**AI Data Ecosystem Development Project**

**◈ Background**

Recently, as artificial intelligence (AI) has rapidly emerged as a central focus in ICT, countries around the world are concentrating their national capabilities on fostering their AI industries. Since the advent of AlphaGo in 2016, the interest in AI has primarily centered around models. Over the years, the global community has strived to overcome the technical challenges of AI through various models based on artificial neural networks, working to develop faster and more efficient models. However, this trend of model-centric AI development has swiftly shifted towards data-centric AI development with the rise of services based on Large Language Models (LLMs), such as chatGPT. At the core of this shift is data. For LLMs to achieve high performance, high-quality, large-scale data is essential. Consequently, leading AI nations are focusing on establishing robust data ecosystems to nurture their AI industries and secure national competitiveness in AI.

Remarkably, given the nature of AI that evolves through continuous learning of new data, establishing a national-level data ecosystem capable of supplying a steady stream of new data is crucial for every country, determining the success of their future AI industry development. Therefore, the Ministry of Science and ICT (MSIT) and the National Information Society Agency (NIA) of the Republic of Korea are launching the "AI Data Ecosystem Development Project" to share Korea's extensive experience and expertise accumulated over decades in building a data ecosystem and to enhance global cooperation in the AI sector.

**◈ Project Overview**

As the importance of data in artificial intelligence (AI) continues to grow, countries around the world are either building or preparing to build data ecosystems for AI. However, many of these countries face two main challenges: first, they do not know which steps to take to establish a data ecosystem, and second, they are unsure of the order to carry out the necessary tasks. Countries facing the first challenge spend so much time determining what needs to be done and planning that they fail to start the actual construction of the ecosystem or miss essential components. Countries facing the second challenge might invest significant budgets and efforts into building a data platform in the wrong order, resulting in an incomplete system.

The project includes status analysis, developing a data classification model and a data platform, and training operators and administrators from 2026 to 2029.

To understand the data ecosystem in Korea, please refer to the World Bank publication "Enabling Data-Driven Innovation—Learning from Korea's Data Policies and Practices for Harnessing AI."

**◈ Project Objective**

This project aims to address these challenges by sharing Korea's experience and expertise in building a data ecosystem comprising a Public Data Platform, Big Data Platform, and AI Data Platform. The goal is to collaborate with partner countries aspiring to develop their AI industries by establishing a suitable data platform based on Korea's proven methodologies. Through this project, we will assist partner countries in effectively and efficiently building their data ecosystems to foster their AI industries.

**◈ Project Description**

**1. Status Analysis & Readiness Diagnosis**

The project will analyze the overall data-related status of the partner countries and assess their readiness to establish the most suitable data platform. To achieve this, we will:

1. Conduct a thorough analysis of the partner countries' general ICT status (e.g., broadband penetration rate, internet speed, number of internet users) and data-related conditions (e.g., market size, enterprises, legal frameworks, organizational structures, public-private data platforms).
2. Based on the results of this analysis and referencing Korea's experience, we will determine the appropriate type of platform to be established in the partner countries.
3. Diagnose the partner countries' readiness for each platform element (e.g., governance, laws and regulations, standards, systems, etc.).

**2. Development of Data Classification Model**

Drawing on the data classification model that underpins Korea's data ecosystem, we will develop a suitable data classification model for the partner countries. This model will consider the function and purpose of the data, including the data producers and managers (public, private), sectors (agriculture, healthcare, education, etc.), and data formats (structured, unstructured, semi-structured). Additionally, we will design the metadata standard considering global standards such as DCAT and Schema.org to facilitate future data exchange and integration across regions, countries, and systems.

**3. Development of A Data Platform Desired by the Partner Country**

Considering the ICT environment of the partner countries, we will design and develop a data platform tailored to their needs. To ensure future scalability and interoperability, the platform will be designed on a cloud-based architecture. We will develop a "modular data platform" that allows for easy separation and combination of functions (or services), facilitating flexible adaptation and growth.

