



Editorial: Data, archives, and tools: Introducing new publication formats on infrastructures and resources for communication and media research

Christian Strippel · Johannes Breuer · Silke Fürst ·
Erik Koenen · Dimitri Prandner · Christian Schwarzenegger

Received: 10 July 2023 / Accepted: 12 July 2023 / Published online: 22 August 2023
© The Editors of the Journal 2023

1 Background and scope

In recent years, a growing number of initiatives in communication and media research have advocated for a “cultural shift” in our discipline toward more open, reproducible, and replicable research practices and better access to infrastructures and shared research resources (cf. Bowman and Spence 2020; Dienlin et al. 2021; Haim and Puschmann 2023; Shaw et al. 2021; Strippel 2021). This special issue aims to contribute to these initiatives by introducing new article formats for publi-

✉ Christian Strippel

Weizenbaum Institute for the Networked Society, Hardenbergstraße 32, 10623 Berlin, Germany
E-Mail: christian.strippel@weizenbaum-institut.de

Dr. Johannes Breuer

GESIS Leibniz Institute for the Social Sciences, Unter Sachsenhausen 6-8, 50667 Cologne, Germany
E-Mail: johannes.breuer@gesis.org

Dr. Silke Fürst

Department of Communication and Media Research (IKMZ), University of Zurich,
Andreasstrasse 15, 8050 Zurich, Switzerland
E-Mail: s.fuerst@ikmz.uzh.ch

Dr. Erik Koenen · Dr. Christian Schwarzenegger

ZeMKI, Centre for Media, Communication and Information Research, University of Bremen, Linzer
Straße 4, 28359 Bremen, Germany

Dr. Erik Koenen

E-Mail: ekoenen@uni-bremen.de

Dr. Christian Schwarzenegger

E-Mail: schwarzenegger@uni-bremen.de

Dr. Dimitri Prandner

Faculty of Social Sciences, Economics Business, Johannes Kepler University Linz,
Altenbergerstraße 69, 4040 Linz, Austria
E-Mail: dimitri.prandner@jku.at

cations dedicated to such resources and related infrastructures, including research software, reusable datasets, databases, and archives.

Looking at the main forums in our field, we see a lack of opportunities to share experiences with such infrastructures and resources. Usually, this information is limited to brief references in articles, documentation scattered across the Internet, and informal exchanges among colleagues. Only a few journals already provide dedicated spaces for in-depth presentations or discussions of tools or datasets, including *Computational Communication Research* (cf. Araujo et al. 2022), *Mobile Media & Communication* (cf. Schoch and Chan 2023; Wedel 2023), and *Medien & Kommunikationswissenschaft* (cf. Haim et al. 2023).¹

However, our field still has little experience in writing, reviewing, assessing, and editing such contributions. As a prominent journal in the field of communication and media studies in German-speaking countries, *Publizistik* has decided to respond to this lack and to the growing calls for the sustainable development and use of research resources (cf. Peter et al. 2020; Hepp et al. 2021) with a new article section. The aim of this section is to encourage and consolidate the exchange about resources and infrastructures within our field (and beyond) through new publication formats, to increase the visibility of both these resources and the work that goes into developing and maintaining them, and, finally, to provide additional incentives for this work by offering a way of crediting scholarly output that is not a publication in the classical sense.

This special issue serves as a kickoff for this new section. Collecting the first examples of articles that are suitable for publication in this section provides an opportunity for testing the new publication formats and discussing the challenges we have faced in organizing this special issue. In the first part of this introduction we, hence, elaborate on these challenges and our discussions about them before presenting the articles collected in this issue in more detail. In doing so, we hope to provide some interesting “behind-the-scenes” insights into the making of this issue and, at the same time, answer some questions for colleagues who are planning to establish similar formats in other scientific journals or want to write articles for the new section at *Publizistik*.

2 Editorial process and challenges

2.1 Abstract submission

We faced the first challenge when we formulated the call for papers for this special issue (cf. Strippel et al. 2022). From the outset, we were aware that authors whose manuscripts do not make it through our review process have limited opportunities to resubmit these manuscripts to other journals, as is usually the case. Therefore, we decided to give interested authors the opportunity to check whether their manuscript

¹ In addition, the “tool demos” of the Computational Methods division of the International Communication Association at their annual conferences have become increasingly popular since they were established a few years ago.

ideas fit the scope of the special issue by first submitting a short abstract. A total of 23 authors or author teams from different countries and disciplines took advantage of this opportunity and submitted an abstract. We were very pleased with this broad interest as it showed us that there is a real need for a journal section on infrastructures and resources, even beyond our discipline.

