From Evidence to Insights: GraphRAG as a Dynamic Knowledge Layer for the Collaboration for Environmental Evidence's Database of Evidence Reviews

Christoph Stade[1], Jodi Schneider[2,3] and Yuanxi Fu[3]

[1]Ludwig-Maximilians-Universität München, Germany
[2]Harvard Radcliffe Institute for Advanced Study, USA
[3]University of Illinois Urbana-Champaign, USA

Abstract

Environmental evidence synthesis plays a crucial role in policy-making, yet policy-makers often struggle to situate individual insights into a broader context of research findings. The Collaboration for Environmental Evidence's Database for Evidence Reviews (CEEDER) provides a collection of evidence, but it remains difficult to navigate. Large language models help to overcome this challenge, doing well at summarization, but often miss critical relational contexts. We propose basing the language model on a knowledge graph that grows with the evidence through the use of Microsoft's GraphRAG. As a dynamic knowledge layer this approach efficiently connects, retrieves, and summarizes relevant research. Unlike traditional manually curated knowledge graphs, GraphRAG dynamically links entities without predefined schemas, adapting to the evolving landscape of environmental science. By enhancing domain-specific query-focused summarization, GraphRAG has the potential to increase data accessibility in settings such as CEEDER where conventional knowledge graphs are not feasible.

Keywords

environmental science, evidence synthesis, generative artificial intelligence, GraphRAG, knowledge graphs, sustainability

1. Introduction

Environmental policy-making depends on the synthesis of research, yet extracting meaningful insights from its vast scientific literature is challenging. The sustainability literature has been doubling every 8.3 years and encompasses disciplines including biology, chemistry, engineering, health sciences, and social sciences [1]. The Collaboration for Environmental Evidence's Database for Evidence Reviews[1] (CEEDER) curates nearly 2,000 systematically assessed reviews to support decision-makers [2], but even these structured resources require manual effort to navigate and interpret.

Traditional search methods like online search or CEEDER's own search string function offer only keyword-based retrieval. They don't provide direct answers and don't connect related findings across different reports. Recent advances in large language models (LLMs) have enabled automated text processing at scale. However, without structured retrieval, LLMs struggle to map conceptual relationships, often overlooking the critical connections necessary for effective decision-making [3].

Knowledge graphs (KGs) offer a promising approach to addressing these challenges by structuring and linking data. In conjunction with LLMs they can make complex webs of evidence navigable, improving discoverability and supporting reasoning over data. However, creating high-quality KGs requires significant involvement from domain experts, which becomes prohibitively expensive at scale [4].

Introduced by Microsoft in 2024, GraphRAG[2] is a novel evolution of the retrieval-augmented generation (RAG) technique, combining automated knowledge graph construction and query-focused summarization to support human sensemaking over entire text corpora [5]. The approach retrieves

KG4S'25: The third international Workshop on Knowledge Graphs for Sustainability, June 01, 2025, Portoroz, Slovenia
christoph-stade@outlook.de (C. Stade); jschneider@pobox.com (J. Schneider); fu5@illinois.com (Y. Fu)
0009-0009-3973-5139 (C. Stade); 0000-0002-5098-5667 (J. Schneider); 0000-0003-2726-0999 (Y. Fu)

[1]https://environmentalevidence.org/ceeder-search/
[2]https://github.com/microsoft/graphrag

contextually relevant information using relationship-based similarity, consequently generating more relevant answers than traditional RAG methods that rely solely on vector similarity of embedded text fragments [5]. Additionally, while RAG is typically used to generate answers based on general knowledge, in this work we begin to see that GraphRAG could excel at connecting across domain-constrained documents with an explicit structure, such as scientific papers and especially evidence reviews. Note that the underlying entity knowledge graph is different from conventional knowledge graphs in that it does not incorporate structured ontologies, which poses a tradeoff we will discuss in section 3.2.

We propose using Microsoft GraphRAG as a dynamic knowledge layer on CEEDER and identify the following three advantages for policy-makers to gain with this approach. Using GraphRAG:

1. Provides provenance links for generated knowledge claims to original text chunks in the dataset
2. Encourages proactive knowledge discovery by making evidence more accessible
3. Continuously integrates the newest scientific documents

2. Related Work

Utilizing graph structures for query-focused summarization was previously suggested by Park and Ko [6]. Their approach uses a graph attention network and a personalized PageRank algorithm to strengthen the relationship between query nodes and document content. Our approach using GraphRAG is advantageous because it can leverage cross-document relationships, integrating multiple pieces of evidence that collectively represent a domain in one unified graph, rather than summarizing one document at a time.