**4. Establishment & Operation of the Project Governance**

To successfully establish a national data ecosystem, the involvement of top decision-makers is paramount. Therefore, we will form a "Data Platform Development Project Governance" centered around the top decision-makers participating in this project. This governance body will include data experts from industry, academia, research, and government sectors in the partner countries. Additionally, we will develop detailed operational plans for this governance structure to ensure effective implementation and oversight.

**5. Education & Training**

After establishing the data platform in the partner countries, we will conduct education and training programs for both users and platform administrators. These programs will be customized based on the roles, levels, and methods (online, offline, hybrid) suitable for the users and administrators. We will design tailored education programs, develop training materials, and deliver personalized education and training sessions to ensure effective utilization and management of the platform.

**◈ Application Process**

**Submit the Project Concept Paper and the official letter with an authorized signature and seal to the Embassy of the Republic of Korea in your country through the Ministry in charge of ODA in your country** **by October 31st, 2024**. The MSIT and NIA will review the submitted proposals and conduct feasibility studies in potential countries in 2025. Based on the results of the feasibility studies and the contents of the Project Concept Paper, the Korean Government approved the project in 2026.

1. **PCP Submission (~October 2024)**- Please submit **the PCP and the official letter from your organization** to the **Ministry in charge of ODA** in your country. If you don’t know which ministry is in charge of ODA in your country, please send us an e-mail.
2. **Approval for ODA project (~October 2024)**- Please get approval for the ODA project from the Ministry in charge of ODA and send three documents **(PCP, an official letter from your organization, and the other official letter from the ministry in charge of ODA)** to the Korean Embassy in your country. Also, please send the PCP and two official letters to us([aides@nia.or.kr](mailto:aides@nia.or.kr)) so that we can select the candidate organizations. If you do not submit one of those documents, you will be disadvantaged in the evaluation.
3. **Evaluation & Selection (~November 2024)**- The MSIT and the NIA will evaluate the PCPs from partner countries and select two candidate organizations to implement the projects in 2026.
4. **Pre-feasibility Study (December 2024 ~ February 2025)**- The MSIT and the NIA will conduct a pre-feasibility study based on the contents of the PCP.

For additional information regarding the project, please send an e-mail to aides@nia.or.kr.

**Project Concept Paper**

**for AI Data Ecosystem Development Project**

**Section I. Basic Information**

*※ Read the instructions and examples written in blue carefully and complete the form in detail.   
Before submitting the form, please delete the instructions and examples. Handwriting is not acceptable.*

|  |  |  |
| --- | --- | --- |
| **Name of country** | *Write the name of the country* | |
| **Ministry**  **in charge of ODA** | *Write the name of the ministry*  *(e.g., Ministry of Finance)* | |
| **Ministry**  **in charge of**  **this project** | **Ministry** | *Write the name of the ministry*  *(e.g., Ministry of Digitalization)* |
| **Name** | *Write the name of the person in charge of this project implementation in the ministry* |
| **Department/division** |  |
| **Telephone number** | *Write a reachable number for communication.* |
| **Mobile number** |  |
| **E-mail address** | *Write a reachable address for communication.* |
| **Position** |  |

**Section II. Platform Selection**

*※ Select one data platform that your country hopes to introduce. Please click or check the box that you want. If you would like more detailed information about each platform, please refer to the appendix below. For even more comprehensive information, please refer to the World Bank's publication "Enabling Data-Driven Innovation - Learning from Korea’s Data Policies and Practices for Harnessing AI."*

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Select** | **Platform** | **Details** |
| **1** |  | AI | Replacing traditional paper notifications sent by mail with smartphone-based electronic notifications to significantly reduce administrative costs and achieve a paperless administration |
| **2** |  | Big Data | Real-time monitoring of fine dust emissions and status, and establishment of a data-driven fine dust forecasting. |
| **3** |  | Open Data | Establish a national health information network by implementing and connecting Electronic Medical Record systems across hospitals nationwide, and provide AI-based medical services utilizing the accumulated data. |

**Section III. Environment Analysis**

**1. Describe your country’s data industry status.**

|  |
| --- |
| *○ Provide a detailed explanation of the data market trends over the past five years, the current status of data companies, the state of the data workforce, and the existing data platforms in operation. This section is critical as it will serve as foundational information to determine which type of data platform is most urgently needed in your country. Therefore, it is essential to include as many specific figures and statistics as possible.* |

