.2.3.1.1. What Data Can Be Processed Through the CDT, and Under What Conditions?

The types of data permitted for processing within the CDT include: (a) relevant website or platform metadata (e.g., URL, account name, account IDs); and (b) a justification for the inclusion of the individual's data in the database (e.g., the classification of the individual as a right-wing extremist actor based on web and social media content). This justification arises from balancing public and individual interest. The third category, (c), comprises supplementary information about the respective individual, such as whether they are categorized as extremist or radical by researchers. This information may overlap with the second category. However, the rationale for inclusion is a legal obligation, whereas the categorization is open to academic debate and elaboration.

As a precondition for inclusion in the database, all individuals must belong to a defined set of relevant ideological categories of actors. These may include those who are part of the right-wing extremist spectrum,

individuals involved in ideologically mixed scenes where extremist ideas are disseminated, or those associated with right-wing conservatism (because they are targeted by extremist agitation). Individuals who cannot be categorized in this spectrum must not be included in the database. The classification of social media accounts may change over time due to shifts in communication practices or political reorientation of the account holder. Within the CDT framework, account labels refer to observed account activity at the time of inclusion rather than to stable attributes of actors. If data subjects request the correction or deletion of their account information, the CDT will update or remove the corresponding records.

As an additional measure, a minimum threshold of discursive relevance for individual accounts to be included in the database has been defined in agreement with data privacy consultants. Accounts or groups classified as belonging to the extremist core may be included if they have at least 100 followers or members on at least one social media platform. For other individuals, a minimum of 500 followers or members on at least one platform is required. A set of rules has been established for the processing of data obtained from the CDT. The data must not be published or shared with third parties, and they will be deleted as soon as the principles of good academic practice allow.

3.2.3.1.2. Who Gains Access and Under Which Conditions?

Access to the CDT is subject to strict conditions deduced from the legal basis. To gain access to the database, researchers must demonstrate that their research purpose is relevant to the field of far-right online research and that they possess the requisite qualifications to conduct scientific inquiry. Researchers are therefore permitted to use the CDT based on the following three criteria:

1. Research purpose: The utilization of account lists is explicitly prohibited for any purpose other than scientific research. The objective is solely to gain knowledge about the social phenomenon of the far-right, which serves the public interest. To gain access to the CDT, researchers must provide a rationale for their research project and demonstrate how it aligns with the purpose of the CDT.
2. Institutional affiliation: To be eligible for consideration, researchers must be employed at a publicly or privately funded research institution. They must also provide appropriate proof of this status. This requirement substantiates the existence of a research purpose in the public interest, validates their qualifications for scientific work, and ensures compliance with the fundamental data protection standards set forth in the GDPR and general standards of IT security.
3. Relevant research background: Access to the database is granted exclusively to those who can demonstrate a relevant background in far-right online research. This criterion ensures that the research is conducted in the public interest and that sensitive data are handled responsibly. Evidence of qualification may include relevant publications, doctoral projects, or affiliation with a research center or institute focused on the subject matter in question.

Once all the approval criteria have been met and formally confirmed by the CDT, each user must partake in a mandatory training session prior to database access. The training program includes an overview of the fundamental principles governing the handling of data collected during the project, along with a practical introduction to the use of the data editor. This introduction aims to ensure the responsible handling of sensitive data within the CDT, thereby protecting the individuals whose data are stored and the researchers involved in the project (see Challenge IV).

One important limitation is that access to the data is restricted to researchers based in EU countries, as the legal framework governing the CDT operates within the scope of the GDPR. This restricts the principle of equal access to data and therefore represents a structural limitation of the CDT model. However, given the sensitivity of the research area and the associated risks to research subjects and researcher safety, such restrictions are necessary to comply with legal and ethical obligations and to prevent potential harm.

3.2.3.1.3. Data Subject Rights

The CDT protects data subject rights according to the GDPR. We developed a workflow that enables individuals to request access to the data stored about them in the database and to request the deletion of their data. For ethical reasons, we agreed that account information will be deleted whenever such a request is addressed to the CDT, even in cases where continued storage of the data would be permissible under the GDPR.

Before providing individuals with information about the social media accounts, the CDT verifies the identity of the requesting person through a verification procedure. First, the requester must demonstrate control over the social media account in question, for example by posting a temporary verification code provided by the CDT. In addition, the CDT may request a secondary verification step, such as confirmation via the email address associated with the account. Information is only disclosed once the identity of the requester has been verified with reasonable certainty, to prevent unauthorized access to personal data.

