Data Sharing and Utilization Technologies for Safe and Secure Value-creation Processes

Tomoaki Washio, Yoshinori Orime, Tetsushi Morita, Koji Chida, Kazuo Morimura, and Yoshihito Oshima

Abstract

To achieve Society 5.0, a vision of a future society that achieves both economic development and resolution of social issues, it is essential to use data across organizations and industries, but this is not as widespread as it should be. This article outlines the risks associated with data utilization and their causes and introduces a new paradigm of data sharing and the platform and key technologies to achieve it.

Keywords: data sharing, security, trust

1. Society 5.0 and cross-domain data sharing

The Japanese government proposed "Society 5.0," as a vision of the future society that achieves both economic development and resolution of social issues [1], and efforts to achieve it are actively being made. In Society 5.0, cyberspace and the physical space will be highly integrated, and by making full use of various data, it will become possible to understand the situation at that time, discover issues, predict the future, and derive optimal solutions, which will lead to economic development and solve social problems. In other words, the success of Society 5.0 depends on how much data across organizations and industries (i.e., cross-domain data sharing) can be used.

2. Barriers to cross-domain data sharing and its causes

However, cross-domain data sharing, especially the sharing of sensitive and rare data, has not progressed as much as expected. The largest barrier is the risk to both the data provider and data user.

Data providers take the risk that the data they provide may be leaked or used in an unexpected way, causing damage to themselves or others. Data users, on the other hand, take the risks associated with confidentiality management of the data provided and the risks associated with the legality of such data.

From another perspective, data users take the risk that the data they receive will not produce the desired results or value or may not meet the conditions imposed for receiving the data (for example, the payment for receiving the data). Data providers take the risk that the data they provide will create more value than expected (i.e, the payment they receive is undervalued) or may not fulfill the purpose of providing data for the data provider (for example, not receiving the obtained results from data utilization) when the expected results are not obtained.

The root cause of the above risks is that traditional methods of data sharing involve passing on or receiving the data. Data can create different values and problems depending on how the data are used; therefore, it is difficult to determine all values and problems beforehand. With all this uncertainty, cross-domain data sharing cannot be actively carried out.

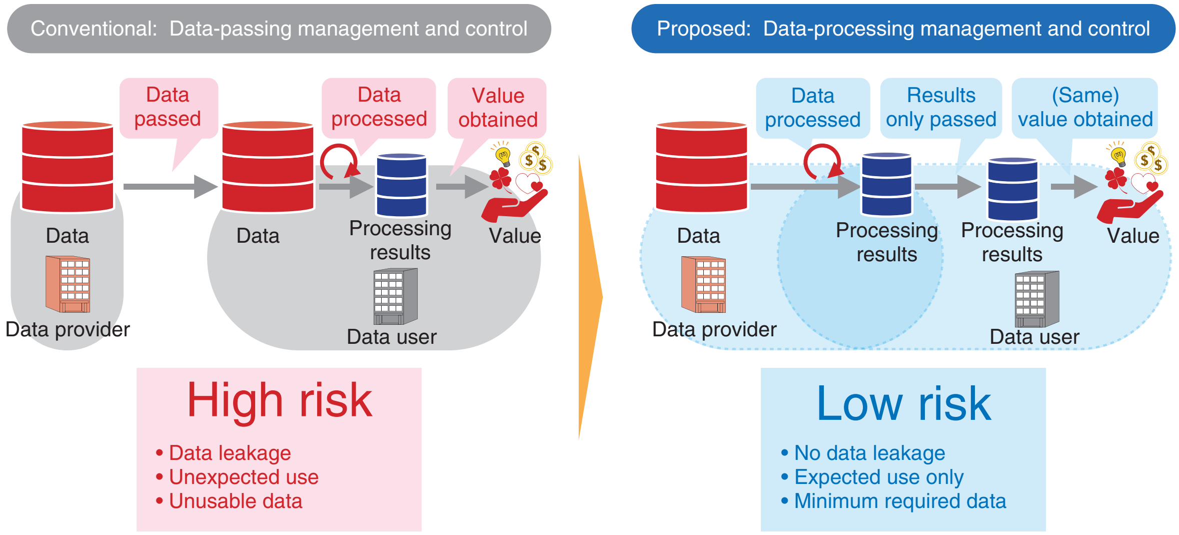


Fig. 1. Paradigm shift in data-sharing methods.

