

Rumadi, Ambar Yoganingrum, Abdurrahman Prasetyadi, Aria Bisri, and Stevry Yushady CH Bissa

From Research to Impact

Mapping BRIN Publications to SDGs Using Topic Modeling

Monitoring institutional research contributions to the Sustainable Development Goals (SDGs) is becoming increasingly crucial for evidence-based decision-making in academic and research libraries. In the era of big data, libraries not only store publications but also support research analysis and visualization to develop institutional strategies. Researchers at the National Agency for Research and Innovation (BRIN)¹ in Indonesia have taken an innovative approach by using topic modeling methods, specifically Latent Dirichlet Allocation (LDA), to develop an application that automatically classifies institutional publications into SDG categories.

The BRIN library can use this application to understand research trends, the interrelationships among topics, and research priorities that support the SDGs and to assist researchers, administrators, and policymakers within BRIN. Thus, the library not only provides access to publications but also provides information on the institution's contribution to sustainable development.

Traditional Challenges in SDG Monitoring

Before the use of machine learning-based methods, mapping publications to SDGs was generally done manually or based on simple keywords. Manual approaches are time consuming, error prone, and difficult to implement for large publication volumes. Keyword-based matching, while faster, often misses context and cross-topic themes, thus not fully reflecting research contributions to the SDGs.

This issue raises the need for more efficient, accurate, and traceable methods, for which topic modeling has emerged as an innovative solution.

The Role of the BRIN Library in the National Research Ecosystem

BRIN, as Indonesia's national research institution established in 2021 through the consolidation of several research institutions, has a mandate to promote basic and applied research, innovation, and collaboration with universities and industry. The BRIN Library plays a strategic role as a research information center, providing a publication repository, standardized metadata, and analytical services to support data-driven decision-making.²

With the increasing number of publications indexed by Scopus,³ libraries face new challenges: how to effectively manage, classify, and present research information to support SDG tracking and institutional strategic planning.

Rumadi is a research engineer at the National Research and Innovation Agency (BRIN), email: ruma003@brin.go.id. Ambar Yoganingrum is a research professor at the National Research and Innovation Agency (BRIN), email: amba002@brin.go.id. Abdurrahman Prasetyadi is a junior researcher at the National Research and Innovation Agency (BRIN), email: abdu059@brin.go.id. Aria Bisri is a research assistant at the National Research and Innovation Agency (BRIN), email: aria005@brin.go.id. Stevry Yushady CH Bissa is a research engineer at the National Research and Innovation Agency (BRIN), email: stev001@brin.go.id.

Uncovering Publication Topics with Machine Learning

Topic modeling is a method that plays a crucial role in the development of this application. This unsupervised machine learning method is used to discover hidden thematic structures within a collection of text. Conceptually, this method can be likened to instructing a computer to read and process thousands of publication abstracts and then automatically group them into specific topics. Unlike traditional cataloging, the computer is not given a predetermined list of subject headings. Instead, the algorithm infers themes from patterns of words that frequently cooccur across documents.

The topic modeling method used in this study is LDA, a probabilistic topic modeling algorithm. The workflow for implementing LDA can be simplified into several key steps that are easy for information professionals to understand:

1. Data preparation (data curation)

The first step is to collect and clean text data. The study dataset consists of abstracts obtained from the Scopus database. This abstract data then undergoes preprocessing, such as removing punctuation, converting to lowercase, removing stop words (common words that lack thematic meaning, such as “a,” “the,” “is,” and “what”), and lemmatization (changing words to their base form, for example, “detection” and “detected” become “detect”). This step aligns with librarians’ core expertise in data curation and metadata management, ensuring that “garbage in, garbage out” is avoided.

2. Applying the topic modeling algorithm

Once the data is cleaned, the next step is to apply the LDA algorithm. This algorithm builds a probabilistic model that assumes each document is a mixture of topics, where each topic is a collection of multiple words.

3. Interpretation

The output of the LDA model is not a human-readable topic label but rather a set of “topics” consisting of words that appear together. This is where we use generative artificial intelligence (AI) to automatically label topics, ensuring each topic has a human-understandable label.

LDA is a widely used unsupervised machine learning technique for extracting topics from a collection of documents.⁴ With it, each document can be represented by a combination of multiple topics, enabling the analysis of cross-cutting themes that are not easily identified by traditional keyword methods.

