https://doi.org/10.26069/greynet-2025-000.766-gn

PISA 2025

Preparing policy and technology for data sharing and reuse In the AI era

Jayhoon Kim
Korea Institute of Science and Technology Information

Abstract

This author proposes follow-up research for the 2024 KISTI-GreyNet joint study on global information repositories. The preliminary study aimed to find shareable knowledge using institutional collaboration and AI, noting respondents' recognition of AI's role in data curation and user service improvement. The author, as a GreyNet member, suggests forming an interest group within GreyNet to discuss AI-driven data sharing and utilization. This group will address practical policy and technology issues, seeking solutions through member case presentations in policy, technology, system, and regulation. The results of the interest group activities will be promoted as specific activities for data sharing cooperation. The results will be delivered at the next GreyNet conference. We ask for your interest and participation. To help readers understand the activity plan of this stakeholder group, the author would like to present data-related issues that need to be addressed in the AI era by citing KISTI-owned data, data processing AI technology, and data utilization cases.

Keywords: Data sharing, Data policy, AI, Artificial Intelligence, Grey Literature, Digital Curation

Introduction

The objective of this paper is to design follow-up research cooperation activities of the global information repository study conducted jointly by KISTI and GreyNet in 2024.

In the previous 2024 preliminary research was to discover shareable knowledge information resources based on institutional cooperation and AI technology. Researchers noticed that the survey respondents fully recognize the need for data curation using AI technology and the need for improving user services using AI. In the AI era, disclosing the status of grey resources held by each institution, sharing data within the scope of license permission, and collecting, indexing, and processing data using AI technology will expand potential data spectrum beyond the data scope of individual institutions and improve the level of user information services greatly.

In the 2024 preliminary study, researchers set survey recipients as 51 GreyNet members. But they got only six institutions' response to the possibility of data disclosure/sharing at the 3rd final survey. The reason for the low response rate is presumed that the purpose of the previous study was not sufficiently recognized by the survey participants, or the results of the previous study did not sufficiently present the value of the study or the influence of the results to each participant. Therefore, this time the author, as a GreyNet member, intends to form an interest group within GreyNet on data sharing and utilization using AI technology as a follow-up to the previous study and to proceed with this discussion in detail.

Background

AI tools can be used to support different aspects of the research process (1) Hypotheses generation based on a given dataset or topic (2) Literature review by analyzing and summarizing

a body of literature on a topic, identifying relevant trends (3) Data Analysis by processing and analyzing large datasets, making it easier to identify emerging trends, correlations, outliers, and other patterns (4) Assist experiment design by suggesting variables, methodologies, and potential outcomes based on existing data (5) Communication assist in drafting, proof-reading, and editing research papers (6) Help collaboration and networking by AI-driven recommendation systems to connect with peers, collaborators, and experts in their field, fostering interdisciplinary collaboration and knowledge sharing. (University of California Irvine Library, 2024)

Recently, Wiley publishing conducted a survey on researchers' use in practice and perception of AI (published it in Nature magazine). The first question asked how researchers use AI tools in their research. The answers, in order of frequency, are as follows: (1) Assistance with writing papers (2) Correction of sentence errors in papers (3) Plagiarism checks (4) Data collection and processing (5) Massive prior research (including translation and summarization) (6) Reference management (7) Unstructured data processing. The second question asked researchers about their concerns regarding using AI in their research activities.

The answers, in order of frequency, are as follows: (1) research ethics issues; (2) information inaccuracy and hallucination; (3) privacy violations; (4) lack of transparency in AI tools; (5) absence of guidelines on whether or not to allow AI; (6) education on the use of AI; and (7) usage fees (Naddaf, 2025).

IFLA AI special interest group suggests three important AI strategies for libraries. The first is to use library AI capabilities to model responsible and explainable applications of descriptive AI. The second is to use librarians' data competencies to enhance organizational AI capability. The third is th promote AI literacy to enhance organizational and societal AI capabilities. The data sharing proposal in this paper relates to the second strategy, data capacity activities (IFLA, 2023).

