Design Issues in Implementation of Decentralized AI-data commons Framework

Young-Ho Suh, Sungpil Woo, Boyun Eom, Dong-Hwan Park, Sunhwan Lim, Chan-Won Park
Autonomous IoT Research Section,
Electronics and Telecommunications Research Institute
Daejeon, South Korea
yhsuh, woosungpil, eby, dhpark, shlim, cwp@etri.re.kr

***Abstract*—In this paper, we propose a system architecture for AI-data commons. An AI-data commons system can form a shared data economy among various stakeholders such as data owners, data providers, and AI/data consumers, computing providers and problem solvers. In the AI-Data Commons ecosystem, various AI modules and data are shared/distributed in a way that ensures the sovereignty and privacy of AI-data owners through decentralized interactions among stakeholders.**

***Index Terms*—AI-data commons, user sovereignty, anonymized data, commons ecosystem**

I. Introduction

AI/data commons [1]–[4] is attracting attention as a new technology. In June 2019, the ITU [5] proposed the concept of AI commons as a collaborative framework to support AI-based problem solving for everyone. In the AI-Data Commons ecosystem [6], various AI modules and data are shared and distributed in a way that the owners' sovereignty and privacy are guaranteed through decentralized interactions among the following stakeholders as shown in "Fig. 1"

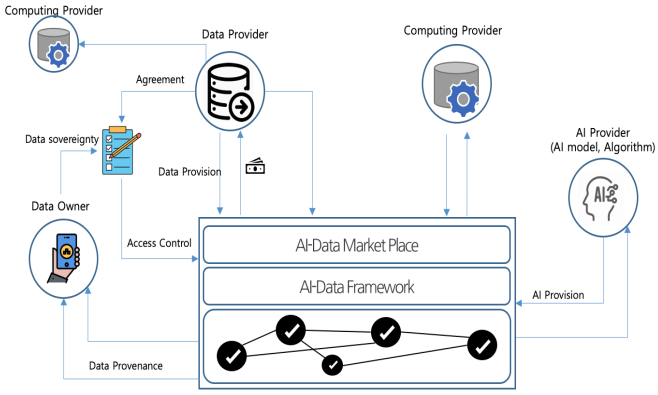

Fig. 1. The conceptual model of the AI-data commons

- Data owners : Data owners are individual users who provide personal data to the AI-Data Commons system. They can directly track and control their own data flow, and are incentivized to provide personal information.
- Data Providers : Data providers receive consent from data owners to provide data, collect and accumulate data from data owners, refine and process meaningful data, register it in the marketplace of AI-Data Commons, and share/distribute it.
- AI providers : AI providers implement AI learning models or algorithms to solve various social problems, register them in the marketplace of AI-Data Commons, and share/distribute them.
- AI/Data consumers : AI/data consumers search the list of data registered in the AI-Commons marketplace and purchase the data they need.
- Computing providers : Computing providers implement a computing environment in the form of a service to execute AI algorithms registered in AI-Data Commons and registers it in AI-Data Commons.
- Problem Solvers : Problem solvers are those who want to obtain a solution by collaborating with various stakeholders belonging to AI-Data Commons, and defines a social problem to be solved and registers it in AI-Data Commons.

II. Commons ecosystem use cases

A. User Sovereignty

The use case for each stage of the sharing personal data based on user consent scenario is as follows.

1) Data Collection: Data providers register apps on the marketplace to collect personal data from data owners. The data owner searches for an app from the marketplace, downloads the app, and installs it. During the app installation process, consent to the use of personal information is obtained based on the contract for the use of personal information provided by the app. Data providers collect personal data from data owners through installed apps.

2) Data Provision: The data providers create schema information about the data storage where the data collected from the data owner is stored. They create metadata about collected data so that data consumers can search for the data they need and create a computing service that can execute AI algorithms by accessing data collected in a virtualized computing environment. Finally they register computing provided services, including storage schema and meta data for collected data, in the marketplace of the AI-Data Commons system.

3) Searching Data Provision: A marketplace provides search functions through various search conditions and search filters. Data consumers search for necessary data provision

services through specific conditions or filtering through the marketplace.

4) Purchasing Data Provision Services: The AI provider presents the data price to the data provider to purchase the searched data. The data provider may accept or reject the offer. If the data provider accepts the offer, the data provider authorizes the AI provider to access the service. The AI provider accesses the data that the data owner has agreed to use through the data provision service and learns in the virtualized execution environment provided by the computing provider. AI providers get AI algorithm execution results.

