

Monitoring Electoral Democracy: Discovering and Reusing Text and Multimedia Data for Political Communication Research

Jakob-Moritz Eberl ¹, Jana Bernhard-Harrer ¹, Mariken van der Velden ²,
Sofia Gil Clavel ², Paul Balluff ³, Johannes B. Gruber ⁴, Noëlle S. Lebernegg ⁵,
Kasper Welbers ², Marvin Stecker ¹, Claes Holger de Vreese ⁶,
Wouter van Atteveldt ², Rens Vliegthart ⁵, and Hajo G. Boomgaarden ¹

¹ Department of Communication, University of Vienna, Austria

² Department of Communication Science, Vrije Universiteit Amsterdam, The Netherlands

³ Department of Computational Social Science, GESIS – Leibniz-Institute for the Social Sciences, Germany

⁴ Department of Data Services for the Social Sciences, GESIS – Leibniz-Institute for the Social Sciences, Germany

⁵ Strategic Communication Group, Wageningen University & Research, The Netherlands

⁶ Amsterdam School of Communication Research, University of Amsterdam, The Netherlands

Correspondence: Mariken van der Velden (m.a.c.g.vander.velden@vu.nl)

Submitted: 14 February 2026 **Accepted:** 7 September 2026 **Published:** 22 October 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

Political communication research relies on a fragmented landscape of media sources, text and multimedia corpora, analytical tools, and research workflows. Researchers face two related barriers: identifying which resources exist and determining whether and under what conditions they can be accessed and reused. This article presents Monitoring Electoral Democracy (MEDem), a distributed European research infrastructure for comparative political communication research, with particular attention to its Service Center for Text and Multimedia Data. The Service Center develops two complementary services: MEDem Meteor provides a curated, knowledge-graph-based discovery layer that connects media sources, datasets, archives, analytical tools, and publications; MEDem Amsterdam Content Analysis Toolkit supports the hosting, controlled reuse, and non-consumptive analysis of text and multimedia corpora. Together, the services address discovery and access as interconnected infrastructural problems. MEDem Meteor makes dispersed resources, relationships, and coverage gaps visible, while MEDem Amsterdam Content Analysis Toolkit makes access conditions transparent and enables reuse without redistributing protected content. This combination can reduce duplicated data collection, support more theoretically informed source selection, and strengthen cumulative and comparative research. Its realization, however, depends on sustained curation, community participation, and long-term institutional commitment.

Keywords

comparative research; content analysis; FAIR data; open science; research infrastructures

1. Introduction

Political communication research relies on a vast but fragmented landscape of media sources, text and multimedia corpora, computational tools, and analytical workflows to investigate public debates, political attitudes, media effects, and democratic processes (Theocharis & Jungherr, 2021). Yet researchers face a two-step hurdle in identifying and using existing resources. First, they need to know that a particular dataset, media source, corpus, or tool even exists. Second, they need to determine whether and under what conditions they can access and use it. What is effectively available therefore varies across researchers and institutions, depending both on awareness and accessibility. Copyright, licensing restrictions, data protection requirements, and platform policies further constrain the sharing and reuse of data at the center of many research questions. As a result, datasets and tools are repeatedly collected or developed in parallel, and some research questions remain unaddressed simply because the technical, legal, or financial barriers are too high. These burdens are also not distributed evenly. A project-based model of data production systematically advantages well-resourced institutions and established research groups that can absorb the costs of data collection, storage, technical development, and legal expertise, while placing early-career researchers, scholars at less well-funded institutions, and researchers working in smaller language communities at a systematic disadvantage (Bahlai et al., 2019; Whitaker & Guest, 2020).

Against this background, the pan-European research infrastructure Monitoring Electoral Democracy (MEDem) is being established to address these structural barriers by treating data and analytical resources as shared scientific goods and making surveys, political texts, mass media and platform content, institutional and contextual data more findable, accessible, interoperable, and reusable (FAIR). MEDem provides a shared infrastructural layer for discovering, harmonizing, linking, and sustaining existing data and analytical resources across countries and over time. Through standardized metadata, machine-readable identifiers, harmonization services, conceptual mapping, and legal and ethical support, it aims to reduce the technical and methodological burdens currently borne by individual researchers and projects (see Boomgaarden et al., 2025).

Within MEDem, the Service Center for Text and Multimedia Data (SCTMD) addresses the particular challenges that fragmented and heterogeneous political communication data, associated legal constraints, and diverse analytical tools pose for communication scholars. Its two central services target complementary bottlenecks. MEDem Meteor provides a curated and community-maintained discovery layer for identifying and connecting relevant sources, corpora, datasets, analytical tools, and expertise. MEDem Amsterdam Content Analysis Toolkit (AmCAT) provides a shared environment for transparent and reproducible research, supporting fine-grained access controls and non-consumptive analysis for the reuse of text and multimedia data within legal and ethical boundaries. Together, and in conjunction with tools developed and maintained by other MEDem Service Centers, they aim to transform dispersed resources and workflows into maintained infrastructures for cumulative and comparative political communication research.

Beyond the SCTMD, additional MEDem Service Centers develop complementary services for survey-data discovery, data harmonization, conceptual mapping, linkage (e.g., MEDem ATLAS), legal and ethical support, and training. These services are intended to interoperate across the infrastructure, allowing researchers to move between different stages and data domains of the research process.

This article briefly presents the current architecture and services of MEDem and then gives particular attention to the SCTMD and its two central components, MEDem Meteor and MEDem AmCAT.

2. What is MEDem?

MEDem is a distributed European research infrastructure for the comparative study of democratic actors, processes, and contexts. Its scope encompasses different empirical domains involved in the democratic process, including citizens and their political attitudes and behaviors, political parties and candidates, media and political communication, policy outputs, and the institutional and societal contexts in which these interactions take place. Research on political representation, responsiveness, electoral behavior, and media effects increasingly requires connections across these domains, as well as across countries and over time. MEDem is designed to provide the infrastructural conditions under which such connections become systematically visible and possible.

MEDem does not primarily collect new data, nor is it intended to replace established surveys, research projects, or archives. Instead, it operates above and between existing research groups and research projects. These include both established data collections and new research communities that retain their scientific identity, methodological autonomy, and governance structures. MEDem's role is to provide them with common services and reference structures that allow their resources to become more findable, accessible, comparable, linkable, and reusable across research contexts. In this respect, MEDem addresses a fundamental coordination problem: Many relevant datasets and analytical resources already exist, but they remain distributed across institutions and projects, use heterogeneous documentation and coding practices, and frequently lack the common identifiers and interfaces required for systematic integration, all of which, in fact, limit their findability, accessibility, interoperability, and reusability.

A central element of this approach is the identification of the real-world entities to which research data refer. Political communication research routinely generates data about countries, elections, parties, candidates, governments, media outlets, and other actors or organizations. Yet, different datasets may refer to the same entity using different names, codes, spellings, or classifications. MEDem seeks to enrich existing data with standardized, machine-readable identifiers that make such equivalences explicit and enable semi-automated linkage across otherwise separate resources. In combination with services for harmonizing variables and measurements, this creates the possibility of connecting data that were originally collected independently and for different research purposes.