3.2.3.2. Technical Solutions

Various technical measures are employed to ensure data security and to enable the trustee to operate smoothly. These measures can be broadly categorized into two groups: (a) solutions concerning deployment infrastructure and network access, and (b) software design decisions.

Most measures in the first category are standard for web applications; however, they are particularly important due to the sensitivity of the data and the political nature of the project. These measures include, but are not limited to: (a) The trustee can only be reached through an SSH connection, separating it from the open internet; (b) the deployment is isolated from other services hosted in the same data center by a containerized software stack and firewall rules; (c) the deployment infrastructure and software dependencies are kept up to date to mitigate vulnerabilities; (d) any irregular access, i.e., not through the provided user interface, is monitored; (e) a tiered backup strategy is employed to prevent data loss; and (f) access is granted only to users whose identity has been verified by two-factor authentication. An external security audit has been conducted to confirm the appropriateness of these safety measures.

The software design decisions are tailored to the requirements of collaboratively working on sensitive digital research data (see Figure 2). The data are divided into different permission groups with varying levels of read and write access based on the research purpose.

When adding a new entity to the database, researchers must provide justification for its inclusion. This justification must include the category of right-wing actors to which the entity belongs, as well as the scientific definitions that underpin its classification. Additional justifications can be submitted later to

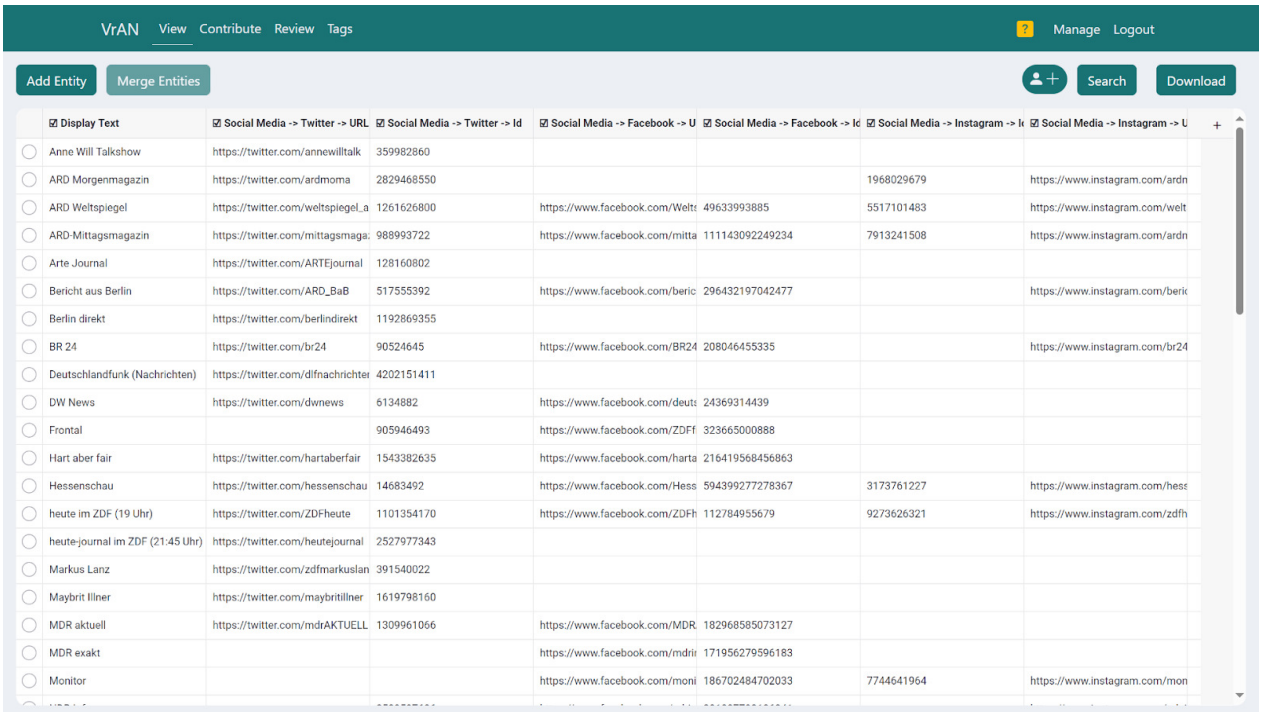

Figure 2. Screenshot of the online editor. Note: For data protection reasons as well as to avoid the amplification of potentially problematic actors, the editor has been filled with German media organization data (Schmidt et al., 2025) for the purposes of this screenshot.

support the continued inclusion of the entity. During bulk uploads of data, a single justification can be attributed to the entire set of new entities to simplify the process.

