

Study Group 'AI governance and its Evaluation'
Report on the Session #2 (Phase II)

1. Introduction

The Japan Deep Learning Association establishes study groups as a forum for deepening knowledge and discussing domestic and international policy trends related to artificial intelligence (hereafter AI) and Deep Learning (hereafter DL). This study group, 'AI Governance and its Evaluation,' defines 'governance' as a system of management and evaluation by various actors, and launched a study group in July 2020 to investigate what forms of governance are possible to help build trustworthy AI systems, and the phase II began in September 2021.

In the second meeting (Nov. 11, 2021), Mr. Kenji Hiramoto of the Digital Agency spoke about the government's data strategy in the first half. In the second half, Mr. Hiroshi Mano from Data Society Alliance spoke on his organization's role. This report is a reconstruction of these topics and a record of the discussion.

2. The Japanese Government's Data Strategy

Mr. Kenji Hiramoto of the Digital Agency spoke about the government's data strategy in the first half of the meeting.

The Importance of Data

The Digital Agency was established in September of 2021, and it formulates the National Data Strategy under the direction of the Prime Minister and the Digital Minister.

The Digital Agency's vision for society is to 'digitalize in a people-friendly way, leaving no one behind,' therefore, it is necessary to provide services that suit each user. In addition to the service's data, data about business operators and personal information is required to make this effective, and the importance of this data is rapidly increasing.

Until now, many government strategies have promoted the 'digitalization of procedures,' but it is necessary to also design data to use it for hundreds of years. This is why we decided to create a data management strategy.

The Digital Agency's National Data Strategy

It is essential to draw out as much value from the data as possible to use it effectively. However, there is a high awareness of personal information in Japan, which prevents all data from being fully utilized. The Digital Agency promotes the development of data

platforms that factor in geospatial data to maximize the available data value. In a more specific sense, the government aims to create a registry of primary data in the digital space, while opening up more government data for use. We will also develop tools such as trust infrastructure, data exchange platform, and usage rules centered around privacy and security to form a seamless package and process.

In Japan, areas that have not been fully developed, such as base registries, geospatial data, data standards, and data quality standards, should be focused on. Additionally, the infrastructure for data exchanges needs to be developed and applied to smart cities and public services to showcase this technology and trust infrastructure or DFFT (Data Free Flow with Trust) needs to be developed to ensure the reliability of the data used. Currently, we are promoting our data strategy in a single package format.

[Currently, the world leader in this is Estonia, which leads the world to maintain base registries of corporate data, property data, among other fields that venture companies can easily access through a speedy digital service. This creates an environment that promotes new business startups. It makes gathering human resources and companies with innovative ideas, developing new services, revitalizing the economy, and nurturing services into a positive cycle.

It aims to revitalize human resources, companies, and investors by creating an environment where everyone can use data easily.

Creating a Data Strategy

The agency is currently making incremental progress towards implementing this strategy. It is presently proceeding with rule development, the development of standards, and the development of tools as the first step of this process. This will enhance the amount and types of data available and improve data infrastructure overall.

The next step will be to review the open data catalogue and implement changes so that it is easier to find data. We will also be studying the mechanism for locating non-open data.

Furthermore, we will improve data quality, provide standard tools, and finally realize the creation of a data ecosystem.

The specific details of this are as follows.

➤ **Data Standards**

When implementing data standards, characteristics unique to the Japanese language should be considered. These include the usage of *Katakana*, unusual

punctuation marks and the regional ways of writing people's names and addresses. We are trying to improve efficiency by creating reference data models that contain common code values and item descriptions. We are using the data models to improve the efficiency of smart cities data, disaster management systems as well as government documents and develop data standards.

➤ **Establishing Rules to Build Trust and Assurance**

We are currently considering mechanisms to improve trust in two areas, specifically, social confidence and technical/institutional trust. We are also developing mechanisms to protect personal information, and solidify transaction rules, and rules relating to data sovereignty, such as data reuse.

