Tentative

Translation

The Conference toward AI Network Society

The Committee on AI Economy

Report 2020

July 2020

Introduction

Global trends toward the implementation of AI in society never stop. GAFA in the US and Alibaba and Tencent in China are ranked high in the top 10 companies of the global market capitalization. These companies have developed AI on their own or provided the basic tools for the development and utilization of AI, and take the initiative in the implementation of AI.

Japan should not be late for these trends. It's important to show the expected future image of AI as well as organize the possibilities, issues and concerns of AI and to show that various entities in Japan must work on the problems as soon as possible. The government of Japan announced "AI Strategy 2019—AI for Everyone: People, Industries, Regions, and government" (Decision of the Integrated Innovation Strategy Promotion Council on June 11, 2019), and outlined the measures to be implemented immediately. Then, these measures were implemented based on the strategy, and some achievements have been already made for the advanced implementation of AI in society. These trends must be further accelerated.

On the other hand, what is necessary for AI to exert its potential is data that is expected as "new capital". The report issued by The Committee on AI Economy in May 2019 clarified the thought that data is positioned as a production factor and the effects of data on the improvement of productivity should be estimated is significant, and showed the necessity of deepening the discussion on fair remuneration in data processing.

Regarding how to utilize data, various measures have been developed based on the various strategies in the government, and some Japanese companies have promoted advanced initiatives such as big data analysis and machine learning, and understanding the role of data in production activities has been deepened. On the other hand, how data is specifically connected with the improvement of productivity and how remuneration should be distributed depending on the effects and value of data have not been established yet not only in Japan but also in the world.

To fully extract the value of data by AI, etc., we need to consider how the value of data should be estimated and how the value should be distributed.

In addition, it can be said that the COVID-19 global pandemic brought a new perspective to the consideration of the ideal implementation of AI in society. People are forced to avoid unnecessary outings and reduce opportunities to contact other people by 80%, and review their economic activities and lifestyle substantially. On the other hand, the importance of communication network and ICT tools such as teleworking and electronic application procedures is attracting attention as a solution to the above situations. There is a possibility that the COVID-19 is getting the cause of acceleration of the implementation of AI in society. In G20 Ministerial Meeting on Trade and Digital Economy in Tsukuba, Ibaraki held in June 2019, the necessity of considering a new social model in the age of AI was shared under the concept of creating new employment and industries by AI, and the necessity might be enhanced globally.

We hope the discussion in the Committee on AI Economy held in this timing will be a clue for future initiatives and strategies for all types of players in Japan and overseas.

Contents

Introduction 2
1. Background (History from the establishment of "The Committee on AI Economy" to the present) 4
1.1 Establishment of "The Committee on AI Economy" 4
1.2 "The Conference toward AI Network Society, the Committee on AI Economy Report Outline (May 2019) 4
1.3 The scope of consideration in this report 6
2. The ideal way to utilize data required for the implementation of AI in society 8
2.1 The current status of Japan's data utilization 8
2.2 The ideal way to utilize data required for the implementation of AI in society 18
3. Data economy policy in the AI era 22
3.1 The functions/roles of data as "new capital" and estimation methods of the effects/value of data 22
3.2 The ideal fair remuneration depending on the effects/value of data 40
3.3 The current status and issues toward marketing and visualization of data transaction 46
4. Future image (the image of "Inclusive AI Economy Society") 58
5. Recommendation 66
5.1 The Ideal Policy toward the Promotion of the Utilization of AI/Data 66
5.2 The ideal data economy policy required for promoting the data-driven economy in the AI era 70
6. Addendum 73
6.1 Questionnaire 73
6.2 Empirical analysis 100

Attachment 1 The Conference toward AI Network Society, The Committee on AI Economy / Members of the Special-Interest Group on Data
Attachment 2 Background of Holding the Conference
Attachment 3 Analysis on Prospect of Ecosystem Formed with Progress of AI Network

1. Background (History from the establishment of “The Committee on AI Economy” to the present)

1.1 Establishment of “The Committee on AI Economy”

In January 2019, the Committee on AI Economy (hereinafter referred to as ‘the Committee’) was established with the aim of considering what kind of social economy should be pursued and how the basic policy and medium-to-long term strategies should be through the promotion of the implementation of AI in society among the items to be considered in The Conference toward AI Network Society. The Committee did consideration, focusing on the following 5 key points and the report was outlined in May 2019.

(1) Directivity to be pursued in AI economy (Realization of ”Inclusive AI Economy Society")

(2) Ideal AI investment for sustainable economic growth and improvement of productivity

(3) How the industrial infrastructure that supports AI economy (Labor/R&D) should be

(4) Japanese enterprises’ global competitiveness over the utilization of AI

(5) Ideal basic policies and strategies on AI economy

1.2 “The Conference toward AI Network Society, the Committee on AI Economy Report Outline (May 2019)

(1) Directivity pursued in AI Economy (Realization of “Inclusive AI Economy Society”)

The tools that make the development/utilization of AI possible (frameworks, learned models, etc.) are widely provided by ICT companies as open source or via cloud/API, so various organizations can utilize them and develop/utilize AI for various purposes, depending on their capabilities. To work on the development of AI, it is also important to understand what kind of data should be collected depending on the purpose or how to combine the collected data and the specific knowledge on the on-site work (domains), etc.

It is important to aim for the economy society (“Inclusive AI Economy Society”) where various subjects (including individuals and SMEs) can be actively involved in social and economic activities in an ingenious way based on their will and judgment and get distribution, fulfillment, increased leisure time, etc. as well as contribute to the improvement in productivity by making the open and inclusive operation of AI possible under the fair competitive environment.

(2) Ideal AI investment for sustainable economic growth and improvement of productivity

It can be said that ICT investment in Japan has not been enough quantitatively, and the effects in the improvement of productivity was limited. In addition, ICT investment in which ICT companies are entrusted with development is the main stream, so there is a possibility that complementary investment such as organizational reform was not enough. It is considered that the system centering on ICT companies being entrusted with development had economic rationality to some extent, but there is a possibility that ICT companies that were excessively entrusted with development can’t do effective and enough AI/ICT investment, so it’s also important for user companies to develop and utilize AI by themselves, going forward.

The thought that data should be positioned as “new capital” and its effects on the improvement of productivity should be estimated as one of the production factors is significant because data can be a source to create value by being analyzed by AI. It is necessary to estimate effects/value after evaluating appropriate value, noting that there are issues in estimation

in digital economy of national economic accounting such as GDP.

In addition, the platforms of some overseas companies are used in many countries in the world, and those companies have accumulated a large amount of data on the Internet, and earned profits. Considering this situation, it is necessary to deeply discuss how data ownership should be and the ideal distribution of remuneration to the providers of data that is a production factor, etc. focusing on the thought that the distribution rate of data as intangible capital has increased while labor's share rate has decreased.

(3) How the industrial infrastructure that supports AI economy (Labor/R&D) should be

Some people predict that the technological unemployment in ICT will occur mainly in "routine tasks" such as manual tasks and clerical work, and "non-routine manual work" such as services/physical work will be the source of employment for unemployed people, but other people predict the function of non-routine manual work as the source of employment will be lowered because the pressure of substitution for "non-routine manual work" will further increase due to the development of AI or robotics. In that case, it is considered that the necessity of personnel relocating and the reeducation for that will increase. There is a possibility that new employment will be created due to the development of AI, but it should be noted that the loss of employment due to the technological unemployment will be a big issue as a whole society until the effects can be visible.

In Japan, there is a possibility that there has been the replacement by non-regular employment instead of the replacement by ICT, so it should be noted that the impact of the development and spread of AI in Japan might be bigger than that in other countries, because the technological substitution including substitution of the "routine tasks" will proceed.

In addition, both the quantity and quality of human resources in AI/ICT are insufficient in Japan. In the medium- to long-term, fostering human resources for AI (promotion of appealing the attractions of AI and data and educational reform from elementary education to higher education (including the contents of education, ideal educational sites and securing excellent AI instructors)) is important. And in the short-term, the perspective of how to promote the introduction of AI in Japan (shifting human resources for ICT from ICT companies to user companies, improving the AI skills of these human resources, and accepting high-level human resources who are familiar with AI from overseas and utilizing outsourcing those human resources overseas) is important.

Regarding the research and development in Japan, the perspective of considering industrial issues by connecting them with science and engineering is insufficient, so it is important to link basic research, applied research and development with businesses. In particular, it is considered to be required that large enterprises enhance the diversity of the processing and employment practices of high-level researchers, etc. and make further efforts to link businesses and research and development. In addition, it is also important for start-ups to have a perspective of focusing on research and development.

(4) Japanese enterprises' global competitiveness over the utilization of AI

Japanese enterprises have correlating issues such as the shortage of Internet connection and data amount, inability to process data, lack of human resources for utilizing data. There are many cases where the status of data isn't figured out in a company, so it is the most important for executives to build effective data strategies after figuring it out. To promote the utilization of data, it is important to build a system where newly collected data don't need to be processed as much as possible and promote fostering human resources who can utilize data.

(5) The ideal basic policy and strategy on AI economy

As digital transformation is proceeding, it is important for the government to take this timing when the structure of industries is being rebuilt as a big chance and promote the policy to accelerate this trend, focusing on the following key points.

(a) In-house development by user companies (utilize the base of Cloud, etc.)
(b) Open innovation with start-ups that are conducting R&D (breaking away from "self-sufficiency", cooperation with enterprises that have specialty in specific technologies and entrepreneurs who have flexible ideas, etc.)
(c) Agile-type development (putting services under development into the market, using them, and improving them successively)
(d) Securing real data through the promotion of IoT (collect data that can utilize the strengths of Japan, centering on manufacturing sites and medical/healthcare fields, etc.)
(e) Complementing data amount (initiatives for open data on public data, sharing various kinds of data beyond enterprises and industries)
(f) Improvement of AI skills of human resources of user companies (internal human resources, external human resources in Japan and overseas)
(g) Reeducation of human resources to respond to the impact from the utilization of AI (planned personnel relocating of human resources for routine tasks and reeducation for that)
(h) Improvement of ICT companies' capability to support user companies (strengthening the support for the improvement of AI skills in user companies and the role to complement the shortage of ICT human resources as well as promoting the development of AI after collecting knowledge on on-site work in industries and enterprises)

To promote the policy, the following points should be especially noted.

(a) How to utilize the strengths of Japan (plan specific measures that can be accepted in global markets, valuing the perspective of adopting AI, and utilizing the superiority that Japan has accumulated)
(b) Upgrading middle-scale enterprises and SMEs (The perspective of what kind of role is required in middle-scale enterprises and SMEs and how effectively it can be incorporated in the whole Japanese economy)
(c) Streamlining the government (The perspective that agile-type development is promoted through the procurement by the government and the system and mechanism that hinder the utilization of AI are reviewed)
(d) Environmental improvement to realize the Inclusive AI Economy Society (aiming to be the state of AI-ready as a whole society, improvement of social acceptance in the utilization of data, data free flow, securing the trust on privacy and security (DFFT: Data Free Flow with Trust))
(e) Promotion of global discussion on data
(f) Monitoring

1.3 The scope of consideration in this report

To conduct further consideration based on the report in May 2019, the consideration was resumed in December, 2019.

And the Special-Interest Group on Data was established under the Committee for the purpose of considering specialized and technical items on the data economy policy among the items to be considered in the Committee.

Based on the contents of further consideration in the Committee that was resumed, the following 3 key points are the scope of the consideration in this report.

(1) Consideration on the ideal way to utilize data required for the implementation of AI in society

(2) Consideration on the data economy policy in the AI era

(3) Consideration on the image of "Inclusive AI Economy Society"

Regarding the key point (1), various entities showed the importance of actively being involved in social and economic activities utilizing AI and data in an ingenious way based on their will and judgment in the report published in May 2019, but further consideration will be conducted on the ideal utilization of data that can be the premise of the implementation of AI in society. First of all, the status of the utilization of data in Japan will be considered. Then, the ideal way to utilize data required for the utilization of AI in society will be outlined. More specifically, the ideal way that all the entities, especially the socially vulnerable such as SMEs and local enterprises can conduct the utilization of data that can be the premise of the implementation of AI in society will be considered. Based on the above, the ideal policy toward the promotion of the utilization of AI/data will be examined.

Regarding the key point (2), the thought that data[1] is positioned as one of production factors and the effects of data on the improvement of the productivity should be estimated is significant, and the necessity of deeply discussing the ideal distribution of remuneration to the data providers was shown in the report in May 2019. To cover these matters, further consideration will be conducted. First, the methods to estimate the value of data and ideal remuneration depending on the effects/value of data will be considered. More specifically, after checking the functions and roles of data, centering on the prior studies, factors to be considered in estimating the value of data will be outlined. And after breaking down the approaches to estimate the value of data, each approach's concept and issues will be checked. Then, the methods to estimate the value of feasible data will be examined, noting securing the versatility and reproducibility applicable to various industries and countries. In addition, we will consider the positioning of data and the degree of remuneration for individuals who produce data and the process of paying remuneration, etc. as one of production factors as ideal remuneration depending on the effects/value of data. Then, based on the above, the conditions for revitalization of data transaction will be outlined and ideal data economy policy required for the promotion of data-driven type economy will be examined.

Regarding the key point (3), the image of "Inclusive AI Economy Society" that can be seen after overcoming the issues shown in (1) and (2) will be viewed. More specifically, among the Analysis on Prospect of Ecosystem Formed with the Progress of AI Network that The Conference Toward AI Network Society has examined, the analysis on the perspective of the utilization of AI will be referred, and the future image of expected "Inclusive AI Economy Society" will be viewed. And some examples of utility scenes classified by individuals, enterprises, and administration will be taken. In addition, human workstyles/how to spend leisure time expected when AI is fully utilized will be introduced.