The objective is not to eliminate methodological heterogeneity or impose uniform measurement standards. On the contrary, much of the scientific value of the field lies precisely in its diversity of theoretical traditions, data types, and methodological approaches. MEDem instead aims to make this heterogeneity transparent and at the same time interoperable: Different operationalizations can remain different while becoming easier to discover, analyze, compare, and, where scientifically appropriate, triangulate, harmonize, or link.

Throughout this process, researchers retain full control over whether and how different data, operationalizations, and methods are combined. This distinction between standardizing interoperability (i.e., providing metadata schemas that enable interoperability of different data) and standardizing research (i.e., restricting the heterogeneity of methods and theories) itself is central to MEDem's infrastructural approach. For example, ideological position may be operationalized using dictionary-based measures, text-scaling models, or supervised classifiers. MEDem can link these tools through the shared concept they seek to measure while documenting their different analytical operations, assumptions, and data requirements, thereby enabling researchers to compare alternative approaches without privileging or prescribing any particular one. To provide these functions, MEDem is organized as a distributed research infrastructure consisting of a central coordination unit and specialized Service Centers that address different infrastructural needs.

3. The SCTMD: Addressing Fragmentation, Discovery, and Reuse

Text and multimedia data are among the field's most important empirical resources, but also among the most difficult to discover, share, and reuse. Relevant (social) media sources, corpora, datasets, analytical tools, and expertise are dispersed across projects, institutions, countries, and languages; metadata and documentation practices vary widely; and copyright, licensing agreements, data protection requirements, and platform policies frequently restrict access to and redistribution of the underlying material. The SCTMD addresses these interconnected challenges through two complementary services. MEDem Meteor focuses on the discovery of existing resources. It is not a general-purpose repository and does not require researchers to upload their datasets or software. Instead, it records structured metadata about such resources and, where available, links to the external locations where they are held (e.g., Dataverse, OSF, GitHub; Balluff et al., 2023). MEDem AmCAT provides an environment for hosting, sharing, and analyzing text and multimedia data under configurable access conditions, particularly where controlled access or non-consumptive analysis is legally or ethically recommended or required (Gruber & van Atteveldt, 2025).

3.1. MEDem Meteor: A Tool for Inventory and Discovery in Political Communication

Research in communication science faces a paradox. The volume and diversity of potentially relevant text and multimedia sources have expanded dramatically, while the ability to systematically identify, compare, and coordinate them across countries, languages, platforms, and media systems has not kept pace. Researchers often rely on ad hoc source lists, project-specific documentation, or commercial archives shaped by institutional licensing arrangements and, therefore, by what they are aware of and can access rather than by theoretically meaningful characteristics of media systems (Esser & Vliegenthart, 2017; Lind et al., 2025). As a result, sampling is often driven by availability, cross-national comparability remains difficult, and the accumulation of knowledge about media landscapes is hindered. In a similar vein, the increasing adoption of computational methods has led the computational political communication subfield—but also the field more generally—to evolve rapidly, making it difficult for individual researchers to keep track of available analytical tools and contributing to duplicated efforts, inefficient use of resources, and limited reuse of both data and software (Stecker et al., 2024).

MEDem Meteor was first developed within the H2020 OPTED project (see Balluff et al., 2023), but the end of the project also brought the original funding and organizational framework for its continued maintenance

and development to a close. Its integration into MEDem now provides a new institutional setting for its further development and long-term operation. Within MEDem, it now serves as the central inventory and discovery layer for political communication resources for text and multimedia analysis. Its architecture combines a user-facing discovery and contribution interface with a knowledge-graph backend, as illustrated in Figure 1.

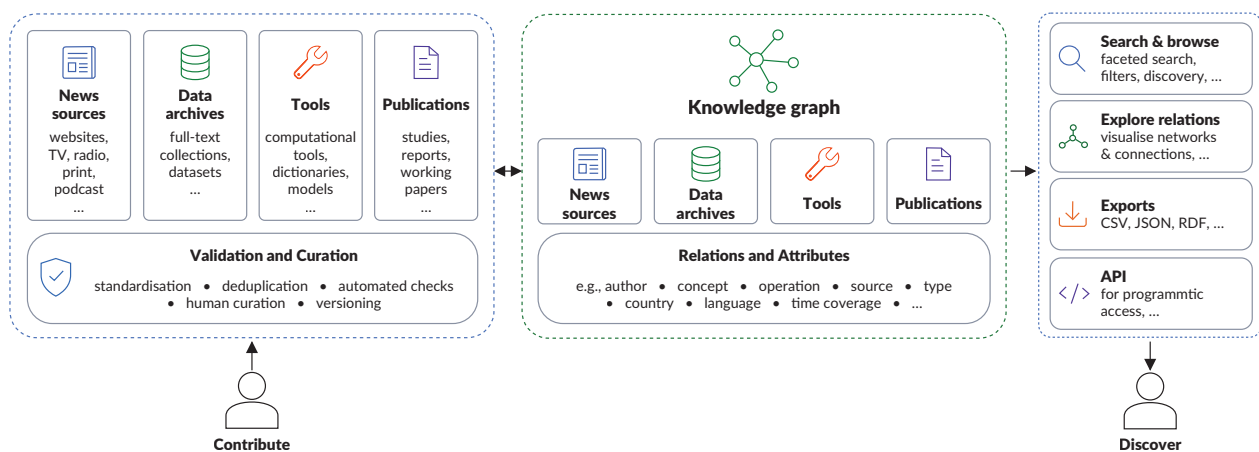


Figure 1. Overview of the MEDem Meteor architecture. Notes: The left panel illustrates the contribution workflow, through which users submit information on tools, news sources, data archives, publications, and related entities via the MEDem Meteor website; submitted entries undergo validation and curation before being integrated into the knowledge graph; the central panel represents the knowledge graph, which stores heterogeneous entities together with their semantic relations and entity-specific attributes; the right panel illustrates the discovery workflow, through which researchers can search and browse the inventory, explore relations between entries, export data and metadata, and access them programmatically through the API.

The technical backend is structured as a knowledge graph, “a large semantic net” (Fensel et al., 2020, p. 6), in which heterogeneous research-relevant entities—including news sources, political parties, analytical tools, full-text archives, datasets, and publications—are understood as nodes, described through entity-specific metadata, and connected through explicit relations. The metadata stored in MEDem Meteor are entity-specific. For news sources, these can include names and aliases, channel, source type, and availability in archives or datasets. For tools, they can include the analytical operation they perform and the concepts they help operationalize (e.g., sentiment). Shared properties establish connections between different types of entities within the knowledge graph.

Rather than functioning as a flat catalogue, MEDem Meteor therefore represents these resources as part of a connected ecosystem. Its discovery workflow, shown on the right-hand side of Figure 1, builds on this knowledge-graph structure and allows users to perform targeted queries as well as move quickly back and forth across entries from any starting point. A researcher can begin with a language-specific tool, move to the datasets in which it has been used, continue to the news sources contained in these datasets, and then move on again to related publications or other tools performing the same operations (e.g., dictionary) or measuring the same concepts. The same process can be reversed just as easily. Additionally, researchers can access this information both through the user-friendly website and programmatically via a free API, allowing them to query and retrieve data and metadata.