Grey literature data have the advantage of being diverse, up-to-date, and able to fill information gaps not found in publications, but they have a critical problem of difficulty in discovery. Using AI technology, GL data can be efficiently collected, metadata built, and searched and discovered. (Savic, 2023)

In the preliminary study, Korea Institute of Science and Technology Information (KISTI) and GreyNet International conducted a joint research project aimed to survey and demonstrate the potential of the data collections that can be expanded through the collaboration of human knowledge and AI technology.

Methods

This study was conducted using triangulation. Firstly, the author analyzed scholarly articles within the recent 5 years (2021–2025). The analysis encompasses changes in research processes, researchers' perceptions of AI utilization, and changes in the role of data managers including librarians in the AI era. Secondly, based on preliminary research, the author expanded the scope of the study to include reciprocal data sharing and utilization methods. Thirdly, the author derived tasks for data collection, sharing, and utilization through KISTI's data curation digital transformation case study.

Results

Based on the results of KISTI and GreyNet's prior research, data managers have the following expectations regarding digital transformation through the introduction of AI technology into data curation work processes.

https://doi.org/10.26069/greynet-2025-000.766-gn

● AI data curation policy impacts on user experience: Policy direction embedded in a strategic plan, where AI technology is incorporated in data processing, is a priority in the curation of data that would improve the user experience.

● Core tasks are data transformation and data innovation: Data conversion and data innovation, such as machine learning and AI services, are among the critical tasks in one's organization.

● Innovation through sharing data and processing technology: Areas that would highly benefit data curation are data sharing and the sharing of AI technology.

● Innovation through digital transformation of user services: When AI technology is introduced in the existing work environment, the greatest expected change is the shift from face-to-face service to chatbot and technical system responses.

However, there are various concerns regarding data management and services utilizing AI technology, and there is no sign of them being resolved yet.

● Issues in securing ownership of data for AI utilization: Data intellectual property rights and licenses for AI utilization are not secured or unclear

● Limitations in introducing AI technology: Technology led by AI specialized companies, high introduction costs, difficulty in securing internal experts

● AI reliability Issues: Hallucinations, bias

● Ethical issues regarding AI utilization

To help readers understand how to address data curation issue in the AI era, the author presents the digital transformation case of KISTI's digital curation work below. KISTI performs research and development of AI technology, promoting digital transformation of data curation works, and expanding data development for AI application. KISTI's directions and activities are as follows.

● Changes in data construction system: Conversion of construction system for information integration construction

○ Process analysis and standardization

○ Adapting AI-human collaboration and crowdsourcing method

○ Component/module-based data processing Plug and Play module development

● Utilization of AI technology in the right place: Development and application of digital transformation technology in data curation process

○ Data processing efficiency technology

○ Service intelligence technology

○ Data identification linkage technology

https://doi.org/10.26069/greynet-2025-000.766-gn

● Expansion and usability of information: Expansion of new types of data, expansion of data through identification linkage