[1] "Data" in this report is defined as currently digitized data or data that can be digitized easily.

2. The ideal way to utilize data required for the implementation of AI in society

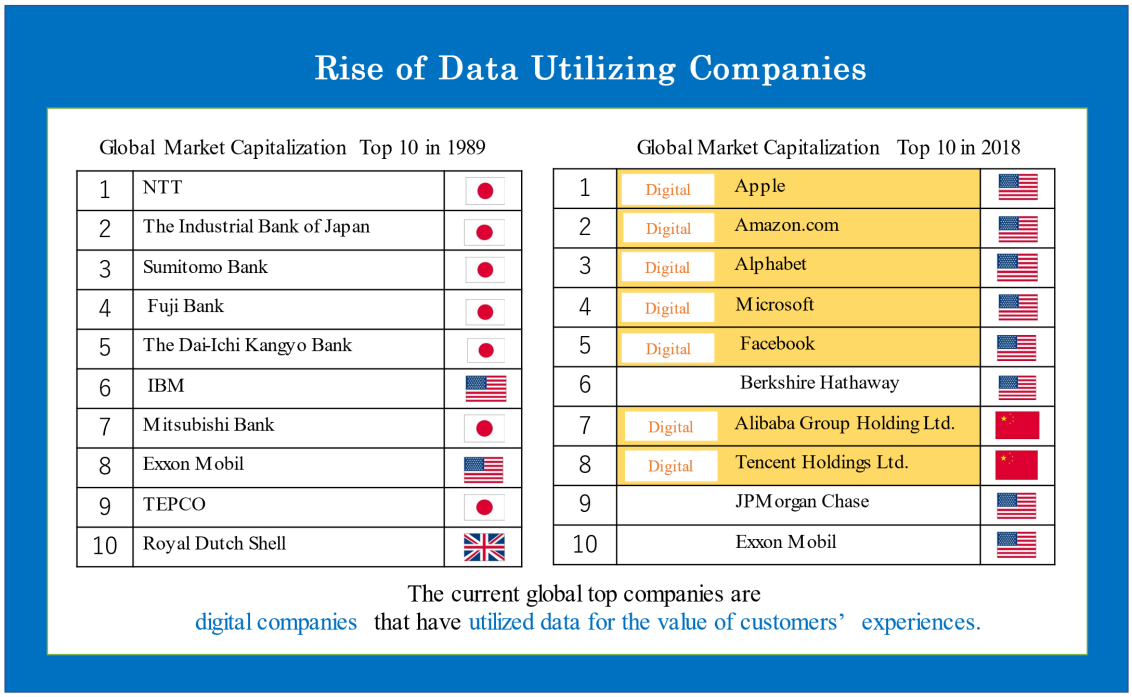
2.1 The current status of Japan's data utilization

First, we will consider the current status of Japan's data utilization to pursue the society where various organizations can be actively involved in social/economic activities, utilizing AI and data in an ingenious way based on their will and judgment.

(1) Changes in global economy

When you see the global market capitalization Top 10 companies ranking, you will notice Japanese companies occupied the top positions, and financial institutions and energy related companies were ranked high in 1989, and GAFA in the US and Alibaba Group and Tencent Holdings that have developed their business based on collected/accumulated data of users were ranked high in 2018 (Table 1). These changes in the global market capitaliztion Top 10 companies show that the global economy has changed from a finance-led economy to the one led by digital companies that utilize ICT and data.

Table 1: Global Market Capitalization Top 10 (1989 and 2018)



Source: SuperMagazine
(https://supership.jp/magazine/seminar-report/2849/)

In addition, in the World Digital Competitiveness Ranking in 2019 published by IMD World Competitiveness Center, Japan was ranked the 23rd out of 63 countries, lowering its ranking by one from 2018 (Table 2). Regarding "business agility", one of the evaluation criteria in digital competitiveness, Japan was rated low and ranked the 41st. Among the components of "business agility" factors, especially in the "use of big data and analysis", Japan was rated the lowest and ranked the 63rd, which shows that Japanese companies are lagging behind in the utilization of big data.

Table 2: World Digital Competitiveness Ranking in 2019 (Ranks 1-30)

Source: IMD World

Competitiveness Center

(https://www.imd.org/wcc/world-competitiveness-center-rankings/world-digital-competitiveness-rankings-2019/)

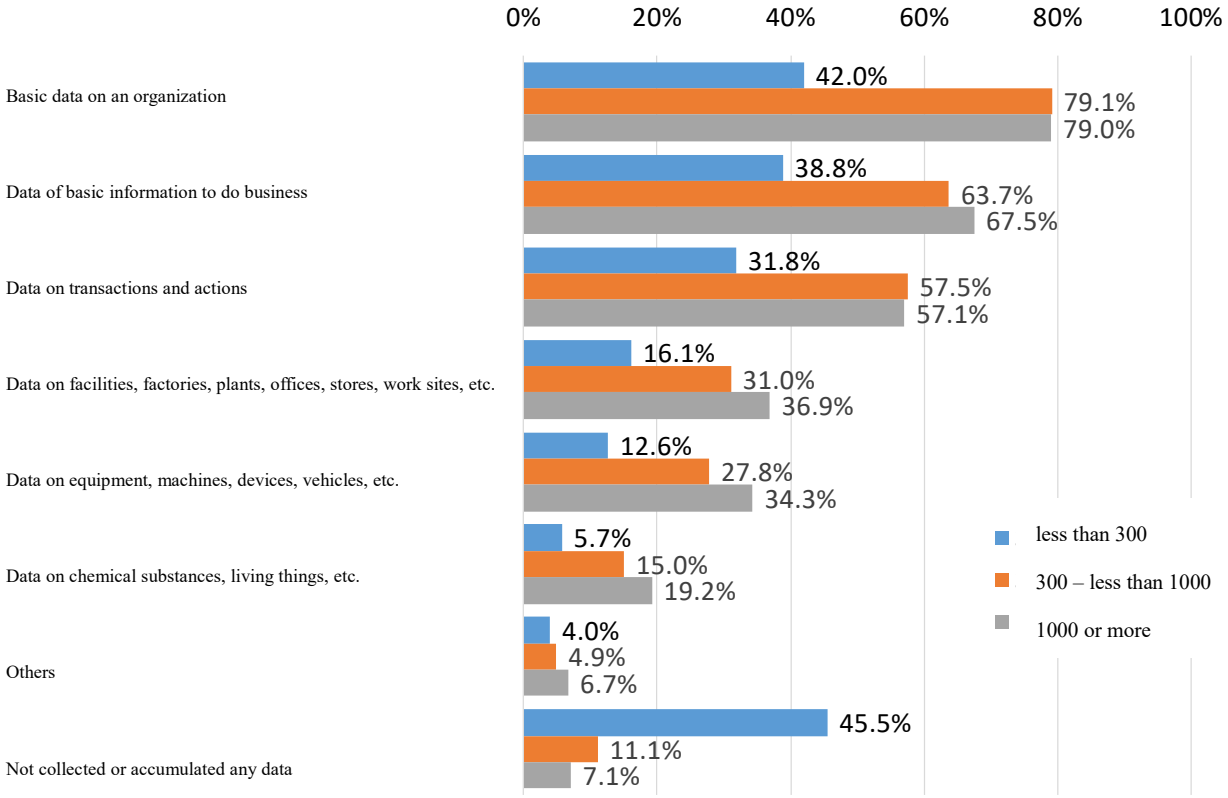
(2) The disparity of the utilization of data between enterprises

Regarding the utilization of data in enterprises, there are the disparity between large enterprises and small and medium-sized enterprises and the disparity between enterprises in urban areas and those in rural areas. In “the Questionnaire on the utilization of digital data” conducted by the Ministry of Internal Affairs and Communications, the percentage of the

enterprises that haven't collected or accumulated any organizational data was 7.1% at large enterprises (with 1,000 or more employees), but it was 45.5% at SMEs (with 300 or less than employees) (Table 3). The result shows that initiatives of utilizing data at SMEs are lagging behind compared to those at large enterprises.

Table 3: The status of collection and accumulation of organizational data

Q: What kind of data among the following organizational data is collected or accumulated in your enterprise? (by number of employees)



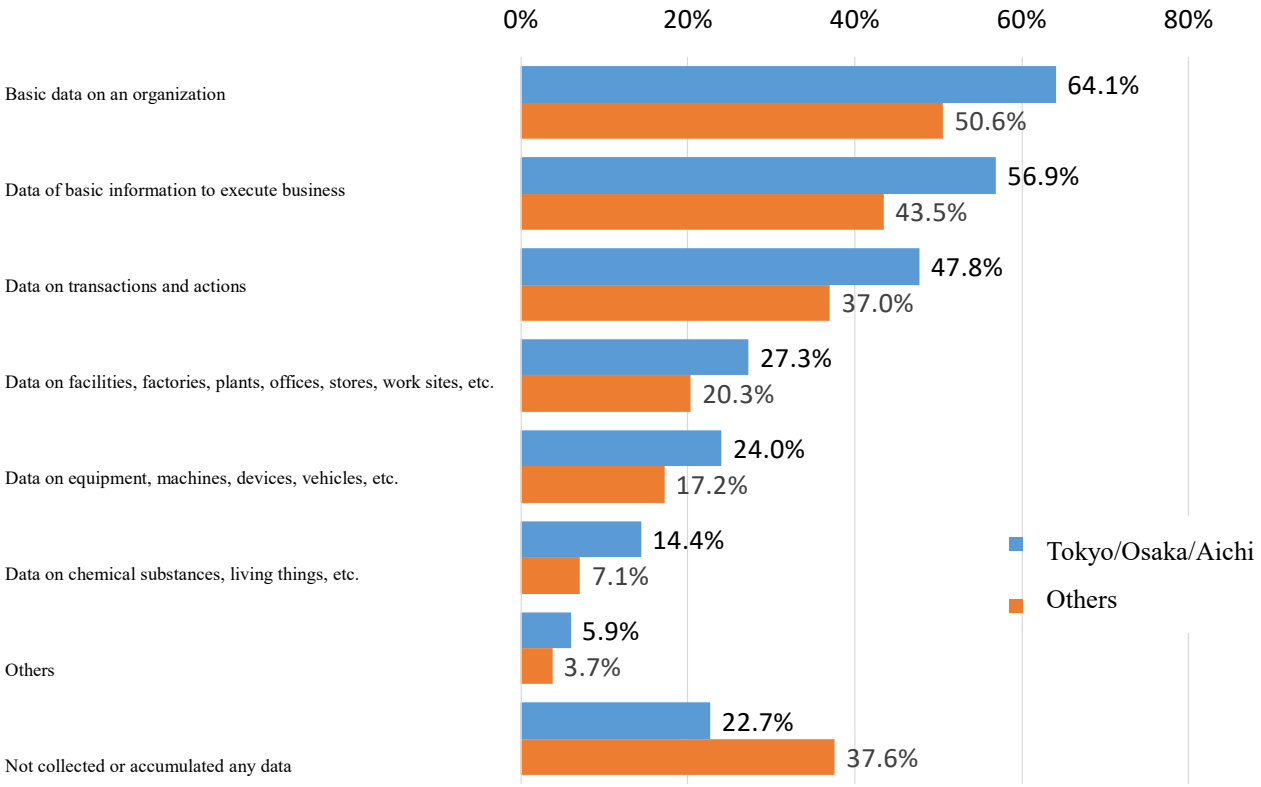
Source: MIC "The questionnaire on the utilization of digital data" (2,003 enterprises responded)[2]

In addition, the same trend could be seen between enterprises in urban areas and those in rural areas. This result shows that initiatives of utilizing data at enterprises in rural areas are lagging behind compared to those at enterprises in urban areas. (Table 4).

[2] This web questionnaire was conducted on 2,003 companies in Japan to check the status of providing/receiving data at enterprises, issues of initiatives on data, etc. from March 19 (Thu.) to March 25 (Wed.), 2020.