3. Toward a new paradigm of data-processing management and control

To solve this problem, it is necessary to shift from the conventional to a new paradigm of data-processing management and control. In this new paradigm, on the basis of the principle of data minimization, data users are only given the results of agreed processing (the processing data users request and the processing data providers approve) of the data, not the original data. By doing this, the value and problems that can arise from the use of those data are limited and predictable, and the risks mentioned above are greatly reduced (**Fig. 1**).

4. Overview of cross-domain data-sharing platform

Our goal is to create a trustworthy[*1] *cross-domain data-sharing platform* based on the above new paradigm so that all data providers and data users participating in the platform can share and use all data with confidence. The platform's three main requirements are (1) to be able to execute the necessary processing for the purpose while protecting the data, (2) to be able to control the processing of the data in accordance with the agreements and laws between the data provider and data user, and (3) to be transparent and accountable about the data handled and their processing (**Table 1**).

We believe the platform should consist of the following three mechanisms (**Fig. 2**):

(1) Data protection and utilization mechanism

- Executes various processes such as analysis or transformation required by the data user on the data while keeping the data (containing derived data[*2]) confidential.
- Ensures that only the processing permitted by the data-processing-authorization mechanism is executed as permitted and proves the facts of data processing, processing history for certain data, and origin of the derived data.

(2) Data-processing-authorization mechanism

- When a request for data processing is made to the data protection and utilization mechanism, it is determined whether to accept the request on the basis of the data usage policy[*3] specified by the data provider, agreement between the data provider and data user, and legality of the request.
- Prior to determining the above, the data provider,

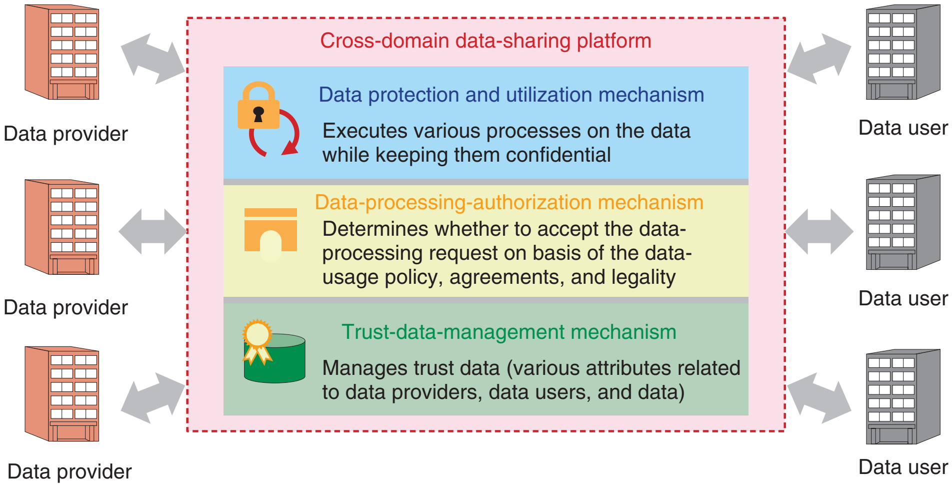
[*1] Trustworthy (platform): To earn the trust of data providers and data users, the platform should be able to demonstrate that it is doing exactly what data providers and data users expect it to do.

[*2] Derived data: The result of processing data entrusted to the platform by the data provider. Includes the results of processing derived data.

