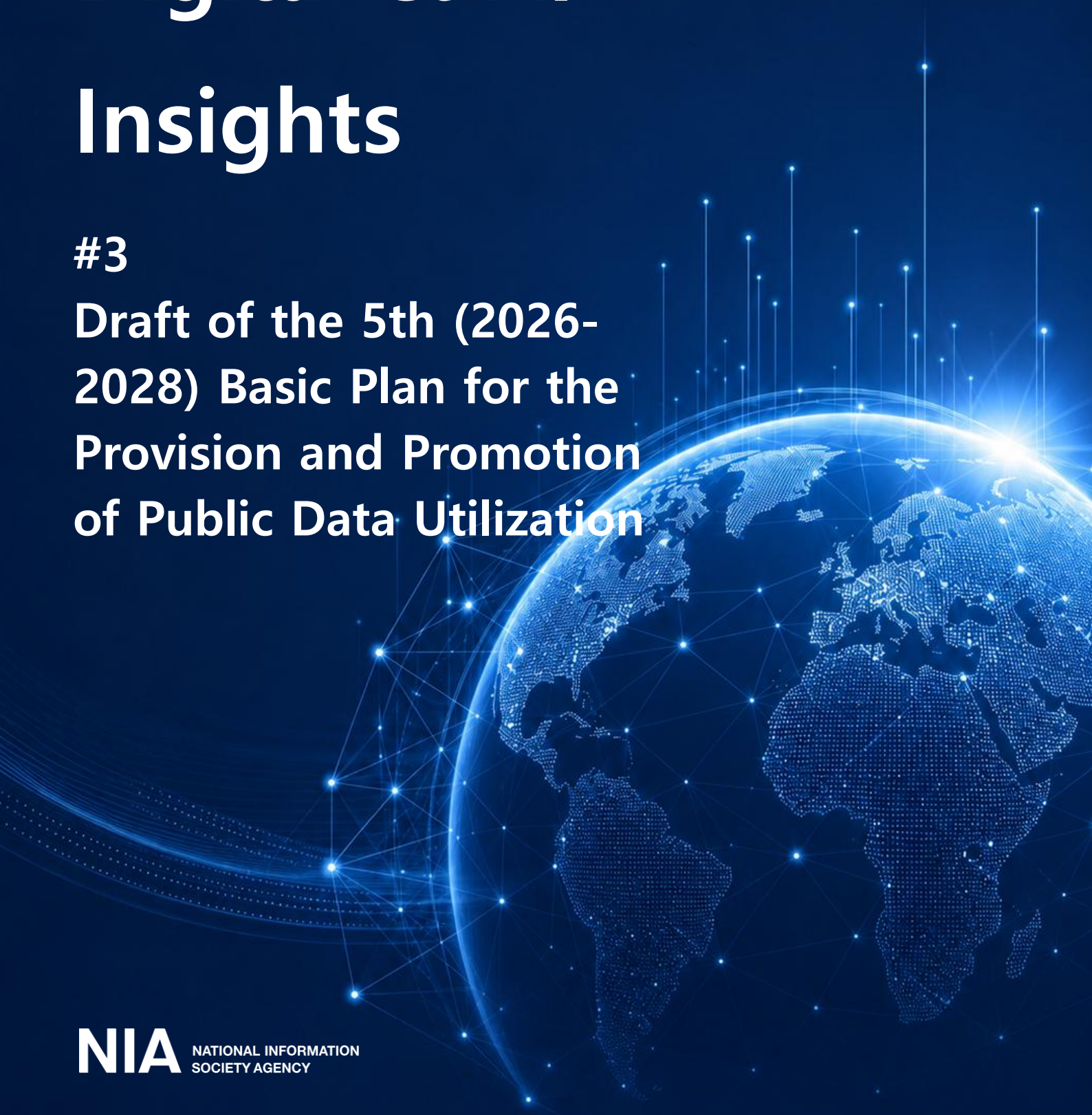
NIA Digital & AI Insights

#3

Draft of the 5th (2026-2028) Basic Plan for the Provision and Promotion of Public Data Utilization



NIA NATIONAL INFORMATION SOCIETY AGENCY

Table of Content

Draft of the 5th (2026-2028) Basic Plan for the Provision and Promotion of Public Data Utilization

Overview

The Government of the Republic of Korea has established the "5th Basic Plan for the Provision and Promotion of Public Data Utilization (2026–2028)" to secure national AI competitiveness and lead the era of AI Transformation (AX). Shifting from a supplier-centric approach to a demand-driven ecosystem, the government will invest 493 billion won over three years to provide high-value, AI-intensive data tailored to businesses and citizens.

The plan centers on four strategic initiatives. First, the government will comprehensively expand the opening of high-demand data by selecting and releasing the "Top 100 AI and High-Value Public Data Sets." This strategy focuses on aggressively increasing the share of unstructured data—such as text, images, and videos—which is critical for AI training, while strengthening support for pseudonymization and the production of synthetic data to ensure privacy.

Second, the government will enhance the user experience by centralizing access. Data catalogs from 125 individual agencies will be integrated into the Public Data Portal, supported by an LLM-based generative AI search feature to provide precise, demand-oriented data packages. To foster innovation, the "Public Data Problem-Solving Support Center" will be established to resolve access difficulties, while startups will receive comprehensive support ranging from technical consulting to cloud infrastructure and global expansion assistance.

Third, public data will be transformed into an "AI-Ready" system. Data will be maintained in machine-readable formats like CSV, JSON, and XML, accompanied by rich metadata to enable seamless AI interpretation. Furthermore, the government will move beyond voluntary management by mandating periodic quality diagnostics for critical databases and standardizing administrative documents to be AI-friendly.

Finally, the plan establishes a robust legal and governance foundation. To encourage proactive data sharing, the government will develop liability exemption guidelines for providers and amend the Public Data Act to mandate data provision. A new control tower led by Chief AI Officers (CAIOs) will integrate decentralized data governance across the public and private sectors. Through these efforts, the government aims to bridge institutional barriers, improve public data literacy, and establish international standards to position domestic companies as global leaders in the AI-ready data market.

#3

Draft of the 5th (2026-2028) Basic Plan for the Provision and Promotion of Public Data Utilization

[I]

The 5th Basic Plan for Public Data

Background

1 Overview and Progress

☐ Overview

- ○ **Plan Title:** The 5th Basic Plan for the Provision and Promotion of Public Data
- ○ **Plan Period:** 2026–2028
- ○ **Legal Basis:** Article 7 of the "Act on the Provision and Promotion of Public Data"
- ○ **Establishing Agencies: 83** agencies (49 national agencies, 17 metropolitan cities, 17 provincial/municipal offices of education)

☐ Implementation Procedure

Ministry of the Interior and Safety	⇒	National and local governments, etc.	⇒	Ministry of the Interior and Safety	⇒	Public Data Strategy Committee
Notification of Basic Plan Guidelines to national and local governments, etc.		Submission of sectoral plans by each agency to the Ministry of the Interior and Safety		Preparation of the comprehensive basic plan and solicitation of opinions from national and local governments, etc.		Deliberation and adoption of the basic plan

☐ Key Contents

- ○ Basic objectives and directions for implementation regarding the provision and promotion of public data utilization
- ○ Matters concerning the formats and methods of public data provision
- ○ Matters related to promoting private-sector use, quality control, and improvements to systems and regulations of public data
- ○ Other matters related to the management, provision, and use of public data

☐ Implementation Progress

- ○ Established the direction for the 5th Basic Plan (Meeting with private-sector members of the Public Data Strategy Committee, July 17th)
- ○ Distributed guidelines for drafting sector-specific plans to agencies (August 1st) and held an explanatory briefing (September 3rd)
- ○ Solicited feedback from stakeholders, including the public, businesses, researchers, and agency officials (September 15th–October 29th)
- ○ Development and submission of sector-specific plans by agencies (October 14th); Solicitation of feedback on the draft 5th Basic Plan (November 27th–December 3rd)

※ Deliberation and finalization of the 5th Basic Plan by the Public Data Strategy Committee: Scheduled for January 6th, 2026

Reference

Agencies Responsible for Formulating Basic Plans

No.	Target Agency	No.	Target Agency	No.	Target Agency	No.	Target Agency
1	Office for Government Policy Coordination	22	Ministry of Culture, Sports and Tourism	43	Rural Development Administration	64	South Gyeongsang Province
2	Broadcasting, Media and Communications Commission	23	Ministry of Agriculture, Food and Rural Affairs	44	Korea Forest Service	65	Jeju Special Self-Governing Province
3	Fair Trade Commission	24	Ministry of Trade, Industry and Resources	45	Office of Intellectual Property	66	Seoul Metropolitan Office of Education
4	Anti-Corruption and Civil Rights Commission	25	Ministry of Health and Welfare	46	Korea Disease Control and Prevention Agency	67	Busan Metropolitan City Office of Education
5	Financial Services Commission	26	Ministry of Climate, Energy and Environment	47	Korea Meteorological Administration	68	Daegu Metropolitan City Office of Education
6	Personal Information Protection Commission	27	Ministry of Employment and Labor	48	Korea Coast Guard	69	Incheon Metropolitan City Office of Education
7	Nuclear Safety and Security Commission	28	Ministry of Gender Equality and Family	49	Seoul Metropolitan Government	70	Gwangju Metropolitan City Office of Education
8	Korea Aerospace Administration	29	Ministry of Land, Infrastructure and Transport	50	Busan Metropolitan City	71	Daejeon Metropolitan City Office of Education
9	Administrative Capital Complex City Construction Agency	30	Ministry of Oceans and Fisheries	51	Daegu Metropolitan City	72	Ulsan Metropolitan City Office of Education
10	Saemangeum Development Agency	31	Ministry of SMEs and Startups	52	Incheon Metropolitan City	73	Sejong Special Self-Governing City Office of Education
11	Ministry of Personnel Management	32	National Tax Service	53	Gwangju Metropolitan City	74	Gyeonggi Province Office of Education
12	Ministry of Government Legislation	33	Korea Customs Service	54	Daejeon Metropolitan City	75	Gangwon Special Self-Governing Province Office of Education
13	Ministry of Food and Drug Safety	34	Public Procurement Service	55	Ulsan Metropolitan City	76	North Chungcheong Province Office of Education
14	Ministry of Finance and Economy (including the Ministry of Planning and Budget)	35	National Data Agency	56	Sejong Special Self-Governing City	77	South Chungcheong Province Office of Education
15	Ministry of Education	36	Overseas Koreans Foundation	57	Gyeonggi Province	78	North Jeolla Special Self-Governing Province Office of Education
16	Ministry of Science and ICT	37	Prosecutors' Office	58	Gangwon Special Self-Governing Province	79	South Jeolla Province Office of Education
17	Ministry of Foreign Affairs	38	Military Manpower Administration	59	North Chungcheong Province	80	North Gyeongsang Province Office of Education
18	Ministry of Unification	39	Defense Acquisition Program Administration	60	South Chungcheong Province	81	South Gyeongsang Province Office of Education
19	Ministry of Justice	40	National Police Agency	61	North Jeolla Special Self-Governing Province	82	Jeju Special Self-Governing Province Office of Education
20	Ministry of the Interior and Safety	41	National Fire Agency	62	South Jeolla Province	83	Board of Audit and Inspection
21	Ministry of Patriots and Veterans Affairs	42	Cultural Heritage Administration	63	North Gyeongsang Province		

2 Achievements to Date

1. Expansion of Open Government Data

○ Government and Public Institutions (1,164) 110,201 (based on the number of entries) Public Data Open Access

- 32,929 (↑42.6%) additional datasets opened compared to the end of the 3rd Basic Plan (2022)

※ Approximately 110,000 datasets have been made available on the Public Data Portal (since 2013) through implementation consulting under the government-wide open data plan and the National Priority Data Open Initiative

○ 242 categories of National Priority Data with high private-sector demand and significant impact have been opened

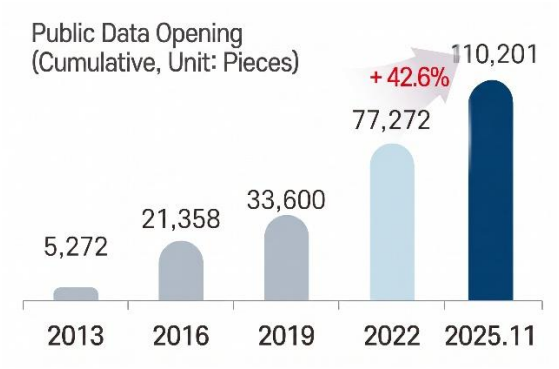
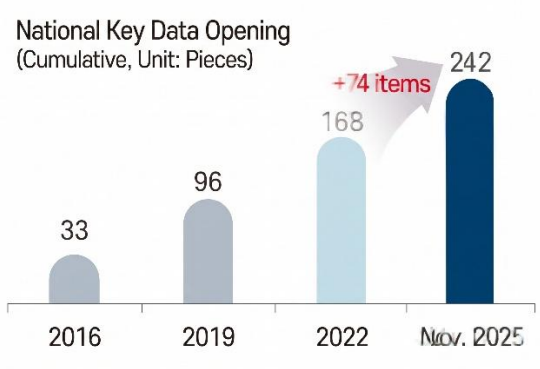
- 74 additional datasets were released compared to the end of the 3rd Basic Plan (2022), achieving 103% of the target (72 datasets) set for the 4th Basic Plan (2023–2025)

※ Cumulative number of national priority datasets released: (2016) 33 → (2019) 96 → (2022) 168 → (2025) 242

※ Constitutional Court precedents, interpretations of laws by central government ministries, special administrative adjudication rulings, real-time Seoul subway information, 119 fire information, medical imaging data, large-scale meteorological data, etc.

