Data Governance for Achieving Data Sharing in the IOWN Era

Katsuhiko Suzuki and Daigoro Yokozeki

Abstract

There is a move around the world to tighten the handling of data, but there is also a move to establish technologies and rules for processing data safely to accelerate data sharing. After surveying global trends that are having an impact on data sharing, this article describes the concept behind data governance for controlling data so that data owners can safely share their data with others and describes the requirements for data governance in the IOWN (Innovative Optical and Wireless Network) era.

Keywords: data sharing, data governance, data trust

1. Background

The spread of Internet of Things (IoT) devices and progress in digital transformation (DX) activities are driving the digitalization of even information that up to now could not be easily handled while dramatically increasing the volume of that data. There is talk about the coming of a totally new Smart World that will ultimately be achieved by fusing the physical space composed of things and people and cyberspace that reproduces society in its entirety as digital twins, analyzing the links between the two, and feeding back analysis results to the physical space (**Fig. 1**). This will require data sharing beyond individuals, business fields, and industries, and since such data are bound to include sensitive information, there will also be a need for a mechanism that can support safe and secure data sharing. We first outline global trends that are having an impact on data sharing. We then describe the concept of data governance mostly from the viewpoint of security as an essential means of supporting data sharing in the Innovative Optical and Wireless Network (IOWN) era.

2. Trends in data sharing

Trends in data sharing include moves to tighten and limit data sharing as well as moves to promote data sharing through the formulation of agreements, etc. (**Fig. 2**).

2.1 Tightening of data management as in personal information protection

Movements to protect personal information are progressing in many countries. The General Data Protection Regulation (GDPR) in Europe is well known as a general regulation covering data protection, but the California Consumer Privacy Act (CCPA) in the United States, and in Asia, China's Personal Information Protection Law, and other personal information protection laws in Korea, Thailand, India, etc. have also been enacted. Therefore, there is a growing demand for stricter management on how information related to individuals is shared.

In Europe, issues related to data collection by hyperscale companies (hyperscalers) are rising to the surface and the need is growing for strengthening the protection of diverse types of data including non-personal data. In the face of this trend, efforts in constructing federated data ecosystems that place particular importance on transparency and trustworthiness are moving forward by organizations such as the International Data Spaces Association (IDSA) [1].

2.2 Appearance of diverse data spaces

Moves toward the construction of data spaces with a focus on Europe have begun as efforts to stimulate data sharing (**Fig. 3**). A data space refers to a

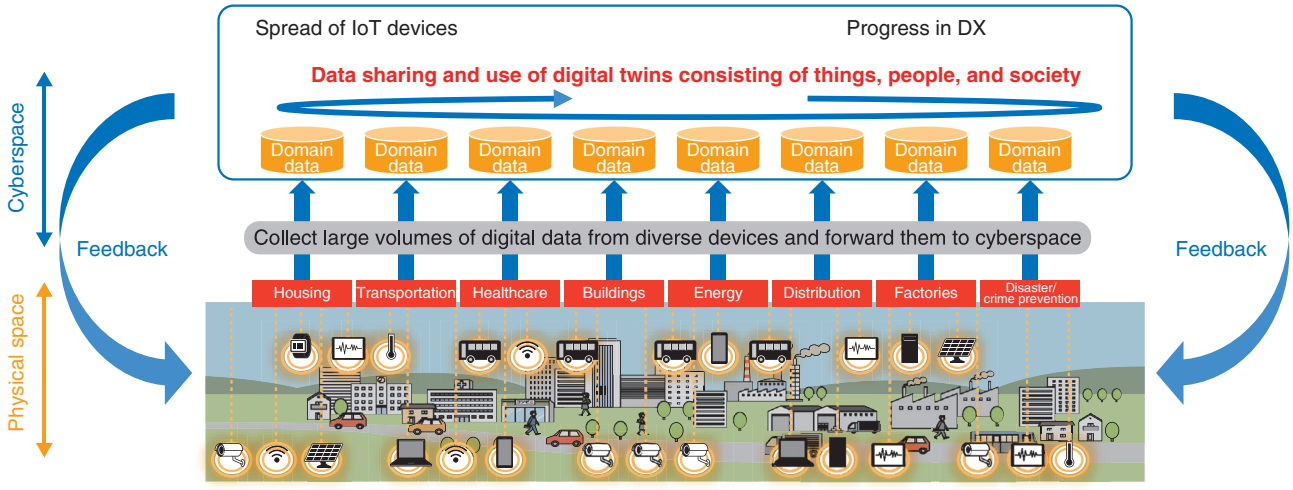

Fig. 1. Data sharing in the IOWN era.

