Interactive platform for the exploration of large-scale ‘living’ systematic maps

Tim Repke
Potsdam Institute for Climate Impacts Research / Potsdam, Germany
tim.repke@pik-potsdam.de

Abstract

Research syntheses, such as systematic maps or evidence and gap maps, provide valuable overviews of the coverage of research in a particular field. They serve as pointers for funders and researchers to identify important gaps in the literature where more research is needed but also to find relevant work for more in-depth systematic reviews or meta-analyses. However, systematic maps become outdated quickly, sometimes even after they are released due to the time it takes to screen and code the available literature and long publication processes. Furthermore, the write-up of the synthesis (in form of a peer-reviewed article) can only serve as a high-level summary—for detailed questions one would need full access to the underlying data. To this end, we developed an interactive web-based platform to share annotated datasets. For some datasets, where automated categorisation passes the necessary scientific quality standards, we also update the data as new research becomes available and thus make them ‘living’.

1 Introduction

The number of scientific publications is continually growing at an exponential rate. For example, more articles on climate change were published during the sixth assessment cycle of the IPCC than during all previous cycles since 1985 combined (Callaghan et al., 2020). Systematic maps of timely topics that are up-to-date are crucial tools to get an overview of a specific field, to identify research gaps, or to identify articles that are relevant for a particular meta-study or review (JPT et al., 2024; Kastner et al., 2016). The sheer amount of potentially relevant literature to consider and the rapid growth make it increasingly prohibitive to conduct systematic maps by hand. Digital evidence synthesis tools can speed up the most time-consuming of a synthesis, particularly screening abstracts in search for relevant articles (Haddaway et al., 2020; Tsafnat et al., 2014). The Covid-19 pandemic has shown the value of so-called ‘living evidence syntheses’ that are continually updated to capture findings from the latest clinical trials and other research strands (Chakraborty et al., 2024; Elliott et al., 2014). The crises of our time require similar up-to-date repositories of evidence to support evidence-based policy-making.

Traditional publication models, however, are not able accommodate requirements of regular and frequent updates (Thomas et al., 2017). By the time a research team submit their initial draft, especially until the final publication, the synthesis might already be outdated. The publication is by definition a high-level overview of the underlying data, for which the authors and contributors have spent a lot of time to compile and annotate. However, this raw data is often not published alongside the article or is no longer available. Even where data is available, it might only be available in a proprietary format or the schema used in a csv file might not be self-explanatory. This makes it hard for other researchers or policy analysts to utilise existing categorisations to find relevant literature of their particular questions.

To this end, we developed an interactive web-platform for sharing the underlying data of systematic maps. The initial prototype[1] hosts four projects: A systematic map of literature on climate policy instruments (Callaghan et al., 2025), a systematic evidence and gap map of literature on carbon pricing (Döbbeling-Hildebrandt et al., 2024), a systematic map of literature on carbon dioxide removal (Lück et al., 2024; Smith et al., 2023, 2024), and a systematic map of literature on climate and health (Berrang-Ford et al., 2021; Romanello et al., 2023, 2024). We are also working on adapting data from past publications and making all maps ‘living’ by building pipelines that automatically retrieve new publications and classify them. Depending on

[1] https://climateliterature.org/

Proceedings of the 2nd Workshop on Natural Language Processing Meets Climate Change (ClimateNLP 2025), pages 260–265
July 31, 2025 ©2025 Association for Computational Linguistics