➤ **Increasing Data Quantities and Types of Available Data**

In addition to establishing base registries, we are establishing new public services by converting existing public data into open data. In the public interest, we are also increasing the disclosure of business data, including in the private sector.

➤ **Initial Targets of the Base Registries**

The base registries work as a digital ledger for society as a whole. It will contain data managed by public institutions and data necessary for community activities. Corporate, administrative and property areas are the areas initially being targeted. Personal information is sensitive and requires more advanced technology for data cleaning, so the application of tools and know-how is necessary.

➤ **Data Quality**

The beta version of the Data Quality Control Guidebook¹ has been released, and public sectors and industries are recommended to read it. It will be necessary for data distribution for quality and traceability to be ensured in the future.

➤ **Improving tools**

Currently, a data exchange platform is being developed to evaluate the quality of data while also developing tools to prevent the inflow of low-quality data

➤ **Creating an Ecosystem**

A mechanism for creating a data ecosystem that accelerates growth and is sustainable that works for the entire cycle between collection and utilization is necessary.

➤ **Smart Cities as Showcases**

We believe that smart cities are a perfect opportunity to showcase these plans and use them as proof of concept for verifying the fusion of the digital world and the real one. This Digital-twin world will enhance administrative services and the

¹ Data Quality Control Guide Book (βversion) (Provided in Japanese),
https://cio.go.jp/sites/default/files/uploads/documents/data_hinshitu_guide_beta_20210831.pdf

environment.

Planning for the Future

Below is our timeline for meeting our 2030 target for essential data and developing a global environment.

- 2022
Development of Data Standards, implementing early projects, preparation of catalogues.
- 2025
Fully Servicing Proceeding Projects, Maintaining Tools, Maintenance of Operation systems.
- 2030
Sustainable data eco-system. It will be seamless environment to the world.

3. Promoting Data Linkages in Infrastructure

Mr. Hiroshi Mano of the Data Society Alliance spoke about his organization in the second half.

What is the Data Society Alliance² (DSA)?

The DSA is a general incorporated association chaired by Professor Koshizuka of the University of Tokyo, with representatives of IT-related companies as its directors and academic experts as its advisors. Currently, 141 domestic companies, including life insurance companies, are participating in the DSA, and local governments' participation has been increasing in relation to smart cities.

In 2017 the Data Trading Alliance (DTA) was established to discuss the development of laws related to the data trading market. In 2018, following the enactment of the Basic Act on the Advancement of Public and Private Sector Data Utilization, the Data Platform Consortium (DPC) was established to hold discussions about the implementation of this law. In 2021, it was seen in Japan data strategy, that there was an urgent need to improve its data infrastructure. In order to unify the discussion, the DTA and the DPC were merged to form the DSA.

The DSA aims to create a world where everyone can use diverse data and continually develop new ideas and innovations. The democratization of innovation will enable everyone to innovate, and therefore, we are developing technologies, infrastructure and international standards to support this.

² The Data Society Alliance, <https://data-society-alliance.org/>, <https://data-society-alliance.org/>

Initiatives and activities of the Data Society Alliance (DSA)

We use a multi-focus approach to resolve common data issues with government, private and academic stakeholders. Committees are created around addressing each issue. Their responsibilities are as follows.

- **Technical Standards Review Committee**
Create a framework and guidelines for data catalogue formats.
- **Operation Standards Review Committee**
Establishing and publishing operational standards that act as rules for the data trading market.
- **Accreditation Committee**
Having an external committee to certify and review the data market.
- **International Standards Promotion Committee**
This committee is responsible for working with the international standardization for Society 5.0 as well as developing the standard specifications of the IEEE's P3800 data trading system.
- **Demonstration and Research Committee**
Develop a testbed environment and conduct demonstrations and experiments. Additionally, this committee will function as a point of contact for local governments and as a place to exchange ideas.

