Asian Journal of Innovation and Policy (2024) 13.3:329-349
DOI: https://doi.org/10.7545/ajip.2024.13.3.329

A Systematic Review on Public Data Sovereign Cloud Application for Large-Scale AI Utilization in the Public Sector

Chang-Hee Yun[*], Seok Yoo[**], Cheonsoo Yoo[***]

Abstract As interest in generative AI has rapidly increased in recent years, there is a growing willingness to adopt large-scale AI in the public sector. However, precise implementation methods remain elusive due to various issues related to the introduction of large-scale AI. This study proposes a public data sovereign cloud to address two main challenges when introducing large-scale AI in the public sector: 1) compliance with information disclosure laws and regulations and 2) operation of closed networks. To this end, we examine the concept of a sovereign cloud and propose a definition and implementation strategy for a public data sovereign cloud. Specifically, we present application strategies focusing on construction methods using public-private partnership (PPP) models, network segregation methods according to data grades, and enhancement of the role of Managed Service Providers (MSPs). The results of this study are expected to contribute to policy formulation for introducing and utilizing large-scale AI in the public sector.

Keywords: Large-Scale AI, Public Data, Sovereign Cloud, Public-Private Partnership(PPP), Network Segregation, Managed Service Provider(MSP)

I. Introduction

With the recent rapid development of generative AI, there is growing interest in adopting large-scale AI in the public sector. The emergence of large language models like ChatGPT has demonstrated the potential for AI utilization across various fields and is presenting new paradigms for public services. While increasing interest in adopting large-scale AI in the public sector, finding clear

Submitted, November 25, 2024; 1st Revised, January 2, 2025; Accepted, February 24, 2025
* Director of AI Policy Research Team of National Information Society Agency, Daegu, Korea; yunch@nia.or.kr
** General Manager, Head of AI development Division, Unidocs Inc., Seoul, South Korea; tobewiseys@gmail.com
*** Vice President, Office of Management and Planning, PCN Inc., Seoul, Korea; linuxyoo@naver.com

implementation solutions remains challenging due to various issues. In particular, data sovereignty concerns are emerging regarding the use of public data. [1] This study proposes a public data sovereign cloud to solve two key challenges in adopting large-scale AI in the public sector: compliance with information disclosure laws and the operation of closed networks. To achieve this, we examine the concept of a sovereign cloud and propose a definition and implementation strategy for a public data sovereign cloud. Specifically, we present application strategies focusing on construction methods using public-private partnership (PPP) models, network segregation methods according to data grades, and enhancement of the role of Managed Service Providers (MSPs). [2] The structure of this paper is as follows. Chapter 2 examines the concept and background of sovereign cloud. Chapter 3 analyzes the definition of public data sovereign cloud and its differentiation from general sovereign cloud. Chapter 4 presents strategies for implementing a public data sovereign cloud, while Chapters 5, 6, and 7 discuss in detail the methods for ensuring operational sovereignty, data sovereignty, and software sovereignty, respectively. Finally, Chapter 8 presents conclusions and policy implications.

II. Sovereign Cloud Concept

1. Definition of Sovereign Cloud

A sovereign cloud refers to a cloud computing architecture that complies with national and regional laws and regulations while granting data sovereignty to the country where the data is generated. The concept of sovereignty can be divided into three categories: operational sovereignty, data sovereignty, and software sovereignty.

Table 1. Components of Sovereign Cloud

Category	Content
Operational Sovereignty	- Grant data access and management rights according to the laws and regulations of the data generation region or country - (Data Localization) Store and process data only within the country or region of origin to comply with current regulations - (Unauthorized Access Prevention) Strictly block data access by unauthorized users - (Operational Transparency) Transparently manage and disclose data storage, processing, and usage processes - For Germany, according to data protection law, data collected within Germany must only be stored/processed domestically, and institutions maintain data control rights through their national data center cloud services - Can be constructed without being affected by the U.S. Cloud ActData Sovereignty
Data Sovereignty	- Data generators retain rights over their data - For example, under GDPR, user data generated within the European Union cannot be transmitted or processed by third parties without user consent Software Sovereignty
Software Sovereignty	- (Software Independence) Users must be able to perform desired operations without being dependent on specific CSPs and should be able to migrate (or transfer) when needed quickly

Source: adapted from https://www.enterprisenetworkingplanet.com/security/what-are-sovereign-clouds/

2. Background

Significant background factor in the emergence of sovereign cloud is the U.S. Cloud Act enacted in 2018. This law stipulates that U.S. government agencies can request data from telecommunication service providers regardless of where the data is stored [3]. In response, the European Union enacted GDPR (General Data Protection Regulation), and various countries began establishing legal measures to strengthen data sovereignty. Under these international trends, domestic and foreign CSP (Cloud Service Provider) businesses have introduced the concept of sovereign cloud as a means to ensure data sovereignty [4].