[*3] Data usage policy: The policy specifies what kind of data usage shall be permitted. The target of authorization is specified using the attributes of the target data and the requesting entity, how to use the data (method for processing), etc. Permitted/prohibited, or conditions to be satisfied to permit, are also specified to each target.

Table 1. Key requirements for cross-domain data-sharing platform.

Classification	Key requirements
Protect data storage and processing	Executes various analysis, transformation and other processing requested by data users while keeping the confidentiality and integrity of data at a high level.
	Also handles the results of data processing (derived data) as described above.
	Restricts platform operators from viewing or modifying data or derived data.
Managing and controlling data processing	Executes only data processing agreed between the data provider and data user.
	Depending on the target data, executes only processes that conform to laws (Personal Information Protection Law, etc.)
	Authorizes after confirming the propriety of data providers, data users, and target data involved in the requested data processing accurately and in detail.
Ensure transparency of data processing	Enables data users to confirm the characteristics and origin of data and derived data.
	Enables data providers and data users to confirm the fact and history of processing performed on data and derived data.
	Enables data providers and data users to confirm the operation of platforms regarding data storage and processing.


Cross-domain data-sharing platform
Data protection and utilization mechanism
Data-processing-authorization mechanism
Trust-data-management mechanism
Data provider
Data provider
Data provider
Data user
Data user
Data user

Executes various processes on the data while keeping them confidential

Determines whether to accept the data-processing request on basis of the data-usage policy, agreements, and legality

Manages trust data (various attributes related to data providers, data users, and data)

*Functions such as data retrieval, matching, billing, and settlement are omitted.

Fig. 2. Cross-domain data-sharing platform architecture.

data user, and target data are confirmed on the basis of accurate and detailed information on them from the trust-data-management mechanisms.

(3) Trust-data-management mechanism

- Manages various attributes related to data providers, data users, and target data (i.e., trust data) and provides them to the data-processing-authorization mechanism and those who need them.
- Manages information that proves the fact and history of data processing by the data protection and utilization mechanism and the origin of derived data as trust data as well and provides those trust data to those who need them.

5. Key technologies for cross-domain data sharing

The following is an overview of the key technologies for implementing the above mechanisms.

5.1 Secure computation

Secure computation is a technology that enables computation while keeping data encrypted. This enables secure use of sensitive data such as personal data or corporate trade secrets. It also should enable the analysis of such data to extract new value by increasing the types and amounts of data to be analyzed. In secure computation technology, not only safety but also performance and diversity of processing are important. NTT Secure Platform Laboratories developed a secure computation system with the world's fastest statistical processing [2]. We are currently developing secure computation artificial intelligence (AI) technology for executing learning and prediction processing in deep learning while keeping the input data and neural networks encrypted [3]. We are also working on increasing the speed of processing and expansion of the AI algorithms needed to process large amounts of data such as high-definition images.

5.2 Enhancing confidentiality and authenticity when data processing

Because cross-domain data sharing processes data in a variety of stakeholder computing environments, it is not easy to ensure data access control (ensuring confidentiality) or the authenticity of data-processing results. However, as mentioned above, cross-domain data-sharing platforms must meet these requirements to be trusted and widely used. Specifically, we believe the following two requirements are particularly important: (1) the data provider can check whether the provided data are being used only within the permitted operations and (2) the data user can check whether the data-processing results are correct as requested.

5.3 Attribute-based authorization

Because large amounts of data are registered continuously on the platform and it is not known in advance who is going to use the registered data, attribute-based authorization is effective for cross-domain data sharing. Regarding attribute-based authorization, data-usage policies are defined in the form of conditions for what kind of subject can use what kind of data using the attributes of the subject and data. When authorization decisions are made, the attribute values of the data users and target data are applied to the conditions and evaluated. For example, a data user's qualification (e.g., Information Security Management Systems Certification) is specified as a required attribute in the data-usage policy.