In line with the principle of data minimization, entities that remain unchanged for an extended period or have outdated justifications are marked as candidates for deletion. Deletion or update decisions are made periodically during the centralized curation process. Our organizational guidelines for good scientific practice ask to keep relevant research data accessible for at least 10 years (Leibniz Association, 2019).

3.3. Challenge II: Incentivizing Participation to Overcome the Collective Goods Problem

3.3.1. The Challenge

Despite the potential benefits of data sharing for the research community, there is limited incentive for individual researchers to share their data because they do not gain direct benefits from sharing. Instead, they may provide a competitive advantage to their peers, who could use their data for career advancement and research funding. This scenario creates a collective goods problem within the research community: Once a researcher contributes data to the CDT, all participants gain access to the data for their own purposes, potentially undermining the original contributor's competitive position.

An effective model for data sharing must, therefore, offer a solution to the collective goods problem. It should ensure that researchers perceive that the benefits of contributing to the CDT outweigh the costs associated with providing their account lists and actively contributing to the database.

3.3.2. Solutions Provided by the CDT

The co-creation showed that two primary aspects must be addressed to incentivize participation. First, it was emphasized that the CDT's greatest asset is access to its database, which provides a significant research advantage for participating researchers who can use the data for their research. This asset can be leveraged to encourage further contributions through a system of reciprocity: Researchers are granted access to the CDT only if they agree to contribute their data to the database. However, a challenge to this system may arise from imbalances between highly active contributors and less engaged participants, leading to concerns that more active contributors might fear sacrificing their competitive advantage.

This is addressed by the second mechanism of visibility (see below). It was suggested that individual contributions should be logged systematically so that contributors can receive academic recognition for their efforts in building and maintaining the database. First, it is necessary to acknowledge everyone's contributions, as this aligns with good scientific practice for publications and data collections. Second, the greater the prominence and relevance of the CDT within the research community, the more valuable individual contributions will become. Being associated with a reputable project can enhance contributors' future career prospects.

3.3.2.1. Organizational Solutions

Regarding reciprocity, all researchers who use the database are required to contribute to maintaining it. The minimum obligations include validating the data flagged for review (see technical solutions below) and reporting any additional changes discovered during the data collection process using the account information from the CDT. The scale of contribution can vary depending on the volume of data retrieved from the pool for research purposes. This commitment to reciprocity is outlined in the code of conduct that researchers must sign before gaining access to the CDT data.

Focusing on the first dimension of visibility—emphasizing individual contributions to the database—the database was designed to be citable, as detailed in the technical tracking below. To achieve this, the content and structure of the database will be documented in a data paper published at regular intervals. Researchers' contributions can then be (collectively) cited when the database is used. This presents a significant incentive, as citations are the primary currency in academia, reflecting the relevance and impact of research.

Listing contributors as authors on the data paper raises the question of the order of authors. The technical tracking and quantification systems detailed below enable the implementation of a meritocratic system, where contributors are listed in descending order based on the number of contributions they have made. While an egalitarian approach—such as alphabetical or random ordering—could have been considered, the meritocratic principle was chosen to enhance the incentive's effectiveness and mitigate potential imbalances related to research advantage (see the discussion about research advantage and reciprocity above).

The second dimension of visibility pertains to the academic reputation of the CDT. In our project, this reputation is enhanced by two key strategies: (a) involving prominent members of the research community as participants and establishing a board of users as part of the governance structure, and (b) documenting the CDT's impact on the research community by tracking and displaying instances of the CDT being cited in

academic works. However, since this aspect may conflict with other challenges, including researcher safety (see Challenge IV), the project primarily focuses on cultivating a strong reputation within the academic community rather than among the general public.

3.3.2.2. Technical Solutions

Focusing on technical solutions to ensure reciprocity, for researchers who use data from the database for their research, 5% of the data are automatically flagged for review. This selection process is based on identifying data points that have not been updated or validated for the longest duration (see Challenge III).