**2. Describe your country’s policy priorities in data.**

|  |
| --- |
| *○ Provide a detailed description of the national data-related policies, including data strategies, master plans, and implementation plans. Given the nature of data, strong government commitment is essential. Therefore, it is important to include information that this project is part of the ongoing or planned government data policies, ensuring it is not a one-time effort but a sustained initiative backed by the government.* |

**3.** **Describe the data governance in your country.**

|  |
| --- |
| *○ Provide a detailed description of all institutions that constitute national data governance, including national-level governance, a coordination body for supporting national-level governance operations, ministry-level governance, a supporting agency for policy formulation and technical expertise in data, and a dispute resolution body for resolving data-related conflicts.* |

**4. Describe data-related laws and regulations.**

|  |
| --- |
| *○ Explain all laws and regulations related to the openness, utilization, and protection of data, such as the Act on Open Government Data, the Act on Promotion of Big Data Utilization, and the Data Protection Act.* |

**5. Describe data-related standards.**

|  |
| --- |
| *○ Describe the national standards governing metadata standards, data sharing standards, open data standards, and other standards related to data set at the national level.* |

**6. Describe data-related systems.**

|  |
| --- |
| *○ Describe data-related systems operated by both public and private sectors. For example, provide a detailed explanation if there is an operational open government data portal. Additionally, if there are services actively utilized by the private sector, please describe them in detail.* |

**Section IⅤ. Project Proposal**

|  |  |
| --- | --- |
| **Objective** | *○ Explain all laws and regulations related to the openness, utilization, and protection of data, such as the Act on Open Government Data, the Act on Promotion of Big Data, and the Data Protection Act.* |
| **Specific Data Platform desired to developed** | *○ Provide a detailed explanation on what platforms will be built for various types of data (public, private), data formats (structured, unstructured, semi-structured), and data domains (agriculture, healthcare, transportation, telecommunications, etc.). It is recommended to provide a detailed description of the current status regarding data in these areas.* |
| **Project Governance** | *○ Given that the strong commitment of the partner country is the most critical factor in successfully establishing a data ecosystem, it is necessary to establish detailed plans for structuring and operating robust project governance that includes heads of the agencies responsible for this project in the partner country, along with participation from experts in academia, industry, and research institutes. This component will play a crucial role in the selection of partner countries in the future.* |
| **Operational Maintenance Plan** | *○ Provide a detailed explanation of how the platform will be operated after its construction.* |
| **Other Helpful Information Related to This Project** | *○ Provide an explanation of other aspects related to this project.* |

**Section Ⅴ. Expected Results**

**1. Expected Outputs**

|  |
| --- |
| *○ Present and explain measurable indicators, such as the number of platform users and the amount of data downloaded, that can demonstrate the direct effects achievable through the implementation of this project.* |

**2. Expected Outcomes**

|  |
| --- |
| *○ Describe the ultimate outcomes achievable through the implementation of this project. For example, explain the national-level effects of building a data ecosystem, such as improvement in the OECD OURDATA Index ranking.* |

**3. Expected Outcomes for Direct/Indirect Beneficiary**

|  |
| --- |
| *○ Describe the project’s anticipated target groups and beneficiaries. Specify who are the main users and their approximate number.*  *○ Describe the expected outcomes and benefits that will be achieved through this Project.* *(e.g., 5% increase in employment rate due to improvement of utilization ability of AI/Data of employees/founders)* |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **This report is prepared by**   |  |  |  |  | | --- | --- | --- | --- | | **Name** |  | **Position** |  | | **Ministry or organization** |  | **Telephone number** |  | | **Department/division** |  | **Mobile number** |  | | **Fax number** |  | **E-mail** |  |   **I confirm that I thoroughly read and understand the project overview, and the information given in this concept paper is true, complete, and accurate.**  *(Signature)*  *(Name of the signee)*  *DD/MM/YYYY* |

