



The Infrastructural and Software Challenges of Independent Search Engine Research

Sebastian Sünkler^a, Phil Höfer^b, Philipp Schaer^c, Aleksandra Urman^d, Dirk Lewandowski^a

^a *Hamburg University of Applied Sciences, Finkenau 35, 22081 Hamburg, Germany*

^b *SUMA-EV, Henniesruh 28D, 30655 Hannover, Germany*

^c *TH Köln, Claudiusstraße 1, 50678 Köln, Germany*

^d *University of Zurich, Andreasstr. 15, 8050 Zürich, Switzerland*

1. Purpose of the Panel

Search engines are the primary gatekeepers of digital information, making their societal, epistemic, and political implications a critical area of study. However, independent research into major commercial search engines (e.g., Google, Bing) faces significant challenges: these platforms operate as proprietary “black boxes”, and official access points (such as APIs) are increasingly restricted, expensive, or discontinued (Sünkler et al., 2025). The purpose of this panel is to move beyond merely outlining the infrastructural challenges of auditing commercial search providers and to actively formulate actionable solutions. This session will critically debate whether emerging regulatory frameworks, such as the Digital Services Act (DSA) (The Digital Services Act | Shaping Europe’s Digital Future, n.d.), serve as a solution for accessing data from commercial search engines, or if the academic community must prioritize building independent, open collaborative infrastructures like the Open Web Search project (Hendriksen et al., 2024). By the end of this 90-minute session, attendees will gain strategies for navigating data access restrictions and a shared action plan for sustainable search engine research.

2. Abstract

As search engines increasingly influence our reality and mediate phenomena such as information disorders and political polarization, the need for independent, society-focused research has a high relevance. Yet, the tools required to conduct this research are incredibly difficult to build and sustain. Search Engine Result Pages (SERPs) have evolved into complex environments dominated by generative components, while commercial search providers heavily restrict data access.

Moderated by Dirk Lewandowski, this panel addresses the acute methodological infrastructure crisis in search engine research. Rather than centering solely on the challenges, this solution-focused session will debate the most viable paths forward for investigating commercial search monopolies.

We will narrow our focus to three critical areas: navigating the tightening of API restrictions imposed by commercial platforms, the economic viability of operating independent search infrastructures, and the dual strategy of evaluation: deploying living labs on open infrastructures to maintain scientific progress, and utilizing advanced user simulations to externally audit the dynamic features of closed commercial SERPs.

SEASON 2026, September 15–17, 2026, Hamburg, Germany

EMAIL: sebastian.suenkler@haw-hamburg.de (A1); mail@phil-hoefer.de (A2); philipp.schaer@th-koeln.de (A3); urman@ifi.uzh.ch (A4); dirk.lewandowski@haw-hamburg.de (A5)

ORCID: 0000-0001-9848-1137 (A1); 0000-0002-8817-4632 (A3); 0000-0003-3332-9294 (A4), 0000-0002-2674-9509 (A5)

DOI: <https://doi.org/10.48441/4427.3733>



© 2026 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



By zooming out from individual tools to the wider ecosystem, we will examine the necessity of open, collaborative infrastructures. Finally, this panel argues that independent software development is a societal imperative to ensure algorithmic accountability, fair search, and the continued viability of independent digital research. Attendees will leave equipped with concrete strategies for sustainable tool development, a critical understanding of how to leverage regulatory acts for research access, and new collaborative pathways to avoid duplicating infrastructural labor.

3. Panelists and Perspectives

- Prof. Dr. Dirk Lewandowski (HAW Hamburg) - Moderator: Dirk Lewandowski is a Professor of Information Research and Information Retrieval and a leading international expert on the societal role of search engines (Lewandowski, 2023; Lewandowski et al., 2024). He will guide the discussion on whether the community should rely on forcing data access from monopolies or building independent public search infrastructures.
- Dr. Aleksandra Urman (University of Zurich): As a computational social scientist researching the societal impacts of AI and digital platforms, Dr. Urman will bring the perspective of conducting large-scale algorithmic impact audits (Ulloa et al., 2022; Urman et al., 2022). She will discuss the high-level challenges of capturing real-time data at a global scale (Urman et al., 2026) and why independent archiving infrastructures are vital to exposing algorithmic biases and information inequalities.
- Sebastian Sünkler (HAW Hamburg): Operating at the intersection of research and software engineering, Sebastian will focus on the methodological and technical realities of extracting data from major commercial search engines like Google. He will address the challenges between academic needs and platform restrictions, and the challenge of building sustainable, open-source environments like the Result Assessment Tool (RAT) (Sünkler et al., 2025) for evaluating search engines.
- Prof. Dr. Philipp Schaer (TH Köln): Specializing in Information Retrieval and Search Engine Technology, Prof. Dr. Schaer will address how the community can pivot when commercial platform data access disappears. Drawing from his DFG-funded STELLA project (Breuer et al., 2019) with partners like GESIS and ZB MED, he will discuss the transition toward open infrastructure for “Living Labs” (Breuer & Schaer, 2021). Furthermore, he will outline how user simulation techniques can evaluate interactive IR systems, while algorithmic auditing (such as extracting and analyzing query suggestions) remains crucial for externally investigating the biases of closed commercial engines like Google.
- Phil Höfer (Suma-ev / MetaGer): Representing the operational trenches of an independent, privacy-friendly search engine Metager², Phil will discuss the structural and economic challenges of maintaining alternative search architectures. He will provide insights into competing against tech monopolies and the pressing need for public, open infrastructures to lower market entry barriers and support independent search.