Attribute-based authorization is not a new concept, but there are technical challenges when applying it to cross-domain data sharing. To make authorization decisions that take legality into account, it is necessary to check and determine the existence or value of specified attributes in accordance with the characteristics of the data, regardless of the data-usage policy, as described later. To avoid making an incorrect authorization decision, it is necessary to strictly determine the existence of the attribute specified in the data-usage policy and its authenticity (whether it has been certified by a third-party organization, etc.). It is also necessary to have flexibility in authorization decisions, such as conditionally allowing data usage in addition to permitting or prohibiting it. For example, the requested data may not be used as is but may be used after being anonymized. We believe that making flexible authorization decisions while coordinating the requirements of both data providers and data users will contribute to expanding data-sharing opportunities (**Fig. 3**).

5.4 Trust-data management

The trust-data-management mechanism collects and manages attributes of platform users (qualifications, achievements, reputation, etc.) and attributes of data processed on the platform (data characteristics such as type, items, collection method, origin, and processing history). It then provides them to the data-processing-authorization mechanism, data providers, and data users as "trust data" that can be relied on when users and data are confirmed (**Fig. 4**).

Because trust data are the basis of trust, it is necessary to be able to verify that such data have been guaranteed by various entities regarding the data's authenticity and correctness. However, trust data may be confidential; thus, it is necessary to appropriately determine when, to whom, and to what extent to disclose trust data. It is also required for the data to be available whenever needed. We will develop trust-data-management technologies and mechanisms that meet these various requirements.

5.5 Implementing legal requirements for trust data and authorization decisions

Data providers and data users must also consider the legality of the data they share. It is difficult for data users to understand how the data were obtained and might not notice that illegally obtained data were included. The Act on the Protection of Personal Information prohibits the use of data that constitute personal information outside the scope of the purpose of use indicated to the person at the time of acquisition and the provision of such data to third parties without the person's consent.

To reduce the burden of legal compliance, we are studying a method of assigning attribute data that

