

01 ABSTRACT

Science and technology (S&T) information institutions build and manage a wide range of data resources, including national R&D reports, scholarly metadata, patent-related data, author and institution identification data, full-text data, and AI training datasets. However, the value of these institutional data resources is realized not through collection and preservation alone, but through **actual use in organizational and service contexts**. This poster analyzes how KISTI's S&T data are used in policy, industry, and research contexts by examining **three representative cases** selected from **66 data requests** registered in the KISTI Integrated Operation System between January 2023 and September 2025. In this system, users state their purpose when requesting data and submit post-use reports after using the data. Based on these records, the study analyzed organization types, requested data categories, and purposes of use, and conducted follow-up investigations and institutional interviews. The three representative cases illustrate different domains of data use. In the **policy domain**, the National Assembly Library used KISTI's national R&D report metadata and abstracts to enhance legislative and policy information services and demonstrate the potential for AI-enabled service expansion. In the **industry domain**, the Korea Invention Promotion Association used KISTI's domestic scholarly metadata and author identification data to enhance patent information services and support inventor identification, patent valuation, and patent analysis. In the **research domain**, KISTI's Large-scale AI Research Center used integrated S&T data to develop KONI, a Korean large language model specialized in science and technology. These cases show that institutional S&T data can create practical value across policy decision-making, industrial innovation, and research services. In particular, **grey literature**, including national R&D reports, can move beyond preservation and access to serve as evidence-based resources for policy support and AI-enabled knowledge services in the AI era.

02 BACKGROUND

S&T information institutions manage diverse data resources. Their value is realized through actual use in organizational and service contexts.

Core Content Holdings (May 2026)

Type	Records
Domestic Journal Papers	1,961,457
Overseas Journal Papers	136,933,767
National R&D Reports	313,157
National R&D Papers	901,405
Author Identification	16,292,623
Institution Identification	233,502
S&T Terminology	19,327
Total	156,655,238

AI Data Holdings (May 2026)

Type	Records
Full-text Paper (HTML)	846,027
Full-text Paper (XML)	1,217,017
Paper Q&A	323,253
Sentence Semantic Tagging	241,119
Figure/Table Description	3,309,903
Paper Metadata Extraction	5,557,300
Reference Attr. Separation	3,815,987
Ref. Type Classification	14,381,544
Topic Auto-class.	1,527,622
Acknowledgment Attr.	20,790
Total	31,340,561

03 OBJECTIVE

Analyze how KISTI's S&T data are used in **policy, industry, and research contexts**, based on 66 data requests in the KISTI Integrated Operation System (January 2023 – September 2025).

04 METHODS

The study analyzed organization types, requested data categories, and purposes of use. Institutional interviews examined three representative cases in depth.



Data source: KISTI Integrated Operation System

Period: January 2023 – September 2025

Records: 66 data requests (23 organizations)

In-depth: 3 cases (Policy - Industry - Research)

Each case examined through request records, post-use reports, and institutional interviews.

Organization Type	Data Category
Research Institute, Company, Academic Society, Public Institution, Library, University, Association	Domestic Journal, National R&D Reports, AI Training Data, Overseas Journal
Request Purpose	Period & Scale
Policy support, Patent services, LLM development, Legislative information etc.	Jan 2023 – Sep 2025 - 23 organizations - 66 total data requests

RESEARCH DESIGN

① **Data Request Analysis:** Organization types, data categories, and purposes of use were systematically coded across all 66 requests.

② **Post-use Report Review:** Outcome reports submitted after data use were reviewed to assess actual use patterns and value generated.

③ **Institutional Interviews:** Follow-up interviews with three selected organizations examined context, application, and impact in depth.

05 CONCLUSION

Core Infrastructure

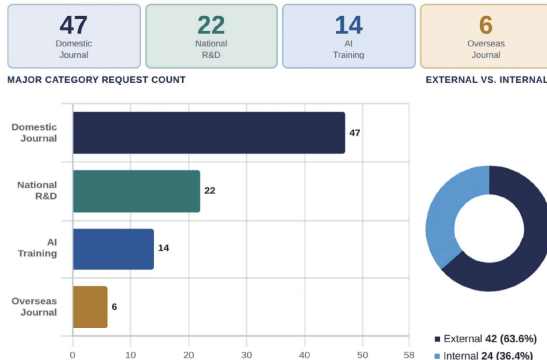
Grey literature, particularly national R&D reports, has transformed from a passive archive into a **core data infrastructure** driving policy decisions, industrial innovation, and AI research. The same data assets simultaneously support legislative reference services, patent analytics, and LLM training.

Strategic Goal

Maximizing the technical usability of data assets to fuel the **AI-driven knowledge economy**. Institutions must evolve from document preservation toward structuring grey literature into machine-readable, reusable data assets.