3. Key Features of Sovereign Cloud

Sovereign Cloud has the following key features [5] [6]:

1. Data Localization: Data storage and processing are limited to within the relevant country or region.
2. Access Control: Data access is strictly controlled according to the laws of

the relevant country.
3. Transparency: Ensures auditability through transparent disclosure of data processing processes.
4. Legal Compliance: Strictly adheres to data protection laws of the relevant country or region.
5. Independence: Enables independent operations without dependence on specific cloud service providers.

4. Expected Effects of Sovereign Cloud

The adoption of a sovereign cloud is expected to bring the following effects:
1. Enhanced Data Security: Enables secure data management through compliance with national regulations.
2. Reduced Legal Risk: Minimizes potential legal risks arising from conflicts between national laws.
3. Data Sovereignty Assurance: A relevant country or agency can exercise actual control over data.
4. Improved Service Stability: Enables customized service provision considering regional characteristics.

The concept of sovereign cloud provides important implications for large-scale AI adoption in the public sector. Notably, the necessity of public data sovereign cloud applying sovereign cloud principles is increasingly recognized in terms of secure utilization of public data and sovereignty assurance. To further emphasize its necessity, the public data sovereign cloud is crucial in enabling large-scale AI's secure and efficient utilization in the public sector. It ensures stable management of massive datasets required for AI training and operation while supporting data sovereignty and compliance with legal regulations. By providing a safe and reliable environment through closed networks and data stratification, the public data sovereign cloud enhances the efficiency of AI deployment, contributing to innovation and improved public services. [7]

5. Sovereign Cloud Implementation Cases

Various countries and organizations are making efforts to implement sovereign cloud. Representative cases include [8] [9]:
1. France's Andromede Project: A national cloud project led by the French government, focusing on data sovereignty assurance
2. Germany's GAIA-X: A cloud infrastructure project to ensure European digital sovereignty
3. Microsoft's Azure Government: Provides a reinforced cloud environment for U.S. government agencies

4. Amazon's GovCloud: A regulatory-compliant cloud service targeting U.S. government agencies and related organizations

These concepts and cases of sovereign cloud provide important implications for large-scale AI adoption in the public sector. Notably, the necessity of a public data sovereign cloud is emerging regarding secure utilization and sovereignty assurance of public data. The next chapter will examine in more detail the concept and characteristics of the public data sovereign cloud.

III. Introduction of Public Data Sovereign Cloud

1. Definition of Public Data Sovereign Cloud

In this research, we define Public Data Sovereign Cloud as follows: It refers to a cloud computing architecture that complies with laws and regulations related to public agencies and grants data sovereignty to public agencies over public data. The main components are operational sovereignty, data sovereignty, and software sovereignty.

Table 2. Components of Public Data Sovereign Cloud

Category	Content
Operational Sovereignty	- (Management according to laws and regulations of public agencies) Grant data access and management rights according to laws and regulations of data-generating agencies or related public agencies - (Data Localization) Store data within public agencies that generate or relate to the data, complying with public agency regulations and preventing unauthorized private access - (Unauthorized Access Prevention) Unauthorized users cannot access public agency data - (Operational Transparency) Public agencies transparently manage how their data is stored and utilized - Data collected within public agencies must only be stored and processed within the public agency or related public agencies. For example, through cloud services operating on data centers of related public agencies within government/administrative networks, agencies maintain complete control over their data Sovereignty
Data Sovereignty	- (Data-generating public agencies retain rights over their data) Implies that regardless of where data is stored, final decision rights over the data remain with the original data owner software Sovereignty
Software Sovereignty	- (Software Independence) Public agencies must be able to perform desired operations without being dependent on specific CSPs and should be able to migrate (or transfer) when needed quickly

Source: Authors' own work

2. Comparison between Sovereign Cloud and Public Data Sovereign Cloud

While both sovereign cloud and public data sovereign cloud focus on the concept of data sovereignty, they differ in their scope of application and key issues. While Sovereign Cloud primarily focuses on data sovereignty issues between nations, public data Sovereign Cloud primarily addresses data sovereignty issues between public and private sectors.

