International FAIR Digital Objects Implementation Summit 2024
Extended Abstracts
https://doi.org/10.52825/ocp.v5i.1182

Published: 18 Mar. 2025

A FDO Approach to SDGs Knowledge Graphs

A Case Study of a Battery-Related Topic Through a Multidisciplinary Approach Using Global Academic Datasets

Mikiko Tanifuji[1,*], Masaharu Hayashi[1], Kazutsuna Yamaji[1], Luca Foppiano[2], and Sae Dieb[2]

[1]Research and Development Center for Open Science, National Institute of Informatics, Tokyo, Japan

[2]Center for Basic Research on Materials, National Institute for Materials Science, Tsukuba, Japan

*Correspondence: Mikiko Tanifuji, tanifuji@nii.ac.jp

Abstract. The G7 Summit in 2023 Promoted Open Science, Leading to the Practical Implementation of FAIRable Research Data Infrastructures and Applications. This Has Increased Researchers' Awareness of Interoperable Digital Transformation. The Focus on Battery Materials Aims to Support Sustainable Development Goals (SDGs), an Important Global Issue. Using Battery Materials-Related Papers, Key Terms Were Extracted with KeyBERT, and Frequent Term Analysis Was Conducted. Author Similarity Analysis Connected Researchers with Similar Interests, Generating Topic Maps. Web Applications Produced Knowledge Graphs of Authors and Topics. The Study Evaluated Various Article Catalogues and Chose Open Alex for Knowledge Extraction. The Availability of More Research Papers for LLM Study, Resulting from Open Science, Marks a Significant Advancement. KeyBERT and d3-Force, with Vue.js, Were Used for Linguistic Analysis and Web Application Development.

Keywords: Open Science, FAIR, Digital Objects DOs, Open Data Catalogue, OpenAlex, Open Repository, KeyBART, SDGs, Battery Materials Knowledge Graph

1. Introduction

Due to the promotion of open science at the G7 summit in 2023, FAIRable research data infrastructures and applications have now progressed to the practice phase. This progress, together with the development of HPCs and new cloud services with generative AI, has increased researchers' awareness of the interoperable digital transformation in day-to-day environments. For example, great efforts are underway in some science domains such as DONA as digital object architecture, NOMAD with digital object compliant approach in materials science and others [1]. They extract elements of infrastructures as FAIRable digital components with PIDs and demonstrate FAIR data space [2, 3]. In the research data alliance community (RDA) there are ongoing discussions on conceptual designs to enable FDO concepts in different infrastructures. In Europe, one of the EOSC projects introduced a study of canonical research workflow for experimental research by analysing digital object architectures across different disciplines [4]. In response to the FDO concept, we were motivated to choose battery materials as a theme that would contribute to worldwide efforts towards sustainable development goals (SDGs), for which there is a high social demand [5].

2. Experiment

We used battery materials-related papers, and LLM processing to extract key terms by using a pre-trained algorithm KeyBERT. In our framework, object A contains the input papers, the selected top key terms, concepts, and the trained algorithm as shown in the figure 1. Having analysed and optimized the process to control the granularity of key terms, we then created weighted key terms for each author and conducted frequent term analysis of the topics according to the number of papers, occurrences, and publication years within the data set. The resulting extracted information analysis in our framework is object B. Object C is an author similarity analysis conducted to connect authors with similar research interests. This analysis contains a list of papers, together with their key terms and the papers' concepts, and the generated topic map information. Object D consists of web application processes which produce knowledge graph of article authors and topics network [6, 7].

In this study, we evaluated open article catalogues, Open Alex [8], Lens [9] and Japan Intuitional Repositories DataBase (IRDB) [10]. We have decided to use Open Alex as an input source for object A, which offers significant advantages in terms of knowledge extraction with scientific concept IDs. For object B, we used the linguistic analysis using BERT-based model and opensource library KeyBERT#. The web applications for knowledge graph were implemented using d3-force to calculate the arrangement of nodes using physical operations, and Vue.js to display the calculation results on a web browser as user interface.

