


German Scholarly Blogs and their Attributes

July 16, 2025


Catharina Ochsner

Humboldt-Universität zu Berlin, Berlin School of Library and
Information Science

 <https://orcid.org/0009-0005-3885-3951>


Heinz Pampel

Humboldt-Universität zu Berlin, Berlin School of Library and
Information Science & Helmholtz Association, Helmholtz Open
Science Office

 <https://orcid.org/0000-0003-3334-2771>


Jonas Höfting

Humboldt-Universität zu Berlin, Berlin School of Library and
Information Science

 <https://orcid.org/0009-0003-4466-1775>

Laura Rothfritz

Humboldt-Universität zu Berlin, Berlin School of Library and
Information Science

 <https://orcid.org/0000-0001-7525-0635>

Abstract

This document describes a dataset that contains information on 866 German scholarly blogs. The data were aggregated through various sources, including scholarly blogging platforms, the German National Library and blogrolls. The dataset contains general data (research field, activity status, language, authors' institutional affiliation), data on ways that scholarly blogs are already being integrated into digital research and information infrastructure (DNB entry, ISSN, DOI, citation proposal, platform used), and data on the efforts of scholarly bloggers that make their blogs accessible and reusable (blog archive, blogroll, Creative Commons, RSS/Atom feed, comments).

Citing the Dataset

Citation proposal: Ochsner, C., Pampel, H., Höfting, J., Rothfritz, L., (2025). German Scholarly Blogs and their Attributes. <https://doi.org/10.5281/zenodo.14905628>

Licenses



This documentation is licensed under Creative Commons Attribution 4.0 International terms.



The data files are licensed under Creative Commons CC0 1.0 Universal terms.

About the Project

Scholarly blogs are an integral part of digital science and scholarly communication. Unanswered questions are for example solutions for the long-term preservation of blog posts and the lack of their integration into libraries and information infrastructure (Pampel & Rothfritz, 2024). The Infra Wiss Blogs project supports blogging researchers and information infrastructure institutions in developing solutions to ensure the long-term accessibility of scholarly blogs in Germany. The project is funded by the German Research Association¹ (DFG) under the project number 528958385.

Motivation

A previously little-addressed research question on the topic of scholarly blogs is that of the long-term accessibility of blogs, as well as the resulting infrastructural aspects. While digital information infrastructures have been organised around other types of scholarly output (e.g. textual publication types such as books and journals) to ensure their long-term accessibility, these processes and infrastructures have not been ensured when it comes to scholarly blogs, which bears the risk of information loss (Burton, 2015; Hank, 2011). Scholarly blogs pose legal and technical challenges for preservation, since they embed dynamic content such as videos, links and comments (Burton, 2015; Hank, 2011). Even though authors of traditional scholarly publications rarely view the long-term accessibility of their content to be their responsibility (Altenhöner & Schrimpf, 2014), the long-term accessibility of blogs has predominantly been discussed by bloggers themselves (Fenner, 2022; Hank, 2011) or projects concerned with building infrastructure to preserve scholarly blogs (Fenner, 2022; Guilleux, 2024; Kalb et al., 2013; Lazaridou et al., 2013). Within the scholarly literature on the topic, there is also no overview of existing scholarly blogs or data on how scholarly blogs are integrated into already existing digital research and information infrastructure. To address this issue, this dataset contains data on scholarly blogs and their attributes.

Methodology

Data Collection

Inclusion and exclusion criteria

Since the scholarly literature on the topic does not provide a clear definition of scholarly blogs, we determined content-related, formal and technical inclusion and exclusion criteria. Table I summarises the criteria that we that we applied to all of the blogs that were included in our sample.