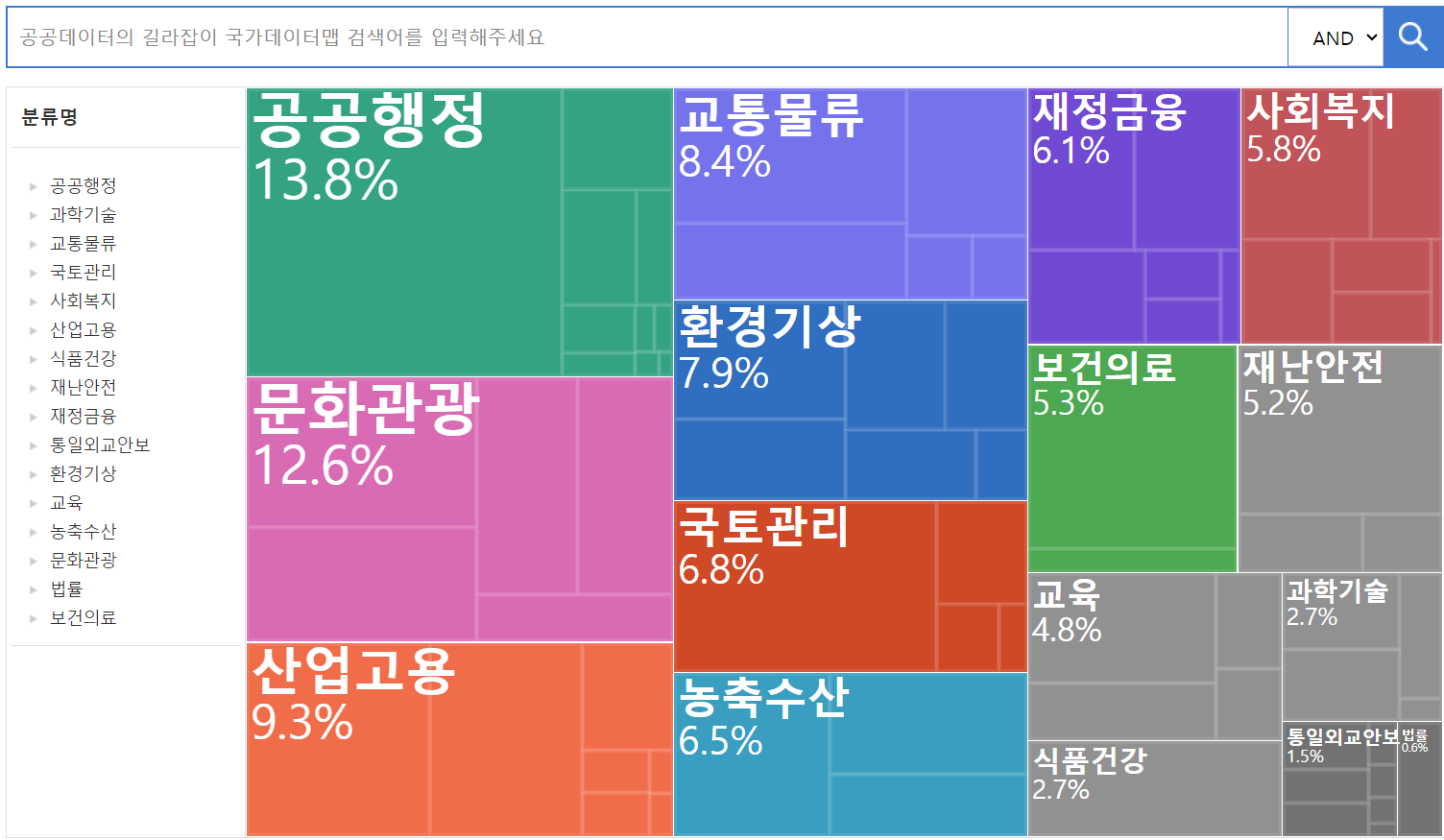
|  |
| --- |
| *(Signature)*  *(Name of the person representing the ministry)*  *(Position)*  *(Name of the* ***ministry*** *in charge of the IAC)*  *DD/MM/YYYY* |

**1. Korean Government Open Data Platform (data.go.kr)**

**◈ Platform Overview**

The Government Open Data Platform, launched in 2011, was established to facilitate the public release of data produced and managed by public entities, including central and local governments, as well as public institutions.