In the context of this application, LDA enables:

1. Identifying key research themes and their distribution across the SDGs
2. Understanding the relationships among research topics that may be relevant to multiple SDGs
3. Providing interactive visualizations to facilitate publication browsing and decision-making

Case Study: Mapping BRIN Research to the SDGs

We attempted to connect BRIN publications to the SDGs framework using this application. The first stage used topic modeling to uncover hidden themes in the BRIN publication

corpus. The second stage mapped the discovered themes or topics to the seventeen established SDGs. This mapping process was automated by matching keywords across topics. The matching was performed based on a query set developed and used by Elsevier. The query set from the Elsevier 2025 SDG mapping dataset⁵ provides a comprehensive list of terms related to each goal.

The dataset includes abstracts of journal articles published by BRIN researchers and indexed in Scopus. The application performed the following steps:

- 1. Data preparation: Abstracts were cleaned of stop words, tokenized, and normalized.
- 2. Topic extraction: LDA generated several topics representing the dominant themes in the publications.
- 3. Mapping to SDGs: A combination of keyword-based matching and expert validation was used to link LDA topics to relevant SDG targets.
- 4. Visualization and reporting: Mapping results are presented through an interactive dashboard that allows for tracking of topics, documents, and research trends over time.

This analysis resulted in a precise mapping of BRIN’s research landscape on the SDGs. Figure 1 shows key findings, indicating an intense concentration of research on several SDGs. The three highest-ranking SDG groups are SDG 6: Clean water and sanitation, SDG 12: Ensure sustainable consumption and production patterns, and SDG 7: Ensure access to affordable, reliable, sustainable, and modern energy for all.

To visualize more complex relationships among topics, we use a knowledge graph, as shown in Figure 2. A knowledge graph represents information as a network of entities (called *nodes*)