Barron et al. explored how KG-assisted RAG can enhance language model performance in highly domain specific settings [7], improving structure and relevance at the cost of significant pre-processing and manual curation. Their integration of KGs with vector stores emphasizes structured ontologies which are explicitly built by human experts-in-the-loop using curated corpora. Although the need for human experts is minimized for scalablility, it can still be a barrier when resources are limited. In contrast, by using GraphRAG to dynamically link entities, our approach enables retrieval while completely eliminating the need for predefined schemas.

Within the sustainability domain, Gupta et al. designed a KG-assisted RAG approach for question-answering about Environmental, Social, and Governance news articles [8]. Their approach uses Rapid Automatic Keyword Extraction (RAKE) for unsupervised keyword extraction from sentences in a news corpus, then creates embeddings of extracted key phrases with phrase-BERT that get saved as entities in a KG. At query time entities are retrieved based on their cosine similarity to the query string. Their KG aims to provide an LLM with factual data about companies, events, and dates. Yet it is limited because the retrieval is based on individual node scoring. However, our envisioned application to policy-making requires abstraction beyond localized facts to answer questions spanning across sub-domains within climate research, not attempted by their approach. GraphRAG is more suitable since it supports high-level, global queries and identifying overarching themes by creating community summaries and combining partial answers.

Our approach integrates a Python package actively maintained by Microsoft[3]. This ensures a readily usable command-line interface deployment, a convenience lacking in similar approaches [7, 6, 8].

3. A Dynamic Knowledge Layer for CEEDER

The CEEDER database's rigorous evidence synthesis methodology creates unique opportunities for graph-based knowledge discovery: Each systematic review's structured review question provides natural anchor points for graph construction. Per individual evidence review, CEEDER's curators assign a review question following the PECO/PICO (Population, Exposure/Intervention, Comparison, Outcome) question format. The review questions serve as a top-level representation of evidence documents

[3]https://github.com/microsoft/graphrag

throughout, describing their contents in the most condensed form, even more so than the review's abstract. We make every review question part of the context fragments input to GraphRAG's indexing. In practice, review questions can guide query optimization as pre-formulated prompts for policy-makers to use as an entry point to a document or topic.