< Example of the Government Open Data Platform Interface >

**◈ Data Availability and Utilization**

Over the past four years (2019-2023), the number of datasets available through the Public Data Platform has increased by 27% annually. As of June 2024, a total of 89,792 datasets are available to the public. Additionally, the utilization of data from the Public Data Platform has seen a significant rise, with an annual increase of 47% over the same period. By 2023, over 60 million instances of data utilization were recorded.

**◈ Public Data Availability and Utilization Statistics**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Category** | **2019** | **2020** | **2021** | **2022** | **2023** |
| # of Datasets Available | 33,600 | 55,139 | 67,441 | 77,272 | 87,682 |
| # of Data Utilization | 13.14 million | 20.84 million | 33.34 million | 46.97 million | 61.17 million |

**◈ International Recognition**

The Government Open Data Platform has been recognized as the world's leading public data platform, achieving the top rank in the OECD Public Data Evaluation for four consecutive assessments in 2015, 2017, 2019, and 2023. This consistent performance underscores the platform's commitment to excellence and its pivotal role in advancing public data accessibility and utility on a global scale.

**2. Korean Big Data Platform (bigdata-map.kr)**

**◈ Platform Overview**

The Big Data Platform was established to consolidate and facilitate the sharing and trading of high-quality data produced and managed independently by private entities, thereby laying the foundation for a robust data economy at the national level.

**◈ Sector-Specific Data Platforms**

To cater to the specific characteristics and needs of various sectors such as finance, environment, culture, transportation, healthcare, retail, and telecommunications, 21 distinct sector-specific data platforms have been developed. These platforms consider the unique legal frameworks, operational systems, and other sector-specific attributes.

**◈ Data Supply and Ecosystem**

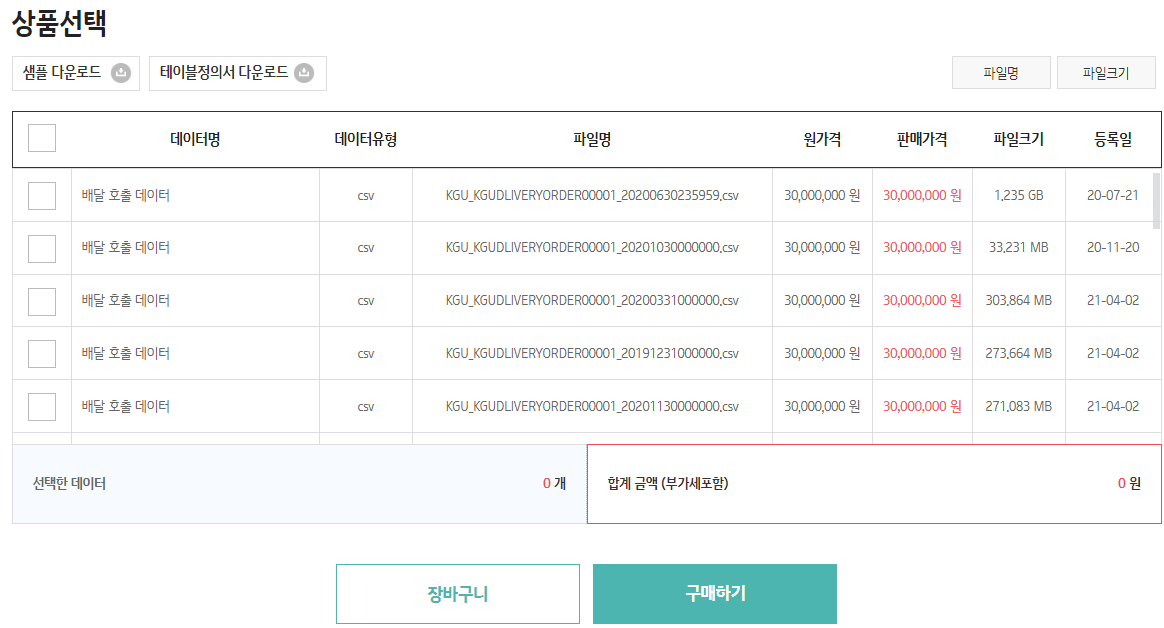
To ensure a continuous supply of data, which is crucial for a thriving data ecosystem, each platform operates under the umbrella of 230 data supply centers. These centers consistently provide new data to the platforms shown as the table below.