○ Expansion of the proportion of open formats (CSV, XML, etc.) for data released via the Public Data Portal (99.5%, 100,000+), Expansion of open API services (cumulative total of 11,846, with an average of approximately 680,000 calls per day)

※ Open format ratio on the Public Data Portal: 2016 69.4% → 2019 88.8% → 2022 97.2% → 2025 99.5%

Number of Public Data Sets Released (Cumulative)	Number of National Priority Data Sets Released (Cumulative)
	

○ In cases where it is difficult to release raw data due to the inclusion of personal information, introduce alternative methods of providing public data, such as the release of synthetic data (simulated data)*

* Virtual data that exhibits statistical similarity to the source data (e.g., public transit card usage data, cancer clinical libraries)

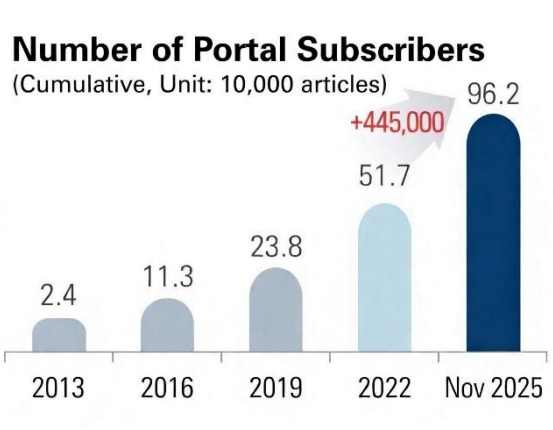
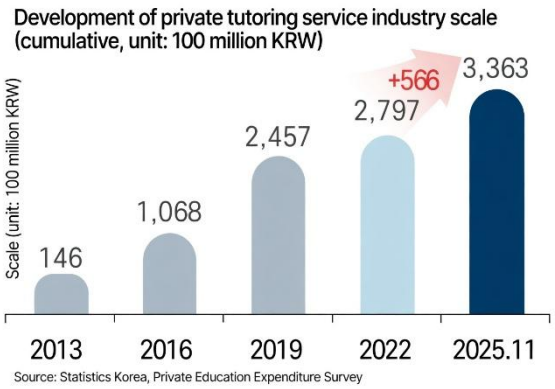
2. Promoting the Use of Public Data and Strengthening Public-Private Collaboration

○ Public Data Portal (data.go.kr) user (registered portal members) exceed 962,000

- An increase of approximately 445,000 (86%) compared to 517,000 at the end of the 3rd Basic Plan (2022)

○ Through tailored support for companies utilizing public data, 3,363 apps and web services with high public impact were created, and startup support and other measures laid the foundation for corporate growth

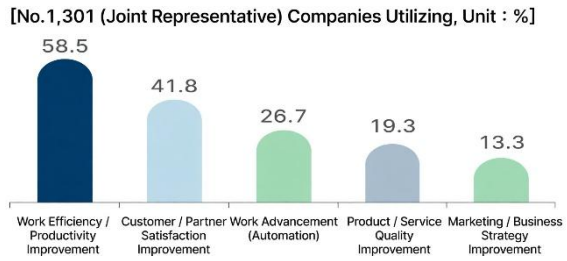
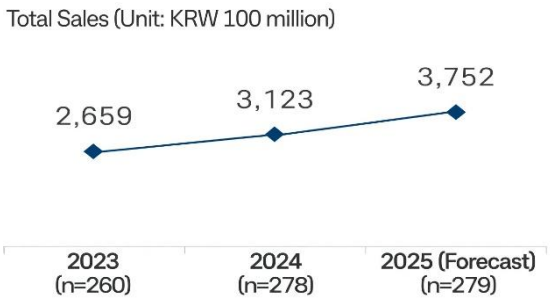
※ 21 companies among the winners of startup competitions were selected for the Technology Incubation Program for Startups (TIPS), a private-investment-led technology startup support program

Public Data Portal Users* (Cumulative)	Public Data Utilization Services (Apps/Web) (Cumulative)
	

<Examples of Services Utilizing Public Data>

ANRISK(Talos) (Grand Prize, 2023 Startup Competition)	ConnectOn (FarmConnect) (Excellence Award, 2023 Startup Competition)	HealingHeart (HealingHeart) (Excellence Award, 2024 Startup Competition)	RiderLog((Let's Go Star-Gazing) (Excellence Award, 2024 Startup Competition)
ANRISK AI CEREBRAL ANEURYSM RISK-ASSESSMENT SERVICE			
• Utilizing health screening data, etc. • AI-based brain aneurysm risk prediction service	• Utilization of weather and crop cultivation environment data • AI and digital twin-based smart farm environmental analysis solution	• Utilization of medical insurance data • Location-based booking platform that connects visually impaired massage therapists with clients	• Utilization of traffic accident data • Motion sensor-based automatic accident reporting service for motorcycles

○ The improvement of operational efficiency and productivity and the enhancement of service quality through the use of public data drive competitive advantage and sustained performance for companies utilizing these data

Key Achievements in Public Data Utilization	Revenue from Public Data Utilization (Total)
※ Multiple Responses Allowed  	

○ By utilizing public data to address local issues, we aim to create social value from the data and strive to improve the quality of life in the community

※ Through Living Lab projects, citizens, businesses, and local governments are working together to identify models for solving problems using public data

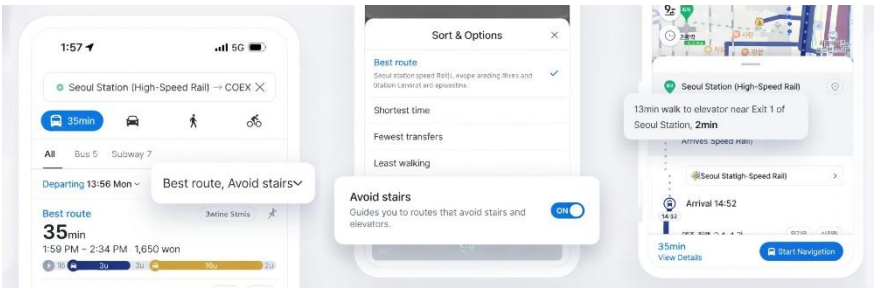
<Key Examples of Living Lab Projects Solving Local Issues Through the Use of Public Data>

Barrier-Free Subway Navigation Service to Ensure Mobility for People with Disabilities (Private Map App)

- Development of a barrier-free route algorithm that provides optimal elevator boarding locations and optimal transfer routes to improve mobility for people with mobility challenges within subway stations

※ Established a foundation for service development through the standardized and optimized release (via API) of public data held by regional transportation corporations and agencies, as well as the collection of new data (such as elevators adjacent to platforms) to enhance route planning

(Naver Maps) Directions: Stair Avoidance Feature and Real-Time Construction Updates

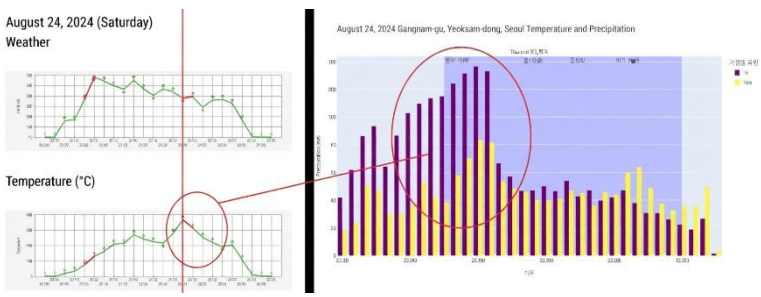


Service for Measuring Crowd Density by Time Slot at Outdoor Festivals to Prevent Safety Incincidents

- Real-time tracking of foot traffic by time slot Algorithm development to prevent crowd crush incidents and support the efficient allocation of resources such as security personnel and trains

※ Utilizing Daejeon Metropolitan City's CCTV dataset for AI training, as well as ticket gate traffic data from the Daejeon Transportation Corporation and transportation big data traffic data, to detect people and moving objects indoors and outdoors and analyze crowd density

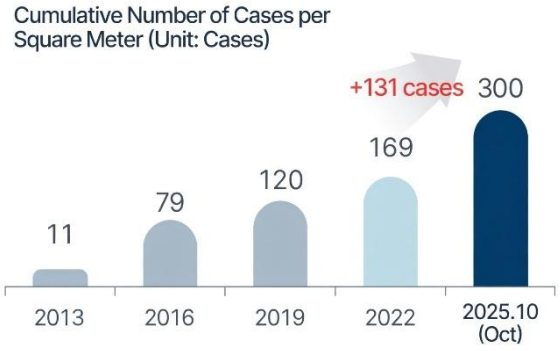
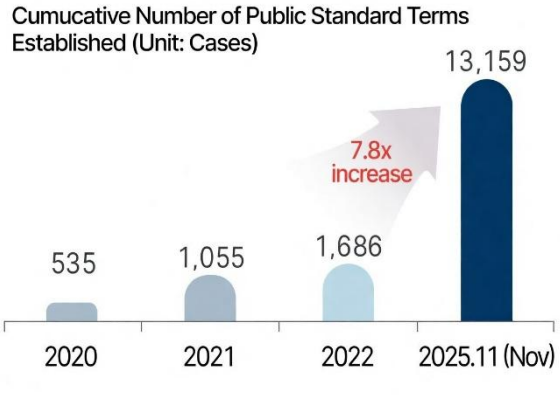
AI Model for Measuring Foot Traffic	Comparative Analysis of Public Transportation Traffic Data and Foot Traffic Patterns



3. Improvements to the data management system, including the standardization of public data

○ Expand government-wide data standardization to establish provision standards* for data jointly held by multiple agencies—which is in high demand among the private sector—as well as to facilitate data linkage and joint utilization

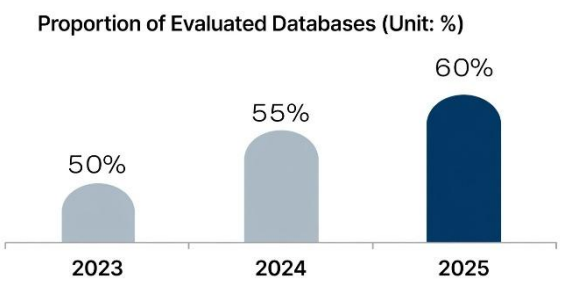
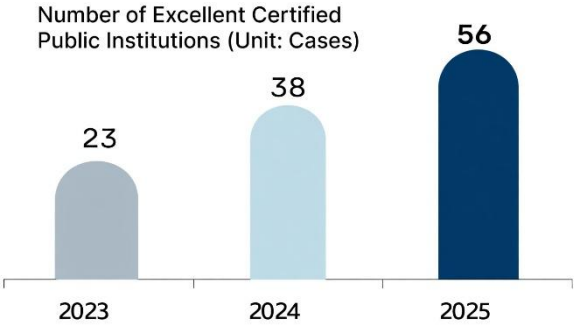
* Additional provision standards will be introduced for information such as local children's centers, shade shelters, and middle and high school class schedules

Public Data Provision Standards (Cumulative)	Common Terminology for Public Data Standards (Cumulative)
	

○ To improve the quality management of public data, we will expand the scope of databases subject to quality management diagnostics and evaluations* and grant certification for agencies with excellent quality**

* (Diagnostics) Gradual expansion of databases subject to evaluation (50% of agency-operated databases in 2023 → 55% in 2024 → 60% in 2025)

** (Quality Certification) Review quality management systems—including data values, standards, structure, and integration—and grant certification to outstanding organizations; (2023) Fair Trade Commission, Korea Customs Service, Korea Gas Corporation, etc.; (2024) Korea Railroad Corporation, Foundation for the Promotion of R&D Special Zones, etc.