DATA-EX³ and the Data Trading Market

For a digital society, data is its life's blood, and the distribution of this data is the circulatory system. Innovation is expected to increase along with the increase in data flow, just as a person's creativity increases when blood flows to their brain.

We hope to connect with various players such as data markets and information banks and use base registries to connect data through governance. At present, we believe that four significant problems hinder data sharing, and we propose that the data trading market is the means to solve them.

1. User and provider anxiety
 - ✓ Using third party mediation to eliminate anxiety.
2. Lack of common value recognition leading to discontentment
 - ✓ Using the marketplace to resolve this.
3. Lack of incentives
 - ✓ Providing incentives through fair transactions.
4. Lack of opportunities
 - ✓ Expanding new opportunities.

³ DATA-Ex is a brand name for the DSA's platform for multifield data linkages.

The data trading market operator is different from other market operators. It does not collect, retain, process, or sell data by itself but only provides a place, a payment function, and credit. The data transaction market makes the existence and value of data visible, ensuring the security of data distribution.

Japan is the first country to advocate the data transaction market model. In Europe and the United States, the focus remains on data collection rather than trading or third-party mediation. For international standardization, we propose that we do not limit the target data. We also note the necessity of unifying the format and vocabulary of the data catalogue and standardizing the data quality evaluation procedure from a technical standpoint.

One crucial issue is to make the originality of the data clear. This means that a map of the data, its terms of use and the fact that it cannot be exclusively owned are attached to and transmitted with the data. Therefore, establishing the sovereignty of the data and integrating this along with its usage conditions into the data's contract so that it can be exchanged is essential.

There is also the concept of data rights that needs to be considered. By linking usage rights and usage conditions to the data when distributed, traceability and the guarantee of original rights can be ensured. By embedding data usage rights into the trading market, we believe that it will be possible to create Data Free Flow with Trust (DFFT).

We hope to create a world where innovation occurs sustainably through data utilization through the above initiatives. This is the DSA's vision.

4. Comments from the Study Group Participants

In the second session, social infrastructure relating to data distribution was discussed. Below are the questions and answers pertaining to this.

Discussion on "The Governments Data Strategy" by Mr. Hiramoto

- Fostering and Protecting Human Resources
 - ✓ It isn't easy to secure talent who not only has experience in data but can collaborate effectively with governments, agencies and the international communities, so we are always recruiting. To combat this lack of people who can think about a distribution environment, we focus on training and collaborating with external organizations. We are constantly recruiting based on ability; however, we have a high percentage of male applicants due to the IT industry's characteristics.

- Regulations relating to data distribution and international collaboration
 - ✓ As data strategy is considered to be global and without borders, we are considering a set of international standards. Rather than being the endeavor of a single country, the current discussions are from a global perspective. While we are not considering restricting open data, we must think about regulations surrounding non-open data for the future.

- What kind of assistance is needed by local governments and the private sector?
 - ✓ We believe there would be a high efficacy in standardizing data reports and showcasing this in smart city initiatives. We also believe that the standardization of case study formats is essential, and we are proceeding to do so while also changing the style we ask questions to be more interactively. This initiative has allowed the Small and Medium Enterprise Agency to collect information effectively. As always, we would like to continue engaging in the Civic Tech initiatives, which collaborates with the people in local communities.

- How user monitoring is conducted
 - ✓ There are many ways to share data. For example, time-limited or temporary access can be granted. How is the given data used and how it should be managed is something the market should decide; however, how it will be controlled needs to be agreed on beforehand.

- Management of comprehensive area data such as for human flow
 - ✓ Active observation and management of user needs and data are necessary, and we believe this is best done in a decentralized format. However, the national government's role is to collect, monitor, and manage metadata. While the cost burden should be assessed on a case-by-case basis, and the government should bear the responsibility for items deemed social infrastructure, there is justification for a user pays system in some areas.