B. Anonymized Data

The use case for each stage of the anonymized data scenario is as follows.

1) Data Collection: Data providers register apps on the marketplace to collect personal data from data owners. The data owner searches for an app from the marketplace, downloads the app, and installs it. The app collects personal data from data owners.

2) Data Registration: The data provider processes the collected data and refines it into meaningful data to form a unique data set. The data provider anonymizes the constructed data set by utilizing the personal data anonymization algorithm provided by the data processor and create metadata about organized data sets so that data consumers can search for the data they need. The configured data set is registered to the marketplace of the AI-Data Commons system based on the generated metadata.

3) Data Search: Marketplace provides a search function through various search conditions and search filters for various anonymized data sets registered in AI-Data Commons. Data consumers search for required data sets through specific conditions or filtering through the marketplace.

4) Purchase and download: The data buyer presents the data price to the data provider to purchase the searched data. The data provider may accept or reject the offer. If the data provider accepts the offer, the data provider authorizes the A-data buyer to access the purchase data. Data purchaser accesses and downloads the purchase data.

III. Conceptual Architecture

"Fig. 2" shows the overall conceptual structure of the AI-Data Commons system. The AI-Data Commons ecosystem is basically a sharing/distribution system in which transactions are made between providers who provide assets and consumers who consume assets. Therefore, the AI-Data Commons system structure consists of three subsystems. The first subsystem is the AI-Data Commons framework for various providers providing assets to be distributed in the AI-Data Commons. The second subsystem is an AI-Data Commons front-end that provides a user interface and management and execution tools for asset consumers. And, finally, there is AI-Data Commons to provide match making between asset providers and consumers and to support decentralized interactions between matched parties.

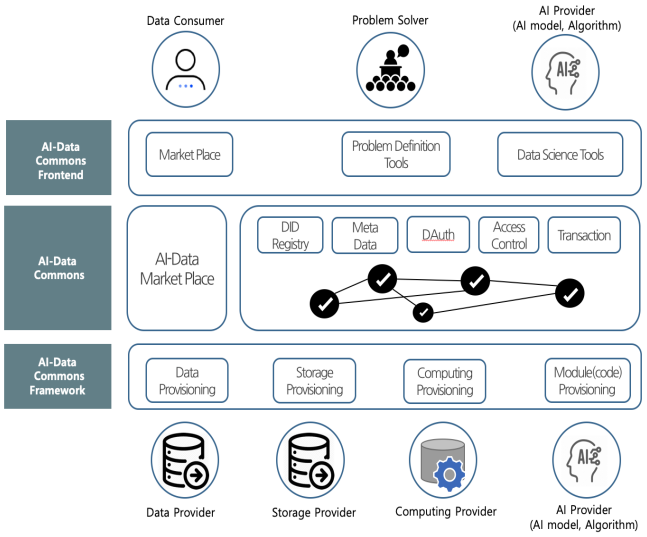

Fig. 2. System Architecture of an AI-data commons framework

A. AI-Data Commons Framework

The AI-Data Commons Framework is a decentralized back-end framework that allows various asset providers to register their assets in the AI-Commons system and circulate in the AI-Data Commons ecosystem.

1) Data Provisioning: Provides a common meta data schema to describe the properties of data so that it can be operated in conjunction with the DID Registry of the AI-Data Commons system. Since the actual data storage is an external storage operated by a data provider, it provides a gateway that can work with various external storages.

2) Storage Provisioning: Unlike Data Provisioning, which organizes a data set and distributes the data set directly, a service that allows private/local access to data is created and the service is distributed. Provides a common metadata schema for services so that it can be operated in conjunction with the DID Registry of the AI-Data Commons system. In order to control storage access based on the consent of the data owner to use personal information, it is linked with DAuth and Access control of the AI-Data Commons system.

3) Computing Provisioning: Provides an environment where verified AI learning modules or algorithms perform learning algorithms based on private data access services provided by storage provisioning. Provides a virtualized container-type execution engine so that it can be operated in various physical computing environments. Provides a command-type tool that manages the deployment and execution of containers in real time.

4) Module (Code) Provisioning: Provides a common metadata schema to describe the properties of the algorithm so that it can be operated in conjunction with the DID Registry of the AI-Data Commons system. Provides a verification mechanism that ensures that the algorithm does not leak the data owner's personal information outside of the private execution space provided by Computing provisioning.