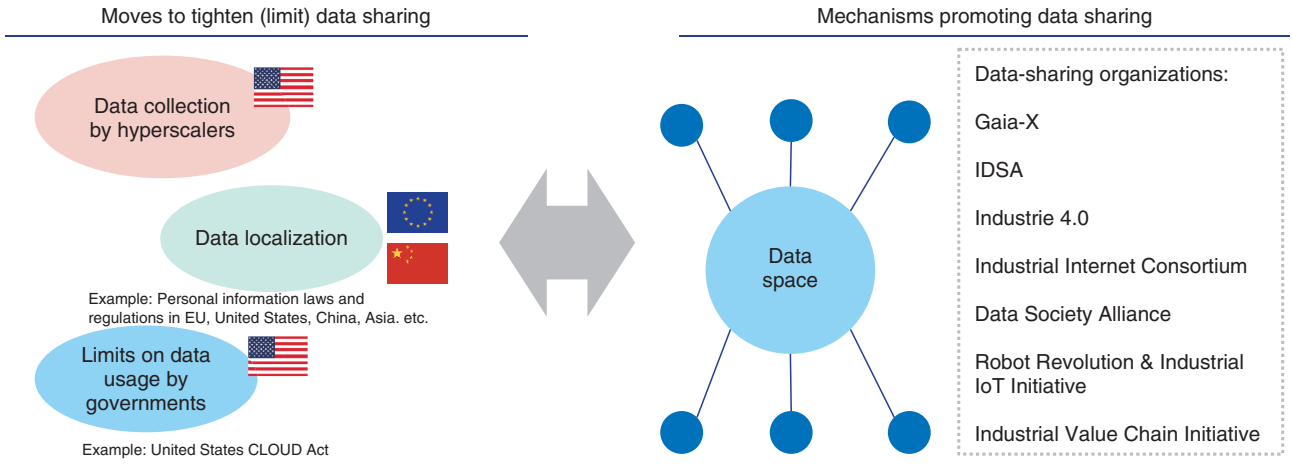

Fig. 2. Global trends in data sharing.

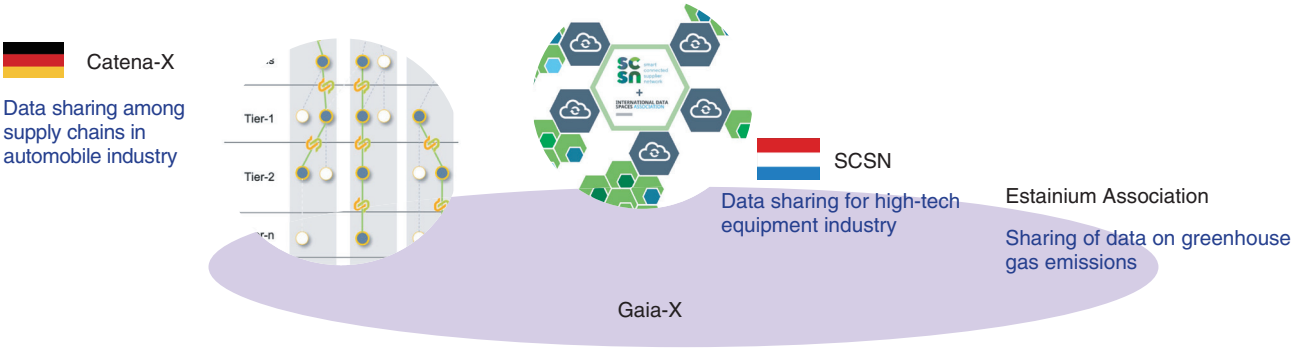

Fig. 3. Appearance of data spaces.

community having systems and rules for achieving highly secure data sharing. In Europe, the construction of industry-specific data spaces on a scale of 1000 companies has begun centered about the manufacturing industry. Companies, as a rule, do not like to release their data, and even when they do, they are apt to set requirements such as the tracing of data-usage history. Thus, against the background of emerging standards for the secure handling of data such as IDSA, many data spaces are being formed. These include well-known data spaces such as Catena-X in Germany that shares data among its entire automobile supply chain with the aim of bolstering the competitiveness of the country's automobile industry [2], Smart Connected Supplier Network (SCSN) that aims to share complex, small-volume, and diverse parts data among high-tech equipment manufacturers in the Netherlands [3], and the Estainium Association that aims to share data on greenhouse gas emissions among companies on supply chains in a cross-industry manner on a blockchain-type open platform [4]. Country-specific and industry-specific data spaces are now appearing, so the above movement is expected to accelerate. We have introduced data spaces centered about companies, but considering the need for exchanging data in a secure manner within social media, within the metaverse, etc., we can think of "data space" in a broader sense.