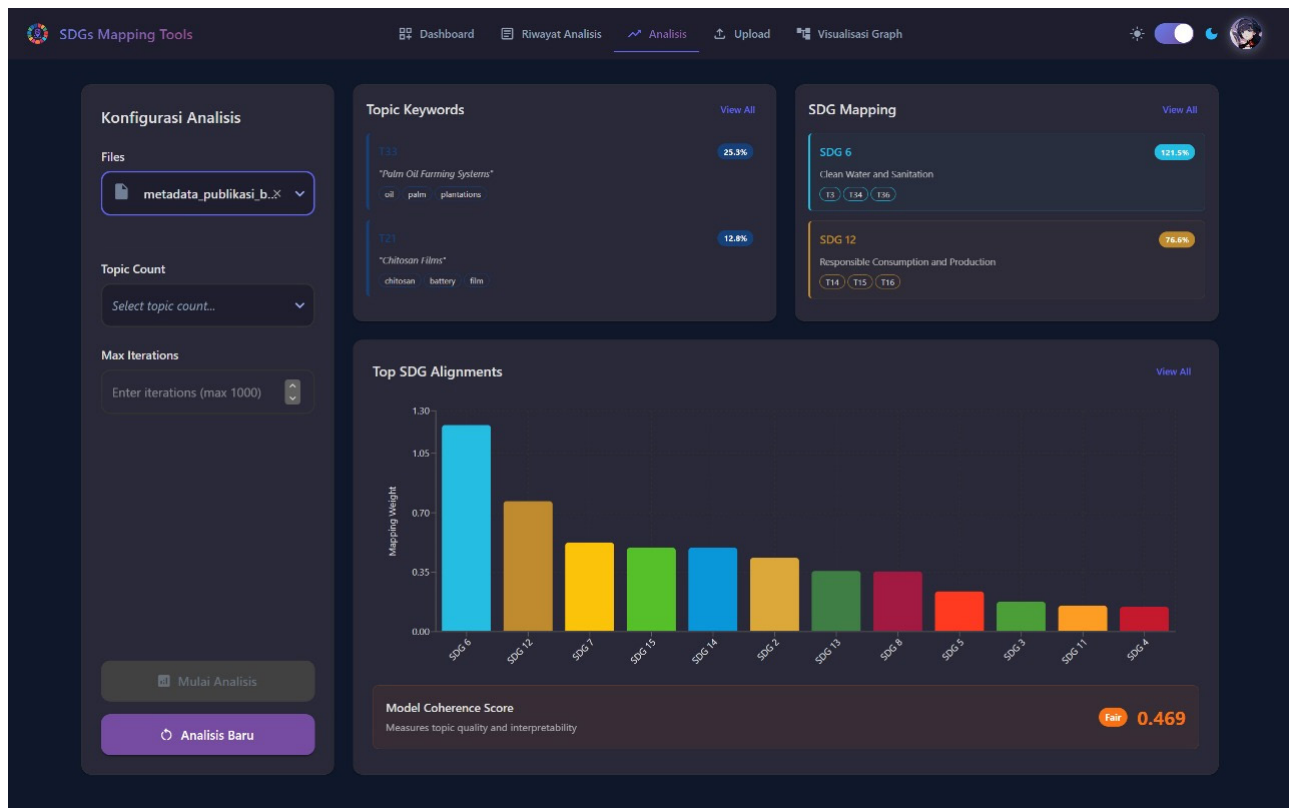


Figure 1. BRIN's research landscape on the Sustainable Development Goals.

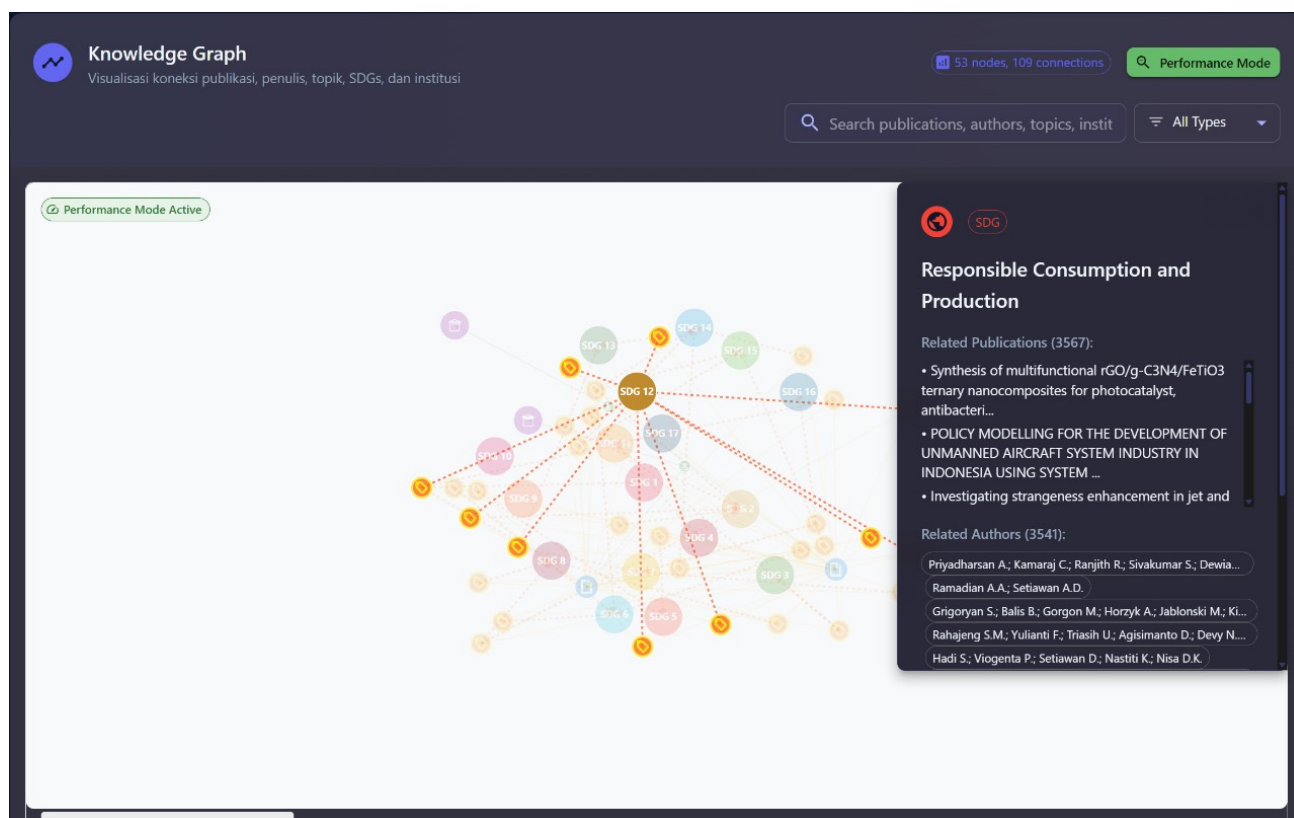


Figure 2. BRIN's research knowledge graph.

and the relationships between those entities (called *edges*). In this context, each identified topic, its key keywords, and each SDG can be a node. Edges then represent relationships, such as “has_topic” (publication-to-topic) or “contributed to” (topic-to-SDG). This interactive visualization allows analysts to explore, zoom in, and filter connections, revealing how different research areas are connected and where thematic gaps or overlaps lie. This approach transforms a static list of topics into a dynamic, explorable knowledge map.