Percentage of Databases Subject to Diagnostic Evaluation Relative to the Total Number of Databases Operated by the Institution (%)	Institutions with Excellent Quality Management (since 2023)
	

○ To support institutional tasks such as data quality and standards management, we will establish and operate a metadata-centric "Integrated Public Data Quality and Standards Management System" (starting in 2024)

* (Current) Manual management of existing database status, data standards dictionary, and database outputs → (Improved) Integrated management of database status through linkage with the IRM-Metasystem, and establishment of a management framework for database standards and outputs

○ Launch of the 'Usability Diagnostic Service'* to support the usability of open data, including quality level, understandability, ease of processing, and interoperability (2024–)

* Mandatory diagnostics upon data registration on the Public Data Portal and restrictions on registering data that fails to meet threshold criteria

Reference

Agencies Responsible for Formulating Basic Plans

Organization Name	Key Points
Personal Information Protection Commission	□ To expand the provision and utilization of pseudonymized data, efforts will be made to clarify the roles and responsibilities for processing pseudonymized data within public institutions (to be submitted to the Ministerial Meeting on National Policy Issues in 2025) □ Expand the release of public data by 435% compared to the 3rd Public Data Master Plan (2020–2022) (14 cases in 2022 → 61 cases in 2025)
Ministry of Employment and Labor	□ Through the Employment and Labor Data Analysis System (ELDS), we provide access to and share raw employment and labor data and offer an online analysis environment to support the implementation of policy initiatives and academic research (10 projects)
Ministry of Science and ICT	□ Release of AI training data, including AIoT anomaly detection data and object status recognition data (59 items) □ Development of a draft guide for the operation and management of public research data and operation of data consultative bodies for common fields (including subcommittees on postal services, radio waves, and science and technology research)
Ministry of Agriculture, Food and Rural Affairs	□ Data Safe Zone: Agricultural business entity data (9 years) and corporate data (3 years) with high demand within the zone will be released after anonymization □ AI training data released (45 items), including images of pests and diseases in greenhouse crops and information on leading smart farm operators
Ministry of Health and Welfare	□ Through the establishment of an anonymous data workflow, provision of anonymous data on reimbursed pharmaceuticals and medical devices to the bio-health industry (1,940 cases) □ Opening of unstructured data from healthcare imaging services, including medical imaging interpretation training data (20,135 cases) and training and raw unstructured data for 9 diseases (27,345 cases)
Ministry of the Interior and Safety	□ Newly released public datasets (654 items) with high utility value and strong private-sector demand, including location data for emergency safety bells, information on rest areas for drowsy drivers, and information on public safety facilities
Gyeonggi Province	□ Expanding the release of high-value unstructured data for AI training, including data for autonomous driving AI training, intelligent CCTV video data, and Intelligent Transportation System (ITS) data

Gwangju Metropolitan City	□ Conducted a comprehensive survey of held data to open previously restricted data (504 datasets), including the status of snow removal bins and the status of victims of rental fraud
Daegu Metropolitan City	□ Identified and promoted the dissemination of public data use cases that directly benefit citizens, such as the development of energy-saving and forecasting models and the prediction of customers at risk of churning in local delivery services □ Opened user-centric, daily life-oriented public data (579 datasets), including drone footage of major tourist attractions and images of cultural heritage sites
Seoul Metropolitan City	□ Improved information accessibility and usability through the “Seoul Data Hub,” a service offering natural language-based interactive public data search and visualization analysis □ Operated the Big Data Campus to enable the safe analysis and utilization of unreleased data

3 Rationale for Development (The Need for a New Public Data Strategy)

[Remarks by the President: First Strategic Meeting on the Rationalization of Key Regulations (September 15, 2025)]

"Public data is a public asset created using taxpayers' money. We will actively open and expand access to the private sector to facilitate the rapid growth of AI-based new industries."

※ (Pledge) Promote the opening of public data to support corporate research and development

(Directive) Actively open public data to contribute to the creation of new industries
(3rd Emergency Economic Review Task Force, July 30, 2025)

☐ South Korea, a digital powerhouse, has weak AI competitiveness

○ While it ranks among the world's best in the OECD's Public Data and Digital Government Assessment, its overall ranking in the Global AI Index* was 6th in 2024, indicating that its AI competitiveness is relatively weak

* An annual assessment conducted by Tortoise Media (UK) based on criteria such as government strategy, infrastructure, operational environment, and development capabilities

☐ The key to AI competitiveness lies in the utilization of large volumes of high-quality data

○ The latest generative AI models (such as GPT-5 and Gemini) are based on vast amounts of refined training data; therefore, securing high-quality data is critical for AI performance

○ To build AI competitiveness, it is necessary to create a virtuous cycle ecosystem of AI and public data by combining private-sector innovation capabilities with high-demand, high-value, and high-quality public data

☐ Current public data policies are insufficient to support the AI era

○ Despite the continuous expansion of public data openness over the past decade, the system has failed to move beyond a structured, fragmented, file-data-centric structure, leading to a widening gap with private-sector demand

※ Unstructured data (text, images, etc.) accounts for only 6% of the open data on the Public Data Portal

○ To ensure the survival of citizens, businesses, and the government in the global AI hegemony competition, a structural transformation is urgently needed to systematically share, open, and utilize public data

○ We must inject vitality into the industrial ecosystem through the bold opening of high-value public data and the creation of public AI demand to support the leap toward becoming one of the world's top three AI powers

※ (National Policy Goal ②) A world-leading innovation economy, (Strategy 1) Leap toward becoming one of the world's top three AI powers, (National Policy Task 24) "Realizing the world's No. 1 AI government"

To leap forward as a nation with world-class AI competitiveness based on public-private cooperation, it is necessary to establish a new paradigm for public data strategy

Reference

Changes in the domestic environment as the transition to AX gains momentum

√ **A Paradigm Shift in Data Policy to Support the New Administration's Goal of "Becoming One of the World's Top Three AI Powers"**

○ Outlines a data policy direction focused on prioritizing the sharing and opening of public data—which has high demand for AI service development—with the goal of realizing the "World's Leading Democratic AI Government"

○ Emphasizes that public data serves as a catalyst to accelerate the AI transformation across all social and economic sectors, and as a key means for "leaping forward to become one of the world's top three AI powers"

※ Agreement reached to actively open public data to facilitate a major AI transformation across all sectors, including businesses, the public sector, and citizens' daily lives (Party-Government Consultation on the New Administration's Economic Growth Strategy, August 20, 2025)

√ **Growing Demand for High-Quality, High-Value Data Suited to the AI Era**

○ With the advent of the AI era, high-quality, high-value data that underpins the performance and reliability of AI models has emerged as a key foundation

※ Public data, public copyrighted works, and industrial data must be made available to foster the development of new industries such as AI and autonomous driving (1st Core Regulatory Rationalization Strategy Meeting, September 2025)

○ There is a need to establish institutional and technical solutions* to enable the opening and utilization of data that was previously difficult to release due to issues such as personal information and copyright

※ Expanding pseudonymization, introducing synthetic data creation methods, and revising the Personal Information Protection Act and the Copyright Act, etc.

√ **Enhancing the data utilization environment to support the spread of AI services**

○ As AI-based services become more complex and sophisticated, the need for linking and integrating data of various types and formats is increasing

※ Advanced AI services require the integrated utilization of heterogeneous data such as text, images, and audio; performance is difficult to achieve with a single type of data alone

○ Due to the proliferation of multimodal AI technologies, there is a need to expand the framework for opening and utilizing unstructured data (such as text, images, and audio) beyond a structured data-centric approach

○ It is necessary to establish an environment where AI can easily find and immediately utilize data through the development of sophisticated metadata and quality control systems, as well as the implementation of RAG and vector databases

Reference | Global Trends in Preparing for the AI Era

√ Securing Future Growth Engines Through the Release of AI Training Data and High-Value Data

Category	Key Points
 USA	< AI Action Plan (July 2025) > • Building high-quality scientific datasets for AI training, providing incentives for dataset publication, and requiring the disclosure of datasets used in AI models during federally funded research and experimentation, etc.
 China	< Public Data Labeling and Utilization Policy (2025) > • Collaborate with the private sector to process (label) public data to make it suitable for AI training, and support its use in the development of large language models (LLMs) and other applications

√ Strengthening Systematic Data Quality Management to Ensure AI Reliability

Category	Key Points
 USA	< Guidelines for Generative AI and Open Data (January 2025) > • Requires the establishment of minimum data quality standards suitable for AI training, as well as systems for data integrity management and automated quality checks
 EU	< Inclusion of AI-Related Data Governance Requirements (AI Act) (August 2024) > • Develops guidelines for the quality management of data used for AI training and mandates compliance with these guidelines to ensure the reliability of public AI systems

√ Strengthening a private-sector-led data ecosystem through support for data-driven companies

Category	Key Points
 USA	< U.S. AI Research Resources (NAIRR) Pilot Project (January 2024) > • Supports researchers and startups in conducting AI R&D by providing secure access to large-scale computing resources and government datasets

 UK	< UK Government's AI Opportunities Action Plan (January 2025) > • Support the resolution of societal challenges by establishing an inter-ministerial data hub to expand data sharing across ministries and promote private-sector data utilization ※ Establishment of the AI Research Resources (AIRR) and the National Data Library (NDL) (2025)

√ Revamping AI and Data-Centric Governance

Category	Key Points
 USA	< AI Action Plan (July 2025) > • Establish and operate the Council of Chief AI Officers (CAIOC) as the central body for interagency coordination and cooperation on AI adoption ※ Four-tier structure: White House OMB (Strategy) GSA (Data.gov Operations) CDO Council (Data Governance) Department-level CDO + CAIO + AI Governance Board (Execution and Coordination)
 France	< Strategy for Modernizing Public Services Based on AI (April 2024) > • Unify the data governance system by establishing a government-wide Chief Data Officer (CDO) body ※ Structure: Prime Minister's Office DINUM (Digital and Data Oversight) Etalab (data.gouv.fr operations and data strategy) State CDO + AMDAC Network (ministry-level CDO and CAIO consultative body)

4 Limitations and Implications of Current Public Data Policies

☐ **(Open Access and Sharing) There is a need to open high-value data for which corporate demand is high, such as for AI service development and to provide data containing sensitive information**

○ (Agency-Centric Open Access) As the focus has been on quantitative open access centered on data that is easy for agencies to release*, there are limitations in reflecting the actual utilization needs of the public and businesses

* In particular, the data release structure remains confined to structured and fragmented file data; the proportion of unstructured data (text, images, etc.) essential for AI training is only 6%.

○ (Lack of Sensitive Data Release) Data containing sensitive information such as personal information* remains largely unused despite its high value for private-sector applications.

* (Examples) Public inquiry consultation information, real-time electricity usage, real-time CCTV footage (parking lots, underpasses, etc.)

○ (Fragmented System of Open Data and Data Sharing) Since data opening and sharing are operated under separate systems (distinct procedures and systems), this hinders inter-agency collaboration and increases the workload of responsible staff

Implications A demand-driven public data opening strategy that considers AI utilization is needed

☞ We must move beyond the supplier-centric opening of structured data and adopt an opening policy that focuses on high-value data—specifically unstructured data with high private-sector demand and data suitable for AI training and service development

☞ Address constraints on data opening and revise legal and institutional frameworks, including guidance on liability exemptions, in response to public data opening

☞ A virtuous cycle system must be established for data shared among agencies, spanning from production to public access

☐ **(Promoting Adoption) Calls for an AI-company-centric support system and improvements to portal usability**

○ (Difficulty in Securing Data) Despite being essential for corporate service development, there have been cases where holding institutions consistently refuse to provide data

Category	Total	Citing a lack of authorization, etc.	Includes non-public information	Requires additional processing	Not Machine-reaedable and others
Number of Refusals (cases)	3,532	2,060(58.3%)	957(27.1%)	316(9%)	199(5.6%)

○ (Lack of Support for Utilizing Companies) While tailored support programs for companies utilizing public data are in operation, there is a lack of AI-related support systems that meet the needs of these companies

※ Among companies participating in the tailored support programs, 97% expressed an intention to apply AI to their products or services, and 60% indicated a need for AI training and support (Survey on satisfaction with tailored support programs)

○ (Lack of Intelligent Search Capabilities) Current portals only provide simple keyword-based search; there is a lack of intelligent search capabilities, such as understanding user intent and recommending related data

※ Satisfaction with data retrieval and search within portals is the lowest, and there is a demand for improvements to search functions (2024 Corporate Survey)

Implications Need for AI-based, customized corporate support and improved accessibility

☞ It is necessary to promote the growth of innovative AI companies by expanding a problem-solving-oriented support system for companies that refuse to provide data, as well as AI- and data-driven, tiered, and customized support

☞ It is necessary to enhance the usability of the Public Data Portal—including intelligent search and recommendation features—and improve data accessibility

☐ **(Quality Control) Fragmented data exists; improvements are needed in data descriptions (e.g., column definitions)**

○ (Constraints on AI Training and Utilization) Since data formats and quality standards vary across agencies, this leads to quality discrepancies at the government-wide level and reduces the efficiency and reliability of AI utilization

○ (Effectiveness of Quality and Standardization) (Quality) Improvement efforts are insufficient due to agency-led self-management and a one-size-fits-all diagnostic approach; (Standardization) There is a lack of standards suitable for AI training and utilization