2.3 Integration and connection of data spaces

The integration and connection of these emerging data spaces can also be seen. For example, there is a movement related to battery regulations in Europe to manage the history of storage batteries from material procurement to recycling in a system called "battery passport" with the aim of reducing greenhouse gas emissions [5]. The tracing of battery history across multiple data spaces in different countries requires the integration and connection of those data spaces. This initiative is already underway in the automobile industry, but expanding it to other industries in relation to batteries is expected to further increase the opportunities for integrating and connecting data spaces across multiple industries.

2.4 Resilient and global supply chain

The COVID-19 pandemic, natural disasters, and international developments have generated many instances of short-term and frequent rearrangements of supply chains. Cases can be found in each industry of delays in procuring materials and decisions to change suppliers, and the impact of an insufficient supply of semiconductors on a wide range of industries is probably still fresh on everyone's mind.

As a trend that runs counter to conventional globalization, there are moves to support domestic production in relation to critical components from the viewpoint of economic security and configure supply chains only among specific affiliated countries.

From the viewpoint of data sharing, there is a need for supporting such rearrangement of supply chains and facilitate dynamic and smooth data sharing with new business partners.

2.5 Advances in information-processing technologies

Finally, we can point out advances in information-processing technologies such as quantum computers, the fifth-generation mobile communications system (5G), 6G, and IOWN. The coming of ultra-wide bandwidth, low-latency, and low-power-consumption networks will drive new demand for remote medicine, self-driving cars, drone control, and other novel services, so there will be a need for mechanisms that can securely handle sensitive data that could not flow on networks in the past. By significantly decreasing data transportation costs, the appearance of new data-processing architecture such as the Internet of distributed datacenters and seamless edge-cloud integration can be expected. Going forward, there is certainly a need for data processing that can achieve both efficiency and security.

3. What is data governance?

Although there is no standard definition of data governance, which is sometimes called "data sovereignty," it refers to following the policy governing the handling of data as specified by the data owner over the lifecycle of that data from the time of its creation to its destruction. For example, a policy may state that no copy of the data is allowed and that the data must always be in encrypted form. Based on the global trends affecting data sharing as described in the previous section, the following describes the requirements envisioned for achieving data governance in the IOWN era (**Fig. 4**).

3.1 Dispersed and distributed data management

As can be seen from the above examples of data spaces, the forecast is for data to be increasingly distributed and managed as such. There are limitations in collecting data by a single company, and it is expected that the inability to collect data at only one