4. Proposed Format (90 Minutes)

- **00:00 - 00:10 | Introduction & Framing:** The Moderator introduces the panelists and sets the stage, outlining the core theme: the infrastructural crisis in independent search engine research and the death of public APIs. He will also introduce a regulatory (e.g., DSA) and corporate compliance perspective to ensure a balanced debate.
- **00:10 - 00:40 | Thematic Opening Statements:** Instead of presenting specific tools, each panelist will spend 5-7 minutes discussing their view on the biggest systemic challenge in their domain (e.g., data volatility, sustainable engineering, corporate gatekeeping, or infrastructure costs).

² <https://metager.de/>



- **00:40 - 01:10 | Moderated Panel Discussion:** A guided, solution-oriented debate. Key questions: Should the community rely on the DSA to compel data access from commercial search engines, or build independent scrapers? How can traditional evaluation methods adapt to generative SERPs using Living Labs and user simulations? How do we financially sustain open-source alternatives to commercial monopolies?
- **01:10 - 01:30 | Audience Q&A:** An open floor for attendees to engage with the panel, share their own infrastructural challenges and solutions, and synthesize the strategies discussed into a shared action plan.

References

- Breuer, T., & Schaer, P. (2021). A Living Lab Architecture for Reproducible Shared Task Experimentation. In C. Wolff & T. Schmidt (Eds.), *Information between Data and Knowledge* (Vol. 74, pp. 348–362). Werner Hülsbusch. <https://doi.org/10.5283/epub.44953>
- Breuer, T., Schaer, P., Tavakolpoursaleh, N., Schaible, J., Wolff, B., & Müller, B. (2019). STELLA: Towards a Framework for the Reproducibility of Online Search Experiments. In R. Clancy, N. Ferro, C. Hauff, J. Lin, T. Sakai, & Z. Z. Wu (Eds.), *Proceedings of the Open-Source IR Replicability Challenge co-located with 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, OSIRRC@SIGIR 2019* (Vol. 2409, pp. 8–11). CEUR-WS.org.
- Hendriksen, G., Hiemstra, D., de Vries, A., Gürtl, S., Nussbaumer, A., Heineking, S., Zelch, I., Potthast, M., Granitzer, M., Dinzinger, M., Zerhoubi, S., & Fathima, N. A. (2024, November). *The OpenWebSearch Hub and the Open Web Index Y2*. Zenodo. <https://doi.org/10.5281/zenodo.14186646>
- Lewandowski, D. (2023). *Understanding search engines*. Springer International Publishing. <https://doi.org/10.1007/978-3-031-22789-9>
- Lewandowski, D., Haider, J., & Sundin, O. (2024). JASIST Special Issue Editorial: Re-orienting search engine research in information science. *Journal of the Association for Information Science and Technology*, 75(5), 503–511. <https://doi.org/10.1002/asi.24845>
- Sünkler, S., Lewandowski, D., Schultheiß, S., & Yagci, N. (2025). Result Assessment Tool (RAT): Empowering search engine data analysis. *PeerJ Computer Science*, 11, e2962. <https://doi.org/10.7717/peerj-cs.2962>
- The Digital Services Act | Shaping Europe's digital future. (n.d.). Retrieved July 14, 2026, from <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act>
- Ulloa, R., Makhortykh, M., & Urman, A. (2022). Scaling up search engine audits: Practical insights for algorithm auditing. *Journal of Information Science*, 016555152210930. <https://doi.org/10.1177/01655515221093029>
- Urman, A., Hannák, A., & Baumann, J. (2026). *GoogleTrendArchive: A Year-Long Archive of Real-Time Web Search Trends Worldwide* (arXiv:2603.21871). arXiv. <https://doi.org/10.48550/arXiv.2603.21871>
- Urman, A., Makhortykh, M., & Ulloa, R. (2022). The Matter of Chance: Auditing Web Search Results Related to the 2020 U.S. Presidential Primary Elections Across Six Search Engines. *Social Science Computer Review*, 40(5), 1323–1339. <https://doi.org/10.1177/08944393211006863>