To ensure that researchers' contributions are visible, a technical system tracks these contributions. The database logs all edits. Researchers can initiate an "edit session," during which additional researcher accounts can be included for logging. The contributions made during such a session will then be attributed not only to the primary author of the update but also to the other participating researchers, thereby acknowledging the collaborative efforts that may have informed the update. This system will facilitate the quantification of the work invested by individuals or groups.

3.4. Challenge III: Ensuring Data Quality

3.4.1. The Challenge

Online phenomena such as far-right movements are characterized by their substantial scale and highly volatile nature, with frequently shifting accounts and platforms. Data sharing is limited due to the sensitive nature of the information, resulting in outdated lists that lack peer review. This contributes to issues such as a lack of consensus regarding the ideological categorization of actors, the absence of consensual definitions for the research area, and difficulties in identifying errors or gaps in the data pools (Kiess, 2011).

The CDT must address two main requirements: (a) It needs to facilitate and ensure a functional and organized collaborative process for the creation and maintenance of these account lists; (b) it requires a database capable of processing, storing, and providing all necessary information related to the creation and maintenance of these account lists.

3.4.2. Solutions Provided by the CDT

3.4.2.1. Organizational Solutions

The CDT's host organization assumes certain data curator functions. For entity management and central data points, such as platform data (ID, account name, link, etc.), the CDT will evaluate whether the edits submitted by users (researchers) can be verified by cross-checking them on the platforms. If they can be verified, the edits will be approved, and the database will be updated accordingly. In cases where the existing data pool overlaps with the new data, new and old data points will be cross-checked to strengthen data integrity.

The 5% review obligation for downloaded data not only establishes a principle of reciprocity between researchers, incentivizing participation (see Challenge II), but also allows for the ongoing review of the entire

database. This review mechanism ensures regular validation of the database's accuracy and therefore addresses the high volatility of the research field and enhances the overall quality of the database. In addition, outdated data might be deleted from the database after an appropriate amount of time (see Section 3.2).

3.4.2.2. Technical Solutions

The CDT is designed to store actors as entities and to connect various types of information associated with each entity (which may differ from one entity to another, such as the platforms where a specific entity holds an account). This flexible database affords the easy addition of new entities, platforms, accounts, or supplementary information. The versioning system relates edits to previous values, ensuring that past states of the database can be accessed by selecting a date, thereby aiding research reproducibility. The database accommodates ambiguity where necessary (e.g., one entity may have several ideological categorizations, reflecting academic debate) while enforcing clarity where expected (e.g., only one platform ID per platform account; however, this does not preclude the existence of multiple platform accounts with several IDs).

The database supports simultaneous access and updates to facilitate collaborative efforts. Data updates can be performed manually through the software interface or through bulk uploads. String similarity metrics are integrated to detect (near-)duplicates, assist in updating efforts, and prevent database pollution. When duplicates are identified, they can be merged; however, the CDT retains the authority to approve or deny these merge requests, thus preventing incorrect updates.

3.5. Challenge IV: Maintaining Researcher Safety

3.5.1. The Challenge

A fourth challenge emerged from the co-creation workshops. Sensitive digital research topics, such as the far-right online, are often highly politicized and may encompass issues like violent extremism and terrorism, which have direct connections to physical violence. Researchers reported that they feel threatened by right-wing extremist networks due to their scientific work. In the past, researchers have been the target of hate campaigns, particularly when their work has gained significant attention in traditional media or on social media platforms (Global Witness, 2023; Nogrady, 2021). Within media and communication research, abuse and threats are prevalent (Obermaier et al., 2024; Seeger, Frischlich, Obermaier, Schmid, Riesmeyer, & Menke, 2024).

Affected researchers may experience anxiety, sleep problems, fear for their safety, or experience actual physical attacks (Global Witness, 2023). On a societal level, the consequences of a withdrawal from public discourse can be observed (Seeger, Frischlich, Obermaier, Schmid, Riesmeyer, & Menke, 2024). This is particularly concerning, as underrepresented groups, such as female scientists, seem to be disproportionately targeted by online abuse (Hodson et al., 2018; Obermaier et al., 2024; Veletsianos et al., 2018). Based on these experiences, researchers may feel apprehensive that their contribution to the data pool of the CDT could expose them to harassment.