Crucially, MEDem Meteor is explicitly designed as a community-curated infrastructure with a platform-level governance model that combines distributed contributions with centralized oversight. The contribution workflow shown on the left-hand side of Figure 1 assigns differentiated responsibilities to contributors, reviewers, and administrators and embeds submissions within a structured validation and curation process. Anyone can create an account and thereby acquire contributor rights. Contributors can propose new entries through a standardized input process or flag errors or omissions in existing ones. Reviewers designated by the SCTMD inspect these submissions, correct or enrich them where necessary, and decide whether they should be accepted into the system. Administrators (i.e., members of the SCTMD) manage permissions and oversee the broader technical and organizational maintenance of the platform. Where possible, MEDem Meteor accesses and integrates external data sources such as Wikidata to enable entity linkage and add contextual information without requiring individual users to enter it manually.

By embedding sources, datasets, tools, publications, and other resources within a shared discovery layer, MEDem Meteor enhances their discoverability and facilitates reuse beyond their original contexts. This can also translate into academic recognition, as researchers who document and share their resources increase the likelihood that these outputs will be used and cited by others. Contributing to the inventory can therefore generate benefits not only for the wider research community, by highlighting heterogeneity in concepts and approaches and reducing duplication of effort, but also for contributors themselves. MEDem Meteor's ability to realize these benefits, however, depends on the breadth, quality, and continued curation of the resources represented in the inventory, a challenge discussed in Section 6.

3.2. MEDem AmCAT: A Tool for Data Annotation and Sharing of (Protected) Text and Multimedia Corpora

The practical ability to share, reuse, and collaboratively analyze large-scale text and multimedia materials drawn from news media, platforms, and other mediated communication environments remains severely constrained. Most news content is protected by copyright and licensing agreements, while much of the material collected from platforms and other digital sources contains personal data protected under the General Data Protection Regulation (Fiesler et al., 2020). Such data cannot be made unconditionally publicly accessible—not because researchers are unwilling to share, but because they are not allowed to (van Atteveldt et al., 2021). As a result, content analysis workflows are frequently fragmented, replication is often impossible in practice, and cumulative research is impeded by the necessity to repeatedly reconstruct datasets that cannot be shared (Assenmacher et al., 2023; van Atteveldt et al., 2019).

AmCAT forms the technological backbone through which MEDem addresses this bottleneck. AmCAT was originally developed in the late 2000s (van Atteveldt, 2008) and has since been continuously maintained and extended within the computational communication science community (van Atteveldt et al., 2023). As a MEDem service, MEDem AmCAT is being developed to host text and multimedia corpora for the research community and rests on a simple idea: Access to a corpus need not be all-or-nothing, but it can be constrained depending on the user and specific use case. A corpus consists of documents with multiple fields, for example, an identifier, a publication date, a headline, the full text, images, and annotations. Legal restrictions typically concern only some of them. MEDem AmCAT therefore implements differential access in which data owners decide, per field and per user, what is visible to whom. Figure 2 exemplifies this with a corpus of news coverage from the perspective of an external researcher: identifiers, dates, headlines, and

annotations are readable, while the copyright-protected full texts and images containing personal data are withheld. A privileged user, such as a member of the team that gathered the material, sees all fields of the very same corpus. Instead of forcing an entire dataset offline because parts of it are protected, differential access confines each restriction to the specific fields it actually concerns; everything else can be shared.

ID 	1	2	3	...
Date 	2026-07-01	2026-07-03	2026-07-03	...
Headline 	Coalition splits	Rain until Friday	Government deal	...
Text 	withheld: copyright-protected			...
Image 	withheld: contains personal data			...
Annotation 	political	not political	political	...

Figure 2. The non-privileged researcher’s view of a corpus stored in MEDem AmCAT (e.g., any external researcher). Notes: Only fields cleared for sharing are readable; protected data (i.e., copyrighted or personal) are withheld; a privileged user would see all fields; MEDem AmCAT withholds fields, not documents—all researchers query the same corpus, while metadata and annotations cleared for sharing remain accessible even when the original data cannot be displayed.

Making these distinctions operational requires more than technical access controls. The conditions attached to each corpus and field must also be documented consistently. Before a corpus is made available to other researchers, SCTMD staff therefore review its accompanying documentation for completeness and consistency, including information on data provenance, field structure, access conditions, and applicable legal or ethical restrictions. This documentation review does not constitute a substantive validation of the data or their scientific quality, but ensures that access and reuse conditions are transparently and consistently documented.

Differential access thus determines what users can see. The principle of non-consumptive research extends what they can do. Non-consumptive research comprises computational analyses of text, image, and audio material that do not require researchers to consume—that is, read, view, or listen to—the raw source material (Gruber & van Atteveltdt, 2025). Because analyses are executed within the infrastructure, where protected fields remain available to the system even when hidden from users, researchers can analyze material they are not permitted to view directly, provided that such computational processing is legally and ethically permissible. Interactively, any user can query the full corpus and retrieve counts, trends, and comparisons—replicating, for example, the query logic of a published keyword analysis of election coverage without the articles ever being displayed. Beyond such queries, entire annotation workflows can be run on protected fields (see Figure 3). An external researcher can, for instance, ask MEDem AmCAT to annotate a corpus with a specific large language model deployed within the MEDem infrastructure for legal and data-protection compliance and a specific prompt. The resulting annotations, together with metadata documenting exactly how they were produced, are added to the corpus as new shareable fields, while the annotated material itself remains protected. Likewise, a researcher can train and validate a supervised classifier and submit the model to MEDem, where it is applied to protected texts the researcher could never read themselves. In both cases, the outputs are derived data. They describe the protected content without reproducing it and can therefore be shared, cited, and built upon even where the underlying texts and images cannot. If the researcher strictly

needs access to the original data for human validation purposes, this access can be restricted to a sample of documents, and limited to document abstractions (e.g., snippets, summaries, n-grams).