- Personal health data
 - ✓ Points that need to be further discussed in the future are not limited to health data but also Personal Data Stores and lifelogs. For Example, such an idea implies that education records belong to academic institutions.
 - ✓ For example, in a smart city where multiple services may be operating using personal data and their services may not be integrated, allowing personal data control to individuals may lead to innovation.

Discussion on Efforts to Promote Data Linkages with Mr. Mano.

- Fostering and Recruiting Human Resources
 - ✓ Due to the division of labor in the software industry at the moment, we only have the knowledge of our own small area to work with presently and this creates limits. Currently, there are few people in Japan with double degrees and triple degrees, which presents an educational problem; we believe a lack of recognition of diversity causes this. There is also the problem of few women in the field in Japan, and while following the rules, we hope to address this gender imbalance. This can be done through macro balancing and creating gender balance ratio rules.

- Regulations relating to data distribution and international collaboration
 - ✓ Unfortunately, while Japan is unlikely to have the data to compete with companies like Google, there is a great benefit for small and medium enterprises in many industries in collecting and sharing data. These efforts need to include China, South East Asia and Europe. However, the involvement of national governments is inevitable, and diplomatic issues such as the Great Firewall of China must be discussed.

- What kind of assistance is needed by local governments and the private sector?
 - ✓ While we tend to look at many example cases in Japan, we need to do more than this ; we need to actually innovate. What was sort of international standardization discussion was use-case scenario, or creating the base scenarios with imagination. We often refer to cases, but it is also essential to be task-driven. A single person can only have one viewpoint, and this is very limited, which is why we should look at a problem from as many points of view as possible, and this will lead to new ideas.

- The impact on AI of changing data collection methods
 - ✓ The mechanism of data trading allows for data to be exchanged by different entities in return for assets. The data exchanged between stakeholders is a partial or complete image of the data set referred to as an object. There are two kinds of data, metadata and incidental data. When making a value judgment, it is essential to consider datasets, usage conditions, and incidental information.

- Monitoring how data is used by the users

- ✓ Data cannot be owned exclusively. While data can be transferred, it is just a copy of the original, and while access can be granted, regulating it is difficult. While technology such as encryption and access rights management programs do exist they have an associated cost.
 - ✓ In the current world, the sale of parts convertible to weapons are forbidden from being exported without permission. This certificate for the possession of weapons corresponds to data usage certificates, and a similar system could be implemented for data. However, a mechanism for public institutions to participate would be needed.
 - ✓ We believe that mediation by the independent institution can also alleviate concerns about black market data trading and serve as a deterrent.
- Management of vast area data such as for human flow
 - ✓ When we consider data as an expression of the environment and humanity, we can standardize it for use. The rights of the person being observed (such as portrait rights) rather than the creator of the data set need to come first. The person who created the data should be responsible for managing it.
- Personal health data
 - ✓ While CT scan data can be deleted, this sort of health data can also be used for research purposes and clinical trials. In this area, the concept of the data belonging to the patient has become the norm and clinics and research organizations need permission before they can use this data.
 - ✓ However, there are two kinds of data present, the patient's individual medical information but also its prescription by the doctor. The idea that is the treatment process is the intellectual property of the doctor implies the medical records may also belong to the doctor. This is one particular instance of data ownership that is unresolved currently.

The discussion of AI Governance domestically and internationally will continue through this study group.

(Written by Rie Ishii)
(Translated by David Shield)

<The 2nd Session of the Study Group>

Date/time : November 11th (Thursday) 10:00-12:00 (On Zoom)

Contents :

- Presentation 1: The Japanese Government's Data Strategy by Mr. Kenji Hiramoto,
Director of Data Strategy, the Digital Agency
- Presentation 2: Introducing the Data Society Alliance by Mr. Hiroshi Mano, Director
General, the Data Society Alliance
- Questions/Discussion