3.1. Methods and Results

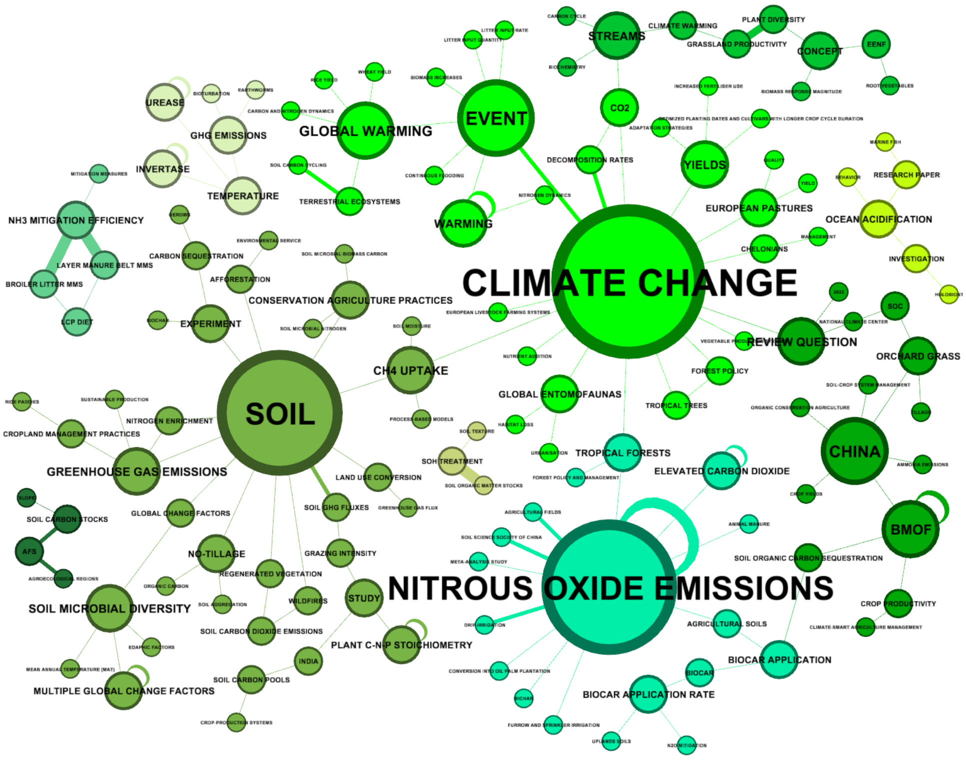

Figure 1: Excerpt of GraphRAG's entity knowledge graph of CEEDER's 354 climate change abstracts[4].

First, we parsed the database dump CSV file (retrieved from CEEDER on August 2, 2024) to extract the Title, Review Question and Abstract columns for each review. To limit the amount of data to process, we restricted ourselves to the 356 reviews in the climate change collection by applying a CEEDER search filter, which was further reduced after removing two duplicate entries. The reviews were formatted as individual .txt files with three lines per extracted column entry and moved to GraphRAG's input directory. Next we ran the GraphRAG index pipeline to create the entity knowledge graph following the documentation[5]. Our configuration differs in pointing to local instances running on ollama `0.6.0`[6] rather than remote OpenAI models, removing dependency on proprietary APIs, cutting cost, and ensuring privacy. For the LLM we chose mistral-nemo with 12B parameters[7] for its strong performance-to-size ratio. Preliminary testing with stronger models yield more populated and denser graphs, suggesting potential for improved performance. The embedding model was nomic-embed-text[8] with high general-purpose embedding quality. Table 1 shows an example global query against all ingested abstracts and a sample review question as user input. Compared to GraphRAG's local search, global search is a resource-intensive method, but it often gives accurate responses for questions that require an understanding of the dataset as a whole. Figure 1 shows the resulting graph visualized using Gephi network analysis software [9] and the GraphRAG visualization guide[9], limited to clusters with

[5]https://microsoft.github.io/graphrag/get_started/
[6]https://ollama.com/
[7]https://ollama.com/library/mistral-nemo
[8]https://ollama.com/library/nomic-embed-text
[9]https://microsoft.github.io/graphrag/visualization_guide/

more than one node and rearranged for improved readability. The indexing process extracted 90,230 entities and 70,895 relationships aggregated to 608 nodes and 250 edges through community-based summarization and pruning (see Figure 1).

Immediate limitations are visible, like the dark green colored cluster to the right with the meaningless artifact "Review question", used to structure the input files. Further, the graph is sparse. There are many clusters with a single concept node (not shown in Figure 1), likely because of the small input size of 354 abstracts with 13.97 sentences on average or because of the limited complexity of the language model used to process the abstracts.

3.2. GraphRAG Compared to Conventional Knowledge Graphs

The entity knowledge graph (EKG) produced by GraphRAG's indexing pipeline does not conform to any single established model of a conventional knowledge graph [10]. Nodes have default types `organization`, `person`, `geo`, and `event` - an LLM prompt can be used to generate alternative types - and are a factor for retrieval. Edges have a weight that counts their normalized occurrences. Both nodes and edges have one or more free-form description strings with arbitrarily rich textual metadata. Beyond the standard attributes of the EKG's GraphML-serialization it is missing a schema resembling an ontology.

In contrast to conventional knowledge graphs, the EKG's descriptions are not exclusively meant for human consumption but treated as the single source for all semantic information about the relation for the LLM to process. For example, the node `Nitrogen Fixation` is described as "Nitrogen Fixation refers to the process of converting atmospheric nitrogen into a usable form for plants." and connected to the node `Plants` via a relation with the description "Plants rely on nitrogen fixation for growth and development." GraphRAG's handling of rich semantic descriptions for both nodes and edges is close to that of a property graph: high complexity in the description enables complex statements to be expressed with relatively little notation, in strings that the LLM can process.