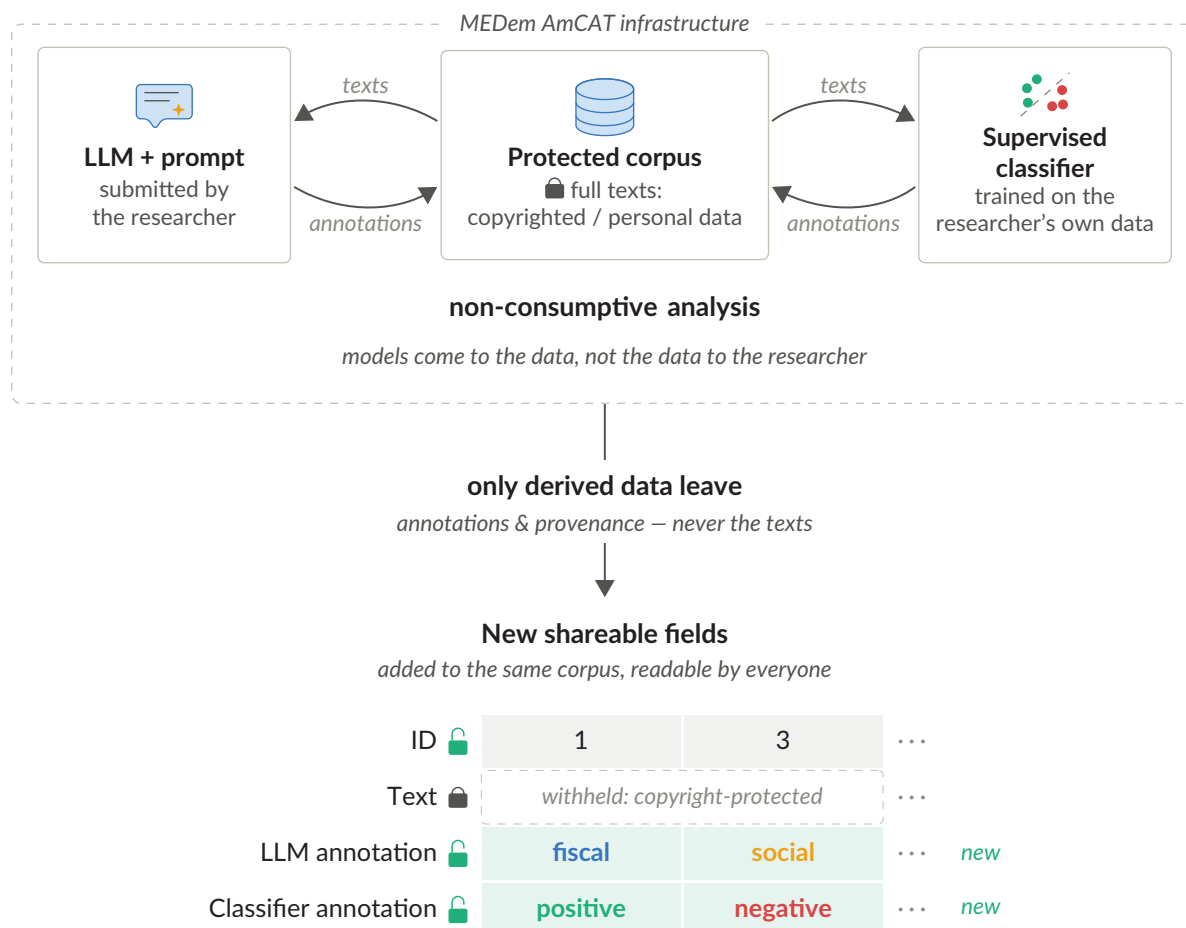


Figure 3. Non-consumptive analysis of a protected corpus in MEDem AmCAT: Instead of giving researchers access to the data, the analysis models come to the data. Notes: An external researcher can submit a prompt for a large language model or a supervised classifier trained on data they do have access to; inside the MEDem infrastructure (dashed boundary), the models read the protected full texts and write their annotations back to the corpus as new shareable fields; only metadata and derived data cleared for sharing are returned to the researcher; continuing the example from Figure 2, the view is filtered to the documents annotated as political (IDs 1 and 3), which are now further annotated by topic (large language model) and tone (classifier).

This design connects MEDem AmCAT directly to the FAIR principles (Wilkinson et al., 2016). FAIR accessibility is often misread as requiring unrestricted openness. Instead, the FAIR principles explicitly allow authentication and authorization where necessary, consistent with the broader principle that data should be “as open as possible, as closed as necessary” (Landi et al., 2020).

If sharing protected corpora thus becomes technically feasible, the question remains why researchers would invest in the effort. The general incentives for data sharing are well documented. Shared data lend credibility to the original research by making it reproducible, attract citations, and promote the research they stem from, as those who reuse a dataset almost inevitably engage with the work it was created for (Dienlin et al., 2021; Gruber & van Atteveltdt, 2025). What has been missing is an infrastructure that makes sharing protected material feasible at reasonable cost (see Section 6). As a maintained service, MEDem AmCAT reduces this

hurdle: Data owners do not need to operate their own servers and can encode and administer access rules centrally, reducing the need for repeated file transfers, bilateral negotiations, and manual access decisions. Data owners can dynamically define the access rules, retain control over the protected fields, and gain visibility through registration in MEDem Meteor. Journals, replicators, and secondary researchers can then be pointed to a durable, citable resource.

By enabling legally compliant sharing of corpora, annotations, analytical outputs, and documented and reproducible workflows across countries and over time, MEDem AmCAT directly addresses key challenges of comparative political communication research. It facilitates cross-national alignment of coding schemes, supports longitudinal monitoring of democratic processes, and allows research teams to build on each other's work without duplicating effort. In doing so, MEDem AmCAT strengthens the capacity of the field to move from isolated case studies toward genuinely cumulative and comparative analyses.

4. Research Scenarios: What MEDem Enables in Practice

The preceding sections have described the main functions of MEDem Meteor and MEDem AmCAT as infrastructural services for finding, accessing, reusing, and analyzing text and multimedia resources in political communication. Their practical value, however, becomes most apparent when considering the concrete problems researchers encounter at different stages of the research process. The following scenarios illustrate several ways in which researchers can use these services to make their own resources more visible, discover existing data and tools, enable the reuse of legally restricted corpora, conduct secondary analyses, and connect different forms of data across domains.

4.1. Scenario 1: Making Analytical Tools More Findable, Accessible, and Reusable in MEDem Meteor

A researcher or research team has developed a tool for computational text analysis, for example, a multilingual classifier for detecting political topics. The tool may have been developed as part of a specific research project and documented in a publication or deposited in a repository. Yet it remains difficult to discover outside the immediate research community surrounding the project. Researchers working on closely related questions may therefore be unaware that an appropriate tool already exists and instead develop similar solutions independently.

By inventorying the tool in MEDem Meteor, its developers can make it discoverable within a structured research ecosystem. The entry's metadata documents what the tool measures, which languages and data types it supports, which methodological approaches it uses, where its code and documentation can be accessed, and in which publications or datasets it has already been applied.

For tool developers, this increases the likelihood that their work will be reused, tested in additional contexts, and further developed by others. For the field more broadly, it reduces unnecessary duplication of effort and contributes to a more cumulative computational research infrastructure in which existing analytical resources become easier to find, evaluate, and build upon.

4.2. Scenario 2: Making a Legally Protected Corpus Accessible for Reuse With MEDem AmCAT

Over three years, a research team collected posts from social media platforms, news articles obtained through media monitoring services and scrapers, and parliamentary records; hundreds of documents were manually annotated, and supervised models were trained to extend these annotations to the entire corpus. The corpus would be a valuable resource for many researchers. Yet the news content is protected by copyright, the social media material contains personal data, and now that the project has ended, nobody on the team has the time to run a server or negotiate individual data-sharing agreements.

Instead of letting the corpus collect digital dust, the team deposits it in MEDem AmCAT and, where needed, defines the access rules with support from SCTMD staff. The accompanying documentation and access conditions are reviewed before the corpus is made available. Metadata and annotations cleared for sharing become publicly readable; the full texts can be queried non-consumptively, returning outputs such as counts, trends, and, where legally permissible, short keyword-in-context snippets or summaries; and fields containing personal data remain restricted to authorized users. Once registered in MEDem Meteor, the corpus becomes findable and citable without requiring its creators to maintain their own technical infrastructure.