A clear dilemma therefore emerged in the co-creation workshops. On the one hand, scientists wish to maintain high visibility within the research community, among policymakers, and potentially even among the general public. On the other hand, researchers may wish to minimize the exposure of their work within right-wing milieus to avoid harassment when involved in the collection and analysis of account and communication data.

Universities and research institutes have been unanimously criticized (Seeger, Frischlich, Obermaier, Schmid, & Schulze, 2024) for lacking structures to support researchers facing harassment. In the worst cases, researchers are left to manage these challenges on their own, which can become a significant psychological burden. In addition, pursuing legal action against insults and threats requires considerable material resources. Initiatives such as SciComm-Support (Germany) and SafeScience (the Netherlands) seek to address these challenges and provide valuable assistance to researchers facing online harassment and other forms of attack. However, a substantial share of the responsibility still rests with individual researchers, rather than being embedded in systematic, institutionally anchored support structures within their respective universities and research institutes.

3.5.2. Support Provided by the CDT

The CDT cannot fully resolve this conflict, but it aims to support researchers in managing it as well as possible. In general, access to the CDT is heavily restricted, both by organizational as well as technical measures. Researchers have the option not to be listed as authors of the CDT (see above). While this choice means they will not be citable or gain visibility for their contributions, they can still access and contribute to the resource for their research. In addition, users can be presented to other users with a pseudonym if they wish. If a legal request is made by individuals listed in the database to access their data, the CDT must comply and release the information, but any details that could reveal the identity of researchers involved in handling that entry will be redacted. A rights management system is in place that protects data uploaded by a user from edits by others. Finally, during the onboarding process for new researchers (see Challenge I), the awareness training provided includes a module on personal security issues, offering guidance and resources on how to protect themselves against potential threats. This advice is based on best practices for dealing with dangerous situations that are increasingly emerging within the research community (Tollefson, 2024). Overall, these measures align with the CDT's general goal of reducing the personal risks and burdens on individual researchers and institutionalizing them as far as possible.

3.6. Challenge V: Platform ToS and Contractual Constraints

In addition to statutory data protection law, digital research practices are shaped by contractual frameworks imposed by platforms through their ToS. These provisions differ substantially across platforms and are subject to frequent modification. They may include restrictions on automated data collection, the reuse of platform data, or the sharing of data with third parties. This creates potential tensions between platform rules and academic research practices.

The CDT does not collect data directly from digital platforms, does not provide scraping tools or API-based harvesting services, and does not circumvent technical protection measures. Its purpose is limited to the structured storage and controlled sharing of actor-level metadata that researchers have collected within their projects.

Responsibility for assessing and complying with platforms' ToSs remains with the individual researchers and their host institution. It is important to note that the legal status and enforceability of ToSs in academic research contexts can vary across jurisdictions. In Germany, it is contested to what extent contractual platform provisions can restrict research activities that are grounded in statutory law and serve the public interest (Golla & Von Schönfeld, 2019). The CDT therefore does not require proof of ToS compliance as a precondition for contribution and does not assume responsibility for enforcing contractual platform rules. During onboarding, researchers must confirm that their data collection was conducted in accordance with applicable statutory law and institutional requirements. The evaluation of potential contractual risks associated with specific platforms is located at the level of the contributing research project, even though such risks may, in practice, primarily concern the suspension of platform research accounts.

4. Conclusion

Our concept of a CDT addresses key challenges in sharing sensitive digital research data, such as in far-right online studies. It offers a collaborative infrastructure for managing and maintaining account lists central to digital data collection. The CDT provides a legally compliant and ethically sound framework, leveraging the EU GDPR research exemption alongside technical and organizational safeguards.

By reducing legal risks and fostering collaboration, the CDT helps overcome data-sharing barriers. A core feature is reciprocity: Researchers contribute data to gain access, ensuring sustainability. Contributions are acknowledged via a citable database paper, providing academic recognition.

Data quality is maintained through continuous updates and peer review, ensuring accuracy in a dynamic research field. Although researcher safety remains a concern—especially in politicized contexts—the CDT includes opt-out mechanisms for public attribution and offers risk-mitigation resources. Still, broader institutional support from universities and research institutes is needed to fully protect researchers, particularly in the form of systematic, institutionally anchored structures such as clear reporting procedures, legal and psychological counseling, IT-security assistance, and support in responding to online harassment or political attacks.