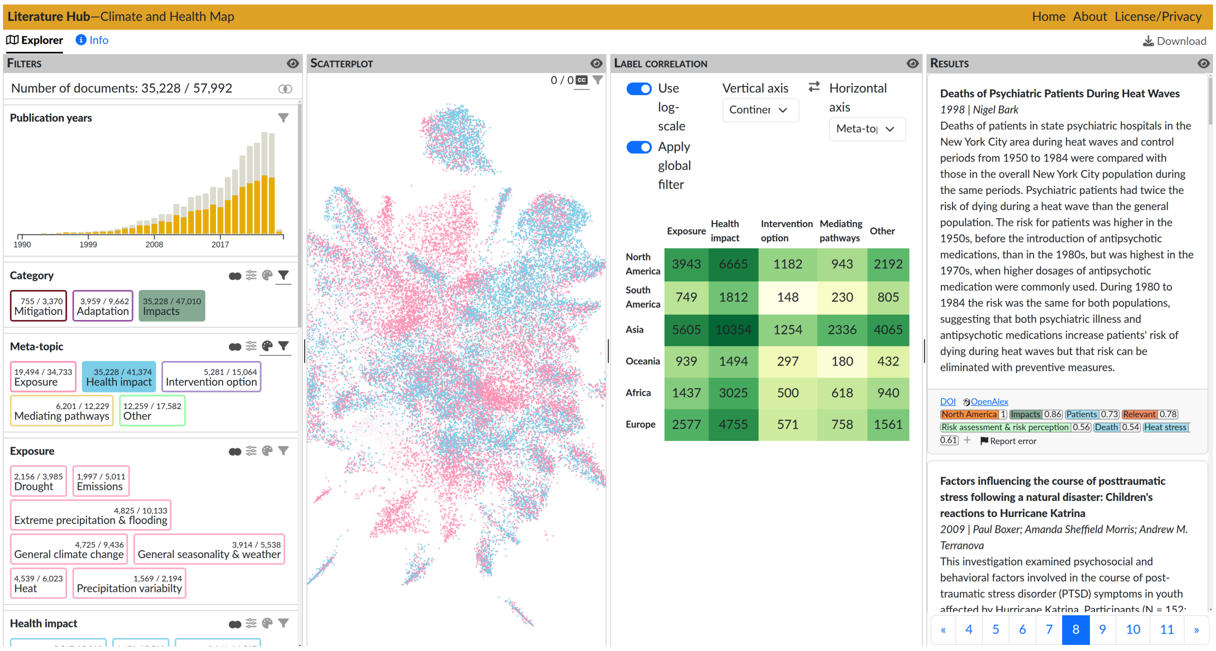

Figure 1: Screenshot of the interactive climate literature hub; Left panel contains all available filters, the second panel shows a scatterplot of the semantic landscape with a descriptive keyword overlay (hidden in the screenshot), the third panel provides a heat-map (or gap-map) to show how different filters or labels correlate, the right panel shows the abstracts that match the current set of filters. Not shown is a panel of a geographic map to see which places are mentioned in abstracts or where authors' affiliations are located.

the purpose of the map and the quality of machine-learning classifiers, this entire process can—with several caveats—be fully automated. With this platform we want to foster open research, transparency, and reusability and make up-to-date evidence easily accessible for anyone. The platform itself is also open-source and available for anyone to adapt or host.[2] The system can easily be adapted to include additional datasets by adding a meta-data and database file using a very basic format (see project repository for details).

Figure 1 shows the main screen of the latest prototype (March 2025) for our map of literature at the nexus of climate and health. The interface is modular, so authors can decide how to best showcase their data. Current components feature various filters for publication year (with a histogram), normal labels (boxes that can be selected), as well as full-text and author search. Furthermore, the dataset can be explored on a 'semantic landscape', a scatter-plot where each dot represents an article and their close proximity indicates high similarity. This explorative visualisation, inspired by Nomic AI (González-Márquez et al., 2024), may be useful to quickly identify clusters of similar works or to see how specific filters cover the topical spectrum. The lasso-selection tool on the landscape also acts as an additional filter. Aside from the semantic landscape, there is also a component to display geographical locations associated with the articles, for example by mentions in the abstract or author affiliations. Regions and location on this map can also be used as filters. The heat-map component provides a quick overview of how labels correlate and is inspired by gap-maps (Snilstveit et al., 2016). Lastly, we implemented a list component to show the most relevant records based on the current global filter.

In the remainder of this article, we describe some of the technical aspects that enable this web-platform to be highly interactive—even for large datasets and many filters. Furthermore, we discuss some challenges of automated updates of datasets shared on this platform and considerations for future work.

2 Real-time filtering

One of the key features of the web-platform is the ability to combine arbitrary filters and receive real-time feedback of how they influence the overall statistics. The platform is built to handle large datasets—tested for a million records. This poses

[2] https://gitlab.pik-potsdam.de/mcc-apsis/living-evidence-maps/literature-hub

a particular challenge for optimising the traffic between the front-end and the server. We address this challenge by only transmitting binary bit-masks for each label during the initial page-load. This allows us to do all computations in the client's browser very efficiently, including updating the visualisations. The required traffic for each bit-mask is around 1.2kb per 10,000 records per filter before additional gzip compression on the transport layer. In the example shown in Figure 1, there are three filters grouped under 'Category'. Publication years and x/y coordinates for the scatter-plot are transmitted as uint16 and float16 in a light-weight batched Apache arrow file amounting to around 120kb per 10,000 records.