Table 3. Comparison between Sovereign Cloud and Public Data Sovereign Cloud

Category	Sovereign Cloud	Public Data Sovereign Cloud
Core Content	- Resolving conflicts between national laws of data-generating entities and CSP's national laws	- Resolving conflicts in laws/regulations/interests between public agencies (data generation/ownership) and various private CSPs
Issues	- U.S. Cloud Act	- Public data access in private cloud environments
Solution Approach	- Ensure operational sovereignty, data sovereignty, and software sovereignty through domestic law application, independent of laws from CSP-providing countries	- Ensure operational sovereignty, data sovereignty, and software sovereignty for public agencies through compliance with public agency-related laws /regulations, independent of CSP providers

Source: Authors' own work

3. Necessity of Public Data Sovereign Cloud

The following fundamental factors emphasize the necessity of a public data sovereign cloud:

1. Secure Utilization of Public Data: This can prevent data leakage and misuse risks when using private cloud services. This is essential for protecting important national information security.
2. Legal Compliance: Cloud services can be utilized while strictly adhering to laws and regulations related to public agencies. This is important for satisfying various legal requirements, including the Information Disclosure Act and the Personal Information Protection Act. [10]
3. Data Sovereignty Assurance: Public agencies can control their data entirely. This strengthens public agencies' rights by maintaining decision rights over data storage, processing, and access.
4. Foundation for Large-scale AI Utilization: This enables the secure utilization of public data for developing and applying large-scale AI. It

serves as an important technical foundation for public service innovation and efficiency improvement.

Based on these various necessities, the next chapter will discuss in detail the specific implementation strategies of the public data sovereign cloud. [11]

IV. Public Data Sovereign Cloud Strategy

1. Overview of Implementation Approaches

To address the challenges in implementing large-scale AI in the public sector, it is essential to identify and resolve key issues beforehand. These challenges include data privacy and security concerns, compliance with legal and regulatory frameworks, financial constraints, lack of technical infrastructure, and ethical considerations. Addressing these obstacles will establish a solid foundation for the effective deployment of the public data sovereign cloud and ensure its alignment with public sector requirements. The approach strategy for implementing a public data sovereign cloud can be summarized as shown in the table below:

Table 4. Approaches to Public Data Sovereign Cloud Implementation

Category	Content	Implementation methods
Operational Sovereignty	- Management according to the laws and regulations of data-generating public agencies - Data localization - Prevention of unauthorized access - Operational transparency	- Construction of public data sovereign cloud center using PPP model - Information system importance level classification - Network segregation for security enhancement - Establishment of public MSP specialist agencies
Data Sovereignty	- Final decision rights over data	- Data grade classification - Security provision according to grade
Software Sovereignty	- Ability to perform desired operations anywhere without dependency on specific CSPs, with quick migration capability when needed	- Provision of technical standards and guidelines - Construction of public sector foundation models - Domain-specific sLLM construction - Enhancement of MSP roles

Source: Authors' own work

For effective implementation of public data sovereign cloud, construction of cloud data centers using a Public-Private Partnership (PPP) model is necessary.

This approach implies cooperation between government and private sectors to build infrastructure and provide public services.

To provide a more detailed discussion on the Public-Private Partnership (PPP) model, we have expanded the explanation as follows [12]:

1. Configurations by Data Sensitivity:
   - Public Cloud: Managed entirely by government agencies, this model is designed for critical and sensitive data that require high security and operational sovereignty.
   - PPP Model 1: This hybrid approach combines public and private infrastructure, allowing medium-level data to be securely managed while sharing responsibilities between public and private entities.
   - PPP Model 2: Utilizing private facilities, this model handles less sensitive data with logical or physical network separation to maintain security while optimizing costs.
2. Practical Example:
   Applying a 70B QLoRA 4bit AI model in a single government agency requires significant hardware investments, such as 8 NVIDIA H200 GPUs to support 150 concurrent users and 15 simultaneous inferences. These hardware costs form a substantial portion of the initial investment. The PPP model allows financial burden sharing between public and private partners while leveraging the technical expertise of private entities for efficient deployment.
3. Implementation Scenarios:
   The G-Cloud infrastructure in South Korea serves as a prime example of how the PPP model can balance security and efficiency. Public agencies manage critical functions, such as data classification and security measures, while private entities contribute their technological capabilities and operational expertise. Hybrid PPP centers, as outlined in the proposed model, enable flexible operation tailored to the sensitivity of data and specific operational needs, providing a scalable and adaptable framework for public data management.

In this research, we propose four types of implementation methods based on the PPP model:

Despite these limitations, a public data sovereign cloud remains a viable and realistic solution for addressing data sovereignty and security issues. Rather than considering alternative solutions, this approach focuses on refining the implementation strategies to mitigate the identified challenges, ensuring a more robust and efficient deployment.