○ Construction of new data derived from existing data

○ New types of information development research

○ Expansion of data linkage through global standard identifier

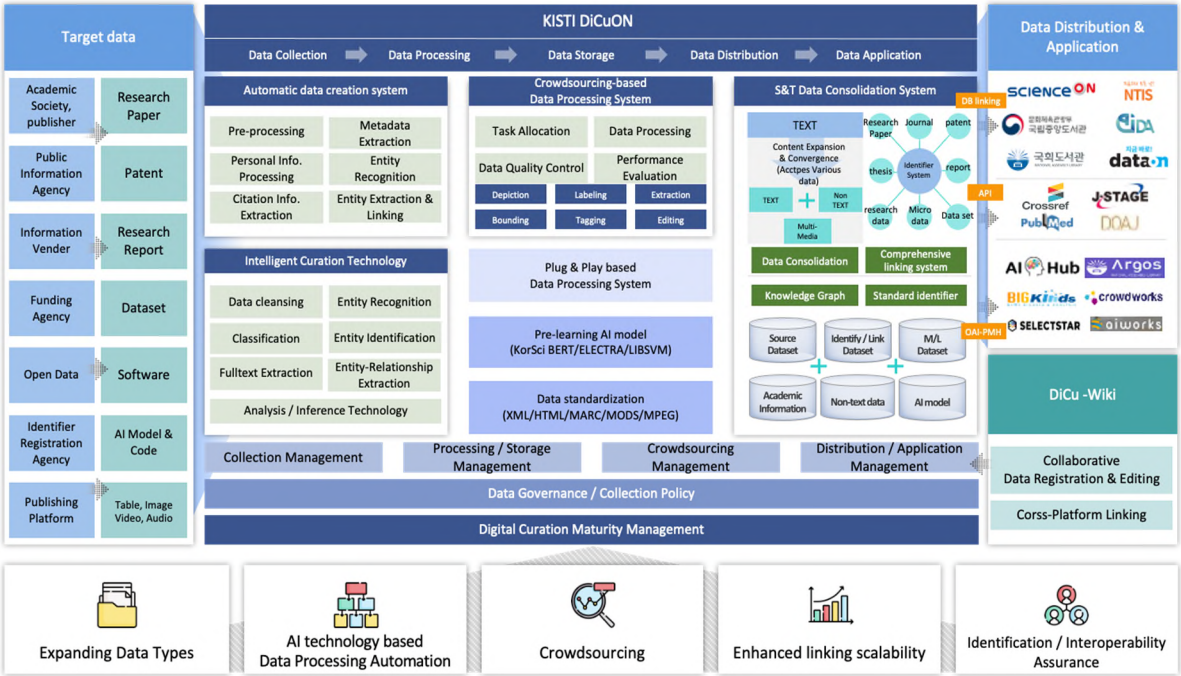

Figure 1: Landscape of KISTI Data Curation System

The IFLA AI Special Interest Group Report can be a useful tool for exploring solutions to data curation challenges in the AI era. Based on the report, the following review agenda is presented.

● Finding data sources in complex information landscapes

● Promoting the value of sharing, openness and interoperability for data

● Explaining the importance of the provenance, validity and quality of that data to understand how it can be appropriately used

● Explaining what data can be used for and what not according to copyright, IPR etc

● Describing data using standards and the value of so doing

Discussion and Conclusion

It is impossible to solve data curation issues all at once in the AI era. It is available by locating data policy, technology, and curation experts together to come up with ideas, develop action plans where possible, and implement them. KISTI and GreyNet's joint research performed last

https://doi.org/10.26069/greynet-2025-000.766-gn

year seems like a useful starting point. The author proposes the formation of the Grey Literature/Data Sharing Interest Group for a mutually beneficial data sharing cooperation planning for AI service. It is a voluntary internal activity of GreyNet member organizations centered on survey participants from the previous year. The interest group formation proposal is as follows:

- KISTI will propose the organization of interest group as one of GreyNet members.
- The subject is "Preparing policy and technology for data sharing and reuse In the AI era".
- This interest group is not for KISTI or KISTI oriented. It is opened and mutually beneficial to all GreyNet members.
- KISTI may facilitate this interest group but any GreyNet member or many members can lead collaboratively.
- This interest group activities may cover 4-5 times of online meeting discussing data sharing policy, technology, etc. and sharing result and develop next plan

This interest group will address policy and technology issues for data sharing and reuse in the AI era in practical ways, by clarifying practical issues of each organization in the four areas of policy, technology, system, and regulation of the AI era and then trying to find practical solutions. Any GreyNet member organization can participate, and organizations wishing to participate need only prepare a brief case presentation for each organization in the four areas above. The meeting will be held online and approximately four times by November of 2025 at the latest. The results of the interest group activities will be promoted as specific activities for data sharing cooperation. The results will be delivered at the next GreyNet conference. We ask for your interest and participation.

References

IFLA Special Interest Group Artificial Intelligence. (2023). Developing a Library Strategic Response to Artificial Intelligence. International Federation of Library Associations and Institutions.

Farace, D., Kim, J., & Lim, S. (2025). Global information Repository Research for STI Development: KISTI-GreyNet Collaborative Research Project. ASFA-FAO, EURASLIC, and GreyNet International Joint Conference.

Naddaf, M. (2025). How are researchers using AI? Survey reveals pros and cons for science. https://www.nature.com/articles/d41586-025-00343-5

Savic, D. (2023). Harnessing the Power of AI for Managing Grey Literature. Conference proceedings on grey literature: GL Conference series, 2023, 119-129.

University of California Irvine, 2024. AI in Research. https://guides.lib.uci.edu/research_ai/home