Depositing the corpus in this way benefits the original researchers as much as the whole field. When a journal later asks for replication materials, the team can point to the corpus and, where the applicable restrictions permit, grant the appointed replicator temporary access to restricted fields. Data that would otherwise remain locked on institutional storage systems thus continue to generate scientific value long after the project that produced them has ended.

4.3. Scenario 3: Planning a Research Project and Discovering Available Media Sources, Archives, and Tools With MEDem Meteor

A researcher is planning a comparative study of news coverage across several countries. At an early stage of the project, however, several practical questions arise: Which news outlets are relevant in each country? Which of them are available through full-text archives or existing research datasets? And which analytical tools already exist for the intended research task?

MEDem Meteor provides a structured starting point for answering these questions. Researchers can search for and compare news sources according to characteristics such as country, media type, ownership, audience reach, or other relevant properties, and examine how these sources are connected to full-text archives, existing datasets, publications, and analytical tools. Rather than relying on ad hoc knowledge, commercial database interfaces, or source lists compiled separately for each project, researchers can obtain a systematic overview of the resources already available for their research question.

This can directly inform the design of the study. A researcher may discover, for example, that a theoretically relevant outlet may not be available for the required period, that an existing dataset already contains much of the desired material, or that a validated tool for the intended analytical task has already been applied to similar data. MEDem Meteor therefore supports more informed decisions about case selection, source sampling, temporal coverage, and methodological design before substantial resources are invested in data collection or tool development.

4.4. Scenario 4: Reusing and Extending Protected Data to Answer a New Research Question With MEDem AmCAT

A PhD researcher studies how political scandals travel between media arenas and wants to test whether the mechanisms described in single-country case studies hold across contexts. Collecting the necessary data from scratch is not a realistic option. The study therefore depends entirely on the reuse of data collected by others.

Searching MEDem Meteor, the researcher discovers several relevant corpora, among them the corpus from Scenario 2, hosted on MEDem AmCAT. As a guest user, the researcher first assesses its relevance non-consumptively: querying keywords, inspecting the metadata and existing annotations, and plotting how attention shifted between platforms and news outlets over time—enough to establish that the corpus covers the research question before further effort is invested. The existing data, however, require additional annotation. The researcher therefore designs an annotation workflow using an open-weight large language model and a prompt that operationalizes the intended coding scheme. The researcher then validates this workflow against the publicly available manual annotations for a small subset of texts that can still be retrieved from their original URLs and asks MEDem AmCAT to run the pipeline on the protected full texts. The new annotations, together with metadata documenting exactly how they were produced, are added to the corpus without the researcher ever reading the copyrighted articles, apart from a few examples accessed through a set of external URLs recorded in the metadata (e.g., direct links to online news articles or posts on social media platforms).

The resulting publication cites the original corpus, the prompt and validation results are shared alongside it, and the contributed annotation layer remains available to subsequent researchers. Reuse thus becomes cumulative. Each study builds on the previous and enriches the shared corpus further, drawing more attention to it and inspiring new research questions.

4.5. Scenario 5: Conducting a Linkage Study by Finding, Accessing, Reusing, and Linking Existing Data Through MEDem ATLAS, MEDem Meteor, MEDem AmCAT, and the MEDem Linking Tool

MEDem ATLAS (De Sio et al., 2026) is currently an internally tested prototype; the MEDem Linking Tool remains under development, and not all transitions between services have yet been implemented. A researcher wants to investigate whether exposure to particular forms of political news coverage is associated with citizens' political attitudes across countries and over time. Conducting such a comparative linkage study requires considerably more than identifying a suitable survey dataset. Researchers must determine whether appropriate media content and survey data exist for the relevant countries and periods (de Vreese et al., 2017; Eberl & Ameseder, 2025).

The researcher would begin with MEDem ATLAS, where existing survey datasets can be searched based on the survey items they contain, including questions on media exposure. Based on the selected survey datasets, the researcher can then determine which media outlets, countries, and time periods need to be covered on the media side of the linkage study. The researcher could then use MEDem Meteor to search for corresponding media sources and to identify existing datasets or corpora that include the relevant outlets and time spans. Where these datasets are available through MEDem AmCAT, the researcher can move directly from discovery to analysis within the MEDem infrastructure.

In MEDem AmCAT, the researcher could analyze the relevant media corpora and construct the content measures required for the study. Depending on the research question, this may involve using existing annotations, applying computational tools, running standardized queries, or generating new indicators. These measures can then be aggregated at the level required for the subsequent linkage, such as outlet, country, day, week, or campaign period.

Finally, the MEDem Linking Tool would support the identification of relevant linkage keys and generate a script for constructing the final analytical dataset. Taken together, these services define the intended end-to-end workflow for secondary-data linkage.

5. Where MEDem and Its Services Stand

At the time of writing, MEDem is transitioning from the design phase to the preparation phase of the research infrastructure lifecycle. In 2026, MEDem was selected for inclusion in the European Strategy Forum for Research Infrastructures (ESFRI) roadmap and thereby recognized as an ESFRI project. In ESFRI terminology, this status does not mean that the infrastructure has already been fully implemented. Rather, ESFRI projects are research infrastructures in their preparation phase that have been selected based on the strength of their scientific and implementation cases and are expected to reach the implementation phase. Roadmap inclusion therefore marks both the completion of an important design and evaluation process and the beginning of a phase in which MEDem's organizational, legal, financial, and technical foundations must be consolidated to achieve a fully functional research infrastructure in 2030/2031.

During the preparation phase, the concepts, partnerships, tools, and prototypes developed to date must be translated into durable institutional arrangements. This includes further developing the service portfolio, establishing governance and legal structures, securing sustained commitments from participating countries and institutions, clarifying responsibilities among MEDem HQ, the Service Centers, and national partners, and preparing viable financial and operational plans. ESFRI roadmap status provides a strategic framework for this process, but does not in itself secure the resources or agreements required for long-term operation.

MEDem is being developed as a modular and continuously evolving ecosystem rather than as a single platform that becomes operational at one particular point in time. Its services therefore differ in their current levels of maturity. Within the SCTMD, MEDem Meteor is currently the most mature service, while still being under active open-source development. As of August 2026, its inventory contains approximately 3,500 news sources and 350 tools for computational content analysis, alongside numerous other types of entries, such as around 150 datasets and 4,000 scientific studies. Current priorities include continued platform development and, in particular, the systematic expansion and curation of the knowledge graph, including its connected metadata. This expansion is pursued systematically by members of the SCTMD, who review major journals in communication and political science and identify studies that develop, use, or describe text datasets made available for reuse or computational tools released for use by other researchers. The main additions are planned to be completed by the end of 2026; this systematic review and curation process also serves to identify and eliminate bugs across the full workflow, from identifying relevant resources to adding and linking them in the knowledge graph. The process does not privilege particular researchers or institutions, as any researcher can also contribute information about their own work at any time through the structured contribution workflow. Following the MEDem roadmap kick-off, planned for

April 2027, MEDem Meteor will be promoted extensively to the research community, accompanied by user-experience focus groups whose findings will inform the development of the next version of the service.