Table 4: The status of collection and accumulation of organizational data (by location of headquarters)

Q: What kind of data among the following organizational data is collected or accumulated in your company? (by location of headquarters)



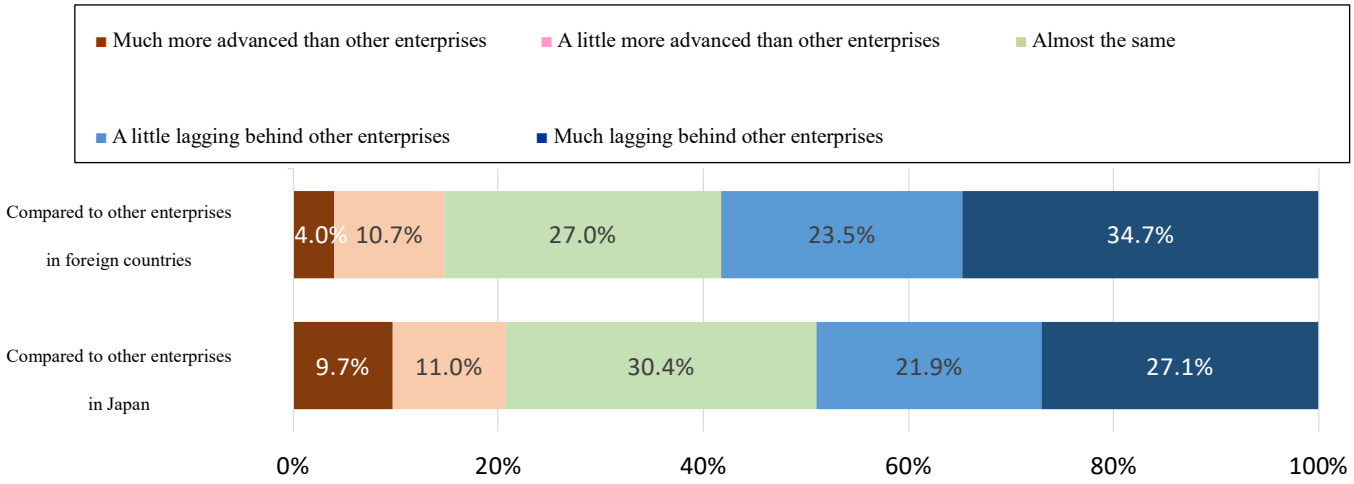
Source: MIC "The questionnaire on the utilization of digital data" (2,003 enterprises responded)

(3) Recognition of lagging behind in initiatives on data

Since an AI system needs to learn from big data, data collection and data analysis are indispensable to the realization of the utilization of AI. However, there are many Japanese enterprises that realize their initiatives on data including data collection and analysis are lagging behind. 34.7% of the enterprises responded that their initiatives on data are much lagging behind those of other enterprises in the same industry in foreign countries, and 23.5% of them responded that they are a little behind other enterprises, so in total, 58.2% or almost 60% of them recognized they are lagging behind other enterprises in the same industry in foreign countries. Compared to other enterprises in the same industry in Japan, 27.1% of the enterprises responded their initiatives on data are much lagging behind other enterprises, and 21.9% of them responded they are a little lagging behind other enterprises, so in total, almost half (49%) of the enterprises recognized that their initiatives on data are lagging behind (Table 5).

Table 5: Awareness on the status of initiatives on data

Q: Do you think your enterprise's initiatives on data (data collection, accumulation, cooperation, analysis/utilization, etc.) are lagging behind compared to other enterprises in the same industry?



Source: MIC "The questionnaire on the utilization of digital data" (2,003 enterprises responded)

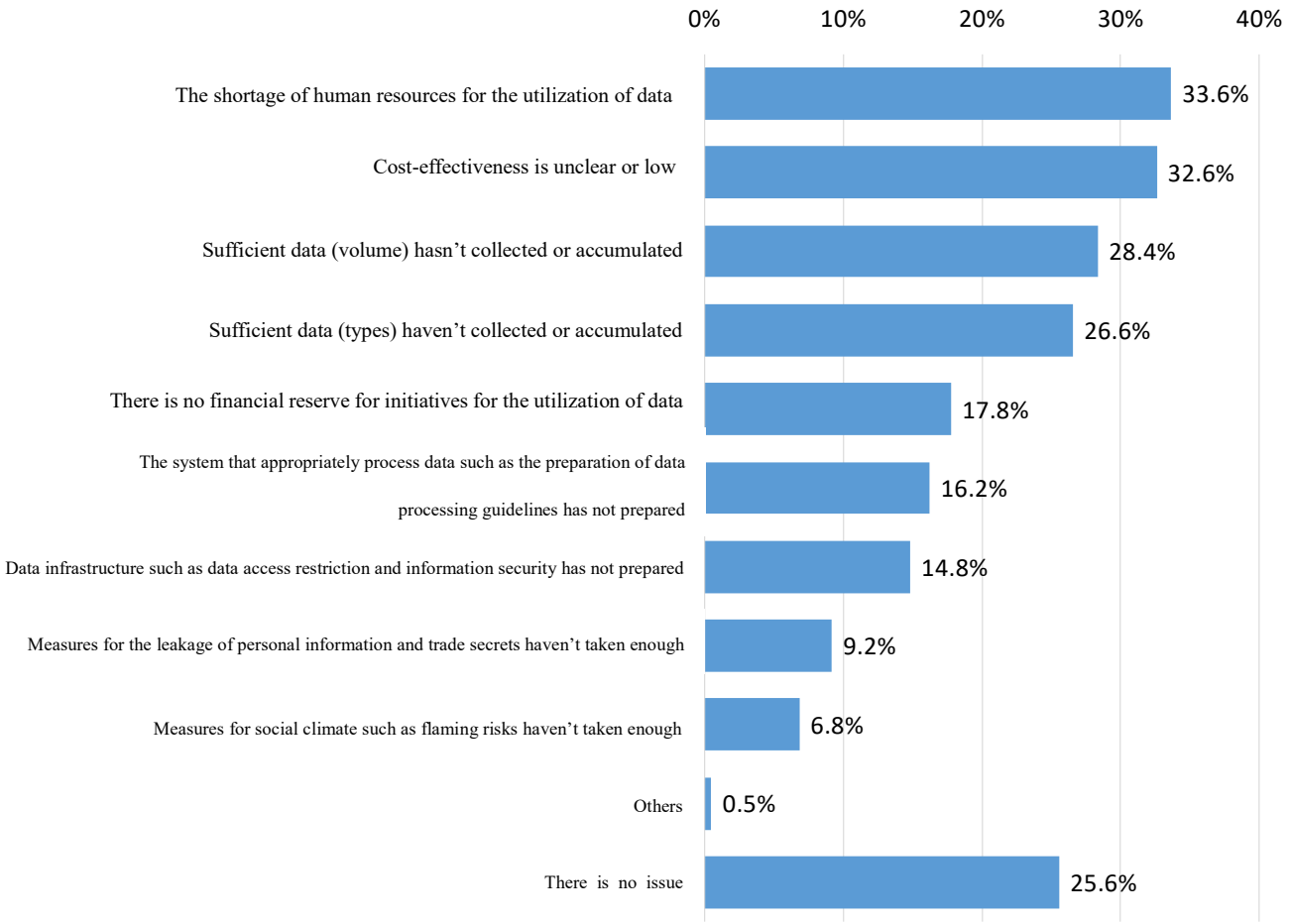
In the background of the status of initiatives on data lagging behind, there is a reality that they have issues such as the shortage of human resources for the utilization of data, unclear or low cost-effectiveness, insufficient data collection/accumulation, even though they are eager to utilize data.

Regarding promoting initiatives on data, "the shortage of human resources who can utilize data" (33.6%) is the most, so the shortage of human resources is the biggest issue in the utilization of data.

In addition, there are quite a few enterprises that have the issue of costs such as "Cost-effectiveness is unclear or low" (32.6%) and "There is no financial reserve for initiatives for the utilization of data" (17.8%). Furthermore, there are many enterprises that are facing issues of the shortage of data such as "Sufficient data (amount) haven't collected or accumulated" (28.4%) and "Sufficient data (types) haven't collected or accumulated" (26.6%) (Table 6).

Table 6: Issues for promoting initiatives on data

Q: What do you think is an issue when your enterprise promotes initiatives on data?



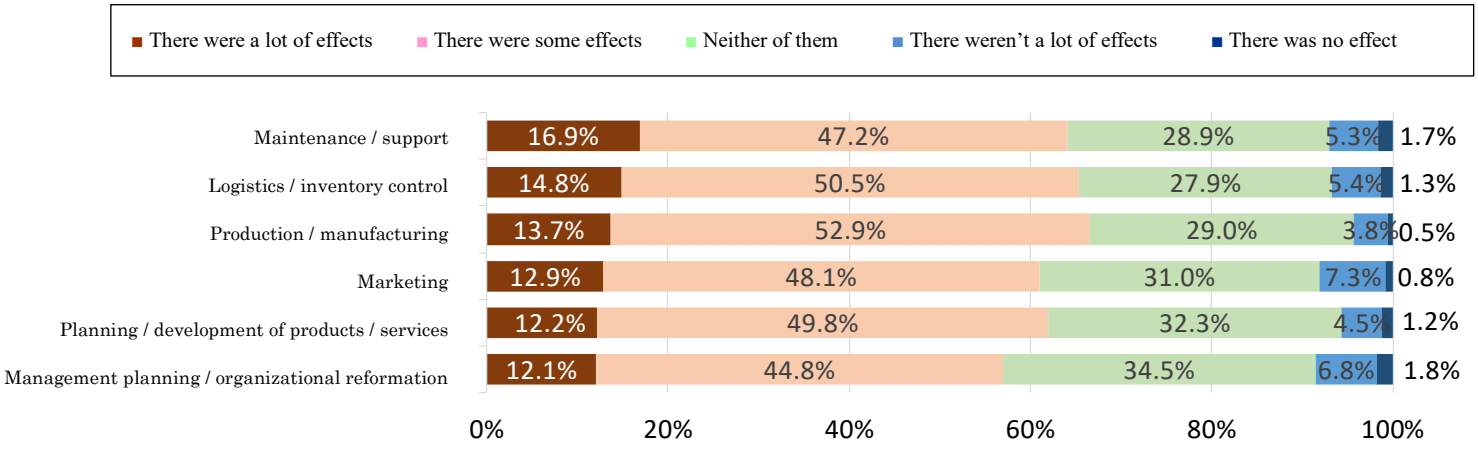
Source: MIC "The questionnaire on the utilization of digital data" (2,003 enterprises responded)

(4) The effects of the utilization of data in enterprises are recognized

On the other hand, roughly 60% of enterprises feel that there have been some effects by utilizing data in their enterprises. The number of enterprises who realized the effects of the utilization of data in the field of "production /manufacturing" was the largest, and 66.6% of the enterprises responded that "There were a lot of effects" or "There were some effects" (Table 7).

Table 7: Effects through the utilization of data

Q: How much effect did your enterprise have by utilizing data in each field?



Source: MIC "The questionnaire on the utilization of digital data" (2,003 enterprises responded)

(5) Careful attitude toward providing personal information[3] and behavior history for enterprises

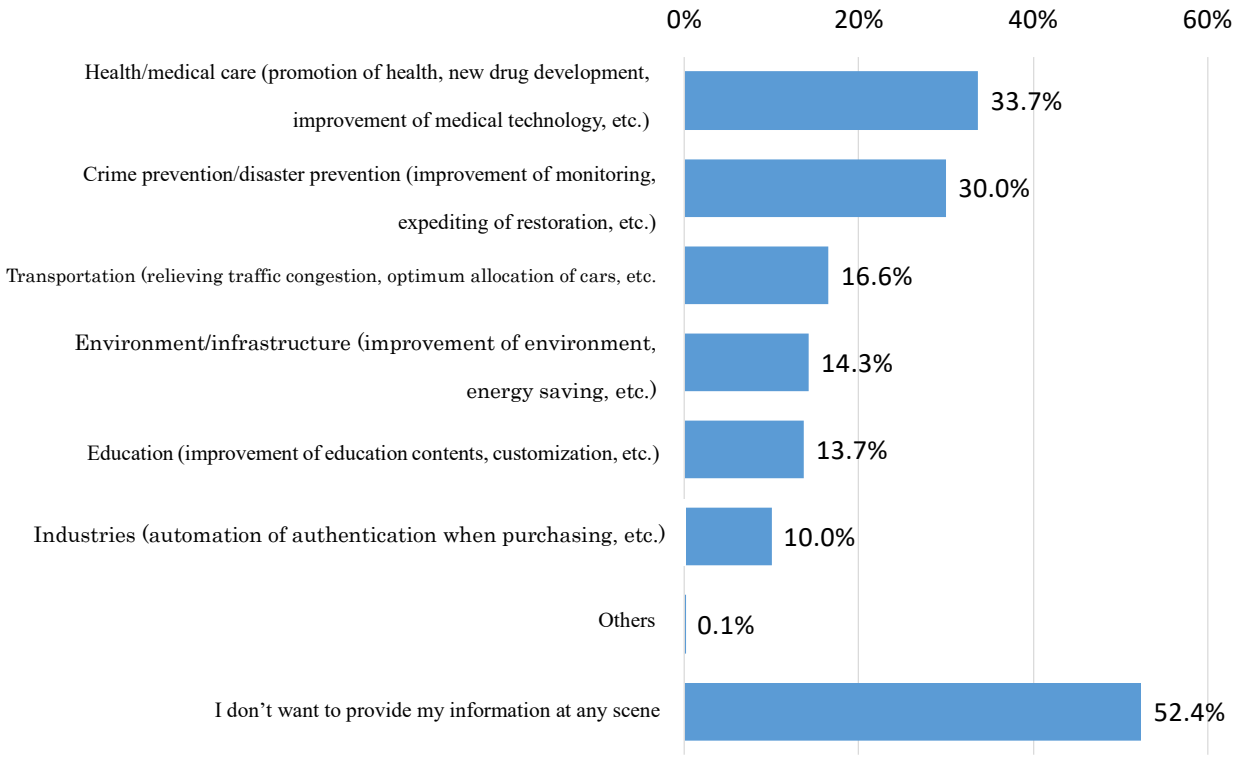
Regarding the issues of consumers who provide data, consumers tend to be careful about providing their personal information and behavior history for enterprises. The percentage of the consumers who don't want to provide their data in any scene was 52.4%, and it was higher than that of those who can provide their data such as their personal information and behavior history even without any reward if their data are helpful in society (Table 8).

On the other hand, the percentage of consumers who responded that they would provide their data for the scene of "health/medical care (promotion of health, new drug development, improvement of medical technology, etc.)" even without any reward was 33.7%, which was relatively higher than the utilization of data at the scenes of disaster prevention/crime prevention, transportation, education, etc.

[3] "Personal information" in this report was used as a general term and it's not used as a legal term in "Act on the Protection of Personal Information (Act No. 57 of May 30, 2003)", etc.