While we have not conducted dedicated performance experiments, we have not encountered any noteworthy lags to hinder any interactive exploration of the data. For example, on a basic laptop from 2021,[3] a dataset with 78k records and 20 filters only takes a few milliseconds to update all counts and has rendered the scatterplot component in under 200ms after a click event on one of the filters is triggered.

This design has the added benefit, that server requirements are very limited. The raw data is stored by sqlite files that contain raw classifier or topic-model scores. Where human annotations are available, they supersede automatically assigned labels and are set as explicit zero or one scores, whereas all others are limited to the range 0.01–0.99. Alternative bit-masks are transmitted if the user sets different thresholds. By setting the thresholds accordingly, users can decide to only show human annotations. For full-text and author search, the same mechanism is used and generates specific bit-masks.

The result set is loaded in the frontend by sending the bit-mask of the current global filter—the combination of all active filters—to the server, which then responds with the records ranked by the sum of the stored classifier scores.

Overall, users have given very positive feedback and were excited to explore our systematic maps in real-time, filter for what they need, and download records including all labels as a csv for the selection they made. At this point, we have not conducted systematic user studies, but plan to do so to inform future developments.

[3]ThinkPad T14s, no dedicated GPU

3 Considerations for 'living' maps

The conventional process for a systematic map follows a linear and very labour-intensive structure. Once an appropriate (boolean) query is developed, the author team would retrieve bibliographic metadata from a search engine like the Web of Science or Scopus and then screen all records by hand for inclusion. Included records are then annotated further, in case of a review additional information is extracted from the full-text. With the help of automation (Thomas et al., 2017), we can speed up the process by prioritising which records to screen and stop early without having to look at all records (Callaghan and Müller-Hansen, 2022). This also means, that we can design more inclusive queries may lead to more complete systematic maps overall. Once enough labels are collected, other categorisations might also be done with machine-learning classifiers to automate future updates once the first version of a systematic map is published. This means, that we can build fully-automated pipelines to reproduce the original study and run this pipeline regularly (for example daily) to also include newly published research. At this point, we did not develop a standardised framework for machine-learning-based classifications and refer to the original publications the respective datasets came from for how automation was developed and how well it performs.

However, such an automated update should also come with a protocol for how the quality is monitored over time. As a research area evolves, keywords that are relevant for the topic of the systematic map may change, which requires updates to the boolean query at the first step of the pipeline. This then may also require additional annotations as the scope changes to ensure a high-quality scientific standard. Depending on the use-case, a systematic map might prioritise inclusiveness over precision. However, that trade-off might be hard to communicate to users, especially with varying levels of classifier performance. The versions of the dataset also need to be clearly marked, for example to distinguish if only new data was added or whether classifiers were re-trained or the process changed in any way.

Furthermore, the research community needs to develop guidelines for the safe and responsible use of automation. Particularly the rise of generative large language models has already found early adopters in research synthesis. The performance of such models are very hard to validate, as they are also shown to suddenly fail. However, more conventional supervised classification models are also rarely perfect, especially with limited data available for training. In these cases, we need clear guidelines when it is acceptable to still use automation or in which use-cases some categorisation have to be omitted from automated updates and rather need to be reviewed by experts before including them in the published dataset.

Another aspect to consider is continuity of a living systematic map that users return to. Visualisations, such as semantic landscapes need to remain relatively stable over time and should not suddenly use a completely different layout, which, to an extent, can be ensured by adding new data to existing layouts and some additional fine-tuning (Poličar et al., 2024; Repke and Krestel, 2021). As mentioned before, classifiers may be updated during the life-time of a living map. In this case, labels for records that were in earlier versions might change, which could have an impact on downstream users.

Initially, the update iterations are just available via additional fine-grained publication date filters. In future work, we are planning to develop newsletters that interested parties can subscribe to that sends them a list of latest publications that apply to their filter settings. Furthermore, we plan to include a dashboard that showcases the latest trends, ideally highlighting semantic shifts.

One major impact on the continuity across versions of a living systematic map is the data availability in academic search engines. Proprietary databases, such as the Web of Science or Scopus are not accessible to all researchers and the number of results may vary based on the institutional subscription. Open repositories such as OpenAlex or SemanticScholar have shown very good interoperability and coverage (Priem et al., 2022; Culbert et al., 2024), but are increasingly sabotaged by large publishers who enforce the deletion of abstracts from these public indices, which renders the database effectively useless for use in automated pipelines and prevents many researchers from doing their work effectively.