Conventional knowledge graphs need additional mechanisms like reification or subgraphs to realize similar contextual information, such as temporal or geographical annotations. And compared to GraphRAG's flexible EKG, conventional KGs are expensive to create and thus impractical when resources are limited. Significant effort in ontology design, data integration, and maintenance is required to create domain-specific knowledge graphs, with domain experts who define taxonomies and update schemas [10]. GraphRAG's indexing pipeline, as a dynamic knowledge layer, eliminates this overhead inherent to knowledge engineering by automatically extracting and structuring relationships as new data is introduced just by re-indexing.

While GraphRAG lacks ontology-driven reasoning and complex SPARQL queryability, it offers compensatory advantages. The flexible EKG enables continuous evidence integration, and scales efficiently without the need for costly manual curation. This adaptability makes GraphRAG a cost-effective alternative in resource-constrained settings, such as the CEEDER database.

3.3. Environmental Policy Impact

Policy decisions are inherently tied to the quality and reliability of synthesized evidence. This is true for both evidence reviews and artificially generated reports, making it essential that the methodologies used to synthesize information are both accurate and comprehensible. In fact, EU policy-makers highlight trust, regulatory uncertainty, and risks associated with AI, including its potential harms and the need for explainability, as key challenges in harnessing AI for policy-driven applications such as climate change mitigation [11]. Ensuring AI transparency is crucial for informed decision-making, as misrepresented or opaque evidence can lead to misguided policies with real-world consequences [12].

Unlike traditional AI models that generate responses without clear sourcing, Microsoft's GraphRAG provides provenance, explicitly linking generated statements back to their original text chunks in the

[9] https://github.com/infoqualitylab/CEEDER_study/blob/main/Spring2025_Christoph/mistral-nemo12b/visualized_merged_graph.pdf

Review Question What are the effects of nitrogen input on greenhouse gas emissions from wetland ecosystems?
Response Nitrogen input significantly impacts greenhouse gas emissions from wetlands, primarily affecting methane (CH_4) and nitrous oxide (N_2O) emissions. Here's a summary of the key findings: 1. **Methane (CH_4) Emissions**: Excess nitrogen can lead to increased methane production and emission in wetlands **[Data: Reports (2, 7, 64, +more)]**. This is due to enhanced microbial activity under high nitrogen conditions. However, the effects of nitrogen input on methane emissions **can vary depending on other factors** such as water table depth and temperature [Data: Reports (4, 6)]. [...] 3. **Plant Community Composition and Greenhouse Gas Emissions**: Nitrogen input may also alter plant community composition in wetlands, potentially affecting greenhouse gas emissions [Data: Reports (3, 5)]. Changes in plant species can influence the amount of carbon sequestered by wetlands, which **indirectly impacts** greenhouse gas emissions.

Table 1
Example review question query and answer pair generated with global search, excerpted for readability.

dataset and relationships in the entity KG. For instance, in Table 1, GraphRAG cites "[Data: Reports (2, 7, 64, +more)]". Consecutive indices are sequential text fragments in the input; larger gaps indicate that GraphRAG makes connections between multiple reviews' findings. By linking research questions and their answers, GraphRAG enables policy-makers to see how findings in different contexts and on different populations relate through a common intervention. For example, consider the review question: "What are the effects of nitrogen input on greenhouse gas emissions from wetland ecosystems?" A separate review addresses: "What are the effects of nitrogen input on soil nitrous oxide emissions and soil ammonia and nitrate concentrations in natural ecosystems?" Both study the effects of nitrogen input but in different settings. A policy designed for nutrient management in wetlands could benefit from lessons learned in natural ecosystems, identifying both shared mechanisms and critical contextual differences. As the CEEDER climate change collection grows, these overlaps will multiply. This will enhance GraphRAG's ability to find consensus, conflicting viewpoints, and trends across multiple evidence reviews and their findings, going well beyond isolated summarizations. Ultimately, GraphRAG could help policy-makers move from narrow, study-specific insights to a comprehensive understanding of the broader landscape of environmental research.