This application can also analyze topics that fall under more than one SDG. Figure 3 shows that BRIN publications cover cross-SDG themes; for example, topic 3 (T3) is relevant to SDGs 6, 14, and 15. Such analysis is complex to perform manually.

Key Findings and Strategic Implications

The LDA analysis of BRIN publications yields several key findings:

- Evolution of research focus: Several research topics have seen significant year-over-year growth, reflecting the institution's strategic priorities.
- Linkages between topics: Topic visualizations clearly show connections among several SDGs, helping research managers understand cross-disciplinary contributions.
- Data-driven action: These results will support libraries in making recommendations for collection development, research progress reports, and institutional strategic planning.

In other words, topic modeling is not only a technical analysis tool but also a strategic information resource for librarians and research administrators. This study serves as a model, demonstrating how librarians, who previously played a traditional support role, can use this application to become strategic partners in their institution's research ecosystem. Libraries

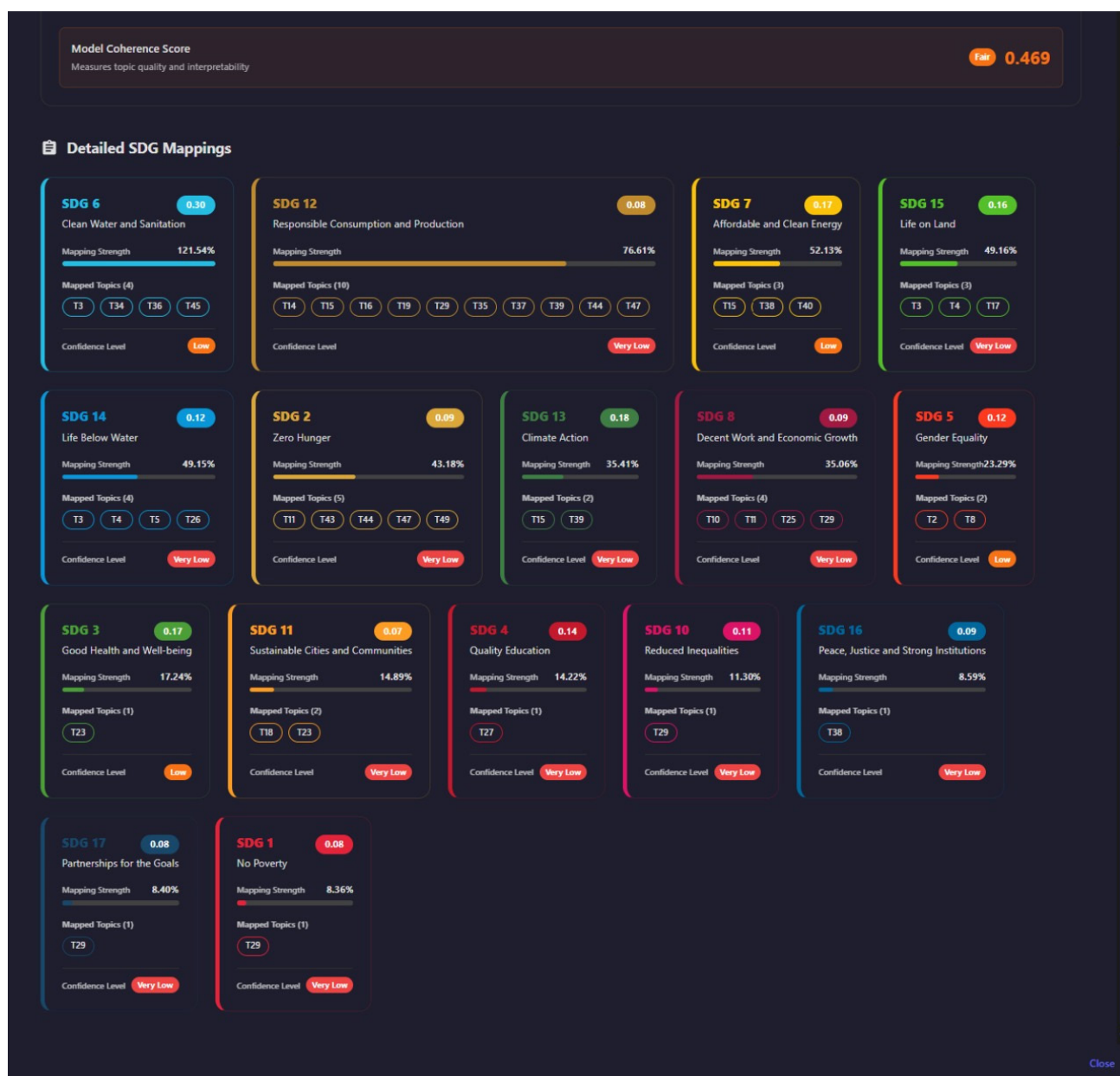


Figure 3. Detailed SDGs mapping.

are well positioned to lead in this academic field; they are neutral entities with deep expertise in scholarly data sources and metadata, and they can provide an institution-wide perspective that specific departments or faculty members lack. By adopting approaches such as topic modeling, libraries can offer research intelligence services that directly support the institution's strategic mission.

Here are some practical scenarios for how this application could be adopted by academic libraries:

- **Supporting institutional strategy:** This type of strategic planning often requires a clear understanding of an institution's research strengths and weaknesses. This is where libraries can play a role. Librarians can create SDG-based thematic maps of all scholarly publications using topic modeling. Institutional publication data can be obtained from institutional repositories or from publication sources such as Scopus and Web of Science.