2.2 Anonymization and double-blind peer review

The next challenge, which we had already discussed in an abstract manner in the context of the call but then encountered in practice as the abstracts were submitted, was the question of anonymizing the texts for the purpose of double-blind peer review. Since the authorship of tools, datasets, and other resources is often public or easy to find online, we wondered whether a double-blind review of the abstracts would be possible at all. However, this is not an exclusive problem of this special issue but a general challenge for our field, given its rather small size. In addition, submissions are not necessarily written by those who have developed the tools discussed in an article.

After weighing the various arguments, we decided not to abandon the process of a double-blind peer review. One guest editor was appointed to collect the submissions, ensure anonymity, and organize the entire review process, but did not review abstracts or articles. The reviewers, on the other hand, were asked not to refuse to review a manuscript if they suspected they knew who the authors were unless they felt biased.

In this way, each of the submitted abstracts was blindly reviewed and evaluated by two guest editors and one out of two colleagues from the editorial team of *Publizistik*, namely Emese Domahidi or Christian Pentzold, who provided us with great support for this issue. The main question to be answered by each reviewer at this point was whether the manuscript idea presented was within the scope we had set for this issue. After discussing the results of our reviews, we invited the authors of 14 of the 23 abstracts to submit a full manuscript. All invited authors responded to our invitation and submitted a full manuscript.

2.3 Short vs. long format

Full manuscripts could be submitted in either a short format (15,000–25,000 characters) or the usual journal article length (35,000–50,000 characters). By offering the short option, we recognized that presenting infrastructures or resources may not require as much space as a full paper (especially if additional documentation of the resources exists elsewhere). We wanted to accommodate those authors who wished to contribute to this special issue but may have been reluctant to write a full-length article to present a resource that they had created or used. Since it is still unclear how articles about infrastructures and resources will be accepted and perceived by the scientific community, it seems plausible to us to be flexible in terms of article format.

2.4 New review criteria

After receiving the full papers, the usual peer review process began. However, the review criteria that we asked each reviewer to follow and apply were not the usual ones. Since we wanted to make room for new text formats with this special issue, we could not apply the traditional review criteria (i.e., theoretical foundation, relevance of the research question, and methods), but, instead, had to develop more appropriate criteria for reviewing articles on infrastructures and resources.

In addition to asking the reviewers to decline a review only in the case of potential bias, we pointed out to the reviewers that only the manuscript at hand was the subject of their review. Accordingly, we left it up to the reviewers to also consider and evaluate the actual infrastructures and resources (e.g., a data set, software tools, or code) described in the manuscripts, as this requires a considerable amount of additional time and potentially also a very high degree of specific expertise.

We then presented seven review criteria with corresponding questions:

- *Usefulness*: Is it made clear how and for what purpose the infrastructure/resource can be used, and if applicable, how it is already being used? Is there a connection made to research questions or problems in communication studies?
- *Assistance*: Does the presentation assist in the future use of the infrastructure/resource?
- *Relevance*: What does the described infrastructure/resource do? How does it differ from other offerings, if applicable?
- *Contextualization*: Is the described infrastructure/resource appropriately placed in the context of previous research or existing offerings? Are these adequately cited?
- *Comprehensibility*: Does the manuscript provide an accurate, detailed, and clear description of the infrastructure/resource?
- *Accessibility*: Are potential barriers to the use (e.g., language, technical requirements, or licenses/terms of use) of the infrastructure/resource displayed? If so, are they adequately addressed and discussed with respect to the implications for research?
- *Presentation*: Is the manuscript concisely and understandably written? Is it well structured, and does it develop a clear argument?

On the basis of these criteria, each submitted manuscript was reviewed by three colleagues: one guest editor, one main editor from *Publizistik*, and an external reviewer. All reviews were then supplemented with a summary review by another member of the guest editorial team, and a final vote was taken. In the case of considering a rejection, all reviews were, once again, presented to all guest editors for discussion, and a final decision was made.

3 Overview of the special issue

This special issue finally contains 10 articles, four of which are short papers, and six of which are full-length articles. Six articles are written in German and four are written in English. The articles vary in style and subject matter. For example, the

issue begins with a more programmatic text but also contains tool presentations, a software overview, and an analysis of existing infrastructures. In terms of the research field, the focus is primarily on media use research (including surveys, tracking, digital trace data, and data donations) and content analysis. Given that this issue is intended to serve as a model for further submissions on infrastructures and resources in *Publizistik*, this diversity is quite fortunate.