4. Limitations and Future Work

The sampled global queries with selected review questions already suggest an improvement over regular RAG. However, to verify, this we will run an empirical evaluation. Since environmental science does not have an existing labeled data set that could be used for validation, we will adopt Edge et al.'s improvement to the LLM-as-a-Judge technique, employing a language model itself as an evaluator for the output of other language models [5, 13]. The goal is to use two different models, one for generating the data, the other for evaluating, to overcome egocentric bias [14]. We will test different configurations through a constructive ablation study to determine the highest contributors, including an LLM with only concatenated abstracts in the context window, an LLM with RAG, an LLM with GraphRAG, and finally, an LLM with GraphRAG with fine-tuned models for sustainability.

Moving forward, we aim to extend GraphRAG's EKG with lightweight ontologies in a hybrid approach to enable additional dimensions to search, such as vertical exploration through evidence hierarchies, utilizing CEEDER's evidence overview collection, horizontal discovery of interdisciplinary connections (e.g., linking the climate change collection to other review collections) and temporal analysis of evolving scientific consensus across review versions. Two sustainability-focused enhancements could be made to the core GraphRAG pipeline. First, we will use ClimateBERT [15], a fine-tuned model trained on 450,000 environmental science abstracts, to improve entity recognition accuracy for sustainability concepts. Second, we will improve the evidence tracking by updating the cryptic "Data: Reports..." (Table 1) provided by GraphRAGs out-of-the-box capabilities to enhance them with publication titles and DOIs.

In the future, we envision the entity knowledge graph facilitating applications such as detecting knowledge gaps and conflicting statements. Missing connections between concepts, sparse graph regions, or dangling nodes may indicate knowledge gaps such as under-researched areas or open questions. By traversing and inspecting the graph, either automatically, manually or both, it may be possible to find conflicting or contrasting statements and compute maximum flow to identify the statement with stronger support [16]. This could provide decision-makers with insight into scientific consensus and uncertainty.

5. Conclusions

Sustainability research urgently needs automated methods to help policy-makers synthesize research outcomes because it is one of the fastest growing fields of research, doubling in unique contributions every 8.3 years, with work from diverse disciplines such as biology, chemistry, engineering, health sciences, and social sciences [1]. Our GraphRAG pipeline improves knowledge discovery in environmental science by structuring relationships between concepts that traditional retrieval models often conflate. GraphRAG furthers LLMs' capabilities for processing vast amounts of text to enable meaningful analysis at scale.

However, challenges remain, especially in the graph construction process. The lack of entity disambiguation and semantic normalization risks missing complex relationships. Further, generative data might not accurately represent the source data. System use should always be accompanied by clear disclosures of AI use and the potential for errors in outputs [5]. Yet, unlike static KGs, GraphRAG continuously integrates emerging evidence, making it well-suited for domains like sustainability, where evolving research shapes policy decisions.

Acknowledgments

This work was partly supported by NSF Grant #2046454 CAREER: Using network analysis to assess confidence in research synthesis. This is a continuation of work Christoph Stade did [17] through a DAAD RISE Worldwide scholarship at University of Illinois Urbana-Champaign while completing his Bachelor's degree at Technische Hochschule Nürnberg Georg Simon Ohm; he is now a Master's student at Ludwig-Maximilians-Universität München. Jodi Schneider is supported in part as the 2024–2025 Perrin Moorhead Grayson and Bruns Grayson Fellow, Harvard Radcliffe Institute for Advanced Study. Thanks to Corinne McCumber, Heng Zheng, and Malik Salami for providing feedback on a draft.

CRediT

Christoph Stade - Conceptualization, Data curation, Formal analysis, Investigation, Software, Validation, Visualization, Writing - original draft, Writing - review & editing; *Jodi Schneider* - Funding acquisition, Supervision, Writing - review & editing; *Yuanxi Fu* - Conceptualization, Methodology, Supervision, Writing - review & editing

Declaration on Generative AI

During the preparation of this work, the authors used GPT-4o for: Drafting content, Paraphrase and reword, and Grammar and spelling check. After using these tools/services, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] L. M. A. Bettencourt, J. Kaur, Evolution and structure of sustainability science, Proceedings of the National Academy of Sciences 108 (2011) 19540 – 19545. doi:10.1073/pnas.1102712108.

[2] K. Konno, S. H. Cheng, J. Eales, G. Frampton, C. Kohl, B. Livoreil, B. Macura, B. C. O'Leary, N. P. Randall, J. J. Taylor, P. Woodcock, A. S. Pullin, The CEEDER database of evidence reviews: An open-access evidence service for researchers and decision-makers, Environmental Science Policy 114 (2020) 256–262. doi:10.1016/j.envsci.2020.08.021.