The next challenge is ensuring the CDT's long-term sustainability while expanding its user base. Currently, trusted participants from co-creation workshops are uploading and sharing data. Continued engagement from this community, along with the integration of new users, is vital.

Researchers not only use and maintain the database but also contribute to its governance. Given the evolving nature of digital research, the CDT must remain adaptable, guided by continuous input from its community. Governance issues—such as breaches of reciprocity—should be resolved through collaborative decision-making. Establishing a scientific advisory board could strengthen community involvement and support ongoing development.

In parallel, securing long-term funding is essential. Many research infrastructure projects rely on short-term grants, which often lapse before ensuring continuity. Sustainable infrastructures like the CDT require mid- to long-term funding commitments from public institutions. Existing research institutes can be motivated to contribute to the long-term funding of CDTs because they either benefit directly through use by their own

researchers or receive third-party funding for developing research data infrastructures. As the operating costs of a CDT are comparatively low—since data curation is largely performed by the community—the model is particularly attractive when institutions aim to support sustainable data governance with limited additional infrastructure and personnel costs. Running a CDT (using existing software solutions) can be achieved with an IT developer for maintaining the database editor and bug fixing (possibly part-time if no substantive improvements are planned), a data curator for internal data review processes and user support, and a scientific community manager responsible for governance issues. AVERA, the community data trustee described in this article, will initially be operated through third-party funding and hosted within DP-REX (Data Portal for Racism and Right-Wing Extremism Research), a research data infrastructure dedicated to extremism research as introduced in Section 3.1. For longer-term continuation, there is the possibility of initiating a structured process to transition AVERA into a regular GESIS service supported by institutional funding. This would require an active user community that engages with the data and demonstrates their scientific value to GESIS governance boards.

Beyond far-right studies, the CDT model can be adapted to other domains of sensitive digital research that rely on account lists—such as political communication, disinformation, and hate speech. These lists are foundational for many digital data projects.

The layer of scalability concerns the localization of the data trustee. For reasons related to data protection and research ethics, our specific data trustee will focus on the German-language sphere and will be accessible only to researchers based in countries within the EU. While this necessarily limits its immediate international applicability, we contend that the general data trustee model we have developed can readily function as a blueprint for adaptation and implementation in other linguistic and regulatory contexts.

By providing a secure, collaborative framework for sharing and managing social media account data, the CDT tackles one of digital media and communication research's biggest obstacles: accessing sensitive data. Its flexibility supports diverse research goals, and its legal and technical foundation allows it to evolve alongside the digital landscape.

Acknowledgments

The authors would like to thank Lina Stille and Daniel Wehrend for their valuable contributions to the development of the CDT concept and infrastructure. We are also grateful to all participants in the co-creation workshops for their constructive feedback, critical discussions, and continued engagement throughout the project. In addition, we thank our colleagues and collaborators from the broader research community who provided comments, expertise, and practical input during the conceptualization and implementation phases of this work.

Funding

This work was supported by the German Research Ministry with grant numbers 01UG2151A&D, 16DTM208A&B, 16DTM404A-E, and also funded by NextGeneration EU. Publication of this article in open access was made possible through the institutional membership agreement between GESIS – Leibniz Institute for the Social Sciences and Cogitatio Press.

Conflict of Interests

The authors declare no conflict of interests.

Data Availability

The community data trustee AVERA is accessible here: https://datenportal-rechtsextremismus.de/en/rex-forschungsdaten/digitale-rex-daten/avera.

LLMs Disclosure

ChatGPT as well as Grammarly have been used for copy editing of the manuscript.

References

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*, Article 971974. https://doi.org/10.3389/fdata.2022.971974

Beuthner, B. E. C., Topci, R., Derks, M., Franke, T., Seelke, S., Puls, M., Schuster, A., Toischer, K., Valentova, M., Cyganek, L., Zeisberg, E. M., Jacobshagen, C., Hasenfuss, G., & Nussbeck, S. Y. (2020). Interdisciplinary research on aortic valve stenosis: A longitudinal collection of biospecimens and clinical data of patients undergoing transcatheter aortic valve replacement. *Open Journal of Bioresources*, *7*, Article 3. https://doi.org/10.5334/ojb.65

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

Castells, M. (2013). *Communication power*. Oxford University Press.