In the first article, *Andreas Hepp*, *Florian Hohmann*, and *Alessandro Belli* address the research community by emphasizing the importance of adopting and normalizing co-creation practices and cultures in academia. In their paper “Let’s build our tools together! Possibilities and limits of co-creating research software for communication and media studies,” they argue that, in order for research environments to become sustainable, durable, or at least less ephemeral, research software, tools, and infrastructures need to go beyond one-off solutions for one project only. Drawing on their own experience of developing qualitative research software, they argue that the development of tool-software-infrastructure must be seen as a community effort to achieve visible, accessible, adaptable, and sustainable research.

The second paper, “Concept and challenges of a social media observatory as a DIY research infrastructure,” by *Gregor Wiedemann*, *Felix Victor Münch*, *Jan Philipp Rau*, *Phillip Kessler*, and *Jan-Hinrik Schmidt* takes a closer look at a specific research infrastructure. The authors discuss ethical and legal considerations as well as implementation and working routines for the Social Media Observatory hosted by the Research Institute Social Cohesion. The article reflects on the decisions and steps to be taken while building such a platform from scratch, balancing the feasible with the achievable, and anticipating the current and future needs of researchers. In the spirit of community orientation and collaboration, the authors present ways in which a curated knowledge base can help preserve the experience and knowledge gained from using the infrastructure. In this way, external researchers can adapt systematic social media observations to their personal research needs.

The short article “Platform research with Instagram data—a review of analytical approaches, digital survey procedures, and research ethical perspectives in times of the APicalypse” by *Yannik Peters*, *Patrick Nehls*, and *Caja Thimm* presents an overview of different methods for collecting data from the popular social media platform Instagram in the “post-API age” (Freelon 2018). They base their discussion and classification of data collection methods on a “media grammar of Instagram” that distinguishes between “surface grammar” and “constitutive property grammar,” which helps to distinguish between different levels of analysis, for example. The systematic comparison between different data collection options takes into account a number of relevant dimensions, such as the type of data obtained or the programming skills required, and provides useful guidance for researchers interested in studying Instagram.

In their German paper “Spotivey—a web application to simplify the use of the Spotify API in online questionnaire studies,” *Matthias Ladleif* and *Steffen Lepa* present an app that tackles the problem of accurately measuring Spotify usage. Once users link their Spotify profile, the Spotivey web application allows online survey tools to combine respondents’ actual Spotify usage with their answers to the questionnaire. As self-reported media use is prone to bias and errors (cf. Cuadrado-

García et al. 2023), such an application provides an interesting solution to this problem. The article illustrates this in detail and provides researchers with relevant insights and inspiration to think about similar application scenarios.

Frank Mangold, Mareike Wieland, Sebastian Stier, and Lukas Otto present two GESIS research tools—AppKit and WebTrack—in their article “New infrastructures for measuring digital media use.” They discuss the changes and relationships between digital media environments and digital media use and derive basic requirements for media use research in the digital age, which, at the same time, are practical challenges for corresponding research infrastructures and tools. The paper emphasizes that the sustainable development and provision of research software for measuring media use is an ongoing complex and resource-intensive task of communication science, which is challenging in the long run and has to be adapted to new media conditions again and again.

In his short article “One app to assess them all. Combining surveys, experience sampling, and logging/data donation in an Android and iOS app,” *Roland Toth* introduces the MART app, which can be used to combine different data sources in smartphone-based studies. The app, which is still under development and will be extended with further functionalities, is an open-source tool for all researchers and can be used for different types of studies and research questions. Since MART allows researchers to easily combine logging and data donation for measuring smartphone usage with experience sampling and surveys, it has wide application potential in different areas of communication research. In addition to introducing the app, the article provides guidance on and examples of how it can be used.

Philipp Dreesen and Julia Krasselt present the Swiss-Applied Linguistics (Swiss-AL) corpus family, which contains texts from the fields of news media, politics/administration, business, science, and civil society in Switzerland. The platform was developed for linguistic discourse analysis and can be used freely. In their short article, “Swiss-AL: Language data platform for the analysis of public communication in Switzerland,” the authors provide basic information on the structure and data basis of the corpora and explain how communication researchers can use Swiss-AL for quantitative content analyses. Swiss-AL is under continuous development and is currently being expanded into an open research data resource according to the FAIR principles². The authors argue for the use of Swiss-AL for data-driven explorations of public communication according to distant reading approaches and as a complement to full-text-based content analyses.