On a similar note, the data shared on such a platform should also be subject to clear licenses where all authors and contributors need to agree to. The license should ideally be very permissive so that other researchers are free to use the annotations for their own work—be it to improve tools for digital evidence synthesis or as a starting point for a systematic review, meta-study, or even a companion map with additional labels or adjusted scope. Finally, as the user base of such a platform grows, they might also identify errors and provide feedback. The maintainers of the living map should to consider how they might want to incorporate the support by a (potentially) global community of experts and laypeople alike. Incorporating feedback and improving the classifications may also mean that annotations for historic data changes. This requires a clear way to reference specific versions of the underlying database for reproducibility, for example by providing daily changelogs.

At this point, we have automated updates for two of the publicly accessible datasets orchestrated by running modified versions of the original studies' scripts and models using scheduled GitLab runners.

4 Limitations

This platform—in its current form—certainly qualifies as an 'expert system'. That means, that some functionality may not be intuitive, especially the combination of filters. We deliberately opted to provide very fine-grained control of how filters can be combined or choosing custom thresholds for classifier scores to adjust the precision/recall trade-offs. Since the development is not directly funded, prioritising usability improvements over additional features or bug fixes is challenging. This highlights another consideration about the sustainability of providing such a platform. The original data should additionally always be published through conventional channels such as zenodo, the OSF, or companion platforms of the publisher.

That said, a systematic evaluation of the usability of the platform should be conducted as part of future work. In particular, such user studies should focus on how each component contributes to a better understanding to contextualise the available evidence and how it can reduce the time required to find relevant evidence for a user's information needs.

5 Conclusion

We have released the climate literature hub, a prototype of a web-platform for sharing data from large-scale systematic maps as we believe in open and transparent research that serves the wider community of researchers and policy-makers. We hope that in exploring our datasets, we inspire 'ecosystems of reviews' in which expert teams can use our maps as starting points for their in-depth analyses on more specific questions or offer a tool for policy-makers to identify relevant research more quickly. Feeding into the platform, we built (semi-)automated pipelines to update the data as new relevant research is published. This submission should serve as a starting point for the wider community how digital evidence synthesis tools can be used responsibly in the future. Future work is needed to develop guidelines for safe use of AI and automation systematic maps feeding into platforms like this, how results can be presented in such a way, that users can make informed decisions if the (possibly imperfect) automated classifications meet their quality needs, and how to address challenges around continuity.

Acknowledgements

Over the course of this project, the authors of this article were supported by the European Research Council (ERC-2020-SyG 'GENIE', grant no. ID951542), the German Federal Ministry for Education and Research ('CDR-SynTra', grant no 01LS2101F), and the Wellcome Trust (PATHFINDER2, grant no. 313586/Z/24/Z and DESTinY, grant no. 227165/Z/23/Z).

References

Lea Berrang-Ford, Anne J Sietsma, Max Callaghan, Jan C Minx, Pauline FD Scheelbeek, Neal R Haddaway, Andy Haines, and Alan D Dangour. 2021. Systematic mapping of global research on climate and health: a machine learning review. *The Lancet Planetary Health*, 5(8):e514–e525.

Max Callaghan, Lucy Banisch, Niklas Doebbeling-Hildebrandt, Duncan Edmondson, Christian Flachsland, William F Lamb, Sebastian Levi, Finn Müller-Hansen, Eduardo Posada, Shraddha Vasudevan, et al. 2025. Machine learning map of climate policy literature reveals disparities between scientific attention, policy density, and emissions. *npj Climate Action*, 4(1):7.

Max W Callaghan, Jan C Minx, and Piers M Forster. 2020. A topography of climate change research. *Nature Climate Change*, 10(2):118–123.

Max W. Callaghan and Finn Müller-Hansen. 2022. Statistical stopping criteria for automated screening in systematic reviews. *Systematic Reviews*, 9.

Samantha Chakraborty, Tanja Kuchenmüller, John Lavis, Fadi El-Jardali, Laurenz Mahlanza-Langer, Sally Green, Ludovic Reveiz, Victoria Carter, Emma McFarlane, Cheryl Pace, Lisa Askie, Fiona Glen, and Tari Turner. 2024. Implications of living evidence syntheses in health policy. *Bulletin of the World Health Organization*, 102:757 – 759.

Jack H. Culbert, Anne Hobert, Najko Jahn, Nick Haupka, Marion Schmidt, Paul Donner, and Philipp Mayr. 2024. Reference coverage analysis of openalex compared to web of science and scopus. *ArXiv*, abs/2401.16359.