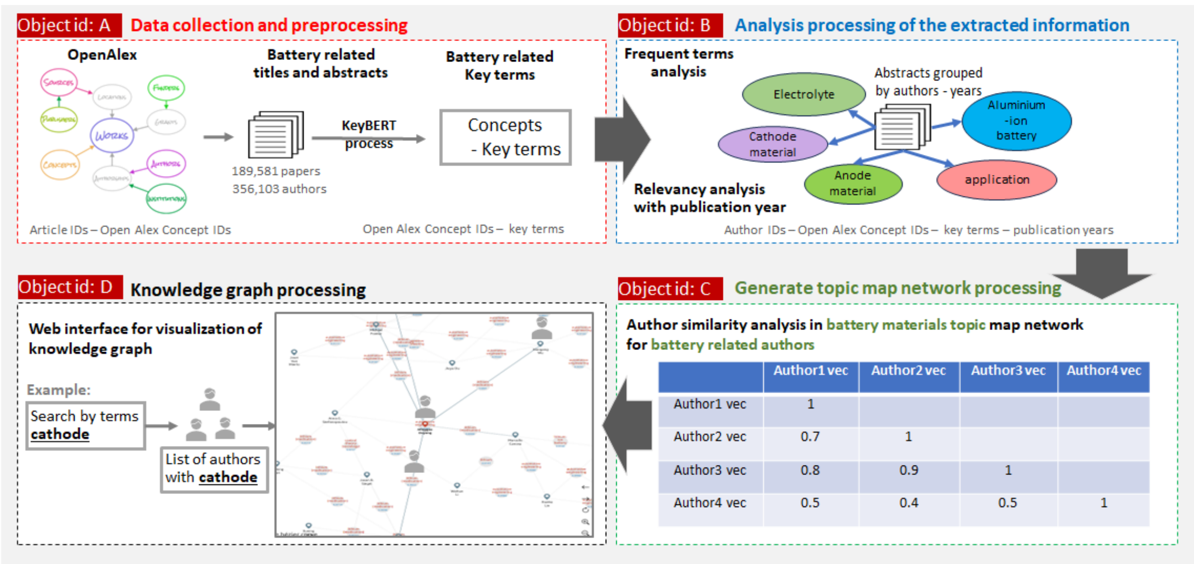

Figure 1. *An implementation of a digital object framework for SDGs knowledge graphs*

3. Conclusions

We assessed the potential FDO concept by examining a combination of four digital objects which produced the SDGs Battery Materials Knowledge Graph. We aimed to produce a knowledge-driven collaborative platform for other topics, with customized digital objects, resulting in interoperable and reusable data spaces. We hope that our FDOs concept will make a useful contribution to open science platforms and share a real-time snapshot of SDG innovations.

Data availability statement

The data set in our experiment is available in an open data catalogue, OpenAlex. The linguistic analysis using BERT-based model, KeyBERT# are available in opensource library.

Underlying and related material

Not applicable.

Author contributions

M TANIFUJI and D SAE were involved in study design and data interpretation. M HAYASHI and K YAMAJI were involved in the data evaluation of IRDB. L FOPPIANO and D SAE are involved training algorithm and analysis for topic maps. All authors critically revised the report, commented on drafts of the manuscript, and approved the final report.

Competing interests

The authors declare that they have no conflict of interest.

Funding

Not applicable.

Acknowledgement

We would like to express our sincere gratitude to Dr. Keitaro Sodeyama from National Institute for Materials Science (NIMS) for his invaluable evaluation of the battery materials data and his insightful advice on the hierarchical tuning of key phrases for battery knowledge graph. His expertise and guidance significantly contributed to the advancement of our research.

References*:*

[1] Koenraad de Smidt, et al: An analysis of scientific Practice towards FAIR Digital Objects, March 2019.

[2] Wittenburg, P. (2019). Digital Objects as Drivers towards Convergence in Data Infrastructures. [DOI]

[3] Wittenburg, P. et al. (2022). FAIR Digital Object Demonstrators 2021 [DOI]

[4] Betz, D. et al: Canonical workflow for Experimental Research, 2022, Data Intelligence [DOI]

[5] Jianmin Ma et al: The 2021 battery technology roadmap, J. Phys. D: Appl. Phys. 54 183001, 2021.[DOI]

[6] DIEB, Sae, FOPPIANO, Luca, SODEYAMA, Keitaro, Tanifuji Mikiko: *Creating a Visual Topic Map for Battery Researchers using a Large Global Open Dataset*. 244 ECS meeting, 2023 October.

[7] DIEB, Sae, FOPPIANO, Luca, SODEYAMA, Keitaro, Tanifuji Mikiko: *Visual Topic Mapping for Battery Related Research using a Global Open Catalogue*. The 84th JSAP fall meeting, 2023 September.

[8] Priem, J., Piwowar, H., & Orr, R. OpenAlex: A fully open index of scholarly works, authors, venues, institutions, and concepts. 2022, ArXiv [DOI]

[9] Jefferson, Osmat A, et al: Mapping the global influence of published research on industry and innovation, Nature Biotechnology. 36, 2019 [DOI]

[10] Nose, M et al: Metadata mapping between disciplinary and general schemas for promotion of data use by a wider community, XXVIII General Assembly of the International Union of Geodesy and Geophysics (IUGG), 2023 [DOI]