MEDem AmCAT likewise builds on an existing and actively maintained technological ecosystem rather than on a purely prospective design. It already supports the storage and management of text and multimedia corpora, differentiated access at the level of users and individual data fields, interactive querying, and the integration of manual and automated annotations. AmCAT is also under active open-source development and is currently supported by an open science infrastructure grant from the Dutch Research Council (2026–2028). To complement the review of corpus documentation and metadata, the SCTMD is exploring the integration of MEDem AmCAT with eduGAIN, the federated authentication and authorization infrastructure used by research and education institutions—or a similar federated access system—until 2028. Together, structured documentation review and federated identity would strengthen provenance and contributor accountability, thereby increasing trust in the conditions under which corpora are made available and reused. Further development, planned from early 2027 and until the end of 2028, will focus on embedding these functions within the broader MEDem service and governance framework and connecting MEDem AmCAT more closely with MEDem Meteor so that researchers can move more directly from discovering a corpus to requesting access or analyzing it.

Other MEDem services remain at earlier stages of maturity. MEDem ATLAS is available as a working prototype that is still undergoing internal testing, whereas the MEDem Linking Tool is currently best described as a conceptual and technical design under development. The workflow presented in Scenario 5 should therefore be understood as the envisioned integration of existing and emerging services, rather than as an end-to-end process in which every transition is already operational or automated.

6. Challenges, Trade-Offs, and Safeguards

The benefits of shared research infrastructures do not eliminate the challenges associated with building, governing, and sustaining them. Rather, infrastructures such as MEDem inevitably introduce dependencies, choices, and trade-offs of their own. Making these explicit is important because decisions about which resources are documented, how they are described and connected, and which forms of access and analysis are supported can shape what becomes visible, comparable, and practically feasible for researchers.

The first challenge concerns community participation. MEDem Meteor can only realize its full infrastructural potential if it is adopted and actively supported by a broad research community. Experience from its development also shows that regular exchange with prospective users is needed already before services become fully operational, as these users can identify requirements and problems that might otherwise become visible only after launch. Such exchange is difficult to establish, however, when an infrastructure is still little known and its future remains uncertain. Its quality, coverage, and continued relevance therefore depend partly on a form of distributed responsibility. Individual researchers and research teams contribute information about sources, datasets, tools, archives, and publications that might otherwise remain visible only within particular projects or scholarly networks. Yet the benefits of visibility and reusability may not always outweigh the time required to prepare and maintain high-quality entries, particularly under the constraints of increasingly precarious academic employment and competing demands on researchers' time.

Community participation presents an even greater challenge for MEDem AmCAT because making a corpus available for controlled reuse requires more than registering its existence. Data owners will need, with support from SCTMD staff where necessary, to document the corpus and applicable copyright, data protection, and ethical conditions, define differentiated access rules, and prepare metadata for secondary use. They may also be reluctant to share resources that they have not yet fully analyzed or fear losing control over how their data are interpreted and reused. MEDem AmCAT can lower these barriers by allowing data owners to retain control over protected fields, manage access dynamically, and make metadata, annotations, or derived outputs available without immediately opening the underlying material.

Nevertheless, low-threshold contribution workflows as well as increased discoverability and reusability alone are probably unlikely to ensure sustained participation. MEDem will therefore prioritize practical support and sustained community engagement, allowing users to articulate their needs, identify barriers, and influence service development. At the same time, the SCTMD will continue to expand and curate the MEDem Meteor inventory systematically rather than relying exclusively on voluntary contributions, while incrementally maintaining and developing MEDem AmCAT.

Community participation also provides an important governance safeguard. Although day-to-day review and administration of MEDem services are carried out by SCTMD staff, these activities form part of broader governance arrangements. Metadata models, reviewing practices, and priorities for further development are discussed with the wider research community through dedicated governance bodies. The research community, therefore, advises the Service Center on its strategic development and is represented on MEDem's Scientific Board. While such arrangements cannot eliminate differences in resources, participation, or influence, they are intended to prevent individual institutions, funders, or intellectual communities from determining curation standards or development priorities unilaterally.

A second challenge concerns coverage. MEDem Meteor can make gaps in documented resources visible, while MEDem AmCAT can preserve and facilitate controlled reuse of selected corpora. Neither service, however, can eliminate inequalities in the underlying research landscape. If certain countries, languages, media systems, data types, or methodological traditions are poorly represented in existing datasets or are less frequently documented in the inventory or deposited in the controlled-access environment, these inequalities may also be reflected in the tools themselves. Their more immediate contribution is to make documented availability and apparent gaps more transparent, allowing researchers to identify where empirical coverage is strong, where it is limited, and where sampling decisions remain constrained by access. Over time, this transparency can also help identify systematic blind spots and inform targeted efforts to document, preserve, or develop resources in areas that remain underrepresented.

A third challenge concerns methodological pluralism. MEDem seeks to standardize metadata in ways that improve interoperability, not to standardize research itself. Different theories, operationalizations, data types, and analytical approaches can coexist within the infrastructure without being forced into a single methodological template. This distinction, however, does not eliminate the risk of infrastructural bias. By design, MEDem focuses on democratic processes and has thus far been developed primarily around quantitative data. Ostensibly technical decisions about metadata schemas, entity definitions, supported formats, search interfaces, or analytical workflows may make some research traditions more visible, accessible, or convenient than others. An important safeguard is therefore not to assume that a research

infrastructure can ever be theoretically or methodologically neutral, but to ensure that its development remains transparent, revisable, and pluralistic. Accordingly, MEDem's metadata models and services are intended to remain extensible and modular and to preserve and expose scientifically meaningful differences between operationalizations rather than collapsing them into a single measure.

A fourth challenge concerns openness and legal or ethical constraints. MEDem AmCAT does not override the legal, contractual, or ethical requirements governing data collection and use. Researchers must still be entitled to obtain and analyze material, whether through licenses, statutory research exceptions, or access frameworks such as those established under the Digital Services Act (e.g., Seiling et al., 2025). MEDem AmCAT can, however, support controlled forms of access and analysis designed to reflect the legal, contractual, and ethical restrictions applicable to a particular corpus. At the same time, legal access does not automatically imply ethical acceptability: Privacy risks, sensitive content, or potential harms to individuals and groups may require additional safeguards. MEDem AmCAT can support such safeguards, but it cannot replace case-specific legal and ethical assessment.

A final challenge concerns sustainability throughout infrastructure development and operation. Work on the SCTMD and its services has already made clear that infrastructure development takes time and requires considerable resources even during the design and preparation phases. Technical development also involves recurring decisions about whether to develop or adapt open-source software in-house, procure existing systems, or rely on external providers. In-house development allows greater control and flexibility but requires appropriate expertise and long-term maintenance; procurement or outsourcing can reduce immediate development demands but may be costly and create dependencies that restrict later adaptations. These decisions therefore affect not only initial development but also the costs, expertise, and dependencies involved in operating the services over time.