In his short article “Text data mining on current newspaper articles from the United States with ProQuest TDM Studio,” *Jens Pohlmann* introduces Pro Quest TDM Studio as a research infrastructure and research resource for communication and media studies. He presents key features and discusses analytical possibilities, ways of handling, and some drawbacks of the cloud-based service. To contextualize Pro Quest TDM Studio, the paper develops a framework to assess the ways in which such a digital resource of textual materials allows for filtering, downloading,

² The FAIR principles are: Findability, Accessibility, Interoperability, and Reusability (cf. Wilkinson et al. 2016).

processing, and sharing data and results. Using this framework, the pros and cons of Pro Quest TDM Studio are discussed in comparison to other options.

In the paper “Use of iLCM as an extensible and adaptable text analysis tool in empirical content analysis,” *Andreas Niekler, Christian Kahmann, Manuel Burghardt, and Gerhard Heyer* give insight into their highly customizable and extensible text analysis tool, which has a wide range of applications in the field of communication science. With the development of the (i)nteractive (L)eipzig (C)orpus (M)iner, the authors hope to fill a gap in the tool landscape between more established yet feature-limited GUI-based software solutions and highly advanced script-based applications that have high entry barriers for their users. The presented introduction to the application may motivate users to try the iLCM for themselves and find out if it does, indeed, close this gap.

Finally, *Annett Heft, Jakob Jünger, Julia Niemann-Lenz, and Daniel Possler* take up the discussion on the further development of research infrastructures for content analysis in their article “The national research data infrastructure—a solution for infrastructural needs of content analysis?”. The article illustrates what kind of infrastructures for content analysis are needed in communication research, to what extent this is already covered by the existing consortia of the National Research Data Infrastructure (NFDI)—i.e., *KonsortSWD, BERD@NFDI, Text+, NFDI4Memory, NFDI4Culture, and NFDI4DataScience*—and which infrastructural needs still exist. Finally, the authors discuss the next steps to better address these needs. This article can be seen as a well-informed and inspiring plea for communication scholars to become more involved in NFDI consortia and in building research infrastructures for content analysis.

4 Outlook and acknowledgments

In addition to presenting and discussing important resources and infrastructures, the articles in this issue are also intended to serve as a pool of examples of how the new section of *Publizistik* on research infrastructures and resources can be used and populated in the future. This section will be officially launched in 2024, and we would like to already encourage and invite all interested researchers to take advantage of this opportunity to present, discuss, or compare research datasets, archives, collections, repositories, software, tools, scripts, or packages. In addition to the articles collected in this special issue, our call for papers can serve as an orientation for future manuscripts (cf. Strippel et al. 2022). There will also be a separate standing call for papers, which will provide all information on formats, requirements, and submission criteria.

For now, we would like to thank all the colleagues who have worked on this special issue and made it possible. This includes all the authors of the articles in this issue and all those colleagues whose abstracts or manuscripts have not (yet) passed the review process. We very much hope that we will soon see their texts in one of the next issues as part of the new journal section.

We would also like to thank the editors of *Publizistik*, who were open to our idea for this issue from the very beginning, generously entrusted us with its realization,

and supported us when it came to the review process. In particular, Emese Domahidi and Christian Pentzold have been very supportive of editing this issue. Without their help and experience, we would not have been able to master the numerous challenges of this experimental issue half as well. We also owe a lot to Gunter Reus. We are very pleased that he agreed to our experiment with this special issue, and have benefited greatly from his experience.

Last but not least, we are grateful to all the reviewers who patiently accepted our modified review criteria and took the time not only to check the quality of the articles in this issue but also to improve it through careful reading, constructive feedback, and helpful advice.

We wish you, dear readers, an insightful reading with many new discoveries and suggestions that may benefit your scientific work in the future. We hope that this issue will help to make working with and on research infrastructures and resources easier, better, and more rewarding. Perhaps it will also inspire you to contribute publications on infrastructures and resources for communication and media research to the new section in *Publizistik*.