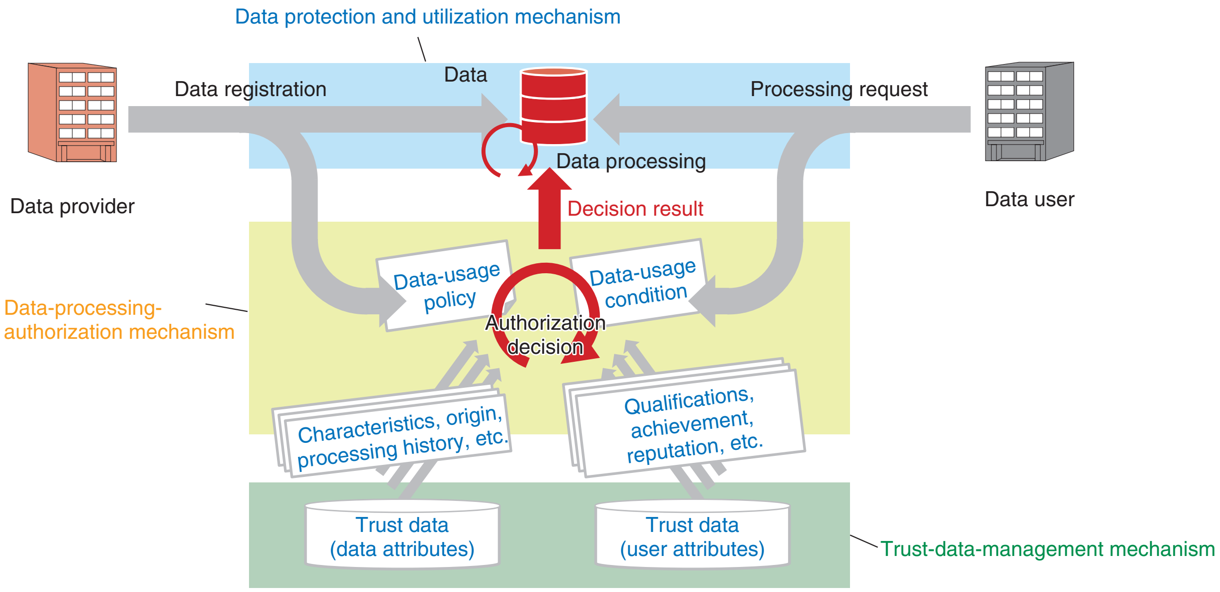

Data protection and utilization mechanism
Data
Data registration
Processing request
Data provider
Data user
Data processing
Decision result
Data-usage policy
Data-usage condition
Authorization decision
Data-processing-authorization mechanism
Characteristics, origin, processing history, etc.
Qualifications, achievement, reputation, etc.
Trust data (data attributes)
Trust data (user attributes)
Trust-data-management mechanism

Fig. 3. Attribute-based authorization.