First, the Public Cloud Center. This operates within public buildings using public facilities and is directly managed by administrative agencies. The G-

Cloud of the National Information Resources Management Service is a representative example. In this model, the center director, management, supervision, and security officers are from administrative agencies, while technical operations can be outsourced to private contractors. From a security perspective, external network connections and remote access are blocked. When necessary, temporary external network connections are only permitted for designated devices under security officer supervision while maintaining isolation from the internal network. In terms of data management, it handles systems and related data classified as "high" importance, as well as high-grade non-public data.

Second, the Public Cloud Center (PPP Type 1). This operates both public and private facilities simultaneously within public buildings. The organizational structure and physical network segregation are identical to the Public Cloud Center. In this model, it processes systems and related data classified as "medium" importance level, and can handle both public data and medium-grade non-public data.

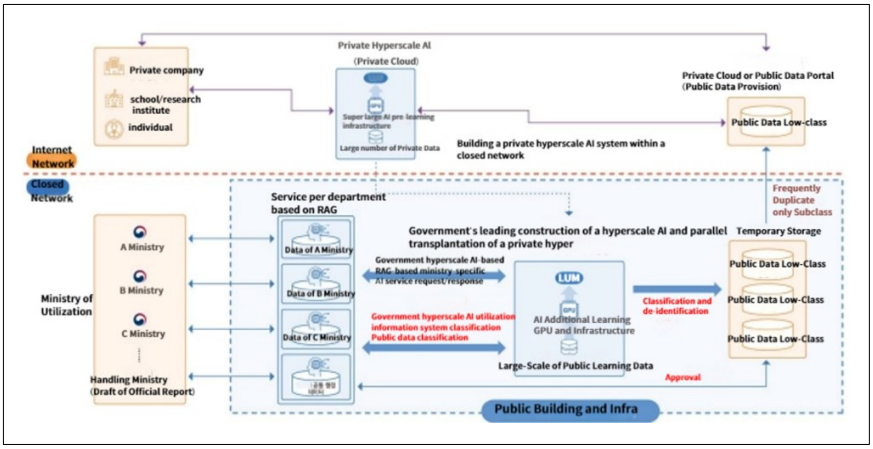

Figure 1. Public Cloud Center (PPP Type 1)

Third, the Public Cloud Center (PPP Type 2) operates both public and private facilities simultaneously within private buildings. Its organizational structure is similar to the previous models, and it can selectively apply either physical or logical network segregation. In this model, it can process public data and low-grade non-public data (such as de-identified data) that operate within systems classified as having a "medium" importance level.

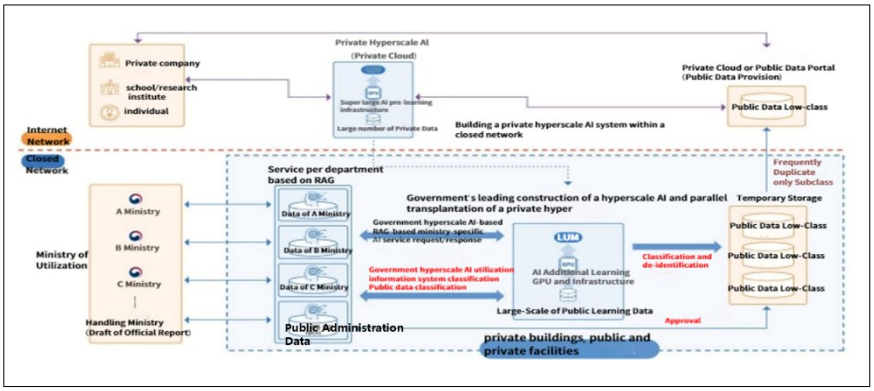

Figure 2. Public Cloud Center (PPP Type 2)

Fourth, the Private Cloud Center. This is constructed and operated within private buildings using private facilities, with Naver Cloud and KT Cloud being representative examples. In this model, all roles, including the center director, management, supervision, and security officers, are fulfilled by private companies. From a security perspective, external network connections and remote access are possible as needed. Regarding data management, it can only process systems and related data classified as "low" importance level, along with public data.