Conway, M., Scrivens, R., & Macnair, L. (2019). *Right-wing extremists' persistent online presence: History and contemporary trends*. International Centre for Counter-Terrorism. https://icct.nl/publication/right-wing-extremists-persistent-online-presence-history-and-contemporary-trends

Delacroix, S., & Lawrence, N. D. (2019). Bottom-up data trusts: Disturbing the 'one size fits all' approach to data governance. *International Data Privacy Law*, *9*(4) 236–252. https://doi.org/10.1093/idpl/ipz014

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 (Eds.), *Internet of things and big data technologies for next generation healthcare* (pp. 199–226). Springer. 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

Franzke, A. S., Bechmann, A., Zimmer, M., Ess, C., & Association of Internet Researchers. (2020). *Internet research: Ethical guidelines 3.0*. Association of Internet Researchers. https://aoir.org/reports/ethics3.pdf

Fürneisen, M., Rau, J., Jungmann, N., Stille, L., Siegers, P., & Wiedemann, G. (2026). *Collaborative social media account editor* (Version 0.6.0) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.18890351

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

Golla, S. J., & Von Schönfeld, M. (2019). Kratzen und Schürfen im Datenmilieu—Web Scraping in sozialen Netzwerken zu wissenschaftlichen Forschungszwecken. *Kommunikation & Recht*, *22*(1), 15–21.

Guhl, J., Ebner, J., & Rau, J. (2020). *The online ecosystem of the German far-right*. Institute for Strategic Dialogue.

https://www.isdglobal.org/wp-content/uploads/2020/02/ISD-The-Online-Ecosystem-of-the-German-Far-Right-English-Draft-11.pdf

Hase, V., Ausloos, J., Boeschoten, L., Pfiffner, N., Janssen, H., Araujo, T., Carrière, T., de Vreese, C., Haßler, J., Loecherbach, F., Kmetty, Z., Möller, J., Ohme, J., Schmidbauer, E., Struminskaya, B., Trilling, D., Welbers, K., & Haim, M. (2024). Fulfilling data access obligations: How could (and should) platforms facilitate data donation studies? *Internet Policy Review*, *13*(3). https://doi.org/10.14763/2024.3.1793

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*, *23*(8). https://doi.org/10.5210/fm.v23i8.9136

Hoffmann, M., & Rone, J. (2023). Interconnected realities: The hybrid dynamics of far-right online and offline mobilisation. In M. Weisskircher (Ed.), *Contemporary Germany and the fourth wave of far-right politics* (pp. 57–75). Routledge.

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

Horst, H. A., & Miller, D. (Eds.). (2020). *Digital anthropology*. Routledge. https://doi.org/10.4324/9781003085201

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

Kiess, J. (2011). Rechtsextrem—extremistisch—demokratisch? Der prekäre Begriff „Rechtsextremismus" in der Einstellungsforschung. In Forum für kritische Rechtsextremismusforschung (Ed.), *Ordnung. Macht. Extremismus* (pp. 240–260). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-93281-1_11

Kreutzer, S., Heimer, T., Nachtigall, H., Pschorn, L., Bauer, F., Blind, K., Martin, N., von Grafenstein, M., 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*. RWTH Aachen University. https://doi.org/10.18154/RWTH-2024-04375

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*, *369*(6507), 1060–1062. https://doi.org/10.1126/science.aaz8170

Leibniz Association. (2019). *Good scientific practice and ombuds services*. https://www.leibniz-gemeinschaft.de/en/about-us/leibniz-integrity/good-scientific-practice-and-ombuds-services

Lindstrøm, M. D., & Kropp, K. (2017). Understanding the infrastructure of European research infrastructures—The case of the European Social Survey (ESS-ERIC). *Science and Public Policy*, *44*(6), 855–864. https://doi.org/10.1093/scipol/scx018

Lorenz-Spreen, P., Oswald, L., Lewandowsky, S., & Hertwig, R. (2023). A systematic review of worldwide causal and correlational evidence on digital media and democracy. *Nature Human Behaviour*, *7*(1), 74–101. https://doi.org/10.1038/s41562-022-01460-1

Miller-Idriss, C. (2020). *Hate in the homeland: The new global far right*. Princeton University Press. https://doi.org/10.2307/j.ctv10tq6km