Implications Improving data quality by converting it into AI-Ready data that AI can easily read and utilize

☞ Improve the data format so that AI can easily utilize it for learning, analysis, and inference by providing rich attribute information such as update history, frequency, source, and summary descriptions

☞ There is a need to strengthen government-wide quality assessments and improvement support from the perspective of AI utilization

☐ **(Law & Governance) Numerous legal and institutional constraints hinder the opening of public data**

○ (Law & Institutions) The absence of a legal basis for pseudonymization within the Public Data Act and the lack of specificity in the exemption clauses regarding liability for data opening have led to a passive attitude among data-holding agencies

○ (Governance) Inadequate cooperation systems between the data governance frameworks of different agencies

Implications Establishing laws, regulations, and a governance framework for public data to lead the AI era

☞ A need to establish an indemnity system to address the passive attitude stemming from the legal basis for public data collection and processing, as well as the excessive burden of responsibility placed on officials at holding agencies

☞ A need to strengthen cooperation among data governance bodies that are currently managed in a decentralized manner

Reference	Gathering opinions from all sectors of society

Category	Key Comments
Citizens	[Online Public Survey (Sept. 2025–Oct. 2025)]] • Need to improve the search functionality and user experience (UI/UX) of the Public Data Portal • A Request to improve data quality (frequency, values, standard terminology, etc.) to enable AI to utilize public data directly • Demand to expand the availability of public data in sectors such as mobility, environment, and healthcare • Request to expand support for training programs to strengthen the public's capacity to utilize public data ☞ Need to improve the convenience, accessibility, and quality of public data searches; expand data openness to enhance AI integration; and provide educational support for the public
Companies	[Survey on the Status of Companies Utilizing Public Data (Oct. 2025), etc.] • Calls for the opening of public-interest-focused data (e.g., infectious diseases, industrial and disaster safety), industry-specific (domain-specific) AI demand-based data, and high-value, highly versatile data (e.g., land and finance) • Need to mandate the opening of research and industrial data funded by the national budget, such as research and development (R&D) and demonstration projects • Need to establish systems to address the passive attitude toward data opening caused by excessive burdens on officials at holding agencies, and to establish specialized data opening organizations and governance structures • Preference for methods that expand data accessibility (data vouchers), match specialized personnel, and provide consulting support ☞ Need to foster an ecosystem for the utilization of AI public data by opening high-value data to drive industrial growth, enhancing agencies' proactive approach to data opening, and strengthening corporate support systems

Researchers	[In-depth interviews with researchers (Sept. 2025), etc.] • Calls to expand accessibility and convenience in utilizing safe zones for the analysis and use of data containing sensitive information such as personal data (e.g., operating hours, access to the latest AI tools and infrastructure, provision of necessary data) • Calls to expand accessibility and convenience in utilizing safe zones for the analysis and use of data containing sensitive information such as personal data (e.g., operating hours, access to the latest AI tools and infrastructure, provision of necessary data) • Need to establish systems—such as measures to protect researchers' intellectual property rights and incentives—to facilitate the open access of data from research funded by the national budget ☞ Need to make the use of safe zones for research purposes more flexible and establish an institutional foundation for opening research outcome data
Institutional staff	[Online Survey for Operational Staff (Sept. 2025–Oct. 2025)] • Need to establish collaborative relationships with companies utilizing the data, such as by expanding projects that match demand and supply and foster collaboration • Need to establish a system for inter-ministerial data sharing and collaboration • Due to unclear accountability regarding data disclosure, liability protection measures are needed, and there is a demand for an incentive system for agencies when data is disclosed ☞ Need to lay the groundwork for breaking down data silos between ministries and establish systems that enable agency officials to actively open data

[Ⅱ]

The 5th Basic Plan for Public Data

Vision and Key Initiatives

Vision Framework of the 5th Basic Plan
(4 Major Areas, 12 Key Tasks)

Vision

Public Data Supporting the Goal of "Becoming One of the World's Top Three AI Powers"

Objective

Opening High-Quality Public Data Focused on the Needs of Businesses and Citizens

4 Major Strategies

1. Opening High-Demand, Hign-Value Public Data for AI
2. Creating a User-Centric Environment for Businesses and Citizens
3. AI-Friendly Data Quality Management
4. Strengthening the Foundation for Promoting

12 Key Initiatives

1. Openness · Sharing

- Prioritizing the Release of Key Data to Meet the Needs of AI Companies
- Diverse Provision and Release of Data, Including Personal Information
- Establishing a Unified System for Data Release and Sharing

2. Promoting Usage

- Strengthening Support for Companies Utilizing AI and Public Data
- Creating an environment where public data can be conveniently accessed in one place
- Strengthening public-private cooperation systems and partnerships for public data

3. Quality Control

- Transition to AI-Ready Public Data Quality Management
- Improving the method of opening public data to make it readable by AI
- Improving data formats to open-source documents designed for AI utilization

4. Law · Governance

- Improving laws and regulations to promote the opening and sharing of public data
- Strengthening cooperation and utilization capabilities among data-related agencies
- Strengthening global leadership and cooperation in public data

1 [Openness・Sharing]
Opening Access to High-Demand, High-Value AI Data

◇ We will identify high-value data necessary for AI service development and make it available and share it in a secure and user-friendly format.

Voices from the Field	Areas for Improvement
□ Urgent Need for High-Value Data Disclosure to Support User-Centric AI Training and Utilization ※ 88.5% of survey respondents (2,050 people) stated that public data necessary for AI training and utilization should be disclosed □ Growing Support for 'Conditional Data Disclosure', Which Prioritizes Balancing Personal Data Protection and Data Utilization ※ 71.2% of respondents emphasized the importance of balancing personal data protection and data utilization	□ Data in high demand by companies, such as for AI development, will be selected through a top-down approach and made available as a priority □ By utilizing pseudonymization and synthetic data, we will expand the availability of high-value data containing personal information □ In the AI era, data will be made available in a format suitable for use

1. Prioritizing the release of key data to meet the needs of AI companies

☐ Selection and Release of the「Top 100 High-Value AI Datasets」

○ (Top 100 Selection) Identification and selection of the「Top 100 AI and High-Value Public Data Sets」based on data with high demand for AI service development or sustained corporate demand

- (Identification) Continuous assessment of demand through various channels, including surveys of AI companies and the general public, consultation with experts, OECD recommendations, and the Public Data Portal
- (Selection and Management) Final selection through the screening of core data by sector with high demand and high value, review of suitability for release, and review by data-holding agencies and private-sector experts*

* Evaluated based on private-sector demand, AI compatibility, economic ripple effects, and alignment with national policy priorities (list finalized and release implementation managed through the Public Data Strategy Committee)

○ (Release) Reflecting the latest annual demand, data is sequentially released and managed under a medium- to long-term plan (3-year cycle), taking into account ease of release, urgency, and impact

※ Medium- to Long-Term Release Plan (Draft): 2025 10 items → 2026 25 items → 2027 30 items → 2028 35 items

<Release·Management Process>

1.Needs Assessment	2.Project Selection	3.Data Releasing Support	4.Implementation Management
Identify key common needs across multiple companies	select priority projects	provide financial and technical support for data opening by institutions	manage performance of data opening implementation through statutory evaluations

< Draft List of the Top 100 AI and High-Value Public Data Sets to Be Released >

Goals	Open fields	Top 100 List (Draft)	Expected Outcomes
Improving public safety and quality of life	Disasters & Safety	Data on industrial accident information and preventive measures Data on sinkhole-prone areas nationwide Data on safety inspections and management of special bridges, etc.	Strengthening the social safety net through the advancement of data-driven response systems
	Health & Medical Care	National Mental Health Counseling Data Nationwide Hospital-Specific Fee Data, etc.	Enhancing the competitiveness of the digital healthcare industry through the utilization of medical and bio data
	Culture & Tourism	National Museum of Korea artifact data Nationwide tourism image and video data 3D data in the cultural sector, etc.	Strengthening the competitiveness of the K-content industry through the convergence of culture and tourism data
Expanding inclusive welfare and creating quality jobs	Employment, Labor & Welfare	Data on the status of childcare service support Data on mobility facilities for people with limited mobility, etc.	Innovating support services for socially vulnerable groups and addressing gaps in the welfare system
Strengthening the Competitiveness of the Innovation Economy and Industrial and Transportation Infrastructure	Land, Transportation & Construction	Data on public transportation card usage Driving data tailored to road environments for autonomous vehicles	Improving industrial productivity through enhanced traffic safety and logistics efficiency

	Finance, Economy & Banking	Tax consultation data Data on housing credit guarantee claims, etc.	Streamlining AI-based risk management and enhancing transparency in public finances
Achieving Carbon Neutrality and the Green Transition	Climate, Environment & Energy	Smart grid operation data from industrial complexes Operation data from power plants nationwide, etc.	Fostering global emerging industries such as the transition to renewable energy and climate response
	Agriculture, Forestry, Fisheries & Oceans	Nationwide agricultural product distribution data Smart farm greenhouse environmental and crop yield research data, etc	Strengthening productivity and food security, and supporting the stabilization of agricultural and fisheries markets
Achieving balanced regional development and a fair public system	Law	Lower court judgment data Central administrative agency legal interpretation data, etc.	Supporting the growth of the AI-based legal tech industry and improving access to legal information
	Balanced Regional Development	Nationwide industry-specific licensing information and daily convenience data Integrated regional currency issuance and management data, etc.	Promoting balanced development through advanced policy-making
	Other	Comprehensive pet data Nationwide academic administration data, etc.	Creating new industries and improving management efficiency through the open access to diverse data

☐ Expanding the Release of Unstructured Data Suitable for AI Training

○ (Unstructured Data) Moving away from a structure centered on structured and fragmented file data, we will establish a management system that takes unstructured data into account and expand its release

※ Demand for unstructured data accounts for 24.8% of total public data release demand (2024 Survey on the Status of Companies Utilizing Public Data)

- (Management System) Develop and disseminate guidelines for unstructured data disclosure, including standard management items and type-specific management systems
- (Priority Disclosure) Prioritize the disclosure of unstructured data held by agencies that is suitable for AI training, such as official documents and public notices

<Proposed Plan for the Phased Release of Unstructured Data>

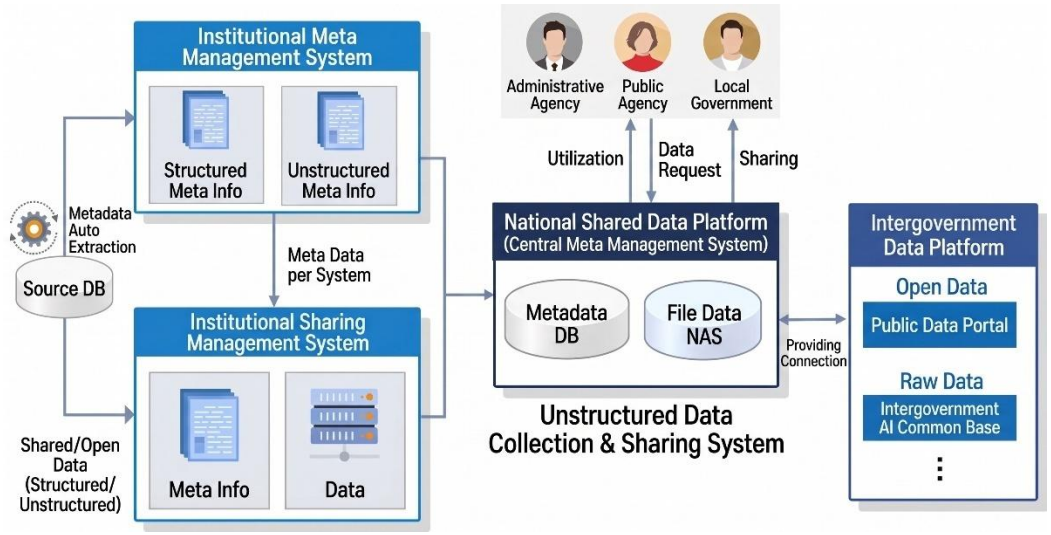
Phase 1		Phase 2		Phase 3
Document and Text Data Release (e.g., public agency reports, public notices, etc.)	⇨	Image and video data release (e.g., CCTV footage, aerial photographs, map data, etc.)	⇨	Audio and Spatial Data Open Access (e.g., meeting minutes, audio transcripts, 3D spatial data, etc.)

- (Expanding Data Access) Through the National Shared Data Platform, we have established an environment for the automatic collection of metadata* and file data for unstructured data, thereby expanding access

* Creation date, data description, sharing/access/privacy status, data/file format and quantity, etc.

Examples: Video (format, playback time, codec, etc.), images (format, resolution, etc.), spatial information (type, map coverage, etc.)