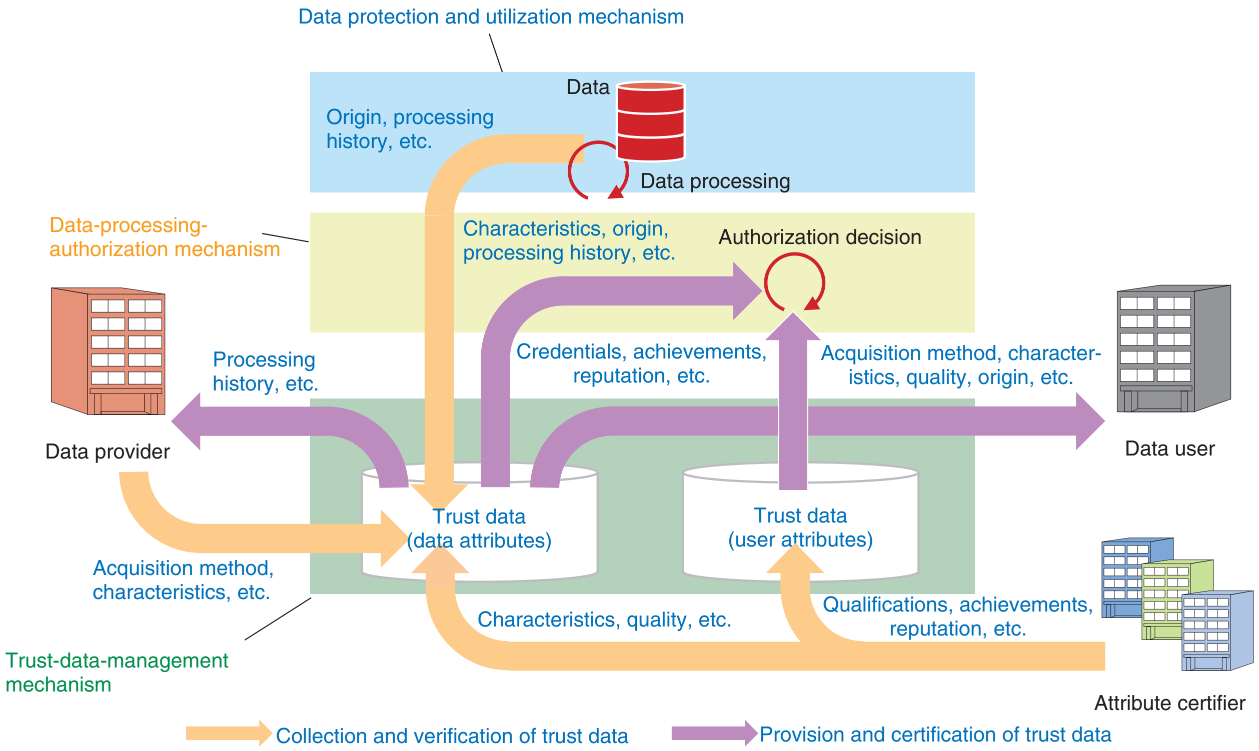

Data protection and utilization mechanism
Data
Origin, processing history, etc.
Data processing
Data-processing-authorization mechanism
Characteristics, origin, processing history, etc.
Authorization decision
Data provider
Processing history, etc.
Credentials, achievements, reputation, etc.
Acquisition method, characteristics, quality, origin, etc.
Data user
Trust data (data attributes)
Trust data (user attributes)
Acquisition method, characteristics, etc.
Characteristics, quality, etc.
Qualifications, achievements, reputation, etc.
Trust-data-management mechanism
Attribute certifier
Collection and verification of trust data
Provision and certification of trust data

Fig. 4. Flow of trust data.

implements legal requirements together with data, managing the data as trust data, and checking the data when making authorization decisions.

6. For the future

We believe that a trustworthy cross-domain data-sharing platform will accelerate collaboration and co-creation among the same or different industries, which were previously difficult, and enable the creation of new value and resolution of major social issues such as those regarding food, health, and the environment.

To achieve this, we will accelerate not only research and development of key technologies but also proof of concept in cooperation with partners aiming at cross-domain data sharing, international standardization, and activities to foster social acceptability.

References

[1] Cabinet Office, Government of Japan, "Society 5.0," https://www8.cao.go.jp/cstp/english/society5_0/index.html

[2] NTT press release, "Trial Service of Secure Computation System San-Shi®," Aug. 8, 2018. https://www.ntt.co.jp/news2018/1808e/180808a.html

[3] NTT press release "Secure Computation for a Typical Training Algorithm of a Deep Neural Network," Sept. 2, 2019. https://www.ntt.co.jp/news2019/1909e/190902a.html

Tomoaki Washio

Senior Research Engineer, NTT Secure Platform Laboratories.

He received a B.E. and M.E. in systems and information engineering from Hokkaido University in 2000 and 2002. Since joining NTT in 2002, he has been engaged in research and development (R&D) of authentication systems and secure data-sharing technologies.

Koji Chida

Senior Research Engineer, Supervisor, NTT Secure Platform Laboratories.

He received a B.S. and M.S., and Dr.Eng. from Waseda University, Tokyo, in 1998, 2000, and 2006. Since 2000, he has been engaged in research on cryptography and privacy-enhancing technologies at NTT. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and Information Processing Society of Japan (IPSJ). He received the IPSJ Best Paper Award in 2012.

Yoshinori Orime

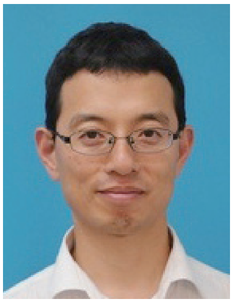
Senior Research Engineer, NTT Secure Platform Laboratories.

He received a B.E. in computer science from Saitama University in 1998 and joined NTT the same year. Since 2020, he has been engaged in research on incorporating legal and social requirements into technical requirements.

Kazuo Morimura

Senior Research Engineer, Supervisor, NTT Secure Platform Laboratories.

He received a B.A. from Aichi University of Education in 1993 and M.E. from Nara Institute of Science and Technology in 1996. He joined NTT in 1996 and is currently conducting R&D on data protection technologies centered on cryptography.



Tetsushi Morita

Senior Research Engineer, Supervisor, NTT Secure Platform Laboratories.

He received a B.E. and M.E. from Kyoto University in 1996 and 1998 and a Ph.D. in engineering from University of Tsukuba, Ibaraki, in 2010. Since joining NTT Software Laboratories in 1998, until 2012, he was engaged in research on information retrieval and personalization systems. His current research interests include secure data-sharing technologies.

Yoshihito Oshima

Executive Research Engineer, Supervisor, NTT Secure Platform Laboratories.

He received a B.E. and M.E. in electrical engineering from Hokkaido University in 1994 and 1996. He joined NTT in 1996 and is currently conducting R&D on secure data-sharing technologies. He is a member of IPSJ.