These various PPP models enable effective and secure cloud data center construction by combining the advantages of both the public and private sectors. Each model can be selectively applied according to data importance and security requirements, enabling secure public data management and efficient utilization. Furthermore, these PPP models are expected to improve public services by supplementing limited government resources with private sector expertise. [13] [14] When implementing the PPP type, the Internet VPC (Virtual Private Cloud) can be configured with an Internet zone for external service provision and a DMZ (Demilitarized zone) zone as a buffer zone between external and internal networks. The internal VPC can be separated into a work zone for internal business system operations and a management zone for system monitoring and log management. Additionally, VPC interconnection and routing control can be structured through Transit Gateway. For security control, IAM (Identity and Access Management) for access permission management, security groups using instance-level firewalls, and network ACLs (Access Control Lists) for subnet-level access control can be established. [15] [16]

2. Role of MSP (Managed Service Provider)

The role of MSP is crucial in constructing a public data sovereign cloud for secure generative AI infrastructure. MSPs act as intermediaries between Cloud Service Providers (CSPs) and customers, managing the entire cloud operating environment, from cloud adoption consulting to migration, operation, and monitoring. MSPs play an increasingly vital role in public data sovereignty assurance, particularly in managing multi-cloud infrastructure for large-scale AI in the public sector. [17] [18]

The key roles of MSP can be summarized as follows:

First, the integrated multi-cloud configuration across public sector domains and hierarchical structure within domains. This is a complex task that involves maintaining overall consistency while reflecting the requirements of various public agencies. MSP effectively coordinates and manages this integration.

Second, cost reduction can be achieved by integrating the common infrastructure of the government and public agencies. MSP identifies infrastructure that can be commonly used across multiple agencies and manages its integration, thereby preventing duplicate investments and increasing cost efficiency.

Third, multi-cloud can be operationally optimized by using multiple private CSPs. An important role of MSP is to optimize overall operational efficiency while meeting public agency requirements by optimally combining services from various CSPs. [19]

Fourth, leading MSP role and Central-edge architecture configuration for domain-specific AI services across government agencies. This means balancing centralized management with specialized service delivery for each domain.

Fifth, minimization of potential security issues. MSP ensures secure management of public data by identifying various security threats in advance and preparing countermeasures.

Sixth, standardizing the design of common foundation models for government large-scale AI utilization, ensuring they can be developed and operated according to the characteristics of each department/agency. This is a key task for effectively introducing large-scale AI technology to the public sector.

These enhanced MSP roles will facilitate cloud and AI adoption in the public sector and enable secure and efficient service provision with public data protection. In particular, the role of MSP becomes more important as it possesses expertise in understanding the uniqueness of the public sector and securely handling civilian data.

Therefore, the government should establish policies to cultivate and support MSPs with such capabilities. Specifically, they can consider:

- Strengthening MSP certification systems

- Operating public sector specialized MSP cultivation programs
- Establishing cooperation frameworks between MSPs and government agencies

Through these measures, stable construction and operation of a public data sovereign cloud will become possible, ultimately contributing to improving public services and enhancing national competitiveness.

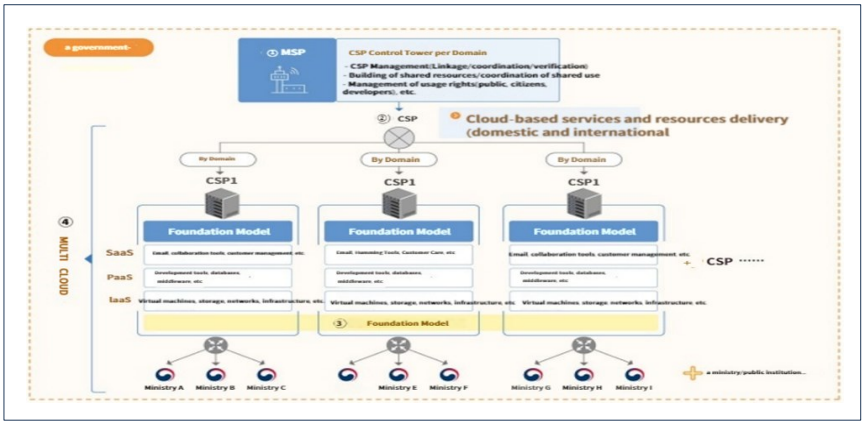

Figure 3. Hierarchical Multi-Cloud Construction for Cross-Government Large-scale AI

① MSP (Control Tower): The role of MSP as a control tower connecting Cloud Service Providers (CSPs) and users is crucial for building secure and efficient generative large-scale AI. It manages the entire cloud operating environment, from cloud adoption consulting to migration, operation, and monitoring.

② CSP (Cloud Service Provider): (Domain-specific Edge Cloud) Performs management and operation of cloud-based Foundation Models for linked departments/agencies, along with data requirements/analysis, metadata, data security, and data quality management. Configures and operates various cloud architectures, including agency-specific public/private clouds and SaaS, PaaS, and IaaS layers.

③ Foundation Model (Basic Model): Serves as a common foundation for cross-government large-scale AI, where foundation models applicable across multiple departments/public agencies are developed and operated.