Long-term sustainability also depends on funding structures and academic incentives. Data, tools, and analytical workflows are often developed within short-term projects, inconsistently documented, and only weakly maintained once funding ends. Such projects can finance individual technical steps but rarely provide sufficient continuity for maintenance, curation, or community engagement. Infrastructure work also tends to receive less recognition in academic evaluation systems than publications and grant acquisition, making it difficult for researchers to devote sustained time to it during the early stages. Research infrastructures can counter this pattern by shifting responsibility for core functions such as technical maintenance, metadata curation, legal support, and coordination from individual projects to collectively maintained systems that persist beyond specific grants (Mustillo, 2025; Strippel, 2021). Yet this does not eliminate the sustainability problem: Continued operation still depends on sustained political, institutional, and financial commitments. Our experience also suggests that the prospect of long-term support matters during development itself, as researchers and institutions are more likely to commit time and resources when they can reasonably expect the services and collaborations to continue beyond the current project.

MEDem's modular structure serves as a partial safeguard during the early stages of infrastructure development because individual services can provide value and develop independently, meaning that delays or funding gaps affecting one service do not necessarily compromise the infrastructure as a whole. At the same time, this structure creates coordination demands because these services must remain interoperable and aligned within a common technical and governance framework. MEDem's inclusion in the ESFRI

Roadmap and existing financial and in-kind commitments for the preparation phase provide an important basis for further institutionalization, but they do not yet secure long-term operation.

7. Next Steps and Conclusion

This article has argued that comparative political communication research faces two interconnected infrastructural barriers. First, relevant resources are often difficult to discover; second, many of the underlying data cannot be openly redistributed. MEDem Meteor and MEDem AmCAT address these barriers through complementary functions. Their contribution is not to make every resource openly accessible or eliminate inequalities in the underlying research landscape, but to make existing resources and gaps in documented coverage more visible, access conditions more transparent and manageable, and reuse possible under applicable restrictions.

Although technically functional, both services remain in a pre-launch stage within MEDem. They already provide most of the technical functionality described in this article and can be used by researchers, but they have not yet been broadly promoted and tested within the wider research community. Their formal launch as working tools is planned for the MEDem Roadmap Kick-Off in spring 2027. Technical functionality and institutional development alone, however, will not ensure long-term success. MEDem's success will also depend on whether researchers use, test, contribute to, and help shape its services. Communication scholars have been found to hold more ambivalent attitudes toward research infrastructures than colleagues in adjacent social science disciplines, a pattern partly linked to limited experience with secondary data reuse and infrastructural modes of research (Prandner, 2023). Despite important initiatives and repeated calls for reform, FAIR data stewardship and open science practices in the field remain uneven and fragmented (Dienlin et al., 2021).

The launch in 2027 should therefore be understood not simply as the public introduction of two technical services, but as the beginning of a broader process of community building and institutionalization. During the following years, MEDem will need to demonstrate that its services function reliably, address genuine and sustained research needs, and can evolve in response to changing community requirements. From around 2029 onward, work will increasingly need to focus on preparing MEDem's transition to the implementation phase, including the further consolidation of governance structures, funding commitments, and operational responsibilities. If supported through sustained participation and political, institutional, and financial commitments, MEDem Meteor and MEDem AmCAT can shift the discovery and controlled reuse of political communication resources away from temporary, project-specific arrangements and toward a shared infrastructure that makes cumulative and comparative research more routinely feasible.

Acknowledgments

We would like to thank Isabella Soares da Cunha, Kleo Berghuber, and Roksana Gloc for their recent support in the development of MEDem Meteor.

Funding

Publication of this article in open access was made possible through the institutional membership agreement between Vrije Universiteit Amsterdam and Cogitatio Press.

Conflict of Interests

The authors declare no conflict of interest.

LLMs Disclosure

ChatGPT was used to assist with language editing and phrasing during the preparation of this manuscript. Figure 3 was created using LaTeX/TikZ code generated with assistance from Claude, based on detailed instructions and iterative feedback from the authors. The authors reviewed and revised all outputs and take full responsibility for the final text and figures.

References

- Assenmacher, D., Sen, I., Fröhling, L., & Wagner, C. (2023). The end of the rehydration era: The problem of sharing harmful Twitter research data. In S. Kleanthous, M. Park, & O. Tsur (Eds.), *Workshop proceedings of the 17th International AAAI Conference on Web and Social Media*. Association for the Advancement of Artificial Intelligence. <https://doi.org/10.36190/2023.56>
- Bahlai, C. A., Bartlett, L. J., Burgio, K. R., Fournier, A. M., Keiser, C. N., Poisot, T., & Whitney, K. S. (2019). Ethics: Open science isn't always open to all scientists. *American Scientist*, 107(2), 78–82. <https://www.jstor.org/stable/26586147>
- Balluff, P., Lind, F., Boomgaarden, H. G., & Waldherr, A. (2023). Mapping the European media landscape—Meteor, a curated and community-coded inventory of news sources. *European Journal of Communication*, 38(2), 181–194. <https://doi.org/10.1177/02673231221112006>
- Boomgaarden, H., Katsanidou, A., Kritzinger, S., Lutz, G., Willmann, J., & Eberl, J.-M. (2025). Open science in democracy research: the research infrastructure “Monitoring Electoral Democracy” (MEDem). *European Political Science*, 24(4), 782–794. <https://doi.org/10.1057/s41304-025-00534-8>
- De Sio, L., Katsanidou, A., & Borghetto, E. (2026). *Notes on the MEDem ATLAS Prototype*. Zenodo. <https://zenodo.org/records/19054067>
- de Vreese, C. H., Boukes, M., Schuck, A., Vliegenthart, R., Bos, L., & Lelkes, Y. (2017). Linking survey and media content data: Opportunities, considerations, and pitfalls. *Communication Methods and Measures*, 11(4), 221–244. <https://doi.org/10.1080/19312458.2017.1380175>
- Dienlin, T., Johannes, N., Bowman, N. D., Masur, P. K., Engesser, S., Kümpel, A. S., Lukito, J., Bier, L. M., Zhang, R., Johnson, B. K., Huskey, R., Schneider, F. M., Breuer, J., Parry, D. A., Vermeulen, I., Fisher, J. T., Banks, J., Weber, R., Ellis, D. A., . . . & de Vreese, C. (2021). An agenda for open science in communication. *Journal of Communication*, 71(1), 1–26. <https://doi.org/10.1093/joc/jqz052>
- Eberl, J.-M., & Ameseder, N. (2025). Linkage studies. In A. Nai, M. Grömping, & D. Wirz (Eds.), *Elgar encyclopedia of political communication* (Vol. 2, pp. 382–385). Edward Elgar. <https://doi.org/10.4337/9781035301447.vol2.00093>
- Esser, F., & Vliegenthart, R. (2017). Comparative research methods. In J. Matthes (Ed.), *The international encyclopedia of communication research methods*. Wiley. <https://doi.org/10.1002/9781118901731.iecrm0035>
- Fensel, D., Simsek, U., Angele, K., Huaman, E., Kärle, E., Panasiuk, O., Toma, I., Umbrich, J., & Wahler, A. (2020). *Knowledge graphs*. Springer.
- Fiesler, C., Beard, N., & Keegan, B. C. (2020). No robots, spiders, or scrapers: Legal and ethical regulation of data collection methods in social media terms of service. *Proceedings of the International AAAI Conference on Web and Social Media*, 14(1), 187–196. <https://doi.org/10.1609/icwsm.v14i1.7290>
- Gruber, J. B., & van Atteveldt, W. (2025). *Sharing is caring (about research): Three avenues for sharing (protected)*