- Enhancing research impact: Research impact services in many libraries still focus heavily on traditional citation metrics. Although important, these metrics reflect only a small portion of the actual impact. With SDG-based topic modeling, librarians can significantly expand their service offerings. Instead of simply providing h-indices to specific researchers or departments, librarians can offer in-depth thematic analyses of their institution's research output.

Conclusion

The use of LDA enables modern libraries to play an active role in supporting the Sustainable Development agenda through intelligent research information management. By combining topic modeling, interactive dashboards, and user-centered design principles, libraries can provide strategic insights to research managers, researchers, and policymakers.

This article emphasizes that libraries are not just repositories of information but also drivers of innovation in data analysis and research management. The application developed by BRIN researchers can serve as a model for other academic libraries seeking to transform data publication into tangible impacts aligned with the SDGs. //

Notes

1. For more information, visit the BRIN website at <https://brin.go.id>.
2. Ambar Yoganingrum et al., "Strategies to Encourage Data Owners to Self-Archive Data Sets," *Library Hi Tech News*, Emerald Publishing (2025), doi:10.1108/LHTN-03-2025-0044.
3. For more information, visit the Scopus website at <https://www.elsevier.com/products/scopus>.
4. S. Ravikumar, Bidyut Bikash Boruah, and Fullstar Lamin Gayang, "Latent Topics Identification from the Articles of Sri Lankan Authors Using LDA," *Global Knowledge, Memory and Communication* 74, no. 1/2 (2025): 186–97, doi:10.1108/GKMC-08-2022-0206; Yaşar Tekin, "Initialization in Gibbs Sampling Implementation of LDA," in 2024 32nd *Signal Processing and Communications Applications Conference (SIU)*, 1–4, IEEE (2024), doi:10.1109/SIU61531.2024.10600919; Yaşar Tekin and Ahmet Cosar, "Abstract or Full-Text in Topic Modeling?" in 2022 30th *Signal Processing and Communications Applications Conference (SIU)*, 1–4, IEEE (2022), doi:10.1109/SIU55565.2022.9864707; Dexin Zhao, Jinqun He, and Jin Liu, "An Improved LDA Algorithm for Text Classification," in 2014 *International Conference on Information Science, Electronics and Electrical Engineering*, 217–21, IEEE (2014). doi:10.1109/InfoSEEE.2014.6948100.
5. Guillaume Roberge and Chris James, "Elsevier 2025 Sustainable Development Goals (SDGs) Mapping," Elsevier Data Repository, version 1 (2025), doi:10.17632/p52c7d3hfs.1.