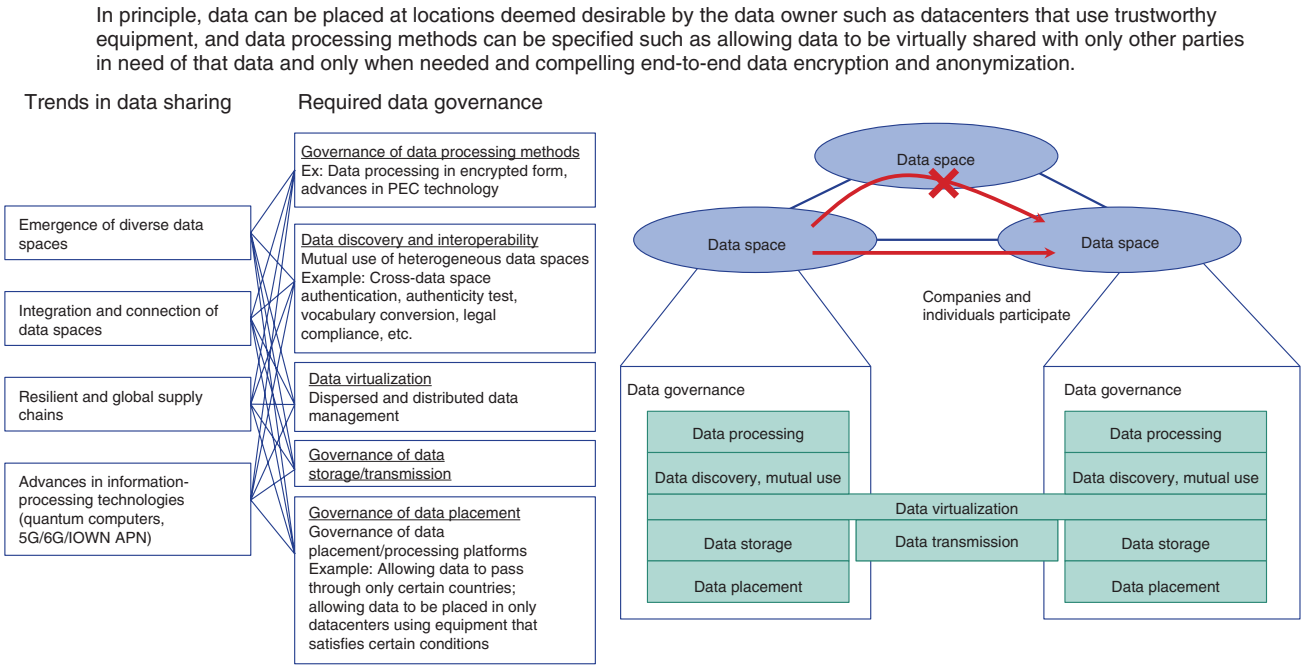

Fig. 4. Targeted data governance.

location will only increase considering differences in legal systems from one country to another. Assuming the widespread of ultra-wide bandwidth, low-latency networks such as the IOWN All-Photonics Network, the cost of moving data should dramatically decrease compared with conventional systems. Data management in the future must, in principle, be able to manage data at locations deemed desirable by the data owner, and when sharing those data with others, be able to do so in a virtual manner according to need. Calculations and processing will be carried out against data collected in a virtual manner after which those data will be destroyed when no longer needed. In contrast to a data lake in which data are collected at one location, technology that makes distributed data appear to be located at one location is called "data virtualization." Data virtualization is currently being used within a relatively small area such as a single datacenter, but going forward, there will be a need for technology that can virtually integrate data in a dynamic manner across countries and data spaces [6].

3.2 Discovery and interoperability of data spaces

The configuration of resilient and global supply chains requires the interoperability of multiple data spaces. Since individual data spaces are generally operated on the basis of various participants and rules, enabling dynamic data sharing with a new party requires a mechanism for mutually connecting and using heterogeneous data spaces. A mechanism for discovering business partners and data is also needed, and once a partner is found, it must be possible to mutually check the authenticity of a partner by some means such as authentication federation between different data spaces. This is because data-space participants, such as individuals or companies, are managed by independent authentication platforms unique to each data space. In the same way, while there are means of checking the authenticity of data within a certain data space using blockchain technology, the need for interoperability between the blockchains of different data spaces should arise. Eventually, there will be a need for even semantic interoperability such as by converting rules and vocabularies used by different communities and regions.

3.3 Governance of data storage and transmission

The prime approach to data protection is encryption by cryptography, which is used as a matter of course when storing and transmitting data to take measures against eavesdropping and other data-related risks. However, there are concerns that

existing cryptosystems such as RSA (Rivest–Shamir–Adleman algorithm) will one day be compromised due to the appearance of quantum computers, so there is a need for cryptography appropriate for the post-quantum cryptography (PQC) era. The National Institute of Standards and Technology (NIST) in the United States is now working on the standardization of PQC technologies. In the near future, there will be a need for data-storage and transmission platforms using those technologies, and the need for making seamless transitions from existing cryptosystems to PQC and for dealing with the compromising of PQC [7].

3.4 Governance of data-processing methods

Although data are encrypted when being stored and transmitted, traditional cryptographic technology requires decrypting those data to plaintext before calculations or other types of processing. A new cryptographic technology that enables calculations on encrypted data without decrypting them has entered a period of practical use. In particular, privacy-enhancing computation (PEC)* is attracting attention as a general term for technologies that protect privacy while processing data, including such new cryptographic technology. The Gartner consulting firm, for example, refers to those technologies in its report. PEC consists of a wide range of technologies including, but not limited to secure computation technology that performs calculations on data in an encrypted form on the basis of homomorphic encryption and secret sharing; confidential computing technology that executes calculations and processing in a trusted area using a trusted execution environment (TEE) enabled by hardware-based memory encryption technology; and data anonymization and differential privacy that reduce the risk of identification and privacy breach of individuals. It can be said that requests made to cloud operators to protect data and privacy formed the background to this growing interest in PEC. We can expect these technologies to enable end-to-end data processing on encrypted data and data utilization without exposing unnecessary privacy information and to be commonly used in the same way as encryption in data transmission [8].