Table 8: Intention to provide data (by useful scene)

Q: For what scene in society would you provide your information even without any reward (including but not limited to monetary benefits, provision of information based on provided personal information and behavior history, etc.) if your data on your personal information and behavior history are helpful?



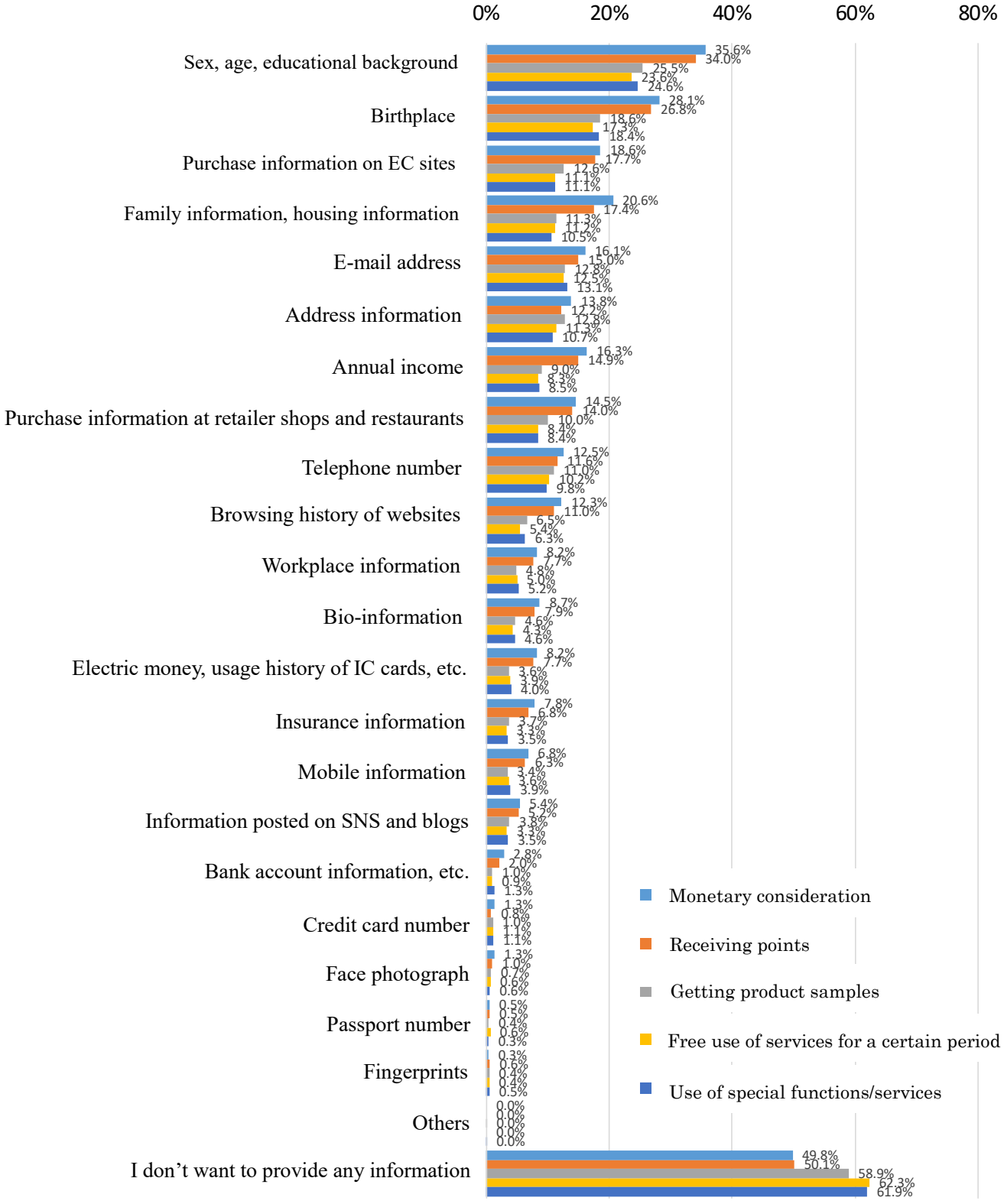
Source: MIC "The questionnaire on the utilization of digital data" (2,155 people responded)[4]

Also, regarding the information for which they would like to provide their data such as personal information and behavior history if they receive monetary consideration or points, more than 50% of consumers responded that they don't want to provide their information at any scene, which was relatively high (Table 9).

[4] The web questionnaire about the awareness on personal information, conditions for providing personal information, etc. carried out for men and women older than 20 years old in Japan during the period of March 24 (Tue.) – March 26 (Thu.), 2020.

Table 9: Intention to provide data (by content of data)

Q: Is there any information that you would provide for a company (except for your company if you work for a company) if you get a reward out of the following data on your personal information and behavior history, etc.?

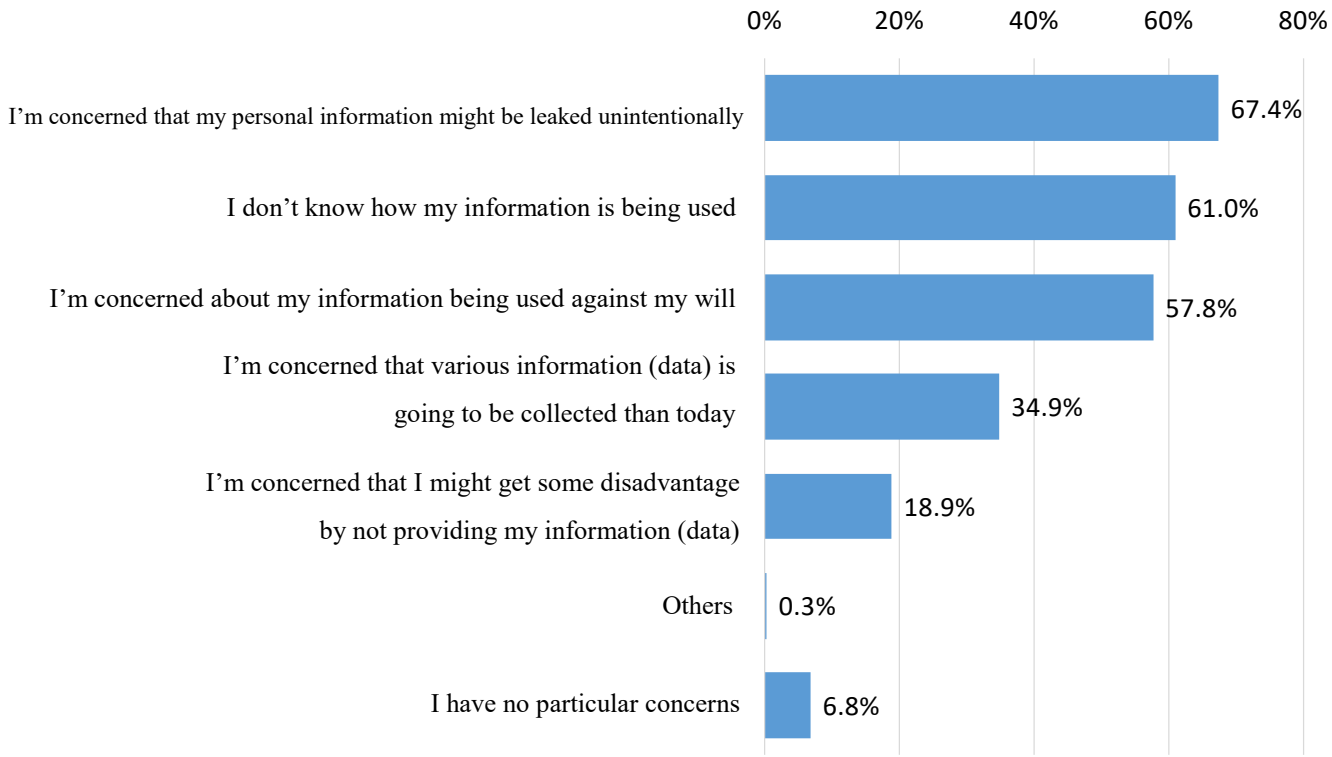


Source: MIC “The questionnaire on awareness on data” (2,155 people responded)

The main reason why consumers take careful attitude toward providing data is that they are concerned about their personal information or behavior history being collected. The percentage of consumers who chose “I’m concerned that my personal information might be leaked unintentionally” about personal information or behavior history being collected and used by a company was 67.4%, which was the highest, and the percentage of consumers who chose “I don’t know how my information is being used” was 61%, which was the second highest, and the percentage of consumers who chose “I’m concerned about my information being used against my will” was the third highest at 57.8%. These results show that there are a lot of concerns about the risk of the leakage of personal information and how their information is used. In addition, the percentage of consumers who feel “I’m concerned that various information (data) is going to be collected than today” was 34.9% (Table 10).

Table 10: Concerns about personal data being used

Q: What kind of concern do you have about your personal information or behavior history, etc. being collected and used by a company (*except for your company if you work for a company)?



Source: MIC “The questionnaire on awareness on data” (2,155 people responded)

2.2 The ideal way to utilize data required for the implementation of AI in society

Based on the situation mentioned in 2.1, the ideal way to utilize data required for the implementation of AI in society can be outlined as follows.

(1) Positioning data/AI in management strategies and business architectures in enterprises

Data is one of management resources, and AI can be tools. You should not consider only data or AI, but consider them in the entire management resources such as humans/goods/money/data (AI) (e. g. putting how humans/goods/money/data (AI) can contribute to corporate P/S and B/S into strategies).

In addition, designing architectures is also required in the implementation of AI in business. It is important to design business related architectures as well as the ICT system. (e. g. value chain (find strengths that can be the core of enterprises), physical architectures (the ideal way of allocating resources), digital architectures (building data platforms), legacy systems (the ideal way of rebuilding)).

(2) Building the collection system of physical data for the utilization of data, the organization system of the utilization of data and organizational data infrastructure in enterprises

Building the collection system of physical data is required for the utilization of data. Quite a few enterprises don't have a digital data collection system or have issues for collecting physical data effectively, though they have accumulated an enormous amount of data resources. It's important to put data of resources into digital servers and build a data collection system.

For example, designing the system of data collection that can detect several billion pieces of baggage in real time is required to allocate human resources such as drivers and sales bases.

In addition, designing an organization and allocation of human resources are essential in the utilization of data. Authorizations and budgets should be granted to an organization that take a lead role in the utilization of data under the understanding and commitment of the enterprises, and it is required to consider that the organization should be independent or located directly under the executive team in designing the organization. At the same time, it is also required to consider designing the organization with a good balance of concentration and decentralization, by scrutinizing the functions broken up into companies/business units.

Furthermore, it is also important to build the data distribution and shared infrastructure in an organization. More specifically, it is required to meet the requirements to respond to various data such as the visualization of data through data brochures in the organization, complicated procedures in rights/intellectual properties/contracts/securities, and structuralizing/non-structuralizing.

(3) Relieving/reducing the concerns of consumers who provide data as well as responding to the existing laws of data of customers, etc. in enterprises

The response to the existing laws of data of customers, etc. is also required. When enterprises conduct structural reform for data-driven management, complete integration of customer data toward the new management system is important. However, there are some enterprises that haven't caught up with the existing law system for their contract. For example, in the field of logistics, personal information listed on slips of parcel delivery services conforms to old

agreements that haven't expected the accumulation/utilization of data, so it is necessary to review the agreements corresponding to related law systems.

In addition, as mentioned in 2.1 (5), consumers who provide information tend to be careful about providing their personal information and behavior history, etc. for enterprises. So, to promote the distribution of data, initiatives for relieving and reducing consumers' concerns about providing data are necessary. For example, in collecting/analyzing information or providing information to a third party, initiatives for confirming the intention of the providers of information, acquiring the agreement from users, clarifying the purpose of utilizing the information are necessary.

(4) Accelerating initiatives of the digital government in administration

Necessary initiatives must be taken in administration so that various organizations can receive benefits of the utilization of data.

The government of Japan pointed out the promotion of collaboration of administration data such as dissemination of standardization of data, open strategies of data held by administration, promotion of solving regional issues by promoting open data in local governments in the "Digital Government Implementation Plan"[5]. The private utilization of public data such as public transportation information, meteorological information and disaster prevention information is expected to contribute not only to the implementation of AI in society but also to the solution to elaboration of the Business Continuation Plan (BCP) of enterprises and social issues Japan and rural areas in Japan have. It is important to accelerate these initiatives, secure the interoperability of data (standardization of symbolic convention of data, universal vocabulary and used characters/letters), promote opportunities to utilize public data such as DATA.GO.JP where API for acquiring data is implemented and e-Stat (statistic GIS function is implemented), and promote using open data that include the design and maintenance of the information system based on the concept of open data by design in local governments.