References

- Araujo, T., Ausloos, J., van Atteveldt, W., et al. (2022). OSD2F: An open-source data donation framework. *Computational Communication Research*, 4(2), 372–387. <https://doi.org/10.5117/ccr2022.2.001.arau>.
- Bowman, N. D., & Spence, P. R. (2020). Challenges and best practices associated with sharing research materials and research data for communication scholars. *Communication Studies*, 71(4), 708–716. <https://doi.org/10.1080/10510974.2020.1799488>.
- Cuadrado-García, M., Montoro-Pons, J. D., & Miquel-Romero, M.-J. (2023). Measuring music-genre preferences: Discrepancies between direct and indirect methods. *Psychology of Music*, 51(3), 1026–1044. <https://doi.org/10.1177/03057356221126200>.
- Dienlin, T., Johannes, N., Bowman, N. D., et al. (2021). An agenda for open science in communication. *Journal of Communication*, 71(1), 1–26. <https://doi.org/10.1093/joc/jqz052>.
- Freelon, D. (2018). Computational research in the post-API age. *Political Communication*, 35(4), 665–668. <https://doi.org/10.1080/10584609.2018.1477506>.
- Haim, M., & Puschmann, C. (2023). Opening up data, tools, and practices: Collaborating with the future. *Digital Journalism*, 11(2), 247–254. <https://doi.org/10.1080/21670811.2023.2174894>.
- Haim, M., Leiner, D., & Hase, V. (2023). Integrating data donations in online surveys. *Medien & Kommunikationswissenschaft*, 71(1–2), 130–137. <https://doi.org/10.5771/1615-634X-2023-1-2-130>.
- Hepp, A., Hohmann, F., Belli, A., et al. (2021). Forschungssoftware in der Kommunikations- und Medienwissenschaft: Stand, Herausforderungen und Perspektiven [Research software in communication and media studies: Status quo, challenges, and perspectives]. https://www.dgpuk.de/sites/default/files/DGPuK%20Positionspapier%20-%20Forschungssoftware%20in%20der%20Kommunikations-%20und%20Medienwissenschaft_0.pdf. Accessed: 6 July 2023.
- Peter, C., Breuer, J., Masur, P. K., Scharkow, M., & Schwarzenegger, C. (2020). Empfehlungen zum Umgang mit Forschungsdaten in der Kommunikationswissenschaft [Guidelines for handling research data in communication science]. *Studies in Communication and Media*, 9(4), 599–626. <https://doi.org/10.5771/2192-4007-2020-4-599>.
- Schoch, D., & Chan, C. (2023). Software presentation: Rtoot: Collecting and analyzing Mastodon data. *Mobile Media & Communication*. <https://doi.org/10.1177/20501579231176678>.
- Shaw, A., Scharkow, M., & Wang, Z. J. (2021). Opening a conversation on open communication research. *Journal of Communication*, 71(5), 677–685. <https://doi.org/10.1093/joc/jqab033>.
- Strippel, C. (2021). Forschungsinfrastrukturen für die Kommunikations- und Medienforschung im deutschsprachigen Raum: Initiativen, Bedarfe und Perspektiven [Research infrastructures for media and communication research in German-speaking countries: Initiatives, requirements, and perspec-

- tives]. *Medien & Kommunikationswissenschaft*, 69(1), 136–157. <https://doi.org/10.5771/1615-634X-2021-1-136>.
- Strippel, C., Breuer, J., Fürst, S., Koenen, E., Prandner, D., & Schwarzenegger, C. (2022). Call for papers—Special issue “Data, archives and tools: infrastructures and resources for communication and media research”. *Publizistik*, 67(2–3), 219–224. <https://doi.org/10.1007/s11616-022-00729-9>.
- Wedel, L. (2023). The R package rtoot. *Mobile Media & Communication*. <https://doi.org/10.1177/20501579231176674>.
- Wilkinson, M., Dumontier, M., Aalbersberg, I., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>.

Christian Strippel is research unit lead of the “Weizenbaum Panel” and the “Methods Lab” at the Weizenbaum Institute for the Networked Society.

Dr. Johannes Breuer is a Senior Researcher in the team “Survey Data Augmentation” at GESIS and leads the team “Research Data Methods” at the Center for Advanced Internet Studies (CAIS).

Dr. Silke Fürst is a Senior Research and Teaching Associate in the Department of Communication and Media Research (IKMZ) at the University of Zurich.

Dr. Erik Koenen is university lecturer (in interim) at the ZeMKI, Centre for Media, Communication and Information Research at the University of Bremen.

Dr. Dimitri Prandner is Senior Scientist (R3) for research data management and empirical social research at the JKU as well as founding member of AUSSDA – the Austrian Social Science Data Archive. He is task lead in the Horizon 2020 Project RIntrainPLUS, researching the skills necessary to manage European research infrastructure.

Dr. Christian Schwarzenegger is currently Professor for Media and Society (interim) at the University of Bremen and the chair of the Communication History Sections of ECREA and DGPuK.