3.5 Governance of data placement/processing platforms

From the viewpoint of economic security, the need must be met for a function that enables the data owner to select data-storage locations and data-processing locations and for a function that enables the range of data distribution to be controlled as in the permissible range of transmission. In this regard, today's Internet is designed with importance placed on efficiency and fault tolerance, which serves as a basis for selecting the data-delivery path. Going forward, however, it should be possible to specify the countries, regions, data spaces, etc. where data storage and processing are permissible. There will be need for a function that specifies the permissible range of data transmission and supports data sharing on the basis of economic security as well as economic efficiency, such as a function that guarantees that transmission range even when using a detour route during a network fault [7].

It should also be possible to place conditions on the cloud that executes data storage and processing and place conditions on the storage equipment, network equipment, and computing equipment to be used as conditions placed on datacenters used by the cloud. An example of the former would be the allowing of processing only on a domestically operated cloud within the country when handling truly sensitive data. Examples of the latter would be the processing of data only on products of certain equipment manufacturers and the allowing of data processing only if the software being used can be tested for any quality problems such as vulnerabilities [9]. Using a software bill of materials (SBOM) for such testing is well known.

4. Future perspectives

Data governance as introduced in this article specifies that data, in principle, can be placed at locations desired by the data owner and that data-processing methods can be specified, such as the sharing of data only with others in need of those data and only when needed and end-to-end data encryption, anonymization, etc. Looking to the future, data governance is expected to evolve toward more secure and flexible control of data and promote an even higher level of data sharing.

* PEC: Generic term for computation technology that fortifies privacy by enabling computations with data in encrypted form, making it impossible to identify the person associated with those data through anonymization, etc.

References

[1] IDSA, https://internationaldataspaces.org/
[2] Catena-X, https://catena-x.net/en/
[3] SCSN, https://smart-connected.nl/
[4] Estainium Association, https://www.estainium.eco/
[5] Ministry of Economy, Trade and Industry, "Data Sharing," July 2022 (in Japanese). https://www.meti.go.jp/shingikai/mono_info_service/chikudenchi_sustainability/pdf/003_03_00.pdf
[6] T. Utahara, T. Fukuhisa, H. Midori, N. Takeuchi, S. Watanabe, and Y. Ikejiri, "Early Deployment and Popularization of IOWN Technology as Targeted by IOWN Product Design Center," NTT Technical Review, Vol. 21, No. 4, pp. 71–76, Apr. 2023. https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202304fa10.html
[7] K. Murakami, A. Taniguchi, F. Kudo, S. Chikara, Y. Kiyomura, A. Mukaiyama, Y. Iijima, Y. Mochida, Y. Sanari, and N. Kimura, "Secure Optical-transport-network Technology in Anticipation of the Quantum Computer Era," NTT Technical Review, Vol. 21, No. 4, pp. 60–66, Apr. 2023. https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202304fa8.html
[8] T. Inoue and T. Morita, "Trusted Data Space Technology for Data Governance in the IOWN Era," NTT Technical Review, Vol. 21, No. 4, pp. 55–59, Apr. 2023. https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202304fa7.html
[9] T. Uehara, Y. Kanemoto, and H. Nomura, "Security Transparency Assurance Technology for Analysis and Visualization of Software Components," NTT Technical Review, Vol. 21, No. 4, pp. 67–70, Apr. 2023. https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202304fa9.html

Katsuhiko Suzuki
Vice President, Head of NTT Social Informatics Laboratories.
He received a B.E. and M.E. in computer science from Chiba University in 1993 and 1995. He joined NTT in 1995 and engaged in the development of a smart card operating system. He is now engaged in the management of the research and development for social transformation and development by combining information science and social science. He is a member of the Information Processing Society of Japan (IPSJ).

Daigoro Yokozeki
Executive Research Engineer / Director, Head of Social Information Sharing Research Project, NTT Social Informatics Laboratories.
He received a B.E. and M.E. in industrial engineering from Waseda University, Tokyo, in 1996 and 1998. He joined NTT in 1998 and conducted research on IT middleware such as database, virtual machine/container, and cloud computing project. He is now in charge of privacy enhancing computation (e.g., TEE/MPC/DP) and data governance to create an innovative future data-sharing world.