(5) The issues we are facing through the outbreak of COVID-19

The outbreak of COVID-19 in Japan and the world showed us how important for enterprises to utilize communication networks and ICT tools in order to continue their business. In Japan, after the first person infected with COVID-19 was confirmed on January 15, 2020, the number of people infected exceeded 17,500 as of June 17, 2020[6]. In particular, many cases of uncertain infection routes were found, and many clusters of infection were also found in medical institutions and nursing care institutions. On April 7, 2020, the government declared a state of emergency over the COVID-19 outbreak to prevent the COVID-19 pandemic and recommended the people should refrain from non-essential and non-urgent going out, and avoid "3Cs" ((1) Closed spaces, (2) Crowded places, (3) Close-contact settings), and the government asked enterprises to promote working at home through teleworking. Because the awareness of enterprises and individuals is expected to change from the current status of the utilization of data in Japan mentioned in 2.1 regarding the data in the behaviors and health conditions of individuals, etc., it is important to consider the ideal ways to manage these data. On the other hand, there are many enterprises, especially SMEs that have issues such as the shortage of ICT human resources

5 Cabinet Office, Strategic Headquarters for the Promotion of an Advanced Information and Telecommunications Network Society (IT Strategic Headquarters) "Outline of Digital Government Strategy Plan (Cabinet Decision on December 20, 2019)" (https://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20191220/gaiyou.pdf)

6 The official website of Ministry of Health, Labour and Welfare (https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000164708_00001.html#kokunaihassei)

and funds for capital investment, so the government, private groups, and business operators proposed several support measures as follows (Table 11). To fully exert the potentials of communication networking including teleworking and ICT tools, it's important to realize the ideal way to utilize data required for the implementation of AI in society mentioned in (1) - (4).

Table 11: Major teleworking related support programs

Responsible organizations/title of measures, etc.	Outline of teleworking related support programs
Ministry of Internal Affairs and Communications "MIC's major economic measures"	・ Preparation of information communication networking that supports studying at home/working at home/online medical examination, etc. ・ Promotion of introduction of teleworking by enterprises and local governments ① Expansion of consultation system by telework managers, etc. ② Regarding the capital investment of SMEs for teleworking, etc., expand the taxation scheme to strengthen SMEs management, and include them for the targets for exceptional measures for corporate tax/income tax[7]. ③ Launch of teleworking support network business collaborated with organizations that support SMEs ④ Building measures for special tax allocation regarding the costs of introducing teleworking for local government workers
Ministry of Economy, Trade and Industry "Project to strengthen the productivity revolution of SMEs" "Project to support digitalization of SMEs, etc."	・ In the "Project to strengthen the productivity revolution of SMEs", capital investment and the introduction of IT of entrepreneurs working on the improvement of teleworking environment will be supported on a priority basis, and "special quotas" with raised subsidy rates will be established. For example, raising the subsidy rate from 1/2 to 2/3 in Project to support the introduction of IT to improve productivity in services and other areas (IT introduction subsidy) (to support SMEs, etc. for the introduction of IT tools to get over the impact of COVID-19)[8] ・ In the "Project to support digitalization of SMEs, etc.", if IT experts provide support services such as building EC sites and consultation services for the introduction of teleworking for SMEs, fixed amounts will be subsidized depending on the unit prices designated per routine

7 Ministry of Internal Affairs and Communications "MIC's Major Economic Measures" (April 20, 2020) (https://www.soumu.go.jp/main_content/000683969.pdf)

8 Cabinet Office "Emergency Economic Measures to Cope with the Novel Coronavirus (COVID-19)～Thoroughly secure people's lives and move toward economic revitalization～" (May, 2020) (https://www5.cao.go.jp/keizai1/keizaitaisaku/2020/20200420_economic_measures_all.pdf)

	work and support achievement, etc.[9]
Ministry of Health, Labour and Welfare "Work Style Reform Promotion Support Subsidy (Teleworking course)"	・ Subsidies (up to 2 million yen) for the costs of introduction/operation of communication devices will be granted for SMEs that newly introduced or continuously utilized teleworking, depending on the status of achievement of targets (the number of employees who did teleworking and times of teleworking)[10]
Japan Telework Association "COVID-19 Measures: Introduction of the support program for urgent introduction of teleworking"	・ Introduction of support programs for urgent introduction of teleworking such as Cisco Webex Meetings (Web-meeting system) 90 days free support program by member enterprises/organization for enterprises that are to urgently introduce teleworking[11]
Microsoft Japan Co., Ltd.	・ Support educational institutions closed due to COVID-19 through initiatives such as free issuance of Office 365 account, free provision of Microsoft Teams that can distribute live events, and free rental of Surface terminals[12]

Source: Prepared from a published document

[9] Ministry of Economy, Trade and Industry "Productivity improvement through supporting SMEs for the implementation of digitalization (promotion of comprehensive digitalization)" (April 2020) (https://www.kantei.go.jp/jp/singi/keizaisaisei/miraitoshikaigi/suishinkaigo2018/chusho/dai9/siryou3.pdf)
[10] The official website of Ministry of Health, Labour and Welfare (https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/koyou_roudou/roudoukijun/jikan/telework_10026.html)
[11] The official website of Japan Telework Association (https://japan-telework.or.jp/anticorona_telework_support/)
[12] Future Classroom －Learning Innovation－ Official website (https://www.learning-innovation.go.jp/covid_19/microsoft/)

3. Data economy policy in the AI era

3.1 The functions/roles of data as “new capital” and estimation methods of the effects/value of data

In the AI era when the implementation of AI in society becomes full-fledged, it is considered that data will be the source to create values and play an important role in contribution to productivity improvement. However, the actual situations such as the degree of value brought by the utilization of data and the process of creating values are extremely hard to understand. Therefore, it’s difficult to clarify the impact of the utilization of data to society/economy and the size of effect brought by the system/technology in the said impact, etc. Based on the above, we clarified data for estimation and the effects of data, and outlined the factors to be considered when estimating the value of data, after organizing the functions/roles of data, centering on prior studies. In addition, we divided approaches to estimate the value of data into three types, and clarified each type’s concepts and issues, then considered estimation methods of the value of data in feasible methods currently available, taking care of securing versatility applicable to various industries and countries and reproducibility.

(1) Targets for estimation

To estimate effects/value of data, data for estimation and the effects of data are needed to be clarified. In doing so, we organized the items to be considered below (Table 12). Since there are various units and values in the estimation of data, targets for estimation are needed to be considered, depending on the estimation method and scope of estimation/purposes of estimation, etc.

Table 12: Targets for estimation

Target for estimation	Items to be considered
Data	・Data in this report is defined as “currently digitized data or data that can be digitized easily”. ・Targets must be considered, taking the estimation method of value into consideration, noting that data can be classified from various aspects, ・Since data has different units depending on types (byte, people, cases, minutes, etc.), it’s necessary to take care of units when processing several different data.
Effects of data	・To add value to data, processing and analysis are needed. As a result, improvements of ideas, innovation, and estimation accuracy are expected. In addition, through these improvements, effects on the entire society (CO2 reduction, etc.), enterprises (sales, profits, etc.), and individuals (benefits, etc.) are expected. ・Various effects can be considered, so it’s necessary to clarify which effect can be the target for estimation. ・The scope of targets and understandability are conflicting (If the target is only individual effects, the results will be easily understood but the entire picture won’t be grasped. On the other hand, if the targets are entire things, it will be hard to understand the results.)

Source: Prepared from a published document

(2) Functions/roles of data

First, when we checked prior studies on the functions/roles of data and intangible capital that includes data (Table 13 - 16), we noticed that data have economic characteristics such as nonrivalness, externality, and partial exclusivity. In addition, it is said that the value of data is determined due to various elements/factors such as the contents of data and market environment.

Table 13: Prior study on functions/roles of data①

Prior study thesis, etc.	Functions/roles of data
Avi Goldfarb and Catherine Tucker (2017) "Digital Economics," NBER Working Paper, No. 23684.	・Saving, estimation, transportation costs of data will be reduced due to digital technologies. As a result, the costs shown below will be also reduced. *The value is more like the value of digital technology than that of data. (1) Search costs: Costs for searching information (2) Replication costs: Anyone can replicate with almost zero costs without deteriorating the quality. (3) Transportation costs: Transportation costs via the Internet. Distribution costs of data are almost zero. (4) Tracking costs: Costs for tracking individuals. Activities on the Internet can be recorded easily and automatically, so it will be easy to individualize data based on past activities. (5) Verification costs: Costs related to the verification of identities and reputations.
Yan Carriere-Swallow and Vikram Haksar (2019) "The Economics and Implications of Data: An Integrated Perspective," IMF Departmental Paper, No.19/16.	・Data has 3 economic characteristics important to public policies. **(1) Nonrivalness** Many people can use the same data. If data such as new ideas are widely shared, more users can use data to enhance the efficiency and make a innovation, and the society will most benefit from data. However, whether it is realized or not depends on the policy and the decision-making of individuals. There is a high possibility that private enterprises don't have any incentive to allow their competitors to access their collected data in the existing policy. So, there is a possibility that social interests gained from data and competitivity in market are restricted. **(2) Externality** Collecting, sharing and processing individual data by an agent can impose costs on others by having an impact on the privacy of others. In other words, there is a possibility that collecting, sharing and processing individual data might cause excessive data collection and lack of privacy in the data market where data collectors can process data in the way they like, which might cause inefficiency. **(3) Partial exclusivity** Since data can be excluded just partially, saving data on the system where many people are connected mutually means that continuous investment is needed to prevent the loss of data by cyber attacks.

Source: Prepared from a published document

Table 14: Prior study on functions/roles of data ②

Prior study thesis, etc.	Functions/roles of data
Yan Carriere-Swallow and Vikram Haksar (2019) "The Economics and Implications of Data: An Integrated Perspective," IMF Departmental Paper, No.19/16.	・Economic value of data from the perspective of demand (2 roles of data): **(1) Data are input in the production of goods and services and contributes to innovation and efficiency** Data has a function as the input of production in properties and services. Data make the creation of knowledge possible, and data will be used for continuous production of properties or the development of new products and services. To extract the value from data as an input, costly processing and analysis are needed, and they will be complementary provided by skilled workers. **(2) Creating information and shifting to information in the entire economy will have an impact on strategic interaction** Since information on the entire economic subjects including consumers and enterprises is included in data, asymmetry of information in the market which these subjects participate in. When the access to data is helpful in reducing the asymmetry of information between buyers and sellers, it will contribute to more efficient economic transaction. For example, sellers who can access data on characteristics on potential consumers' interests and purchase habits, etc. can provide products and serivces for more individuals. In a similar manner, customers who have data on characteristics of potential products can make a decision of purchase based on more information. Heterogeneity of data: Data are heterogenious in terms of subjects, timing, formats, quality, etc., which will affect economic characteristics. Collection costs of data: As the collection amount of data increases, the avarage costs for data collection will be reduced (There is a possibility that economy of scale will be created).

Source: Prepared from a published document

Table 15: Prior study on functions/roles of data ③

Prior study thesis, etc.	Functions/roles of data
Yan Carriere-Swallow and Vikram Haksar (2019) "The Economics and Implications of Data: An Integrated Perspective," IMF Departmental Paper, No.19/16.	・Factors to determine the prices of data: **(1) Contents of data** Prices of data are not homogeneous, and they are determined based on many attiributes such as when, where, and by whom they were collected, how the structure of data is, if they can be integrated with other data. And they greatly change as time goes by. **(2) Market environment** Prices are determined by many factors such as how much information asymmetrically different from customers' and competitors' exists, the degree of influence on market and size of market. Evaluating the value of personal data is pretty difficult even if the data has a direct incentive. For example, advertisers have invested much funds on personal data about online users, etc. on the premise that sales will increase by displaying advertisements, focusing on targets. However, though the profits gained from targeted advertisements seem significant, the causal effects on sales are small, which are smaller than the expenditure for targeted advertisements. (Marotta, Veronica, Vibhanshu Abhishek, and Alessandro Acquisti. 2019. "Online Tracking and Publishers' Revenues: An Empirical Analysis".).
Charles I. Jones, Christopher Tonetti (2019) "Nonrivalry and the Economic of Data," NBER Working Paper, No.26260.	・Same data can be used by several companies due to nonrivalness of data, which will bring great social profits. However, companies might hesitate to sell data due to the concern of creative destruction. ・Data are byproducts of consumption and new data will be produced every time products are consumed, which will contribute to the improvement of productivity. Data can be considered as the quality of ideas. While data is a production factor, a new idea is a new function for production. ・Ideas and data are kind of information. An idea is a series of orders to make economic profits and a part of information, and it may include other ideas. Data shows other forms of information. Data itself is not an order to create good products, but data might be helpful in the production process including the production of new ideas. Therefore, an idea is a production function, and data is a production factor.

Source: Prepared from a published document

Table 16: Prior study on functions/roles of intangible capital

Prior study thesis, etc.	Functions/roles of intangible capital
Corrado, C., C. Hulten, and D. Sichel (2006) "Intangible Capital and Economic Growth," NBER Working Paper, no.11948. Cambridge, MA; National Bureau of Economic Research. Corrado, C., C. Hulten, and D. Sichel (2009) "Intangible Capital and U.S. Economic Growth," Review of Income and Wealth, 55, pp.658-660.	・Classification of intangible capital (an example): **(1) IT capital** Softwares and databases **(2) Revolutional capital** R&D in science and engineering fields, resource search rights, copy rights/licenses, etc., development and design of other products, etc. **(3) Economic competitiveness** Brand capital, corporate special human resources, organizational change
Jonathan Haskel, Stian Westlake (2018) "Productivity and secular stagnation in the intangible economy".	・Economic characteristics of intangible capital: **(1) Marginal costs are low** The business size in intangible capital can be more easily expanded than that in tangible capital. (Example) While Uber can provide services for more customers, using intangible capital (softwares), taxi companies need to buy tangible capital (cars) to expand the size of their business. **(2) Spillover effects are high** Intangible capital can be widely shared. Other people's tangble capital can't be used, but other people's intangible capital can be potenatially used. **(3) Sunk costs are high** It's difficult to resell intangible capital in most cases. **(4) Synergy effects are high** It's possible to create big value by combining intangible capital with another intangble capital. (Example) iPhone became a profitable product by combining R&D and its design, and the organizational design and brand in the supply chain.