05 FINDINGS — INTEGRATED SYSTEM DATA REQUEST ANALYSIS (66 REQUESTS · JAN 2023 – SEP 2025)

1) OVERVIEW

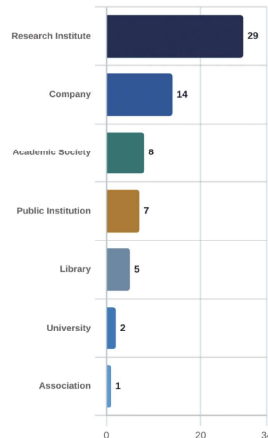


Key Insights

- Domestic Journal: 71.2% of all requests
- AI Training: 8% (2023) → 33% (2025), a 4x increase
- External requests: 63.6% · Research Institutes: 43.9%

3) ORGANIZATION

REQUESTS BY ORG TYPE (TOTAL 66)



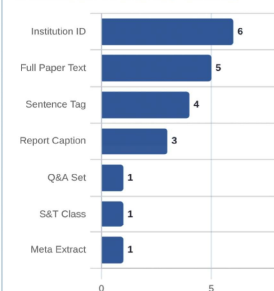
ORGANIZATION TYPE × DATA CATEGORY (EXTERNAL 42 REQUESTS)

Type	AI	Dom.	R&D
Company	1	14	2
Academic Society	1	8	1
Research Institute	1	4	1
Public Institution	2	2	2
Library	0	2	4
University	0	2	0
Association	0	1	0

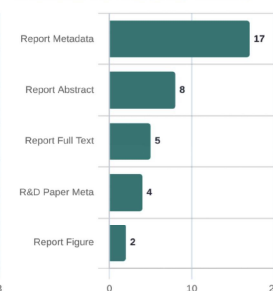
* Darker = higher proportion. Companies focus on Domestic; Libraries lean toward R&D.

2) DATA TYPES

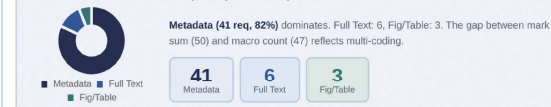
AI TRAINING SUB-TYPES (14 REQ · 21 MARKS)



NATIONAL R&D SUB-TYPES (22 REQ · 36 MARKS)

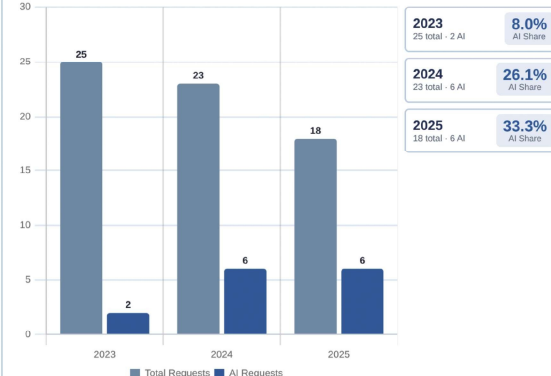


DOMESTIC JOURNAL SUB-TYPES (47 REQ · 50 MARKS)



4) AI TREND

ANNUAL AI TRAINING DATA REQUEST TREND



Trend Summary — AI training data requests surged from 8% (2023) to 33% (2025) — a 4x increase. Primary driver: growing demand for large language model (LLM) development.

06 THREE REPRESENTATIVE CASES



1. POLICY

National Assembly Library of Korea

Used KISTI's national R&D report metadata and abstracts as verified reference resources for legislative and policy information services, especially for detailed R&D-related issues in science and technology.

Data	National R&D report metadata & abstracts
Context	Legislative and policy information services
Value	Verified reference resources for R&D-related S&T issues



2. INDUSTRY

Korea Invention Promotion Association

Used domestic scholarly metadata and author identification data to enhance patent information services and support inventor identification, patent valuation, and analysis.

Data	Domestic scholarly metadata & author identification
Context	Patent information services
Value	Inventor identification, patent valuation & analysis



3. RESEARCH

KISTI Large-scale AI Research Center

Used integrated S&T data to develop KONI, a Korean large language model specialized in science and technology, trained on papers, R&D reports, and patents.

Data	Integrated S&T data (papers + R&D + patents)
Context	Development of KONI (Korean S&T LLM)
Value	Korean LLM specialized in science & technology

07 IMPLICATIONS: GREY LITERATURE IN THE AI ERA

Institutional S&T data creates value across three distinct modes:



Grey literature — particularly national R&D reports — is no longer merely a preservation concern but a **core infrastructure** underwriting both human-curated knowledge services and AI-enabled applications. Critically, the same institutional data assets can simultaneously serve as verified reference materials for policy work, as analytic inputs for patent intelligence, and as training corpora for specialized language models. Institutions must evolve from document preservation toward structuring grey literature into **machine-readable, reusable data assets** that serve human and AI users alike.