[3] R. Logan, N. F. Liu, M. E. Peters, M. Gardner, S. Singh, Barack's wife Hillary: Using knowledge graphs for fact-aware language modeling, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 5962–5971. doi:10.18653/v1/P19-1598.

[4] M. Hofer, D. Obraczka, A. Saeedi, H. Köpcke, E. Rahm, Construction of knowledge graphs: State and challenges, Information 15 (2024). doi:10.3390/info15080509.

[5] D. Edge, H. Trinh, N. Cheng, J. Bradley, A. Chao, A. Mody, S. Truitt, J. Larson, From local to global: A Graph RAG approach to Query-Focused Summarization, 2024. doi:10.48550/arXiv.2404.16130.

[6] C. Park, Y. Ko, QSG Transformer: Transformer with Query-attentive Semantic Graph for Query-Focused Summarization, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '22, 2022, p. 2589–2594. doi:10.1145/3477495.3531901.

[7] R. C. B. et al., Domain-specific retrieval-augmented generation using vector stores, knowledge graphs, and tensor factorization, 2024. doi:10.48550/arXiv.2410.02721.

[8] T. K. Gupta, T. Goel, I. Verma, L. Dey, S. Bhardwaj, Knowledge graph aided LLM based ESG question-answering from news, in: Proceedings of the 2nd International Workshop on Knowledge Graphs for Sustainability, volume 3753, 2024. URL: https://ceur-ws.org/Vol-3753/paper6.pdf.

[9] M. Bastian, S. Heymann, M. Jacomy, Gephi: An open source software for exploring and manipulating networks, Proceedings of the International AAAI Conference on Web and Social Media 3 (2009) 361–362. doi:10.1609/icwsm.v3i1.13937.

[10] A. Hogan, E. Blomqvist, M. Cochez, C. D'amato, G. D. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, A.-C. N. Ngomo, A. Polleres, S. M. Rashid, A. Rula, L. Schmelzeisen, J. Sequeda, S. Staab, A. Zimmermann, Knowledge graphs, ACM Comput. Surv. 54 (2021). doi:10.1145/3447772.

[11] F. af Malmborg, Narrative dynamics policy—Sensemaking, agency construction, and anchoring, Review of Policy Research 40 (2022) 757–780. doi:10.1111/ropr.12529.

[12] J. Gurevitch, J. Koricheva, S. Nakagawa, G. B. Stewart, Meta-analysis and the science of research synthesis, Nature 555 (2018) 175–182. doi:10.1038/nature25753.

[13] L. Zheng, W. Chiang, Y. Sheng, S. Zhuang, Z. Wu, Y. Zhuang, Z. Lin, Z. Li, D. Li, E. P. Xing, H. Zhang, J. E. Gonzalez, I. Stoica, Judging LLM-as-a-judge with MT-Bench and Chatbot Arena, 2023. doi:10.48550/arxiv.2306.05685.

[14] D. Li, B. Jiang, L. Huang, A. Beigi, C. Zhao, Z. Tan, A. Bhattacharjee, Y. Jiang, C. Chen, T. Wu, K. Shu, L. Cheng, H. Liu, From generation to judgment: Opportunities and challenges of LLM-as-a-judge, 2024. doi:10.48550/arXiv.2411.16594.

[15] N. Webersinke, M. Kraus, J. Bingler, M. Leippold, ClimateBERT: A pretrained language model for climate-related text, in: Proceedings of AAAI 2022 Fall Symposium: The Role of AI in Responding to Climate Challenges, 2022. doi:10.5167/uzh-235046.

[16] P. Shiralkar, A. Flammini, F. Menczer, G. L. Ciampaglia, Finding streams in knowledge graphs to support fact checking, in: 2017 IEEE International Conference on Data Mining (ICDM), 2017, pp. 859–864. doi:10.1109/ICDM.2017.105.

[17] C. Stade, J. Schneider, Y. Fu, Making sense of the environmental science landscape: An exploration of the CEEDER database, in: METSTI 2024: Workshop on Informetric, Scientometric, and Scientific and Technical Information Research, 2024. doi:10.5281/zenodo.13895517.