Niklas Döbbeling-Hildebrandt, Jan Minx, and Tim Repke. 2024. Protocol: Carbon pricing outcomes – an evidence and gap map of the ex-post literature.

Julian H. Elliott, Tari Turner, Ornella Clavisi, James Thomas, Julian P. T. Higgins, Chris Mavergames, and Russell Lindsay Gruen. 2014. Living systematic reviews: An emerging opportunity to narrow the evidence-practice gap. *PLoS Medicine*, 11.

Rita González-Márquez, Luca Schmidt, Benjamin M. Schmidt, Philipp Berens, and Dmitry Kobak. 2024. The landscape of biomedical research. *Patterns*, 5.

Neal R Haddaway, Neal R Haddaway, Max W. Callaghan, Alexandra Mary Collins, William F. Lamb, Jan C. Minx, James Thomas, and D. John. 2020. On the use of computer- assistance to facilitate systematic mapping. *Campbell Systematic Reviews*, 16.

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, and Welch VA. 2024. Cochrane handbook for systematic reviews of interventions version 6.5.

Monika Kastner, Jesmin Antony, Charlene Soobiah, Sharon E Straus, and Andrea C Tricco. 2016. Conceptual recommendations for selecting the most appropriate knowledge synthesis method to answer research questions related to complex evidence. *Journal of clinical epidemiology*, 73:43–49.

Sarah Lück, Max Callaghan, Malgorzata Borchers, Annette Cowie, Sabine Fuss, Oliver Geden, Matthew Gidden, Jens Hartmann, Claudia Kammann, David P Keller, et al. 2024. Scientific literature on carbon dioxide removal much larger than previously suggested: insights from an ai-enhanced systematic map. *ResearchSquare*.

Pavlin G Poličar, Martin Stražar, and Blaž Zupan. 2024. opentsne: a modular python library for t-sne dimensionality reduction and embedding. *Journal of Statistical Software*, 109:1–30.

Jason Priem, Heather A. Piwowar, and Richard Orr. 2022. Openalex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *ArXiv*, abs/2205.01833.

Tim Repke and Ralf Krestel. 2021. Robust visualisation of dynamic text collections: measuring and comparing dimensionality reduction algorithms. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, pages 255–259.

Marina Romanello, Claudia Di Napoli, Carole Green, Harry Kennard, Pete Lampard, Daniel Scamman, Maria Walawender, Zakari Ali, Nadia Ameli, Sonja Ayeb-Karlsson, et al. 2023. The 2023 report of the lancet countdown on health and climate change: the imperative for a health-centred response in a world facing irreversible harms. *The Lancet*, 402(10419):2346–2394.

Marina Romanello, Maria Walawender, Shih-Che Hsu, Annalyse Moskeland, Yasna Palmeiro-Silva, Daniel Scamman, Zakari Ali, Nadia Ameli, Denitsa Angelova, Sonja Ayeb-Karlsson, et al. 2024. The 2024 report of the lancet countdown on health and climate change: facing record-breaking threats from delayed action. *The Lancet*, 404(10465):1847–1896.

Stephen Smith, Oliver Geden, Matthew Gidden, W Lamb, G Nemet, Jan Minx, Holly Buck, Josh Burke, Emily Cox, Morgan Edwards, et al. 2024. The state of carbon dioxide removal.

Stephen M Smith, Oliver Geden, Gregory F Nemet, M Gidden, WF Lamb, C Powis, R Bellamy, M Callaghan, A Cowie, E Cox, et al. 2023. The state of carbon dioxide removal–the state of carbon dioxide removal.

Birte Snilstveit, Martina Vojtkova, Ami Bhavsar, Jennifer L. Stevenson, and Marie Moland Gaarder. 2016. Evidence & gap maps: A tool for promoting evidence informed policy and strategic research agendas. *Journal of clinical epidemiology*, 79:120–129.

James Thomas, Anna Noel-Storr, Iain James Marshall, Byron C. Wallace, Steve McDonald, Chris Mavergames, Paul P. Glasziou, Ian Shemilt, Anneliese J Synnot, Tari Turner, and Julian H. Elliott. 2017. Living systematic reviews: 2. combining human and machine effort. *Journal of clinical epidemiology*, 91:31–37.

Guy Tsafnat, Paul P. Glasziou, Miew Keen Choong, Adam G. Dunn, Filippo Galgani, and Enrico W. Coiera. 2014. Systematic review automation technologies. *Systematic Reviews*, 3:74 – 74.