④ Multi-Cloud: Provides cloud integration (IaaS) for each domain within the government closed network. Configures hybrid public/private clouds within departments/public agencies and performs integration with other agencies. [20]

3. Expected Effects

Implementation of public data sovereign cloud is expected to bring the following effects:

1. Integrated Operations

   Realizes integration and hierarchical structure across government domains for large-scale AI

   Maintains overall consistency while reflecting the requirements of various government agencies

   Systematically coordinates complex tasks
2. Budget Savings

   Enables cost reduction through integration of government and public agency common infrastructure

   It prevents duplicate investments and improves cost efficiency by identifying and managing infrastructure that can be commonly used across multiple agencies
3. Operational Optimization

   Enables service optimization through the utilization of multiple private Cloud Service Providers (CSPs)

   Improves resource utilization while meeting specific public agency requirements
4. Control Tower

   Establishes control tower role for AI services across government domains

   Builds specialized service systems for each government domain through Central-edge configuration
5. Enhanced Security

   Ensures secure management of public data by identifying various security threats in advance

   Prepares countermeasures for potential security issues
6. Model Standardization

   Unifies design to enable the development and operation of common foundation models for government-wide large-scale AI utilization

   Can tailor models to the characteristics of each agency

4. Data Sovereignty Assurance

We propose the following measures to ensure data sovereignty:

First, Data Grade Assignment: When public agencies want to convert and utilize their managed information in private clouds, data must be classified into four levels according to the "Information Disclosure Act" and national information security basic guidelines, as follows:

Table 5. Data Classification Levels

Classification Major	Classification Minor	Content	Method
Non-public C, high (Classified)	Level1	- Information corresponding to Article 9, Paragraph 1, Items 1,2,3,4 of the Information Disclosure Act - Information classified as 'C' in national information security basic guidelines	Utilize in Public Cloud Center - No external access
Non-public S, middle (Sensitive)	Level2	- Information corresponding to Article 9, Paragraph 1, Items 5,7,8 of the Information Disclosure Act - Information classified as 'S' in national information security basic guidelines	- Utilize in Public Cloud Center (PPP Type 1)
	Level3	- Information corresponding to Article 9, Paragraph 1, Item 6 of the Information Disclosure Act - Information classified as 'S' in national information security basic guidelines	- Utilize in Public Cloud Center (PPP Type 2) - Can perform non-public information de-identification
Public O, low (Open)	Level4	- Information classified as public through public classification work - Information classified as 'O' in national information security basic guidelines	- Can be utilized in a Private Cloud Center

Second, Security Requirements by Data Grade: Define and apply differentiated security requirements for each grade level. For example, higher-grade data requires more stringent encryption and access control policies.

5. Software Sovereignty Assurance

We propose the following strategic measures to ensure software sovereignty:

1. Standards and Guidelines Development

Provide standards and guidelines for public agencies to build cloud-based large-scale AI utilization systems using public data.

2. Public Agency Data Management System Construction

Establish methods for public agencies to manage data collection/ processing/refinement processes without dependency on specific CSPs

3. Public Sector Foundation Model Development Strategy

Ensure foundation models are built within closed networks in public cloud centers or public cloud centers (PPP type), excluding private cloud centers in the PPP model

4. Foundation Model Performance Measurement and Maintenance Strategy

Establish strategies for continuous model updates, model efficiency techniques, and domain adaptation

These detailed implementation strategies are expected to enable the practical construction and operation of a public data sovereign cloud.

V. Policy Recommendations for Public Cloud Implementation

This research has analyzed the concept and implementation strategies of public data sovereign cloud for large-scale AI utilization in the public sector. The key implications derived can serve as important considerations for public cloud implementation. In particular, we present the following policy recommendations to achieve both data sovereignty and large-scale AI utilization simultaneously:

1. Cloud Utilization for Large-Scale AI

Cost-effective information resource management through strategic use of cloud technology is essential, as large-scale AI requires extensive high-performance computing resources and data management.

2. Establishment of Public Data Sovereign Cloud Concept

Beyond international data sovereignty issues, the legal and institutional establishment of the "public data sovereign cloud" concept is required to guarantee public agencies' sovereignty over public data in their relationship with private CSPs. To ensure the effective implementation of a public data sovereign cloud, it is essential to account for differences in national contexts, including

legal, technical, and cultural aspects. The following considerations highlight the importance of tailoring the approach to specific regional requirements:

Legal Contexts: In South Korea, laws such as the Personal Information Protection Act (PIPA) and the Information Disclosure Act are critical for defining the framework for public data management. These regulations mandate strict compliance with data sovereignty principles, including data localization and secure handling of sensitive information, ensuring that public agencies retain complete control over their data.