<Draft Framework for the Collection and Sharing of Unstructured Data>



2. Various methods of providing and sharing data containing personal information

☐ Promoting the Provision of Pseudonymized Data Containing Personal Information

○ (Support for Pseudonymization) Coordination and support for specialized pseudonymization agencies (designated by the Personal Information Protection Commission) that provide one-stop services covering everything from data selection to pseudonymization, adequacy reviews, and post-processing management

< One-Stop Support Procedure for Pseudonymization (Personal Information Protection Commission) >

Users of pseudonymized data	Public institutions	Specialized agencies for pseudonymization
1) Request for the provision of pseudonymized information from public institutions →	2) Request for one-stop support from the Personal Information Protection Commission (specialized agency) →	3) Perform pseudonymization of personal information in accordance with the intended purpose ↓
6) Receive pseudonymized information and use it for the intended purposes (e.g., statistics, scientific research) ←	5) Approve the provision of certified pseudonymized information to users ←	4) Conduct an adequacy assessment to verify that the pseudonymization process was performed correctly

○ (Providing a Supportive Environment) Personal Information Innovation Zone* (Reviewed and Designated by the Personal Information Protection Commission) to expand the safe provision of pseudonymized data

* A space where the pseudonymized data processing environment is securely managed, enabling more flexible use of pseudonymized data (7 locations in operation as of 2025)

<Examples of AI research conducted using the "Personal Information Innovation Zone">

- Survey on the Prevalence and Survival Rates of Cardiotoxicity in Patients with Stomach, Breast, and Colorectal Cancer, and Development of an AI Model for Risk Prediction (National Cancer Center)
- Development of an AI Model for Predicting Mortality-Related Prognosis in Korean Breast Cancer Patients (National Cancer Center)
- Development of an AI Model for Re-evaluating and Predicting the Effectiveness of Social Welfare Policies for the Elderly (Korea Social Security Information Service)
- Development of an AI model for identifying and reporting vehicles in violation of traffic laws (Korea Expressway Corporation)
- Study on the current status of beneficiaries of combined employment and welfare policies for child-rearing support (Korea Social Security Information Service)

○ (Special Provisions for Data Utilization, etc.) Regulatory sandbox for unstructured data*, and special provisions for data utilization under the Personal Information Protection Act**, among others: expansion of the scope of application (Personal Information Protection Commission) and support for utilization***

* (Existing) Autonomous driving ⟶ (Expanded) Disaster prevention, public safety added

** (Existing) Up to 4 years ⟶ (Expanded) Period necessary for research

*** Survey of corporate demand for regulatory sandboxes, etc. ⟶ Matching with relevant agencies ⟶ Linkage and provision of Data Safe Zones (Ministry of Science and ICT)

☐ Diversifying Methods for Opening Data Where Raw Data Cannot Be Disclosed

○ (Synthetic Data) Support for the creation and release of synthetic data* to enable the safe release and utilization of data without exposing personal information

* Virtual data with patterns similar to raw data, which is advantageous for both data protection and utilization

- (Support ①) Support for the processing and utilization of synthetic data through the Personal Information Innovation Zone

- (Support ②) Support for personal information security assessments of synthetic data through linkage and utilization of the One-Stop Support System for Pseudonymization (Personal Information Protection Commission)

※ Provision of resources such as a pool of experts to support verification of the risk of re-identification of personal information

○ (Authentication Services) Expand the provision of authentication services for data where there is high private-sector demand but opening source data is difficult due to sensitive information, etc.

3. Establishing a unified system for openness and sharing

☐ Establishment of a Government-wide Data Pipeline

○ (Platform) Establish a "Government-wide Data Pipeline*" system to ensure that data held by agencies is shared seamlessly across ministerial silos and made available to the private sector

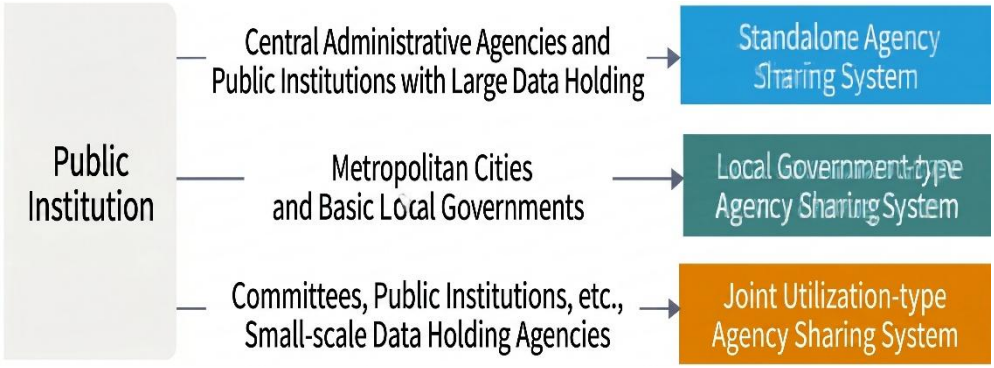
* Agency Source System ↔ Agency Shared Data Management System ↔ National Shared Data Platform ↔ Public Data Portal

- (National Shared) Establish the 'National Shared Data Platform*', which enables the searching, linking, and utilization of vast amounts of data held by the major systems and data platforms of administrative and public agencies

* A foundation for sharing source data necessary for expanding AI training data and developing customized AI services

- (Institutional Sharing) Establish and expand the 'Institutional Sharing System*', which generates and manages shared data for each agency, and link shared data to the National Shared Data Platform

* A system that collects and processes an agency's source data to generate and manage shared data for inter-agency sharing and utilization; the Ministry of the Interior and Safety has developed a standard system for distribution and expansion



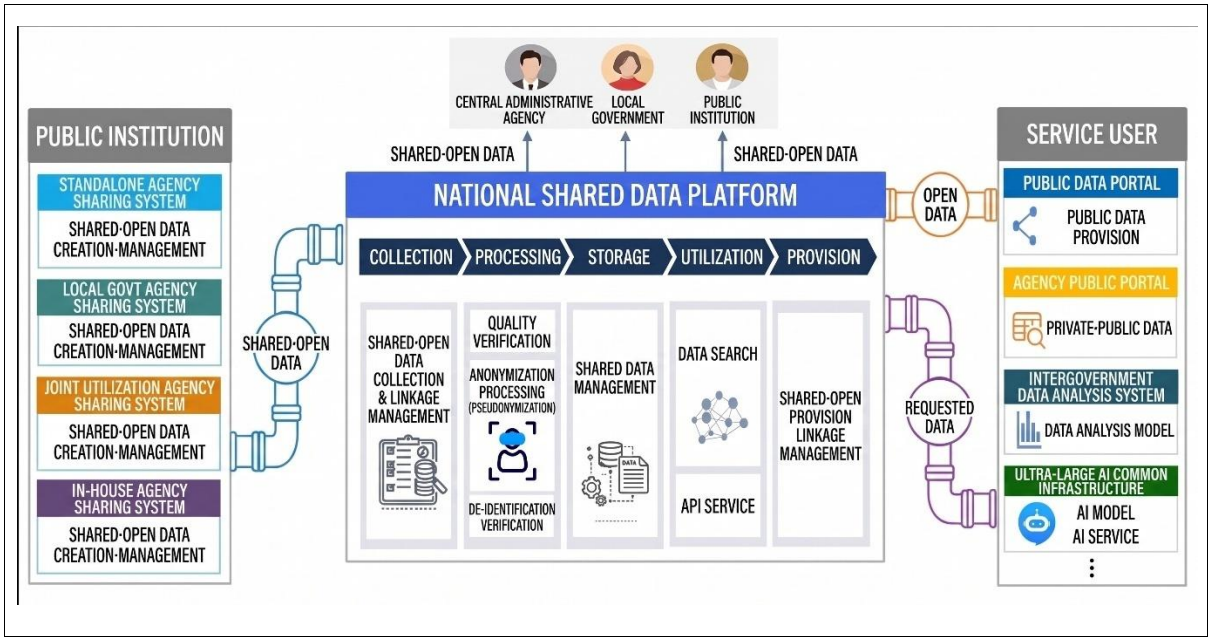
(Standard solution or In-House development)

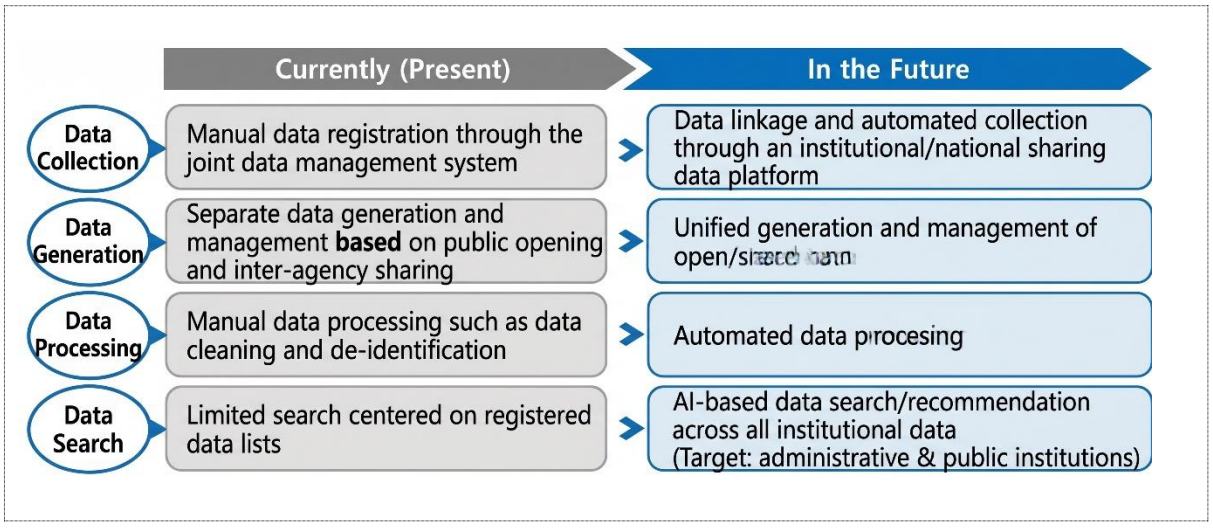
(Development and release in the first half of 2026)

(Development and release in the first half of 2026)

○ (Integration and Utilization) Integration and utilization of all data through public government-wide data analysis systems and a common foundation for super-large AI, as well as private public data portals* integration

* By unifying and integrating the National Shared Data Platform and the Public Data Portal, data from the shared data pool that is eligible for release will be automatically registered on the Public Data Portal





2 [Promoting Usage] Creating a User-Centric Environment for Businesses and Citizens

◇ We will create a tailored data acquisition and utilization environment where companies can easily find public data at any time and receive support for processing and utilizing it.

Voices from the Field	Areas for improvement
□ There is a need for practical support and consulting to help companies find and utilize the data they need □ Similar data is scattered across multiple portals, and applying for access requires separate negotiations with each agency, which is inconvenient ※ Focus Group Interview (FGI) with companies utilizing public data and AI (Oct. 2025)	□ Customized support ranging from acquiring the public data businesses need to processing it and consulting on utilization strategies □ Easy and convenient access to all open data from various agencies in one place on the Public Data Portal

1. Strengthening Support for Companies Utilizing AI and Public Data

□ Support for Problem-Solving by Companies Utilizing Public Data

○ (Support Center) Operation of the ‘Public Data Problem-Solving Support Center*’ to match and support public data holding agencies with companies that have a continuous demand for public data

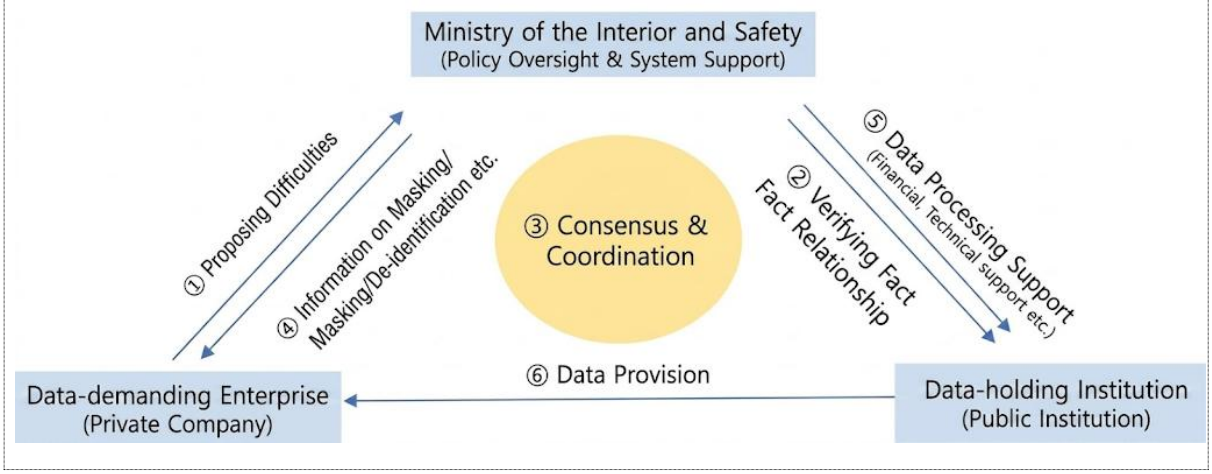
* Established within the Public Intelligence Data Division of the National Information Society Agency (NIA) (July 2025)