Source: Prepared from a published document

In Table 17, we organized the functions/roles of data to be considered when considering the estimation of data including the factors referred in the prior studies.

Table 17: Items to be considered on the functions/roles of data

Functions/roles of data	Items to be considered
Complementary factors required for adding value to data	・Various subjects utilize data, so even if each entity uses the same data, the added value might not be necessarily the same. ・It is necessary to consider complementary factors (human resources, investment, organizations, business models, market environment, etc.) required for adding value to data.
Changes in the value of data over time	・In single data, the effects of data utilization on business are different depending on the point of time (today's data, data 1 month ago, data 10 years ago, etc.). On the other hand, if data are utilized by combining data and old data are included, the accuracy of estimation might be improved. ・There is a possibility that the values of data are different depending on the purpose of utilization and method of analysis, so how to process data needs to be considered.
Negative economic value	・By holding data, negative risks such as management costs for security measures and impression drops due to information leakage will arise (negative economic value). ・How much the negative economic value should be considered in measuring the value of data needs to be considered.
Relationship between data and ICT	・ICT (hardware and software) is required for data analysis, so the value of data can be considered by replacing it with the value of ICT in a broad sense. ・It is necessary to consider how clearly the differences between the value of data and the value of ICT should be shown.
Relationship between data and intangible capital	・(Part of) data can be considered to be included in intangible capital, and there are some common characteristics between data and intangible capital, so there is a possibility that referring to analysis methods on intangible capital is useful, considering the similarities and differences regarding functions/roles, etc. ・There are prior studies on the impact of patent on the performance of companies as the application of analogy, so it is necessary to consider if referring to them is useful for measuring the value of data.

Source: Prepared from a published document

Regarding the functions/roles of data, characteristics of data are that data can be used many times without losing its value, data are replicable with almost zero marginal costs (additional costs)[13], and so on (Table 18).

[13] As the case where additional costs arise, a case where purchase costs of hardware/software, etc. associated with adding ICT infrastructure and strengthening the processing capacity arise is expected.

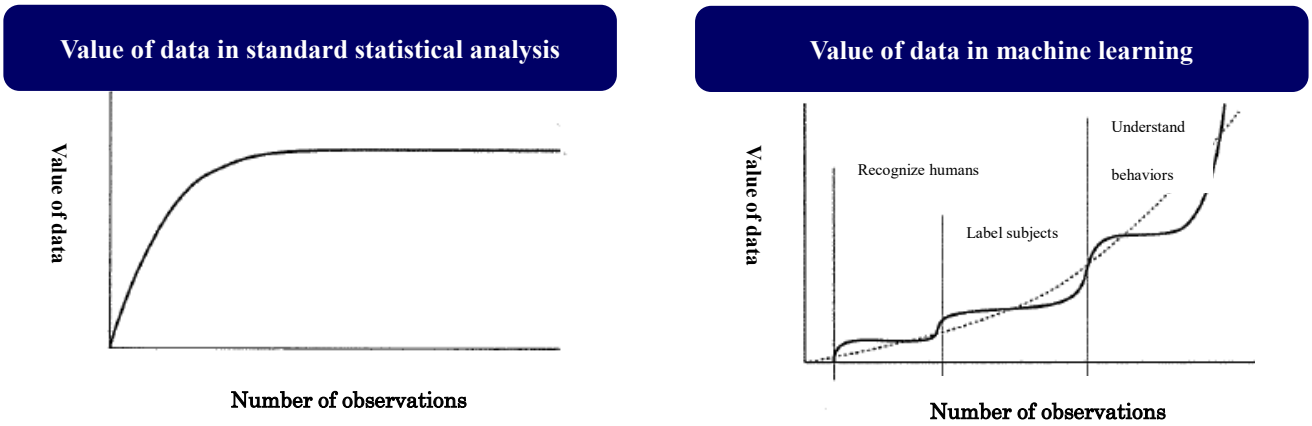
Table 18: Characteristics of data

Characteristics of data	Outline
Marginal costs (replication costs) are almost zero	・Data can be replicated with almost no costs. That is, it is easy for several enterprises to utilize the same data technically.
Value is created only when data are utilized or accumulated	・Collecting data itself isn't always worthwhile, and value is created by utilizing various meanings and knowledge acquired from data.
The value and how to process data are different depending on the contents of data	・The classification of contents of data are for example, personal information, industrial information, open data, etc. Strict management is required for processing personal data, but value through usage is considered to be significant.
When data amount (the number of cases) is not much, the effect of analysis (value of data) is small	・When data amount is small, the notice acquired from that is also small, so the effect through analysis is limited. ・However, it's hard to consider that effects of data can be infinitely big in proportion to data amount.
There is a risk of data leakage	・In case personal information is leaked, enterprises will produce negative value.
Even if the same data is used, the value (effect) of data is different depending on the region and usage	The effects of utilizing data on actual business are different depending on the time when data was collected (today's data, data collected 1 month ago, data collected 10 years ago).

Source: Prepared from a published document

The number of observations and the value of data are considered to be different depending on the methods of analysis. In a standard statistical analysis, if data are estimated with a certain number of observations, the marginal value of data will suddenly diminish. On the other hand, in machine learning, there are several tasks to do, so the more complexity of target is required for solving a problem, the more value will be added to data. However, even in machine learning, if data are estimated within a certain kind of tasks, the marginal value of data will diminish as in a standard statistical analysis, when the number of observations reaches a certain point (Table 19).

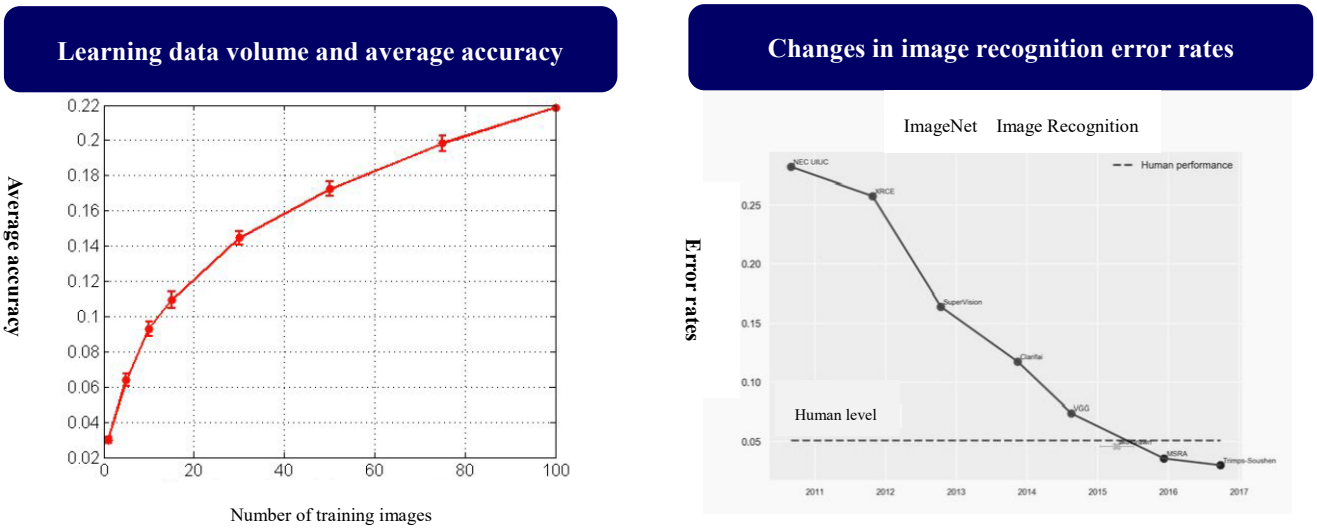
Table 19: The number of observations and value of data ①



Source: Eric A. Posner and E. Glen Weyl (2018) "Radical Markets"

It is shown that in machine learning, as learning data amount increases, the progress rates gradually decrease, though the expected average accuracy increases. In image recognition competition in recent years (the learning data amount is fixed in the period), error rates have declined significantly, and there is an opinion insisting that other factors such as the improvement of algorithm, hardware, and expertise are much more important than learning data amount (Table 20).

Table 20: The number of observations and value of data ②

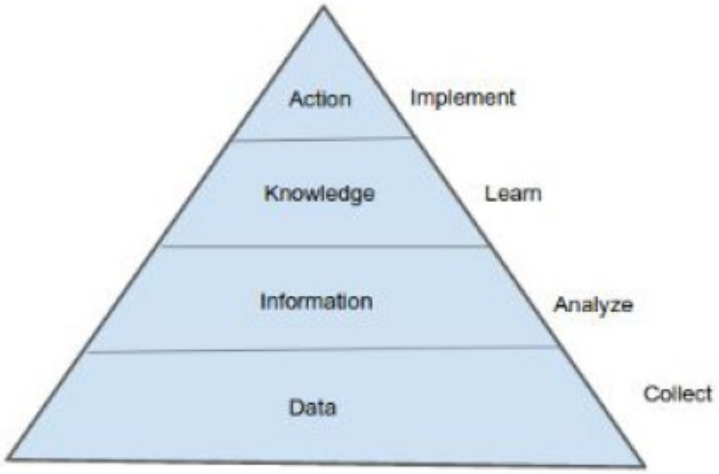


Source: Hal Varian (2018) "Artificial Intelligence, Economics, and Industrial Organization," *NBER Working Paper*, No.24839.

With regard to the roles of data when producing value, there is a concept of "data pyramid" consisting of data, information, knowledge, and action in information science (Table 21). The concept is based on the idea that data are converted into information by collecting data and organizing/analyzing data, and the insight from the information is embodied to turn into knowledge, which is linked to action. In addition, while there are fully developed markets and regulatory environment with regard to information (publications, articles, websites, music, video) and knowledge (labor market, consultants), the data market (unprocessed collection of bits) has not developed. This is because that data greatly

depend on contexts and are not helpful very much until data are converted into information.

Table 21: The concept of the "data pyramid" in information science



Source: Hal Varian (2018) "Artificial Intelligence, Economics, and Industrial Organization," *NBER Working Paper*, No.24839.

(3) Estimation methods of effects/value of data

Data has various functions, roles and characteristics, so it's difficult to define or express the value of data uniquely. Therefore, the attempt to estimate the value of data from several approaches is underway, and methods to estimate the value of data (or data-related things) can be roughly divided into three based on prior studies, etc. (Table 22). Each approach has issues.

Table 22: Estimation Methods of the Value of Data

Approaches	Methods	Tools	Limitations	Issues
Cost-based Approach	・Estimation method which pays attention to costs related to the creation, management, and utilization of data (personnel, and time, etc.). It is based on the assumption that the value of data is worth the cost.	・Calculating personnel expenses ・Calculating time costs ・Calculating security measure costs	• This approach can't take into account quality and productivity. • Since prices or quality of ICT equipment are diverse, the estimated value of data may be affected by other factors. • It is difficult to intuitively envision a relationship between cost and value.	・Which aspect of cost can be appropriately ascertained in order to define the value of data?

Market-based Approach	・Estimation method which is based on the market prices of similar products or of willingness of users to pay for those products	・Market prices (data transactions, etc.) /Representation preference method ・Case studies of M&A, negative effects ・Experimental works	・Lack of data for estimates ・In the case of stated preference method, the result can be affected by the user's bias. ・The value could be perceived differently between the business side and the consumer side.	・How is data in general estimated (not individual data)? ・Which point of view is more appropriate for data estimation, the business side or the consumer side?
Impact-based Approach	・Estimation method where the effects (productivity, and revenue, etc.) of utilizing data is estimated	・Econometric analysis (production function analysis, and regression analysis, etc.) ・Experimental works	・Statistical estimates are needed. ・It is difficult to differentiate between the effects of one type of data and the effects of other types of data.	・It is necessary to clarify subjects for which the value of data is to be estimated (the definition and value of data) ・How can a good analysis model be established based on consideration of complementary elements which are necessary for data to generate value?

Source: Prepared based on a published document

Information and knowledge acquired from data usually can't be targets in marketing, so various estimation methods are proposed in International Monetary Fund (IMF) (Table 23). Currently, there is no fixed method, because there are various issues depending on the approach and there is a possibility that a different value is indicated depending on the approach.

Table 23: Estimation of digital economy in the macroeconomic statistics

Approaches	Concept	Issues
Market-based Approach	・Valued based on market prices of products that can be compared in markets	・Products that can be genuinely compared in market don't exist, excepting the products sold in markets. ・The value of "unprocessed data sold in a market" and that of "processed data suitable for adjusting, cleaning and using" are different, and the prices on the stock market are different from the intristic value.

Cost-based Approach	・Valued based on costs that produce information and know-how acqured from data	・Setting of the range of costs (the entire data assets need to be expanded, considering the costs of data scientists). ・If the costs of platforms are considered as costs for acquiring data, there is a possibility that profits other than the profits that platforms get from users might be ignored.
Incomet-based Approach	・Valued by estimating future cash flow acquired from data	・The discounted present value to be aquired from data (that shows how much value the value to be acquired in the future has currently) might be different from the value in cost-based approach. ・In 2008SNA, causion is needed when income is used in the valuation of assets (It's difficult to decide appropriate assumption on durable years of assets, future cash flow, and discount rates). ・In most cases, a platform's income is attributable to other tangible and intangible capital such as technical and business capability, relationship with customers, network externality, and innovation.

Source: Marshall Reinsdorf, Jennifer Ribarsky (2020) "Measuring the Digital Economy in Macroeconomic Statistics: The Role of Data."

Some specific initiatives related to the estimation of the value of data based on prior studies will be introduced below.