Technical Contexts: South Korea's technical environment requires advanced infrastructure, including high-performance hardware such as NVIDIA H200 GPUs optimized for large-scale AI deployment. Additionally, secure network segmentation strategies and compatibility with existing government systems are necessary to support operational efficiency and data protection.

Cultural Contexts: South Korea's cultural emphasis on trust, transparency, and accountability aligns with global best practices in public data management. While cultural nuances may influence stakeholder engagement and policy adoption, these values are broadly consistent with international standards and require no additional localized adjustments.

Addressing Legal and Ethical Challenges: It is essential to consider specific regulatory and ethical frameworks to comprehensively address the legal and ethical challenges associated with implementing a public data sovereign cloud. Compliance with data sovereignty principles, such as those mandated by South Korea's Personal Information Protection Act (PIPA) and Information Disclosure Act, is critical. These regulations enforce strict data localization requirements and protect sensitive information, ensuring public agencies retain complete control over their data. Additionally, cross-border data transfers must align with international frameworks like the EU's General Data Protection Regulation (GDPR) to mitigate conflicts between national and international laws. [21] [22]

Ethically, public data utilization for AI raises concerns about bias, transparency, and accountability in AI models. Adopting ethical governance frameworks, such as South Korea's AI Ethics Guidelines for the Public Sector, is crucial to address these issues. [23]

These guidelines emphasize fairness, accountability, and transparency, ensuring responsible use of sensitive data. Regular legal audits, the integration of ethical principles into AI and data policies, and mechanisms for enhanced transparency and accountability are recommended to overcome these challenges. By addressing these legal and ethical concerns, implementing a public data sovereign cloud can ensure compliance, foster public trust, and support the responsible development of large-scale AI systems.

3. Systematization of Public-Private Partnership Model

When building a public data sovereign cloud using the PPP model, detailed operational guidelines and governance frameworks are needed to operate by distinguishing facilities, operating entities, and management systems into public cloud centers, public cloud centers (PPP type), and private cloud centers.

4. Network Segregation Specification by PPP Model

To build a public data sovereign cloud, policies for data classification and network segregation according to the PPP model must be established to balance security and efficiency.

5. MSP Role in Public Data Sovereign Cloud

To secure public data sovereignty, clear definitions of MSP roles in multi-cloud-based domains and establishing education and certification systems to enhance expertise are necessary.

6. Data Classification

Legal provisions state that public data can be classified as public or non-public according to the Information Disclosure Act, so it is necessary to restrict data grades and establish differentiated management policies for each type of cloud center (public, PPP-type, private).

7. Software Sovereignty Assurance

Public sector foundation models must be built to ensure the software is not dependent on specific CSPs for public data sovereign cloud.

Suppose these policy recommendations are systematically implemented with comprehensive consideration. In that case, both objectives—secure and efficient utilization of large-scale AI and public data sovereignty assurance in cloud environments—are expected to be achieved simultaneously.

VI. Conclusion

This study presented implementation directions for a public data sovereign cloud that utilizes large-scale AI in the public sector. Public data sovereign cloud is emerging as a crucial concept that enables effective utilization of large-scale AI while guaranteeing data sovereignty for public agencies, and its importance is increasingly being highlighted. Several prerequisites must be met to establish government-wide large-scale AI infrastructure and secure AI sovereignty. This includes resolving data sovereignty issues through government-led efficient implementation, preventing private enterprise monopoly, and establishing measures to secure public interest. [24]

Significant are: government-wide resource consolidation to overcome limitations in implementing large-scale AI across government agencies, establishing preemptive measures for legal and ethical issues that may arise in government-wide public data utilization, and expanding development and utilization of domestic AI chips to reduce dependence on foreign companies. If these policy recommendations are systematically implemented comprehensively, we can successfully establish a government-wide AI ecosystem that can safely and efficiently utilize large-scale AI. This will ultimately serve as a foundation for advancing national AI operation and utilization capabilities, enabling the internalization of various rapidly evolving AI technologies to meet our government's requirements. Furthermore, this experience and technology have the potential to develop new export models similar to the Korean e-Government Wave. A limitation of this study is the lack of empirical analysis based on actual implementation cases. Therefore, future research needs in-depth discussion on specific technical and institutional measures for the actual implementation of public data sovereign cloud, and empirical research should be conducted on the impact of public data sovereign cloud adoption on public service quality and efficiency. The results of this study are expected to serve as fundamental material for policy formulation regarding the introduction and utilization of large-scale AI in the public sector. Additionally, the concept of a public data sovereign cloud presented in this study will provide an important theoretical foundation for future national AI strategy development and ultimately contribute to strengthening national competitiveness. This study systematically reviewed the data sovereign cloud model for large-scale AI utilization in the public sector and derived key policy implications. Future research should validate its effectiveness through empirical case studies, refine legal frameworks to balance data protection with AI training and explore sustainable public AI infrastructure strategies in the global AI landscape. These efforts will further enhance the effectiveness of public AI utilization and strengthen data sovereignty.