¹<https://www.dfg.de/en>

Table I: Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
Websites/Blogs with reversed chronological feeds	Microblogs, newsletters, news portals, podcasts, static websites
German (based on address, content, or author affiliation)	Non-German blogs or those with no clear country affiliation
Blogs by active scholars, science journalists, doctoral students and supervised students	Blogs by news journalists or practitioners that are not communicating research or research- related content
Blogs from archives, libraries, or museums communicating research	Blogs from academic or infrastructure institutions that are not communicating research or research- related content
Content about research, academia, and science journalism	Journalistic news or practical content that are not communicating research or research- related content

Blog collection

To collect the sample of blogs, we started out by including blogs listed by the annual award “Wissenschaftsblog des Jahres” (eng: Scholarly Blog of the Year) hosted by the blog “Wissenschaft kommuniziert” (eng: Science communicated/Science communicating) (Wissenschaft kommuniziert, 2022). From 2011 to 2022 the blog published a list of varying numbers of scholarly blogs as part of the award. Using this process, we identified 34 German scholarly blogs across various disciplines that met our inclusion criteria. As a second step we included seven scholarly blogs that were mentioned in the project proposal for the project Infra Wiss Blogs (Pampel & Rothfritz, 2024).

We then expanded the sample by including blogs from the research centres Helmholtz (15) and Jülich (16), as well as 10 relevant blogs from the Staatsbibliothek zu Berlin (eng: State Library Berlin). Additionally, we included blogs from several blogging portals, such as de.hypotheses (446), ScienceBlogs (88) and SciLogs (24). To collect blogs from de.hypotheses, we used the OpenEdition blog catalogue and limited the results to blogs where the country of publication was Germany and included all the results that fit our inclusion criteria. We started this process on the 11th of July 2024, and updated the list until the 14th of October 2024.

Furthermore, we researched scholarly blogs that are hosted by the German blog portal ScienceBlogs. This included their active and archived blogs, which resulted in 88 blogs from ScienceBlogs. Lastly, we included 24 blogs hosted by SciLogs, a German blogging platform that hosts interdisciplinary scholarly blogs. Even though we were provided with a list of blog names from the SciLogs editorial office, finding inactive blogs was difficult, since these were not listed on the SciLogs blog portal. Additionally, we were provided with a complete list of catalogued

blogs in the German National Library (up until the 30th of July 2024). This list contained 707 blogs, however due to duplicates (especially from de.hypotheses), inaccessible blogs and blogs that didn't meet our inclusion criteria, we ended up including 76 blogs into our sample.

Then we added 14 blogs that were listed in the blog section of Rogue Scholar, a scholarly blog archive (Fenner, 2023b). We also researched blogs hosted by German universities and ended up including 41 blogs. Lastly, we expanded the sample using the snowball method to include blogs that were linked to by the blogs in the initial sample. We only considered blogs that were linked in the starting pages of the blogs of our initial sample, like blogrolls or recommendations. We found another 96 blogs using this method. Since many blogs were mentioned in multiple sources, we avoid duplicate entries by only counting each blog once. Our final sample therefore consists of 866 blogs. Table II summarises the steps we took to collect the blogs for our sample.

Table II: Blog collection process

Step	Outcome
Nominees of the "Scholarly Blog of the Year" (2011-2022)	34 blogs
Project proposal	7 blogs
Helmholtz blogs	15 blogs
Jülich blogs	16 blogs
SBB blogs	10 blogs
Open Edition	446 blogs
ScienceBlogs	88 blogs
SciLogs	24 blogs
DNB entries	76 blogs
Desk research	41 blogs
Rogue Scholar	14 blogs
Blogroll	95 blogs
Final sample size	867 blogs

Data analysis

After collecting the blog sample, we developed three sets of analysis criteria in coherence with our research questions. We then examined each blog for pre-determined criteria, which we list and define in the following. We have further summarised our analysis criteria, answer options and their definitions in table III. The categories feed, license, citation proposal and persistent identifiers (ISSN and DOI) were based on the guidelines for scholarly blogs, a set of recommendations for scholarly bloggers to support the long-term archiving, discoverability, and citation of blog contents (Fenner, 2023a).