The first is a study that estimated the value of personal information, targeting American platformers. It adopted the approach that estimates the differences between the amount companies pay for targeted advertisements based on personal information and the amount companies pay for untargeted advertisements as the value of personal information, focusing on the advertising revenue that platformers acquire. More specifically, it analyzed the data from AdChoices, a self-regulatory program in the Internet industry, in 2018, and disclosed the advertisement costs that weren't based on personal information were 52% lower than those based on personal information. Based on this, 52% of advertising revenue was regarded as the value of personal information, and estimated approximately 57.7 billion dollars as the value of personal information in advertising platforms in 2018, by multiplying advertising revenue of platformers who utilized personal information by 0.52 (Table 24).

Table 24: Values in personal information in US digital advertisements (Unit: Million dollars)

Platform	2016	2017	2018	Increase
Google	$15,303.6	$18,132.4	$21,453.5	40.2%
Facebook	$6,432.4	$9,344.4	$11,882.0	84.7%
Amazon	$582.4	$920.4	$2,397.2	311.6%
Microsoft	$1,736.8	$1,944.8	$2,339.4	34.7%
Oath (Verizon)	$1,830.4	$1,872.0	$1,917.8	4.8%
Twitter	$707.2	$608.4	$728.2	2.9%
Other	$10,951.2	$14,180.4	$17,045.7	55.7%
Total	**$37,544.0**	**$47,002.8**	**$57,763.9**	**53.9%**

Source: Robert Shapiro and Siddhartha Aneja (2019)
"Who Owns Americans' Personal Information and What Is It Worth?".

The second is a study that investigated the impacts of the leakage of personal information on stock prices. 28 companies that were listed in New York Stock Exchange (NYSE) and had more than 1 million cases of the leakage of personal information were investigated. To get rid of various impacts from political economy, changes in stock prices compared with those in NASDAQ (If the stock price in NYSE increases by 1% and that in NASDAQ decreases by 1%, that would be regarded as an increase by 2%) were observed. When you see them from several categories, you will find the impacts from confidentiality of information were bigger than those from the number of the leakages and the impacts in old days were bigger than those in recent years (There is a possibility that investors have been used to the leakages of personal information that frequently occur and they don't respond strongly) (Table 25).

Table 25: Changes in stock prices due to the leakage of personal information

Category	Classification	6 months before the leakage	6 months after the leakage
Time when the leakage occurred	Before 2011	-15.71%	-3.73%
	2012 - 2015	+9.99%	+0.99%
	After 2016	-9.26%	+4.11%
Industry	Finance	-6.42%	-4.71%
	Technology	+6.79%	-4.48%
	e-Commerce, SNS	-6.1%	+9.87%
	Retailing	-4.26%	-0.47%
	Healthcare	+4.76%	+2.97%
Number of leakage cases	More than 100 million cases	-2.15%	+11.74%
	10 - 99.99 million cases	-1.34%	-1.12%
	1-9.99 million cases	-0.36%	-5.90%
Leaked information	Highly confidential information (credit card numbers, etc.)	-1.74%	-3.52%
	passwords, etc.	-8.86%	+11.02%
	Address, phone number, e-mail address, etc.	+3.57%	-0.11%

Source: Prepared based on Comparitech "How data breaches affect stock market share prices"

The third is a case study on the amount of damage compensation regarding the leakages of personal information in Japan. When personal information leaked out, companies tend to voluntarily pay compensation (including cash vouchers, etc.), so the amount of compensation can be regarded as the value of personal information. When you check the past cases, you will find there are many cases of compensations of 500 yen - 1000 yen per person (Table 26). Also, there are some cases where a lawsuit is filed in addition to voluntary compensation. Sometimes there are some different leading cases even in the same leakage cases, so currently there are no established business practices on the cause of a compensation (if a leakage itself is the cause of a compensation) and how to determine the amount of compensation.

Table 26: Examples of leakages of personal information and compensation in Japan

Year	Organizations that leaked information	Scale of leakage (people)	Average amount (per person)	Total compensation amount (million yen)
1998	University A	1,400	5,000 yen	7
1999	City B	220,000	10,000 yen + attorney's fee 5,000 yen	3,300
2002	Company C	37,000	30,000 yen / 17,000 yen (depending on the secondary damage)	1,110
2002	Company D	560,000	500 yen	280
2002	Company E	183,000	1,000 yen	183
2002	Company F	132,000	5,000 yen	660
2003	Company G	1,150,000	5,000 yen	5,750
2003	Company H	180,000	1,000 yen	180
2004	Company I	4,517,000	500 yen compensation 5,000 yen + attorney's fee 1,000 yen in the lawsuit	2,259
2004	Company J	478,000	500 yen	239
2005	Company K	122,000	500 yen	61
2007	Company L	8,640,000	500 yen	4,320
2008	Company M	123,000	1,000 yen	123
2009	Company N	49,000	10,000 yen	490
2009	Company O	18,000	10,000 yen / 3,000 yen (depending on the leakage of credit card information)	180
2009	Company P	149,000	500 yen	75
2014	Company Q	28,950,000	500 yen	14,475

Source: Prepared based on Mizuho Chuo Law Firm “Cases, standards and market prices of civil legal responsibility (the amount of compensation) of the leakages of personal information”,
Tomoaki Sato, “The leakages of personal information and liability for damages” and afterwards”, etc.

The fourth is the referring that how the importance of data should be considered when we consider the merger and acquisition (M&A). Currently, specific methods of estimating data are not clearly indicated, though the necessity of estimating the value of data and data related risks and the perspective of estimation are referred (Table 27).

Table 27: The value of data and data related risks to be considered in M&A, etc.

Value of data	Risks of legal aspects	Risks of technical aspects	Risks of other aspects
・Value of data as capital ・Value of data through usage ・Expected future value of data	・Existence of legal regulations that restrict the transmission or use of data ・Existence of contractual use restriction ・Existence of data-related great responsibility	・Cyber security risks ・Forms of data storage ・Management methods of data ・Quality of data	・Risks of hollowing-out of human resources ・Risks of losing intellectual property ・Data protection/privacy policies

Source: Prepared from a published document

The fifth is the estimation method of value of consumer data that are provided for business operators in regulations of California Consumer Privacy Act in the US. Though a specific estimation formula isn't mentioned, enterprises are required to consider at least one of the following items[14].

1. Marginal prices for the businesses of sales, collection and deletion of consumer data
2. Average prices values for the businesses of sales, collection and deletion of consumer data
3. The total prices of businesses of selling, collecting, and deleting consumer data divided by the total number of consumers
4. Profits produced from businesses through selling, collecting and holding personal information of consumers
5. Costs related to selling, collecting, and holding personal information of consumers
6. Costs related to presenting, providing and requesting the differences between monetary incentive/prices and services
7. Profits produced from businesses through selling, collecting, and holding personal information of consumers
8. Other practical, reasonable and reliable estimation methods

[14] TEXT OF MODIFIED REGULATIONS [CLEAN VERSION] TITLE 11. LAW DIVISION 1. ATTORNEY GENERAL CHAPTER 20. CALIFORNIA CONSUMER PRIVACY ACT REGULATIONS PROPOSED TEXT OF REGULATIONS (https://oag.ca.gov/sites/all/files/agweb/pdfs/privacy/ccpa-text-of-second-set-clean-031120.pdf)
The final drafts of these regulations are being examined with other regulations as a package by the California Office of Administrative Law (OAL).

(4) Empirical analysis on the estimation of the value of data

After figuring out holding data/use of data through questionnaires for enterprises and analyzing the relationship between financial information and holding data/ use of data, we positioned “Capital”, “Labor”, “Data”, “Others” as the factors to produce output of enterprises, and analyzed the relationship between holding data/use of data and output through empirical analysis. When doing that, we estimated a production function that is the impact-based approach as a feasible method currently available to estimate the value of data, noting securing versatility/reproducibility applicable to organization of the functions/roles of data and various industries/countries, etc.

Since the actual status of utilization of data in enterprises needs to be grasped in detail in consideration of the method, we conducted the following questionnaire survey for enterprises.

<Questionnaire survey on the utilization of data>

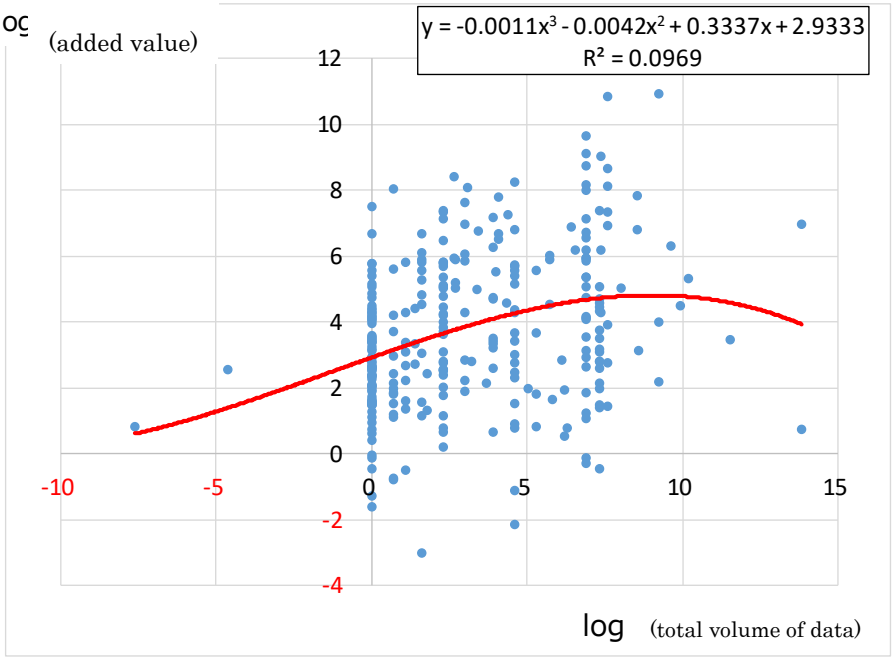
① Target of the survey: All the listed enterprises (3,819 enterprises)＋Non-listed companies (467 enterprises): In total: 4,286 enterprises[15]

② Period of the survey: February 7, 2020 (Fri.) – March 27, 2020 (Fri.)

③ Content of the survey: Data (amount, diversity, etc.) enterprises utilize and the status of analysis (organizations, methods, frequency, etc.) etc.[16]

First, when you see the relationship between the “total amount of data” accumulated by enterprises and “added value[17]” (Table 28/29) by matching the results of the questionnaire survey and financial information of enterprises, you will find that the enterprises that have bigger total amount of data have bigger added value, which suggests that data have a positive impact on production activities.

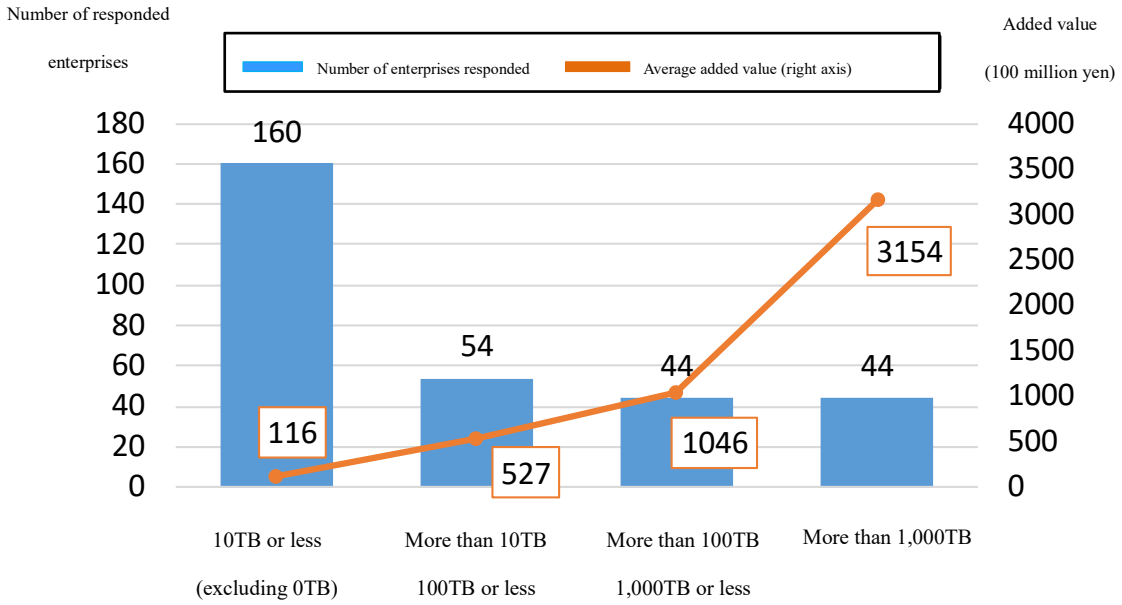
Table 28: The relationship between total amount of data (logarithmic value) and added value



Source: Prepared based on MIC, “Survey on the utilization of data”

[15] Enterprises that were confirmed to have disclosed financial information were targeted.
[16] Regarding the detail of questionnaire, please refer to 6. Addendum.
[17] It was defined as “Net sales”－“Sales costs”－“SG & A”＋“Depreciation costs”.