|  |  |  |  |
| --- | --- | --- | --- |
| Domain | Organization | Domain | Organization |
| Finance | Nota, Nielsen, Vibe | Culture | Yanolja, RedTable, RedTie |
| Environment | GreenESCO, Novacos, Eco&Partners | Telco | BC Card, Openmate |

**◈ Data Trading and Services**

Each of the 21 Big Data Platforms focuses on facilitating data transactions (sales and purchases) between companies and institutions, aiming to stimulate the data economy.

< Example of the Big Data Platform Interface >



**◈ Data Trading and Services**

Beyond data trading, the Big Data Platforms offer a range of data-related services including: Basic and advanced data analysis, Data visualization, Showcasing exemplary use cases of data utilization.

**3. Korean AI Data Platform (aihub.or.kr)**

**◈ Platform Overview**

The AI Data Platform is designed to be a central hub (AI Hub) for the construction and public dissemination of AI training data, which is a crucial resource in the AI era.

**◈ AI Training Data Construction**

The AI training data includes source data in text, video, image, and audio formats. This data is transformed into labeled datasets that models can accurately recognize and learn from.

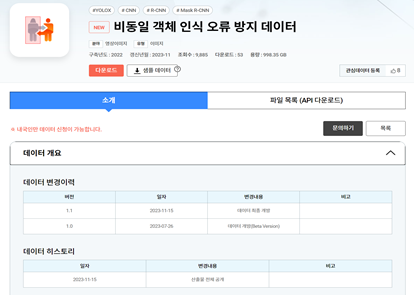
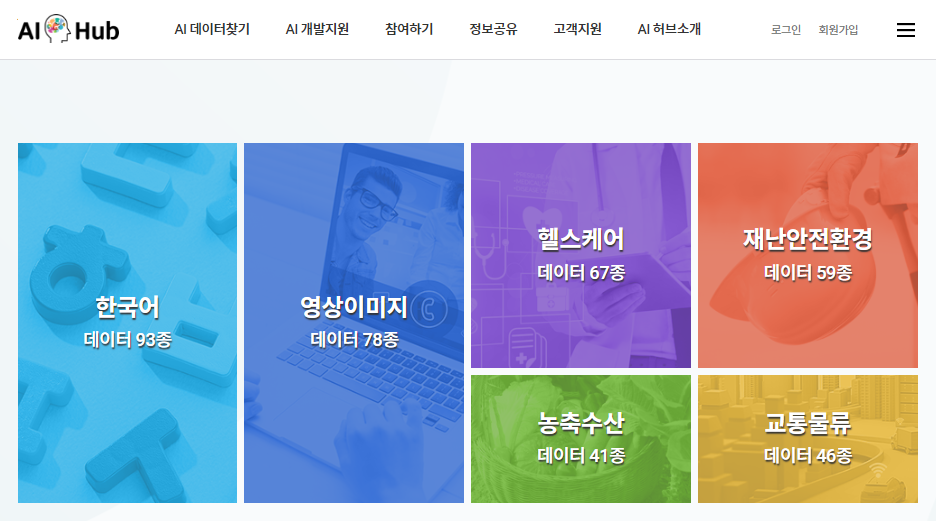
< Example of Different Types of Labeled Data >

|  |  |  |
| --- | --- | --- |
| **Audio/Text**  **(Korean Speech Audio)** | **Image** | **Video**  **(CCTV Anomaly Detection)** |
|  |  | 텍스트, 하늘, 길, 도로이(가) 표시된 사진  자동 생성된 설명 |

**◈ AI Data Portal (AI Hub)**

The AI Data Portal (AI Hub) has been developed to allow citizens and companies to easily search, download, and utilize the AI training data. This portal is organized based on a systematic data classification framework.

< Example of the Big Data Platform Interface >



**◈ Current Status**

Currently, the AI Data Portal offers 706 AI training datasets. On average, around 250 datasets are downloaded daily for learning purposes.