- text collections and the need for non-consumptive research. SocArXiv. https://osf.io/preprints/socarxiv/gz3xf_v1
- Landi, A., Thompson, M., Giannuzzi, V., Bonifazi, F., Labastida, I., da Silva Santos, L. O. B., & Roos, M. (2020). The “A” of FAIR—As open as possible, as closed as necessary. *Data Intelligence*, 2(1/2), 47–55. https://doi.org/10.1162/dint_a_00027
- Lind, F., Song, H., Boomgaarden, H. G., Kathirgamalingam, A., Ali, K. P. S., & Vliegenthart, R. (2025). Research practices in comparative communication research: Visibility, topical and geographical disparities, and their longitudinal patterns. *International Journal of Communication*, 19, 1103–1128. <https://ijoc.org/index.php/ijoc/article/view/23510>
- Mustillo, T. (2025). Building a FAIR, linked, and open data ecosystem: Innovation cascades in political science. *European Political Science*, 24(4), 810–825. <https://doi.org/10.1057/s41304-025-00520-0>
- Prandner, D. (2023). Do not only share—But also care: Das komplexe Verhältnis zwischen österreichischer Kommunikationswissenschaft und Forschungsinfrastruktur. *MedienJournal Zeitschrift für Medien- und Kommunikationsforschung*, 47(3), 23–45. <https://doi.org/10.24989/medienjournal.v47i3.2626>
- Seiling, L., Keller, C. I., Ohme, J., Klinger, U., & de Vreese, C. (2025). *Data access for researchers under the Digital Services Act: From policy to practice* (Weizenbaum Policy Paper No. 14). Weizenbaum Institut. <https://hdl.handle.net/10419/326736>
- Stecker, M., Balluff, P., Lind, F., Dinhopl, C., Waldherr, A., & Boomgaarden, H. G. (2024). Tools of the trade—When are software tools mentioned in computational text analysis research? *Computational Communication Research*, 6(1), 1–21. <https://doi.org/10.5117/CCR2024.1.6.STEC>
- Strippel, C. (2021). Forschungsinfrastrukturen für die Kommunikations- und Medienforschung im deutschsprachigen Raum. Initiativen, Bedarfe und Perspektiven. *M&K Medien & Kommunikationswissenschaft*, 69(1), 136–157. <https://doi.org/10.5771/1615-634X-2021-1-136>
- Theocharis, Y., & Jungherr, A. (2021). Computational social science and the study of political communication. *Political Communication*, 38(1/2), 1–22. <https://doi.org/10.1080/10584609.2020.1833121>
- van Atteveldt, W. (2008). *Semantic network analysis: Techniques for extracting, representing, and querying media content* [Unpublished doctoral dissertation]. Vrije Universiteit Amsterdam.
- van Atteveldt, W., Althaus, S., & Wessler, H. (2021). The trouble with sharing your privates: Pursuing ethical open science and collaborative research across national jurisdictions using sensitive data. *Political Communication*, 38(1/2), 192–198. <https://doi.org/10.1080/10584609.2020.1744780>
- van Atteveldt, W., Gruber, J. B., & Welbers, K. (2023). *AmCAT 4: A toolkit for corpora storage, sharing and pre-processing* (Deliverable 7.4). Observatory for Political Texts in European Democracies. https://opted.eu/fileadmin/user_upload/k_opted/OPTED_Deliverable_D7.4.pdf
- van Atteveldt, W., Strycharz, J., Trilling, D., & Welbers, K. (2019). Computational communication science | Toward open computational communication science: A practical road map for reusable data and code. *International Journal of Communication*, 13, 3935–3954. <https://ijoc.org/index.php/ijoc/article/view/10631>
- Whitaker, K., & Guest, O. (2020). #bropenscience is broken science: Kirstie Whitaker and Olivia Guest ask how open ‘open science’ really is. *The Psychologist*, 33, 34–37. https://www.mpi.nl/publications/item_3286863
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blimberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . & Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1), Article 160018. <https://doi.org/10.1038/sdata.2016.18>

About the Authors



Jakob-Moritz Eberl is a senior scientist at the Department of Communication and a part of the MEDem Headquarters team at the University of Vienna. He studies how information environments shape citizens' attitudes and perceptions, and how norms and trust surrounding media, science, and democratic institutions are formed and contested.



Jana Bernhard-Harrer is a postdoctoral researcher at the University of Vienna, working on the validation of computational text analysis methods in political communication as part of MEDem's SCTMD. She is actively involved in the development of both MEDem Meteor and MEDem AmCAT.



Mariken van der Velden is a university professor of political communication at Vrije Universiteit Amsterdam, where she serves as co-director of the Resilience, Security and Civil Unrest (ReSCU) Lab. She also leads the SCTMD and serves on MEDem's Board of Directors.



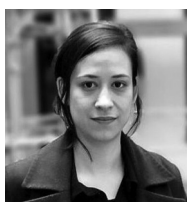
Sofia Gil Clavel is a postdoctoral researcher and lab manager of the Societal Analytics Lab at Vrije Universiteit Amsterdam. She is involved in developing and testing AmCAT. Additionally, she writes user guides, blog posts, and installation resources on how to use, set up, and run AmCAT.



Paul Balluff is a postdoctoral researcher at the Department for Computational Social Sciences at GESIS. He was the lead developer of MEDem Meteor within the OPTED project and remains an active contributor.



Johannes B. Gruber is a senior researcher and team lead of the team Metadata Standards and Interoperability at GESIS. His work focuses on digital political communication and open science, using computational text and image analysis. He began contributing to AmCAT during the OPTED project and remains an active contributor.



Noëlle S. Lebernegg is a postdoctoral researcher in the Strategic Communication Group at Wageningen University & Research. Her research focuses on computational social science and political communication. She is also a collaborator within MEDem's SCTMD.



Kasper Welbers is an associate professor in the Department of Communication Science at Vrije Universiteit Amsterdam. He is co-director of the Societal Analytics Lab, chair of the Computational Methods Division at the International Communication Association, and an active open-source developer of AmCAT and other research infrastructures.



Marvin Stecker is a pre-doctoral researcher at the University of Vienna. His research focuses on political communication, social identities, and computational methods. He served as project manager of the OPTED Project, contributed to MEDem, and remains involved with MEDem Meteor.



Claes Holger de Vreese is a university professor of AI and society, specializing in media and democracy. He is director of the Digital Democracy Centre and chair-in-residence for 2026/27 at the Danish Institute of Advanced Studies. He is also a permanent collaborator within MEDem's SCTMD.



Wouter van Atteveldt is a university professor of computational communication science and political communication at Vrije Universiteit Amsterdam. He is co-director of the VU Societal Analytics Lab, founding editor-in-chief of *Computational Communication Research*, and the original developer of AmCAT.



Rens Vliegenthart is a university professor in strategic communication at Wageningen University & Research. He is one of the Dutch National Coordinators for MEDem and a permanent collaborator within the SCTMD.



Hajo G. Boomgaarden is a university professor of methods in the social sciences at the Department of Communication at the University of Vienna, where he is co-leading the Computational Communication Science Lab. He is also co-lead of the MEDem headquarters and the current chair of the board of directors of MEDem.