- (Demand Assessment) Establish and operate a permanent online submission portal within the Public Data Portal to gather feedback on corporate demand for public data

- (Problem Solving) Support for resolving difficulties*—such as negotiating the scope and method of data release with holding agencies, and processing and releasing data—through matching and collaboration between demand-side (companies) and supply-side (holding agencies, data processing companies)

* Data not released due to agency non-cooperation despite repeated requests, inclusion of personal information, or lack of budget for processing and release

<Framework for Supporting Public Data Problem-Solving>



☐ Tailored Growth Support for Companies Utilizing Public Data

○ (Data Processing) Support for processing data for AI training through the creation of a pool of domestic data processing specialists and matching with companies requiring tailored support

※ Support for matching with specialized companies that meet AI utilization objectives, including data cleansing, standardization, and quality verification

○ (AI Company Support) Providing tailored consulting for AI companies on AI service planning, data integration and utilization strategies, technical support, and commercialization

○ (Growth Support) Through collaboration with companies and institutions possessing sector-specific expertise, connect prospective entrepreneurs and startup companies* to support programs and assist with participation

* Entrepreneurs (Companies): Individuals who have started a business within the past 7 years from the date of business commencement (Article 2 of the 「Small and Medium Enterprise Start-up Act」)

No.	Program (Partner Organization)	Details of Support
1	Technology Gurantee Support (Korea Technology Finance Corporation)	• Support for securing startup capital through guarantee, technology evaluation, and certification systems
2	Credit Guarantee Suppport (Korea Credit Gurantee Fund)	• Support for securing startup capital through credit guarantee systems
3	Employment Support (Korea Student Aid Foundation)	• Operation of a work-study student support program for companies utilizing public data* * Companies with 5 or more employees, including the CEO, that have been in business for at least one year
4	Overseas Expansion Consulting (Korea Trade Insurance Corporation)	• (Trade-sure Consulting) Provides consulting services to companies requesting advice on overseas expansion, legal matters, finance, etc. • (Free credit checks for importers) Up to 5 free checks per year
5	Intellectual Property Application Support (Korea Invention Promotion Association)	• Linked support for the “Excellent Idea Intellectual Property Conversion Service” (IP Stepping Stone*) * Limited to prospective entrepreneurs or those who started a business on or after January 1 of the current year
6	AI and Cloud Support (Microsoft Korea)	• Support for data utilization, analysis, visualization, and AI (MS MPP*) training* * MPP(Microsoft Professional Program): AI expert training for beginners to intermediate and above • Discounts on Cloud service (Azure) usage fees
7	Credit Inquiry and Evaluation (Korea Evaluation Data)	• Support for mobile services providing corporate and credit information, as well as technology credit evaluation services
8	Data Utilization Support (KB Kookmin Card)	• KB Kookmin Card data utilization support (including access to the Data Open Lab analysis space)

9	Customized Consulting (Korea Accelerator Association)	•Provision of 1:1 customized business growth consulting on a 3–6 month basis (business plan preparation, pitching, and digital marketing, etc.)
10	Credit Inquiry, Evaluation, and Promotion (NICE Credit Information)	•Discount support for corporate credit information inquiry services •Discount support for fees related to credit evaluation, technology evaluation, and investment evaluation for public institutions •Provision of an Excellent Enterprise Certification and promotional opportunities on the startup data platform "Forest of Innovation"
11	Medical Data Support (Health Insurance Review and Assessment Service)	•Access to the HIRA Big Data Analysis Center •Provision of medical imaging verification lab data •Support for utilizing customized research and analysis (industry) services

2. Creating an environment where public data can be accessed easily and conveniently in one place

☐ Integrated Provision of Public Data Through Inter-Portal Linkage

○ (Inter-Portal Linkage) Integrate data scattered across individual agency portals into the Public Data Portal to provide a unified service that allows users to conveniently access data in one place

- (Existing Portals) Through the Public Data Portal Enhancement Project, automatically link the public data catalogs of 125 individual portals via API

※ Link public data to One Window (data1window.kr, Ministry of Science and ICT) to enable viewing of both private and public data in one place

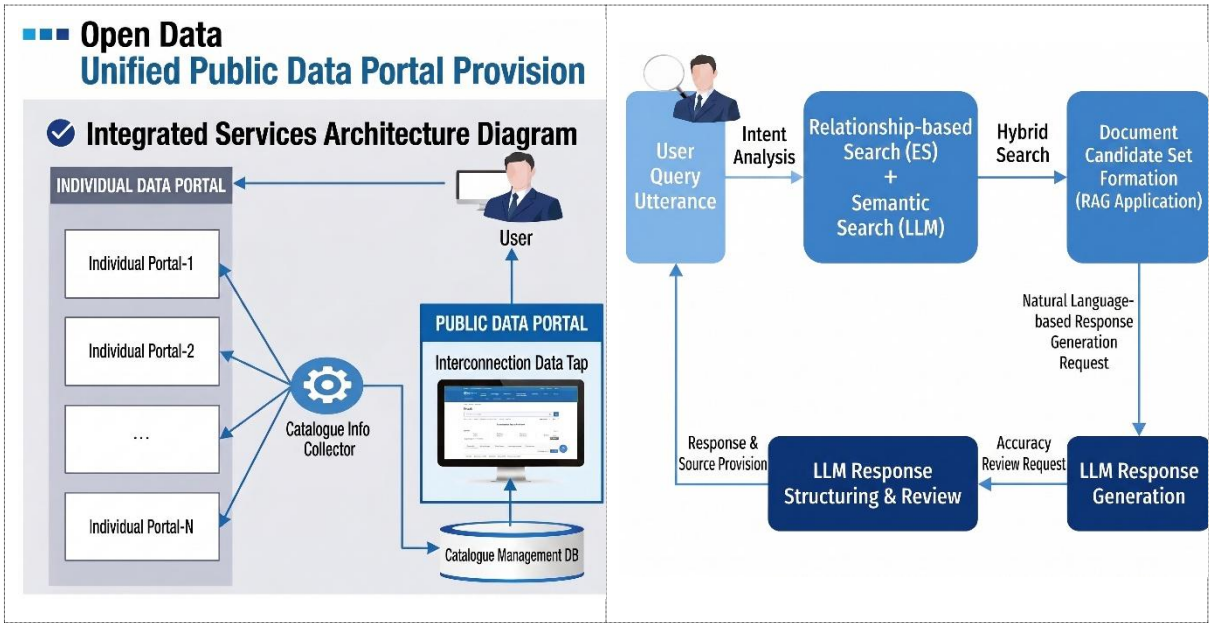
- (New Portals) Promote pre-consultation and inspection for IT projects to ensure that new or upgraded individual portals can be automatically linked to the Public Data Portal

☐ Implementing an environment that intelligently provides the data users want

○ (Data Search) By introducing an LLM-based search service, natural language queries are analyzed to conveniently provide the exact data users want

※ Implementation of generative AI-based search and recommendation services through the Public Data Portal Enhancement Project

○ (Data Recommendations) Recommended data is provided based not only on the data the user searched for but also on past search history, usage patterns, and correlations between data



※ When searching for "Seoul tourism information," recommend a package of related data such as "Current Status of Seoul Tourist Attractions," "Seoul Public Transportation Information," "Current Status of Seoul Accommodations," and "Restaurant Recommendations"

Reference

Trends in Public Data Portals in Major Countries

Opening High-Value Datasets (HVDs) and Laying the Groundwork for AI Utilization

Nation	Key Points
EU	• In accordance with the establishment of the “Open Data Directive,” drive the selection of “High-Value Datasets (HVDs)” with significant socio-economic impact and mandate their provision via APIs
UK	• Provide practical guidelines such as the “Generative AI Framework” and “AI Playbook” • Mandate the “Algorithm Transparency Standards (ATRS)” for AI-based decision-making
FRANCE	• Designated a portal as the single official point of access for High-Value Datasets (HVDs) • Approximately 85% of High-Value Datasets have been made publicly available (as of 2024)

Enhancing Portal Features to Maximize User Experience

Nation	Key Points
EU	• Provides a “Metadata Quality Dashboard” to assess compliance with data standards within the portal • Provides a “Licensing Assistant” tool to verify data licenses
SINGAPORE	• Introduces a “Write API” to automatically publish data to the portal, enhancing data timeliness • Provides user-friendly features such as integration with “Google Colab”
FRANCE	• Operate ‘Topic-Specific Portals’ by specializing data by subject • Provide the ‘Real Estate Transaction Price Explorer’ to view real estate transaction histories on a map

Data Standardization and Enhanced Interoperability

Nation	Key Points
 USA	• Application of the “FAIR principles” and the “DCAT 1.1 international standard for datasets” to ensure the reliable opening and linking of data • Promoting the transition to ‘DCAT 3.0’ to enhance the interoperability of federated data
 UK	• Building ‘Linked Data’ and Semantic Web infrastructure by applying web standard technologies such as RDF and SPARQL, along with the ‘5-Star Open Data’ model
 JAPAN	• Providing guidelines for the ‘Government Interoperability Framework (GIF)’ • ‘Standardizing metadata description languages based on the ‘DCAT-JP’ standard

Data Governance and Quality Management Centered on a Dedicated Team

Nation	Key Points
 USA	• Development of the draft “DCAT-US 3.0” standard and management of data standards, led by the U.S. CDO Council (Chief Data Officer Council)
 GERMANY	• Oversight of portal operations and technology development, led by the “IT Planning Council (IT-Planungsrat)” under the Federal Ministry for Digital and National Modernization (BMDS) • Automatic assessment of data standards using the “Automated Metadata Validator”
 UK	• Defining the data quality framework and overseeing portal operations, led by the “Government Digital Center” under the Ministry of Science, Innovation, and Technology (DSIT)

3. Strengthening Public-Private Cooperation Systems and Partnerships Based on Public Data

☐ Establishing a Public-Private Partnership Governance Framework for the Open Access to Court Rulings

○ (Implementation Framework) Establishment of a public-private open consultation body comprising the Ministry of the Interior and Safety, the Ministry of Justice, the courts, the National Assembly, and legal tech companies

< Proposal for the Formation of a Task Force on the Public Release of Court Rulings >

- (Ministry of the Interior and Safety) Support* for the release of court rulings in machine-readable formats and via open API

  *National Priority Data Release Project (administrative, financial, and technical support available)

- (Ministry of Justice) Amend relevant laws and regulations, including the Criminal Procedure Act, Civil Procedure Act, Administrative Litigation Act, Family Relations Act, Patent Act, and Public Official Election Act
- (Office of Legislation and Legal Affairs) Promote the provision of linked legal provisions and case law through the National Law Information Center
- (Courts) Amend relevant provisions within Supreme Court Rules (e.g., Rules on the Inspection and Copying of Criminal Judgments) regarding non-anonymization, fee imposition, and file formats for provision
- (Personal Information Protection Commission) Pursue the relaxation of regulations on the collection, use, and provision of personal information under the Personal Information Protection Act
- (Ministry of Economy and Finance) Provide new budget support to replace court fee revenue and facilitate early de-identification processing
- (National Assembly) Review and propose relevant bills to expand the public access to court judgments
- (Private Sector) Demand for open access to court judgments from legal tech companies and academia

○ (Seeking Social Consensus) Building public consensus through initiatives such as the Open Data Forum (ODF) hosting open seminars and publishing opinion pieces

☐ Strengthening Public-Private Partnership Content and Restructuring Systems Based on AI and Corporate Needs

○ (Content Enhancement) Encourage companies to actively utilize AI technology when conducting citizen-participation living lab projects aimed at resolving social issues

※ Provide benefits such as priority selection when applying for the “Tailored Support for Companies Utilizing Public Data” program

○ (System Restructuring) Integrate the operation of similar projects to enhance the efficiency of public-private partnership initiatives and generate tangible results

※ Select social issues currently facing agencies as Living Lab projects and solicit participants

☐ **Strengthening Networking for Companies Utilizing Public Data**

○ (Sharing Achievements) Bring together past winners of startup competitions to create a forum for exchange and cooperation aimed at utilizing public data, developing services, and discovering new business ideas

※ Conduct follow-up surveys of winning companies to assess business performance and identify needs for follow-up support related to public data

○ (Growth Support) Identify challenges in utilizing public data, organize business meetings with industry stakeholders, invite relevant experts, and facilitate connections with data-holding agencies and venture capital firms

3 [Quality Control] AI-Friendly Data Quality Management

◇ We will ensure the quality of all data with AI service development in mind and strengthen standards and quality control throughout the entire data lifecycle.