Acknowledgment

This paper was written using various research materials and literature from the National Information Society Agency. We would like to express our gratitude to those who granted permission to use the cited materials and reviewed this paper.

References

[1] N. Kushwaha, P. Roguski and B. W. Watson, "Up in the Air: Ensuring Government Data Sovereignty in the Cloud," 2020 12th International Conference on Cyber Conflict (CyCon), Estonia, 2020, 43-61
[2] Jung-Bo Kim, Jung-In Kim, A Study of Application Development Method for Improving Productivity on Cloud Native Environment, v.23 no.2, 2020, 1-3
[3] Young Jin Song, "The Passage of the U.S. CLOUD Act and Its Implications for Extraterritorial Data Access," Journal of Criminal Policy Studies, v.29, no.2, 2018.
[4] https://velog.io/@rlawogus73/ITCEN: Why Do Online Lectures Use the Cloud?, CSP-MSP-2
[5] https://m.ddaily.co.kr/page/view/20240315170921900 25,http://www.itdaily.kr/news/articleView.html?idxno=214309, https://brunch.co.kr/@b047a588c11b462/64
[6] https://www.oracle.com/kr/cloud/sovereign-cloud/
[7] Michels, J.D., Walden, I., & Millard, C., (2025). Storm Clouds are Building: Surveillance, Sovereignty, and State Interests. Sovereignty, and State Interests (February 03, 2025). 50-51
[8] https://azure.microsoft.com/en-us/explore/global-infrastructure/government
[9] https://devblogs.microsoft.com/azuregov/microsoft-publishes-secure-isolation-guidance-for-azure-and-azure-government/
[10] Sharma, P., Martin, M., Swanlund, D., Latham, C., Anderson, D., & Wood, W. (2024). A cloud-based solution for trustless indigenous data sovereignty: Protecting Māori biodiversity management data in Aotearoa New Zealand. Transactions in GIS, 28(4), 837-838.
[11] Cordes, A., Bak, M., Lyndon, M., Hudson, M., Fiske, A., Celi, L. A., & McLennan, S., (2024). Competing interests: digital health and indigenous data sovereignty. NPJ digital medicine, 7(1), 178. 17, 30, 50]
[12] https://news.zum.com/articles/72936079
[13] Ma, J., & Li, X., (2024). Performance Evaluation of Education PPP Projects in the Operation Stage Based on Limited Cloud Model and Combination Weighting Method. Available at SSRN 4916345, 97-100.
[14]Li, X., & Guo, Z., (2024). Research on Public-Private Partnership Models in Digital Government Construction. In E3S Web of Conferences (v.565, p.03004). EDP Sciences, 20-22.
[15] Financial Network Segmentation and Electronic Financial Supervision Regulations (2017)
[16] Electronic Financial Supervision Guidelines, Financial Supervisory Service (2017)
[17] Jung-Bo Kim, Jung-In Kim, A Study of Application Development Method for Improving Productivity on Cloud Native Environment, v.23 no.2, 2020, 1-3
[18] https://www.lgcns.com/blog/cns-tech/cloud/32946/
[19] Borra, P., (2024). An Overview of Cloud Computing and Leading Cloud Service Providers. International Journal of Computer Engineering and Technology (IJCET) v.15, 130-131.

[20] Borra, P., (2024). Comparison and analysis of leading cloud service providers (AWS, Azure and GCP). International Journal of Advanced Research in Engineering and Technology (IJARET) v.15, 268-269.

[21] Michels, J.D., (2025). Sovereign Cloud for Europe: Independent Research Report prepared for Broadcom. Available at SSRN 5146122, 837-839.

[22] Michels, J.D., Millard, C., Walden, I., & Wuermeling, U., (2024). Cloud Sovereignty and the GDPR, Part One: US Government Access to European Data. Queen Mary Law Research Paper, 8-10, 15-17.

[23] Michels, J.D., (2025). Sovereign Cloud for Europe: Independent Research Report prepared for Broadcom. Available at SSRN 5146122, 840-841

[24] Rone, J., (2024). 'The sovereign cloud' in Europe: diverging nation state preferences and disputed institutional competences in the context of limited technological capabilities. Journal of European Public Policy, 31(8), 5-7, 10-12.