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*. Open Science Framework. https://osf.io/cn4e3/files/jswav

Piller, F. T., Ihl, C., & Vossen, A. (2010). *A typology of customer co-creation in the innovation process*. SSRN. https://doi.org/10.2139/ssrn.1732127

Pirro, A. L. P. (2023). Far right: The significance of an umbrella concept. *Nations and Nationalism*, *29*(1), 101–112. https://doi.org/10.1111/nana.12860

Rau, J., Münch, F. V., & Asli, M. (2021). SOCRATES: Social media research assessment template for ethical scholarship. Leibniz-Institute for Media Research; Hans-Bredow Institute (HBI) https://leibniz-hbi.github.io/socrates

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

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

Schmid, U. K., Schulze, H., & Drexel, A. (2024). Memes, humor, and the far right's strategic mainstreaming. *Information, Communication & Society*, *28*(4), 537–556. https://doi.org/10.1080/1369118X.2024.2329610

Schmidt, J.-H., Merten, L., & Münch, F. V. (2025). *Die „Datenbank Öffentlicher Sprecher:Innen" (DBÖS)* [Data set]. Open Science Framework. 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*. Open Science Framework. https://osf.io/yu6w2

Seeger, C., Frischlich, L., Obermaier, M., Schmid, U. K., & Schulze, H. (2024). *Hate Speech und Angriffe auf Wissenschaftler* innen. Ein Forschungsüberblick*. Berlin-Brandenburgische Akademie der Wissenschaften. https://wissenschaft-im-dialog.de/documents/217/TransferUnit_Forschungsueberblick_Hatespeech.pdf

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

Sinnott, R. O., Li, Q., Mohammad, A., & Morandini, L. (2023). The Australian Digital Observatory: Social media collection, discovery and analytics at scale. In C.-H. Hsu, M. Xu, H. Cao, H. Baghban, & A. B. M. Shawkat Ali (Eds.), *Big data intelligence and computing: International conference, DataCom 2022* (pp. 316–327). Springer. https://doi.org/10.1007/978-981-99-2233-8_23

Theocharis, Y., & Jungherr, A. (2020). Computational social science and the study of political communication. *Political Communication*, *38*(1/2), 1–22. https://doi.org/10.1080/10584609.2020.1833121

Tollefson, J. (2024, September 20). Harassed? Intimidated? Guidebook offers help to scientists under attack. *Nature*. 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

Weller, K., & Kinder-Kurlanda, K. E. (2016). A manifesto for data sharing in social media research. In *WebSci '16:*

Proceedings of the 8th ACM Conference on Web Science (pp. 166–172). Association for Computing Machinery. https://doi.org/10.1145/2908131.2908172

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

About the Authors

Jan Rau is a researcher at the Media Research Methods Lab at Leibniz Institute for Media Research | Hans-Bredow-Institut (HBI) in Hamburg. His research centers around issues like the far-right online, disinformation, and platform governance. As the project coordinator of AVERA at the HBI, he co-developed the community data trustee.

Nils Jungmann is a data specialist at Tern ZE GmbH in Cologne. Holding an MA in political communication (Düsseldorf) and a BA in political science (Mannheim), he previously spent eight years at GESIS, contributing to the projects GLES and Com-DTM. His research explores the psychology of voting behavior.

Moritz Fürneisen is a software developer at the Media Research Methods Lab at Leibniz Institute for Media Research | Hans-Bredow-Institut (HBI) in Hamburg. He developed the technical infrastructure of the AVERA community data trustee.

Gregor Wiedemann (PhD) is a senior researcher in computational social science at the Leibniz Institute for Media Research | Hans-Bredow-Institut, where he heads the Media Research Methods Lab. As principal investigator of Com-DTM, and Com-DS, he led the development of the technical infrastructure of the AVERA community data trustee.

Heidi Schulze is a research associate at the Department of Media and Communication at LMU Munich. In her research, she focuses on online radicalization dynamics, radical/extremist (group) communication on alternative social platforms and fringe communities, as well as characteristics and audiences of alternative and hyperpartisan news websites.

Pascal Siegers (PhD) is team lead for national studies at the Department of Survey Data Curation at GESIS – Leibniz Institute for the Social Sciences in Cologne. As one of the principal investigators of DP-REX, Com-DTM, and Com-DS, he led the projects funding the development of the AVERA community data trustee.