Voices from the Field	Areas for Improvement
□ Data structures, variable names, and release cycles vary by agency, making it difficult to integrate, link, and utilize data □ There is a need to establish a mid- to long-term data quality management system to improve quality with a focus on “usability” and “accuracy” ※ Focus Group Interview (FGI) with public data and AI-utilizing companies (Oct. 2025)	□ Improve data quality to a format suitable for use in AI training, data analysis, and service implementation □ Conduct usability assessments on all open data

1 Transition to AI-Ready Public Data Quality Management

☐ Establishing an AI-Ready Public Data Management System

○ (AI-Ready) Improving the management system to ensure public data is maintained in an AI-friendly format (AI-Ready) so that AI can easily learn from and utilize it immediately

<Concept and Standards for AI-Ready Public Data>

▪ Concept : Public data that has been refined and processed to facilitate its use by AI for tasks such as learning, analysis, and inference

▪ Criteria : Public data that meets appropriate format and quality standards to facilitate AI search and utilization

① (Search) Provides format and attribute information that AI can easily read and understand

- (Format) Provided in open formats (such as CSV, JSON) that are easy for AI to utilize

- (Attribute Information) Provide rich descriptions in metadata—such as update cycles, change history, and summary statistics—that AI can reference when utilizing the data

② (Utilization) Ensure appropriate quality and standardization to facilitate AI use for training and other purposes

- (Quality) Enhance data integrity by minimizing missing values, errors, and inconsistencies; meet quantitative and qualitative requirements—such as accuracy, consistency, and timeliness—necessary for widespread use in AI training

- (Standard Codes) Standardize identifier values (e.g., addresses, administrative codes) to facilitate data linkage

- (Management System) Development of guidelines* for managing ‘AI-Ready Public Data,’ creation and revision of standard guidelines, and establishment of an evaluation model

* Includes detailed criteria by category (format, attribute information, quality, standard codes), implementation plans for agencies, and specific examples of AI-Ready management by data type

- (Pilot Implementation) AI-Ready standards will be applied and expanded, starting with data that has high AI utility, such as AI and high-value public data, national priority data, and standard-compliant data

- (System Development) Each agency will develop a system to perform quality diagnostics and improvements, format conversion, and the subsequent opening and sharing of public data in accordance with the ‘AI-Ready’ standards

※ Build quality assessment and improvement functions into the government-wide data pipeline (agency sharing → national sharing → Public Data Portal)

☐ **Enhancing the Effectiveness of Public Data DB Quality and Standard Management**

○ (Quality Management) Shift from a gradual expansion of evaluated databases to mandatory selection of critical databases, such as those for open and utilized data, thereby enhancing the effectiveness of quality management

- (Diagnostic Evaluation) Revise the database selection method* to ensure that all databases within each agency undergo quality diagnostic evaluation on a 1–3 year cycle

* Current: Selected at the discretion of each agency ⟶ Future: Mandatory selection of critical databases and prioritization of undiagnosed databases

- (Quality Certification) Conduct reviews of critical databases that are open and utilized, such as National Priority Data, and grant quality certification (on a per-database basis)

* (Scope) Current: Selected at the discretion of each agency from all databases ⟶ Future: Limited to critical databases such as open and utilized data

○ (Standardization) Analyze column names and domains in open data to identify common standard terminology suitable for AI utilization, and expand and establish provision standards that include rich metadata

* Current standardization for system construction and integration ⟶ Future standardization for AI training and open data utilization

☐ **Assessment of Open Data Usability and Quality Improvement**

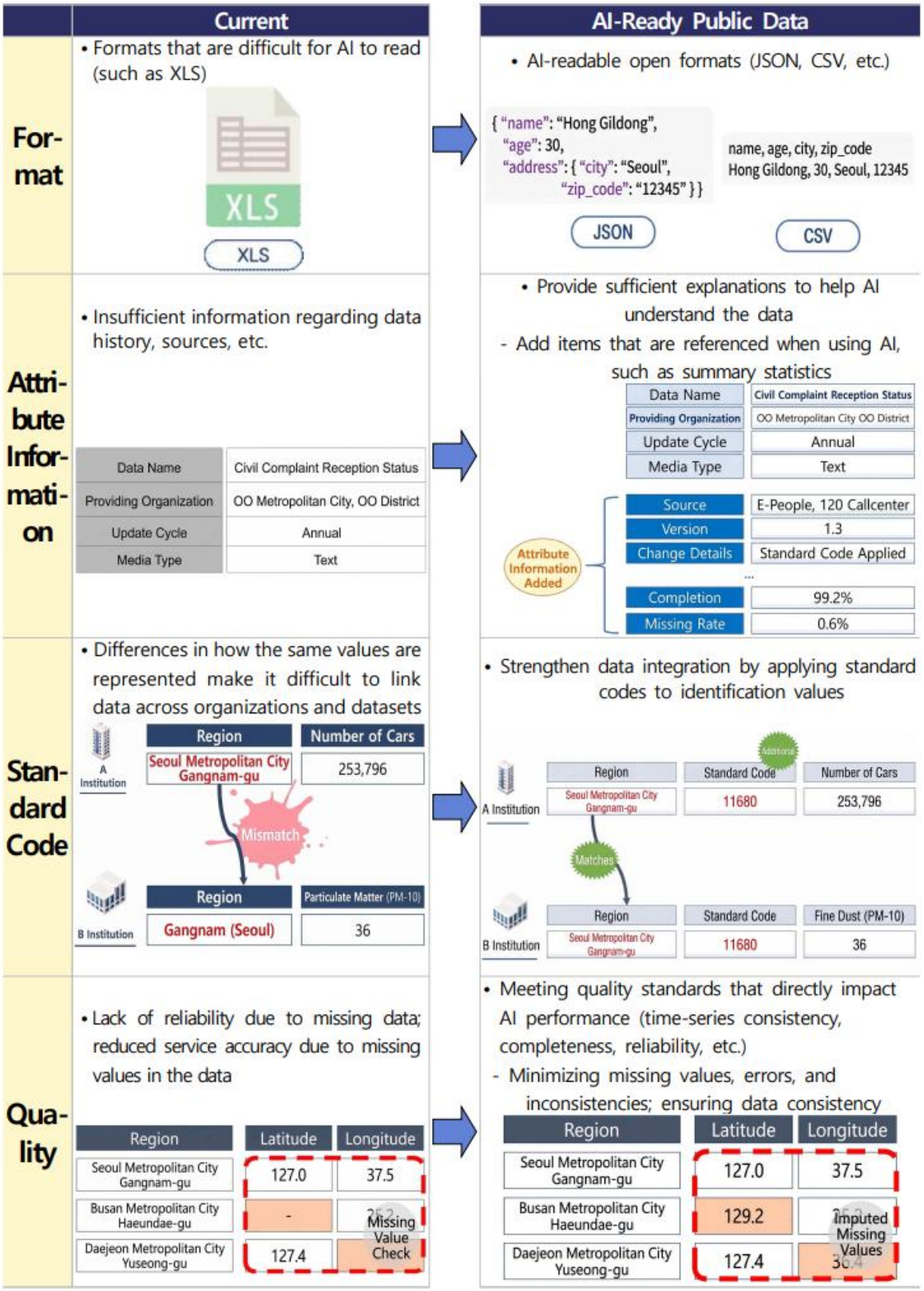
○ (Mandatory Assessment) Require an ‘Usability Assessment*’ for all data on the Public Data Portal and improve functionality to measure the level of data with high AI usability

* Improve and apply usability assessment indicators based on ‘AI-Ready’ standards to the Public Data Portal (starting in 2026)

○ (Enhanced AI Explanations) Expand data explanations using LLMs to help users better understand open data and utilize it appropriately

Reference

AI-Ready Public Data



2. Improving the method of opening public data to make it readable by AI

☐ **Full Conversion of Public Data to Open Formats and Release of AI-Compatible Data**

○ (Full Conversion to Open Formats) Non-open format data will be fully converted to open formats*

* Open formats: Machine-readable data formats that can be processed by at least one software application without cost or other usage restrictions (e.g., CSV, JSON, XML, ODT, etc.)(CSV, JSON, XML, ODT 등)

Type	Step 1	Step 2	Step 3	Step 4	Step 5
Machine-readable format	Non-compliant format (Cannot be registered on the Public Data Portal)	Minimum Sufficient Format	Open Formats		
Example	PDF	HWP, XLS	CSV, JSON, XML	RDF	LOD

- (Public Data Portal) Initiating conversion of non-open-format data (approx. 1.3%)
- (Individual Portals) Conducting a comprehensive survey of non-open-format data (HWP, image-based PDFs, etc.) across 125 individual portals to support conversion to open formats

※ To promote data conversion by agencies, support will be provided for the use of OCR software to convert image-based PDFs

【 Data Scheduled for Conversion to Open Format 】

- [Ministry of Economy and Finance] Status of tax treaty signings, key indicators for the service industry, etc.
- [Anti-Corruption and Civil Rights Commission] Collection of Administrative Appeal Rulings, Manual for Handling Special Public Complaints, Collection of Cases on Mediation and Settlement of Grievances, Collection of Authoritative Interpretations of the Anti-Graft Act, Collection of Best Practices for the Code of Conduct for Public Officials, etc.
- [Korea Consumer Agency] Discount information on daily necessities, price trends by product, etc.

○ (AI-understandable) The “Top 100 AI and High-Value Public Data Sets” and other datasets are released in a AI-friendly* format that facilitates AI training and application and allows for an understanding of the relationships between data points

* Data that meets the criteria of time-series nature, completeness, reliability, and versatility, making it suitable for AI training

<AI-understandable (machine-readable) data >

▪ Rich metadata* and linkage information are provided together, enabling machines to understand and interpret the meaning, context, structure, and relationships of the data so that it can be appropriately utilized in various situations

* Variable definitions, sources, licenses, data dictionaries, structural descriptions, etc.

3. Improved to an open-format document designed for AI integration

☐ Transition to an AI-Friendly Administrative Document System

○ (Regulatory Amendment) Reflected in the Basic Principles for Document Processing under the 「Regulations on the Operation and Innovation of Administrative Affairs」 to take into account the use of AI when processing administrative documents

Amendment to Article 5 (Basic Principles of Document Processing) of the “Regulations on the Operation and Innovation of Administrative Affairs”

② The head of an administrative agency shall process documents in accordance with the criteria set forth in the following subparagraphs in order to enhance the convenience of citizens’ daily lives, systematically manage electronic documents, and utilize them for artificial intelligence.

○ (Onnara System) Phased functional improvements* to allow only open document formats to be attached, along with the ability to attach PDF documents**, to support AI utilization

* Agencies that have configured the system to allow only open document formats to be attached (as of October 2025): 65 (21.5%)

	Total	Central government agencies (75)	Local governments (227)
Organizations that attach only open-access documents	65(21.5%)	62(82.7%)	3(1.3%)

** Rollout following a pilot program involving select central government agencies

○ (Pilot Implementation) Situation reports, regular reports, and other simple documents: Use and promote the standard-format-based document editor of the ’Intelligent Work Management Platform‘ instead of the Hangul document editor

○ (Guidelines) Reflect open document attachment standards in the Digital Government Service Guidelines

Proposed Amendments to the 「E-Government Website Quality Management Guide」	「Digital Government Service UI/UX Guidelines」 Proposed Amendments
□ Web Accessibility ○ Attachments must be provided in at least two file formats, including open formats, to accommodate users in various environments and facilitate AI utilization ※ (Example) Open format*file, PDF file * JSON, XML, ODT, etc.	Usability Guidelines <New> ○ Attachments must be provided in at least two file formats. (Open format file, PDF file)

* Add diagnostic items and content to the criteria for Attachments Designed for AI Use

○ (Criteria) Establish criteria for the format and metadata management items of open documents

- (Format) Prioritize the use of open formats that are widely compatible (e.g., JSON, XML, ODT)

- (Metadata) Establish criteria for management items such as unique ID, title, source, and document type

4 [Law·Governance] Strengthening the Foundation for Promoting the Opening of Public Data

◇ We will address the barriers to data openness and utilization and establish a robust implementation framework to build a leading AI nation based on public data.