B. AI-Data Commons

When various assets (data, storage, computing, and algorithms) are registered in the AI-Data Commons based on the AI-Data Framework, the asset owners' sovereignty and privacy is shared/distributed in a way that is guaranteed. The AI-Data Commons system is a system for providing match making between asset providers and consumers and supporting decentralized interactions between matched parties.

1) AI-data market place: Marketplace provides an interface to register assets in AI-Data Commons in conjunction with DID registry and MetaData. Marketplace provides an interface to access assets registered in AI-Data Commons in conjunction with DID registry and MetaData. Marketplace provides the ability to search and filter assets registered in AI-Data Commons through search queries in conjunction with the DID registry and MetaData.

2) DID Registry: DID Registry is a registry for registering and managing assets in AI-Data Commons based on decentralized identifiers. Deployed on the blockchain, it is provided in the form of a smart contract to verify and verify the identity of assets in a decentralized manner.

3) Meta Data: Provides a metadata format for assets based on DDO (DID Document Object) schema linked with DID registry. Distributed on the blockchain, it is provided in the form of a smart contract that can store and verify the distributed attribute information of the identifiers of assets in a decentralized manner.

4) DAuth: The data provider records and manages the consent obtained from the data owner to provide data access services through the storage service. Controls access to data access services according to entitlement policies based on user contracts in conjunction with access control processors. It is distributed on the blockchain and provided in the form of a smart contract that can store and verify user consent details in a decentralized manner.

5) Access control: Access control processing controls access rights according to the qualification policy for assets registered in AI-Data Commons. In conjunction with DAuth, the access control method is operated according to the data owner's consent to use personal information. It is distributed on the blockchain and provided in the form of a smart contract that can store and verify qualification policy details in a decentralized manner.

6) Transaction: Transaction processing is a distributed ledger-based decentralized transaction system that is provided in the form of a smart contract and exchanges tokens between transaction parties. Upon completion of the transaction in which the asset consumer delivers the token to the asset provider, the AI-Data Commons system grants the consumer access to the purchased asset. After access to the asset is completed, records of access and use of the asset are stored in the blockchain, and the history can be notified to the asset owner.

C. AI-Data Commons Front End

The AI-Data Commons Front End is a user interface and management and execution tool that corresponds to client applications for consumers of assets (data, storage, compute, and algorithms).

1) Marketplace: The Marketplace provides a front-end to the AI-Data Marketplace of the AI-Data Commons system. Asset providers enter metadata about the assets they want to register and register the assets through the GUI provided by the marketplace. Asset users search for, filter, and purchase the assets they need through the GUI provided by the marketplace.

2) Problem definition tool: The problem definition tool registers problem definitions for social problems that problem solvers want to solve through AI-Data Commons. The problem definition tool connects the assets registered in the problem solver AI-Data Commons to DAG (Directed Acyclic Graph) to configure a workflow for problem solving and provides a GUI to control the execution.

3) Data science tools: Data science tools provide a front-end for AI experts developing AI-algorithms to evaluate the AI algorithms they develop. The data science tool works with the marketplace to provide a GUI to search for and purchase storage services needed for problem solving. AI-Data Commons is a tool that allows you to purchase and run based on purchased storage services and compute services.

IV. Conclusion

In this paper, we present high-level design issues of an AI-data commons framework. In the AI-Data Commons ecosystem, various AI modules and data are shared/distributed in a way that ensures the sovereignty and privacy of AI-data owners through decentralized interactions among stakeholders.

Acknowledgment

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT)(2020-0-00048, Development of 5G-IoT Trustworthy AI-Data Commons Framework)

References

[1] https://datacommons.org.

[2] Fensel D. et al. (2020) Introduction: What Is a Knowledge Graph?. In: Knowledge Graphs. Springer, Cham. https://doi.org/10.1007/978-3-030-37439-6_1.

[3] "Home - schema.org". schema.org. Retrieved 2019-04-01.

[4] Ramasubramanian, Sowmya (21 September 2020). "Google's open source data to study impact of COVID-19". The Hindu. Retrieved 14 October 2020.

[5] ITU News," Introducing 'AI Commons': A framework for collaboration to achieve global impact".Artificial Intelligence — Emerging Trends. Retrieved 7 June 2019.

[6] S. Lim, Y. -H. Suh, D. Park, S. Woo and C. Park, "Design of SW Framework for Trustworthy AI-Data Commons," 2020 International Conference on Information and Communication Technology Convergence (ICTC), 2020, pp. 1883-1885, doi: 10.1109/ICTC49870.2020.9289370.