Overview of scholarly blogs: To examine how scholarly blogs vary across disciplines in terms of activity, authorship, and language, we categorised blogs by discipline and analysed their

activity status, institutional affiliation, and the languages they used. To categorise blogs into subjects, we used the first layer of the subject classification of the German Research Foundation (e.g. humanities and social sciences, life sciences, natural sciences and engineering sciences (German Research Foundation, n.d.)). Multiple answers were possible if blogs could not be sorted into only one discipline. Since the platform de.hypotheses only hosts scholarly blogs from the humanities and social sciences, we automatically sorted the de.hypotheses blogs into the humanities and social sciences. In order to determine the blogs' activity status, we researched the dates of first and last posts of blogs. We considered blogs to be active when they posted a year before we started our research, so as we started our research on the first of July 2024, all blogs that posted at least once from the first of July 2023 until the 14th of October 2024 (when we concluded data analysis) were considered active. All blogs whose last post was dated before the first of July 2023 were considered inactive. We defined blogs with an institutional affiliation to include blogs that are explicitly affiliated with research performing organisations e.g. through research programs or researchers that explicitly blog about their work. We also included blogs by archives, libraries and museums. Blogs with no institutional affiliation are independent blogs, by scholars (who may work at a research performing organisation but whose blog is not explicitly connected to the research they performed at their institution), associations, clubs, or publishers. Lastly, we noted the language used in blogs. When recording multilingual blogs, we only included blogs that consistently posted in multiple languages and did not include blogs that only sparsely published posts in another language.

Integration into digital research and information infrastructure: To assess how scholarly blogs and their content are integrated into existing digital research and information infrastructures for long-term accessibility and preservation, we evaluated platform and software choices and checked whether the blogs are indexed by the German National Library with an International Standard Serial Number (ISSN). The ISSN is an identification system for periodical publications (Deutsche Nationalbibliothek, n.d.). Additionally, we checked if bloggers are using Digital Object Identifiers (DOIs) and citation proposals for their posts. We determined blogging platforms by checking URLs and the presentation layer of blogs and its integration into other websites or infrastructure. We knew prior to the data collection that the platforms de.hypotheses, ScienceBlogs and SciLogs used the WordPress Software. For other blogs we checked for software information on the blogs or used the browser add-on *Wappalyzer*². We noted entries in the German National Library using the list of catalogued blogs the library provided for us. Checking other libraries or collective library catalogues for entries was unfortunately beyond the scope of this study. Once again, we relied on ISSNs provided by the German National Library through the blog list they provided for us and also added entries if we found ISSNs on the blogs themselves. Furthermore, we checked the blog posts for the provision of DOIs and citation proposals.

Efforts by Bloggers: To explore the strategies bloggers use to enhance the accessibility and reusability of their content, we searched for a blog archive, a blogroll, a Creative Commons (CC) license, and an RSS/Atom feed. We only collected archives that were integrated into the blogs

²<https://www.wappalyzer.com/>

themselves, not whether the blogs had an entry in an institutional archive. Furthermore, we only checked the landing pages of blogs for blogrolls and made an entry only if we noticed that a blogroll was provided there. We checked starting pages, posts and imprints for information on license and only included blogs that licensed their whole blog content, as opposed to individual posts. Lastly we noted whether or not blogs provided feeds directly on their website and – in case they didn’t – noted if we could access feeds through feedreaders. Lastly we checked if the comment function was enabled for blog posts.