Voices from the fields	Areas for Improvement
□ Public data involving issues such as copyright, personal information, and portrait rights leads to reluctance to open data due to the risk of legal liability in the event of problems ※ Excerpt highlighting limitations in the development of sector-specific plans by each agency	□ Comprehensive overhaul of relevant laws and systems and establishment of liability guidelines □ Policy leadership centered on AI and data leaders with expertise and governance, as required in the AI era

1. Improvements to Laws and Regulations to Promote the Opening and Sharing of Public Data

☐ Amendments to Laws and Regulations to Promote Data Openness and Sharing

○ (Public Data Act) Pushing for amendments to promote openness, including the introduction of an assessment of factors hindering the provision and use of public data and the establishment of a legal basis for providing pseudonymized data containing personal information

※ Disclosure of all public data catalogs, excluding non-public information (pushing for amendment of Article 19, starting in 2026)

○ (Data-Driven Administration Act) Pushing for amendments to establish the concept of data sharing, mandate sharing and utilization, and create provisions for pseudonymization of personal information

○ (Liability Exemption Guidelines) Developing practical liability exemption guidelines* so that public data providers and others can provide data without worrying about liability issues

* Includes types of exempted liability, criteria for determining negligence, procedural requirements for applying exemptions, similar domestic court precedents, and examples of proactive administration

☐ Revision of Laws and Regulations to Remove Barriers to Data Access

○ (Personal Information Protection Issues) Revise laws and regulations to allow the use of original personal data for AI technology development purposes when necessary for the public interest or social benefit, and eliminate barriers to data provision and access, such as difficulties with de-identification measures

○ (Licensing) Systemic reforms to address the issue of restricted access caused by agencies holding only a portion of the copyright to public works* under Article 17 of the Public Data Act

* Public Works: Works for which public institutions or similar entities hold all or part of the copyright

- When generating or constructing public data through outsourced services, establish a system to mandate the acquisition of a "Confirmation of License for Copyright and Other Rights" during the contracting process (Public Data Management Guidelines, etc.).

※ Currently, this is only specified in the Public Data Operational Manual, resulting in minimal enforceability

○ (Individual Laws Hindering Data Release) Promoting amendments to individual laws that hinder data release by agencies, such as provisions prohibiting use for purposes other than those specified or restrictions on provision to third parties

※ Led by the Office for Government Policy Coordination (OGPC), plans are underway to form a task force involving ministries responsible for individual laws such as the Framework Act on National Taxation, the Statistics Act, the Personal Information Protection Act, and the Medical Service Act

【 Legal and Institutional Reforms 】

▌[Basic Tax Act] Push for amendments to the relevant laws to revise the purposes of use and disclosure methods for income tax sample data, and to enable the extraction and release of corporate tax sample data as public data (National Tax Service)

▌[Childcare Support Act] Push for amendments to relevant laws to allow agencies responsible for implementing IT projects to retain and utilize data generated during such projects (Ministry of Gender Equality and Family)

▌[Statistics Act] Promote amendments to the Statistics Act regarding the opening and utilization of statistical data, the use of standard classifications for administrative data, and the standardization of microdata and metadata (National Data Agency)

2. Strengthening Inter-agency Cooperation and Data Utilization Capabilities

☐ Strengthening the Cooperation Framework for Data Governance

○ (Establishing a Control Tower) Establish a public data control tower, with the National CAIO (Chief AI Officer) and Ministerial CAIOs (Vice Ministers) serving as the National and Ministerial Chief Data Officers (CDOs)

○ (Strengthening Governance Cooperation) Strengthen the government-wide national data governance cooperation system by consolidating data governance currently managed in a decentralized manner across the private and public sectors and relevant ministries

- Strengthen coordination and mutual cooperation through joint meetings between the "Data Subcommittee" under the National AI Strategy Committee and data-related committees*

* National Data Policy Committee (Ministry of Science and ICT), Public Data Strategy Committee (Ministry of the Interior and Safety), National Statistics Committee (National Data Agency)

☐ Strengthening Public Data Literacy and Capabilities

○ (Strengthening Civil Servant Capabilities) Conducting training to enhance civil servants' capabilities in public data disclosure and management, including AI-Ready public data quality control

❙ [Rural Development Administration] Training to Strengthen AI and Data-Based Research, Analysis, and Application Capabilities (Online or In-Person)

❙ [Korea Coast Guard] Training of AI Facilitators to Expand Field-Centered AI and Data-Driven Operations

❙ [Korea Regional Information Development Institute] Public Data Training on AI Utilization by Level, Applicable to Regional Digital Administrative Environments

❙ [Korea Housing Finance Corporation] Establishment of a Comprehensive ICT Training Plan to Cultivate AI and Data Analysis Specialists

○ (Strengthening Support for Small and Medium-Sized Enterprises) Enhancing consulting services to support AI adoption by small and medium-sized enterprises and operating programs to foster specialized talent

❙ [Daejeon Transportation Corporation] Creating local jobs and social value through AI and big data industry-academia collaboration field training

❙ [Korea Institute of Industrial Technology Planning and Evaluation] Providing data, training on data utilization, and consulting support to companies that apply for data access

❙ [Korea Broadcast Advertising Corporation] Operating customized group training sessions and providing guidance on data utilization for companies using AiSAC data

○ (Tailored Education for the Public) Promoting targeted AI and data literacy training for specific groups, such as young people and seniors, by utilizing public education platforms like the Digital Learning Center

- [Seoul Metropolitan City] Supporting digital employment and entrepreneurship for middle-aged and older adults through customized big data training programs (Big Data Campus)
- [Korea Post Financial Development Institute] Training on public data concepts and institutional data utilization for data-vulnerable groups
- [Korea Student Aid Foundation] Hosting briefings on public data concepts, utilization methods, and case studies for young (prospective) entrepreneurs
- [Seocho-gu, Seoul] Promoting the cultivation of practical talent focused on cutting-edge technologies such as AI and big data for young people (AI College)

○ (Ethical Guidelines for AI and Public Data) Developing ethical guidelines for the use of AI and public data in the public sector, and promoting regular training and institutionalization

- [National Information Society Agency] Establishment of internal guidelines to apply ethical principles throughout the entire data lifecycle
- [Korea National Oil Corporation] Establishment of ethics principles centered on private-sector AI utilization to promote awareness of proper AI use
- [Local governments] Establishment of ethical guidelines regarding the development of institutional AI services and data utilization

4. Leading the Way in Public Data and Strengthening Cooperation

☐ Strengthening Global Cooperation, Including Discussions on AI-Ready Public Data Standards

○ (Global Cooperation) Expanding global cooperation through the Organization for Economic Cooperation and Development (OECD), the Open Government Partnership (OGP), the Global Public Administration Cooperation Group, and Digital Nations

- Promoting joint research projects with international organizations and consulting projects for foreign governments
- Promote projects to strengthen the stakeholder participation ecosystem (e.g., ODF) related to the opening and utilization of public data in Korea through cooperation with the Open Government Partnership (OGP)

○ (Standard Leadership) Promote discussions on standards and the globalization of AI-Ready public data opening methods with international organizations (e.g., OECD)

☐ Support for Overseas Expansion of Public Data Utilization Companies

○ (Support for Overseas Expansion) Support the overseas expansion and promotion of domestic companies in the data sector

❙ [Ministry of Science and ICT] Support for data companies to secure global competitiveness and expand overseas (DATA-Global)

❙ [Ministry of SMEs and Startups] Support for promising AI startups to expand globally (Vertical AI Startup Challenge)

❙ [Gyeonggi Province] Provides opportunities for dispatch to Silicon Valley, participation in trade shows, and investor relations (IR) (AI Global Challenge Program)

☐ Government-wide Response to the OECD Open Data Assessment

❙ [Purpose] To assess the current status, achievements, and areas for improvement of public data policies by country (biennial cycle)

❙ [Scope] OECD member countries and candidate countries (Brazil, Croatia, etc.)

❙ [Evaluation Content] 2 years of public data policies and performance

❙ [Evaluation Categories] 3 Areas (△Data Availability, △Data Accessibility, △Government Support) and High-Value Data Open Access

○ (Opening High-Value Data) Expanding the release of high-value datasets recommended by the OECD—such as declarations of interests by senior public officials and corporate ownership information—and enhancing accessibility

○ (Availability, Accessibility, and Government Support) Improving the ease of access to open data and related policies (open formats), enhancing data portal functionality, and strengthening collaboration with public data stakeholders

[Ⅲ]

The 5th Basic Plan for Public Data

Implementation Schedule

1 Project Timeline

Key Initiatives	Implementation Schedule			
	'25	'26	'27	'28
1. [Openness·Sharing] Opening Access to High-Demand, High-Value AI Data				
○ Selection and release of the 「Top 100 High-Value AI Datasets」	Top 100 selection	Phased release		
○ Expansion of the release of unstructured data suitable for AI training	Phased release and expansion			
○ Promotion of the provision of pseudonymized data containing personal information	Reflection in evaluations	Linkage with related projects		Implementation/ expansion
○ Diversify methods for opening data that is difficult to release as raw data	Establish release methods		Implementation/ Expansion	
○ Establish a government-wide data pipeline	Government-wide expansion		Operation/Improvement	
2. [Promoting Usage] Creating a User-Centric Environment for Businesses and Citizens				
○ Support for problem-solving by companies utilizing public data	Opening of a support center	Operation/Improvement		
○ Tailored growth support for companies utilizing AI and public data	Provision and expansion of consulting and support			
○ Creation of an environment for convenient, one-stop access to public data	Phase 1 implementation	Phase 2 implementation	Operation/Improvement	
○ Strengthening public-private cooperation systems and partnerships for public data	Strengthening public-private cooperation and partnerships			
3. [Quality Control] AI-Friendly Data Quality Management				
○ Establish an AI-Ready Public Data Management System	Develop guidelines		Implementation/Expansion	
○ Enhance the effectiveness of public data database quality and standard management	Implementation/Expansion			
○ Assess the usability of open data and improve its quality	Mandatory	Operation/Improvement		
○ Convert all public data to open formats and release AI-understandable data	Implementation/Expansion			
○ Transition to an AI-friendly administrative document system	Regulatory revisions		System improvements, pilot implementation	
4. [Legal·Governance] Strengthening the Foundation for Promoting the Opening of Public Data				
○ Amendments to laws and regulations to promote data openness and sharing	Amendment of laws		Supplementary amendments	
○ Revision of laws and regulations to remove barriers to data openness	Inter-ministerial consultation		Promotion of amendments to relevant laws and regulations	
○ Strengthening the cooperative framework for data governance	Consultation and amendment		Operation/Improvement	
○ Enhancing public data literacy and capabilities	Implementation/Expansion			
○ Strengthening global leadership and cooperation in public data'		Preparation and response for the 2027 OECD Public Data Assessment		

2 Proposed Budget

☐ Total Budget: 493 billion won

※ (2026) 180.6 billion won → (2027) 165.5 billion won → (2028) 146.9 billion won

☐ Scale by Institution

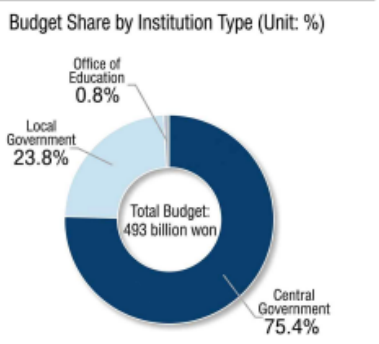
○ (Scale) 83 institutions; the total budget for 2026–2028 is 493 billion won

○ (Share) Central government account for 75.4%, local governments account for 23.8%, and office of education account for 0.8%

<Annual Budget Allocation>

(Unit: 100million won)

Type	2026	2027	2028	Subtotal
Central	1,372	1,241	1,102	**3,715** (75.4%)
Local Government	420	400	353	**1,173** (23.8%)
Office of Education	14	14	14	**42** (0.8%)
Total	**1,806** (36.6%)	**1,655** (33.6%)	**1,469** (29.8%)	**4,930** (100%)



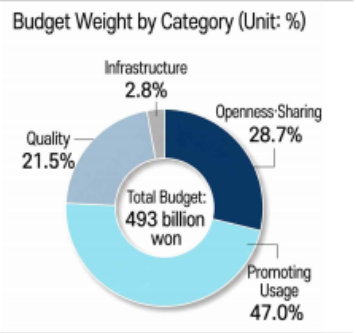
☐ Scale by Category

○ Promoting Usage (47.0%), Openness and Sharing (28.7%), Quality (21.5%), Infrastructure (2.8%) in that order

<Budget allocation by sector>

(Unit: 100million won)

Tyoe	Openness ·Sharing	Promoting Usage	Quality	Infra-structure	Subtotal
Central	1,118	1,566	937	94	**3,715** (75.4%)
Local Government	288	743	96	46	**1,173** (23.8%)
Office of Education	9	9	24	0.09	**42** (0.8%)
Total	**1,415** (28.7%)	**2,318** (47.0%)	**1,057** (21.5%)	**140** (2.8%)	**4,930** (100.0%)



NIA Digital & AI Insights Series

Draft of the 5th (2026-2028) Basic Plan for the Provision and Promotion of Public Data Utilization

Published by

Relevant Korean Ministries, led by the Ministry of the Interior and Safety (MOIS)

Original Authors

Department of Public AI & Data, NIA

English Edition Prepared by

Department of Global ICT Cooperation

Yoon-seok Ko, Vice President

Hyang-nae Jeon, Director

Sora Jeong, Senior Manager

Doyoung Park, Manager

LLM(Large Language Model) technology was partially utilized in the translation process of this report.

NIA NATIONAL INFORMATION SOCIETY AGENCY