Table III: Components of the dataset

Criteria	Options	Definition
Discipline	Humanities & Social Sciences, Life Sciences, Natural Sciences, Engineering Sciences	Assigned using the German Research Foundations classification.
Institutional affiliation	Yes/No	<i>Yes</i> : Blogs affiliated with research performing organisations (e.g. universities, colleges, research institutions) archives, libraries, museums. <i>No</i> : Independent blogs, by scholars, associations, clubs, or publishers.
Blog archive	Yes/No	Integrated archive of blog posts.
Blogroll	Yes/No	List of recommended blogs on starting page.
Comments	Yes/No	Applies to blog posts.
ISSN	Yes/No	Provision of an ISSN.
DOI	Yes/No	Applies to blog posts.
Citation proposal	Yes/No	Indicates if blog posts provides a citation proposal.
German National Library entry	Yes/No	Indicates entry in the German National Library catalogue.
Feed	Yes/No (or only via reader)	Indicates availability of an RSS/Atom feed.
Activity status	Active/Inactive	<i>Active</i> : Last post after June 30, 2023. <i>Inactive</i> : Last post before this date.
Creative Commons	License/NA	Applies to full-text content.
Platform	Open field	Blogging platform used.
Software	Open field	Underlying software.
Language	Open field	Languages of the blog.
First post date	Open field	Date of the first blog entry.
Last post date	Open field	Date of the most recent blog entry.

References

- Altenhöner, R., & Schrimpf, S. (2014, December). 11.2 Bestandserhaltung und Langzeitverfügbarkeit digitaler Ressourcen: Strategie, Organisation und Techniken. In *11.2 Bestandserhaltung und Langzeitverfügbarkeit digitaler Ressourcen: Strategie, Organisation und Techniken* (pp. 862–884). De Gruyter Saur. <https://doi.org/10.1515/9783110303261.862>
- Burton, M. (2015). *Blogs as Infrastructure for Scholarly Communication* [Doctoral dissertation, University of Michigan].
- Deutsche Nationalbibliothek. (n.d.). *Nationales ISSN-Zentrum für Deutschland*. Retrieved January 23, 2025, from https://www.dnb.de/DE/Professionell/Services/ISSN/issn_node.html
- Fenner, M. (2022). Building an archive for scholarly blog posts. <https://doi.org/10.53731/br9f5xa-a556w2t>
- Fenner, M. (2023a). Guidelines for Scholarly Blogs. *Front Matter*. <https://doi.org/10.53731/a0d9m3n-n7r8h0m>
- Fenner, M. (2023b, April). The Rogue Scholar is now open for business. Retrieved February 21, 2025, from <https://doi.org/10.53731/z9v2s-bh329>
- German Research Foundation. (n.d.). *DFG Classification of Scientific Disciplines, Research Areas, Review Boards and Subject Areas (2020-2024)*. Retrieved January 6, 2025, from <https://www.dfg.de/resource/blob/175336/46d0a8c5c14d0f6530da28b4d25bf589/fachsystematik-2020-2024-en-grafik-data.pdf>
- Guilleux, C. (2024). *La pérennité de vos contenus sur Hypothèses*. <https://doi.org/10.58079/w9ne>
- Hank, C. (2011). *Scholars and their blogs: Characteristics, preferences, and perceptions impacting digital preservation* [Doctoral dissertation, The University of North Carolina at Chapel Hill].
- Kalb, H., Lazaridou, P., Banos, V., Kasioumis, N., & Trier, M. (2013). BlogForever: From web archiving to blog archiving. In *Informatik 2013 informatik angepasst an mensch, organisation und umwelt* (pp. 536–549). Gesellschaft für Informatik e.V.
- Lazaridou, P., Kalb, H., Pinsent, Y. K., Banos, V., Ross, S., Kopidaki, S., & Berninger, L. (2013). BlogForever: D3. 2 Interoperability Prospects.
- Pampel, H., & Rothfritz, L. (2024). Kooperative Informationsinfrastruktur für wissenschaftliche Blogs (Infra Wiss Blogs). <https://doi.org/10.5281/zenodo.13234482>
- Wissenschaft kommuniziert. (2022). *Die Wissenschafts-Blogs des Jahres 2022 sind gewählt! Fakten gegen Querdenken*. Retrieved January 9, 2025, from <https://wissenschaftkommuniziert.wordpress.com/2023/01/31/die-wissenschafts-blogs-des-jahres-2022-sind-gewahlt/>