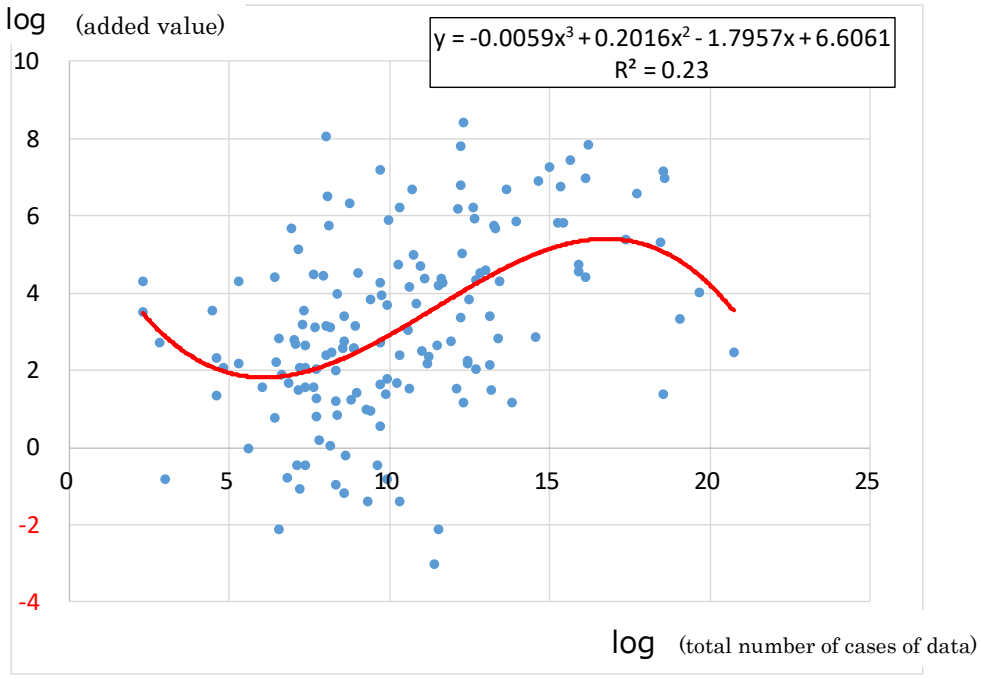
Table 29: Relationship between the total amount of data and added value



Source: Prepared based on MIC, "Survey on the utilization of data"

Likewise, we analyzed the relationship between the "total number of cases of data[18]" accumulated by enterprises and "added value" (Table 30/31). In this survey, the enterprises that have larger total number of cases of data have bigger added value, as the same as the enterprises that have bigger total amount of data.

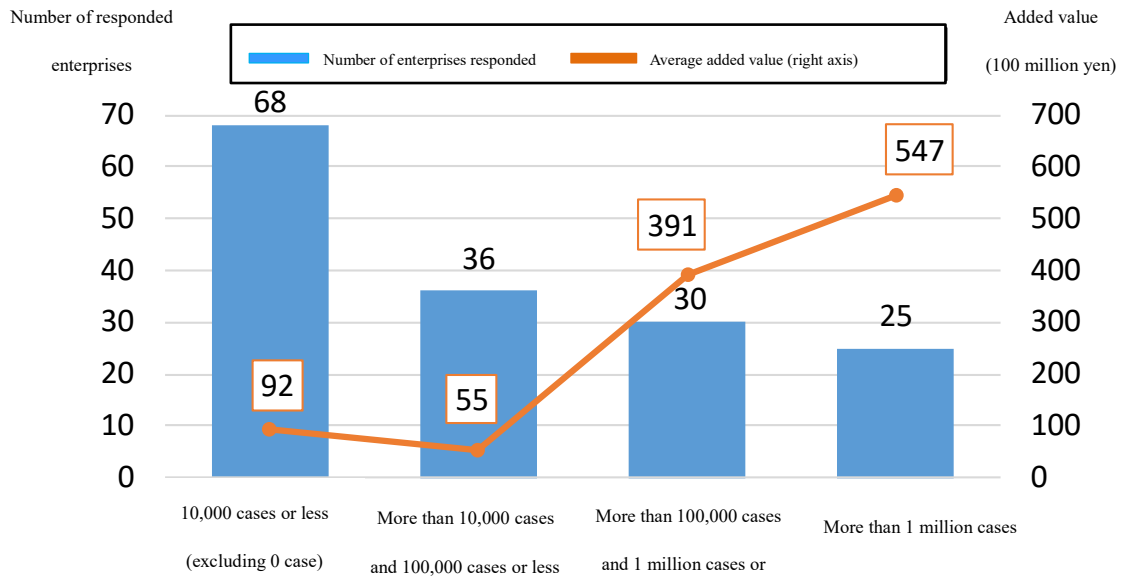
Table 30: The relationship between the total number of cases of data (logarithmic value) and added value



[18] The total of the number of the targets of data (based on units such as number of people, companies, customers, etc.)

Source: Prepared based on MIC, “Survey on the utilization of data”

Table 31: The relationship between the total number of cases of data and added value



Source: Prepared based on MIC, “Survey on the utilization of data”

Since it is assumed that data affects production activities of enterprises, the study positioned data as a production factor, as with other production factors (capital and labor), and conducted empirical analysis with the Cobb-Douglas Production Function modified to include data as a factor (linear homogeneity was not assumed).

$$V = A_o K^{\alpha} L^{\beta} Data^{\gamma}, \qquad \log(V) = \log A_o + \alpha \log(K) + \beta \log(L) + \gamma \log(Data) + Industry\ Dummy^{*}$$

*Classified into manufacturing and non-

Here, V stands for added value, K stands for capital (tangible fixed assets ＋ intangible fixed assets), L stands for labor (the total number of employees[19]), and Data stands for data variables. As estimated analysis results are shown in Table 32, the amount/number of cases of utilized data have a positive impact on added value as well as other production factors (capital, labor), and in the analysis conducted using the amount/number of cases of utilized data in 3 years ago, the amount/number of cases of utilized data had a positive impact on added value. Due to the fact, it can be presumed that the utilization of data has an impact on the increase in added value, even if you consider the timeline. In addition, an increase in the amount/number of cases of utilized data by 1% resulted in an increase in added value by 0.05%. This result is a reflection of the current status of the utilization of data, and it should be noted that the increase in the amount/number of cases of utilized data is not necessarily connected with the increase in added value and the average results of the enterprises that have enhanced effects and the enterprises that have not enhanced effects are shown in this study.

[19] The total number of employees was set as ‘the number of employees at the end of the term ＋ the number of temporary employees,’ and when the number of temporary employees was deficient, it was treated as zero.

Table 32: Empirical analysis results[20]

Data variable	Sample size	Adj R^2	*K* (Capital)	*L* (Labor)	*Data* (Data)
Utilized data amount (= total data amount × ratio of data used for analysis)	258	0.8343	0.44 ◎	0.50 ◎	0.05 ◎
Utilized data number (= total number of data × ratio of data used for analysis)	135	0.8157	0.34 ◎	0.55 ◎	0.07 ◎
Utilized data amount (FY2015)	258	0.8332	0.44 ◎	0.51 ◎	0.05 ◎
Utilized data number (FY2015)	135	0.8136	0.34 ◎	0.56 ◎	0.06 ○
Externally acquired data amount (= total data amount × ratio of data acquired from outside)	267	0.8401	0.47 ◎	0.47 ◎	0.05 ◎
Externally acquired data number (= total number of data × ratio of data acquired from outside)	140	0.8193	0.35 ◎	0.51 ◎	0.07 ◎
Internally owned data amount (= total data amount - externally acquired data amount)	267	0.8379	0.47 ◎	0.48 ◎	0.05 ○
Internally owned data number (= total number of data - externally acquired data number)	140	0.8134	0.35 ◎	0.55 ◎	0.06 ○
Total data amount × data utilization rate (use area/type/processing method)	261	0.8334	0.44 ◎	0.49 ◎	0.05 ◎
Total number of data × data utilization rate	137	0.8150	0.32 ◎	0.58 ◎	0.05 ○
Total data amount × data diversity (diversity of data acquisition and provision)	173	0.8537	0.53 ◎	0.42 ◎	0.04 ○
Total number of data × data diversity	85	0.8576	0.41 ◎	0.49 ◎	0.06 ◎

(Note) ◎: Significance level 1%, ○: Significance level 5%, △: Significance level 10%

When what the above empirical analysis means applies to actual business scenes, the following situation can be assumed. Having said that, however, it's just an "assumption" and it should be noted that the assumption isn't based on the grasp of initiatives of enterprises via hearing, etc.

① Manufacturing industry: The enterprises that grasp the operational status of facilities and detect abnormities of facilities through the visualization of data of sensors, etc. tend to have realized stable production.

② Retailing industry: The enterprises that have finely classified the types of customers through customer attributes and past purchase data, etc. and have taken optimum marketing measures based on data tend to have increased their sales.

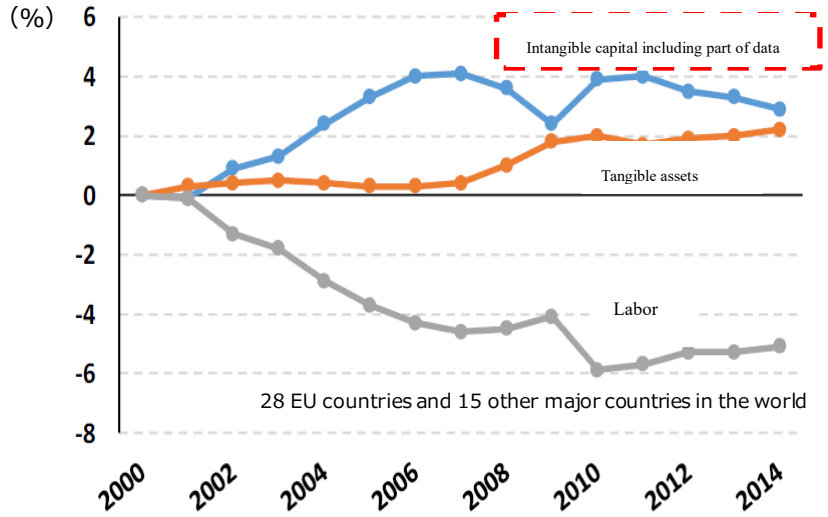
[20] The data variables without the description of FY2015 are values in FY2018. In addition, the proportion of the data obtained from external was corrected based on the acquisition status of data from the external.

③ EC sites: The EC sites that are providing services for recommending optimum products for customers based on customers' favor and past behavior pattern data tend to have increased sales.

3.2 The ideal fair remuneration depending on the effects/value of data

To enjoy the benefits of effects/value of data in the entire society, fair remuneration should be distributed to appropriate subjects. When you see the distribution to capital/labor in recent years, you will see the distribution has shifted to intangible capital including part of data (Table 33). If there is a possibility that most of the wealth produced from data is distributed to enterprises that collect and utilize data, it should be considered what the ideal fair remuneration is depending on the effects/value of data that is expected to occupy an important place in economic activities.

Table 33: Distribution ratio of intangible capital, tangible capital, and labor (cumulative variation width)



Source: Chen, Los & Timmer (2018) "Factor Incomes in Global Value Chains: The Role of Intangibles," *NBER Working Paper* 25242.

Based on the above, we will consider the issues on the approach of "paying remuneration to users who produced data" after reviewing the concept of approaches that take data as capital or labor as a production factor. Though data is generally treated as capital, there is an opinion that data should be treated as labor and monetary rewards should be given to individuals who produce data, considering the characteristics of AI that depends on the active involvement of individuals who produce related data. "Data as Capital" is a concept that values the utilizers of data (AI companies/platformers/data scientists, etc.) as the contributor to the value created from data. On the other hand, "Data as labor" is a concept that values the producers of data as the contributors to the value created from data from the perspective that recognize the contribution of individuals who produce data as general labor.

Table 34: Major characteristics of "Data as Capital" and "Data as Labor"

	Data as Capital	**Data as Labor**
Contributors to creating the value of data	Value utilizers of data	Value the producers of data
Incentive	Entrepreneurialism	Usual contribution (equal to labor)
Future of Labor	Universal Basic Income	Data Labor
Expected concerns	・AI's contribution to improving productivity might be thwarted because AI depends on active involvement of individuals who generate related data. ・Market dynamism might become stagnant if the data generated by individuals are concentrated in certain companies.	・Tragedy of the anticommons (possibility that assets to be shared are subdivided and privatized, then utilization of resources useful in society can be hindered) might arise. ・The original motivation for the activity might be lowered due to the monetary consideration.

Source: Prepared by referring to Imanol Arrieta-Ibarra, Leonard Goff, Diego Jiménez-Hernández, Jaron Lanier and E. Glen Weyl. (2018) "Should We Treat Data as Labor? Moving Beyond 'Free'," *American Economic Association Papers & Proceedings*, Vol.108, pp.38-42, other reference information, and discussion in the Special-Interest Group on Data

With regard to the approach that values data producers as contributors to the value created from data, EU proposed data producer's right as an item of "A future EU framework for data access" presented by EU in 2017, about non-personal data collected via IoT devices/connected cars, etc. that are exempt from the protection based on General Data Protection Regulation: GDPR[21](Table 35). After that, concerns about establishing a new proprietary right on data were expressed by many stakeholders in public comments, and the necessity of securing the access to data shared by enterprises was rather claimed, then the proposal was excluded from policy options[22].

[21] In May 2018, it was fully enforced, and unified rules on the protection of personal data in EU were enacted. The right to data portability (it's the right to retrieve his/her personal data that they have provided for a controller in a machine-readable format, and it allows for data subjects to directly transmit their personal data to another data controller.) is also included in this regulation.

[22] The perspective of sharing data among enterprises referred to the said framework is inherited (such as support for dealing with problems concerning the commonly generated data (in industrial IoT, etc.) in the plan concerning the Data Act (scheduled to be enacted in 2021) in the European strategy for data (February 2020) and clarification of the legal responsibility for data sharing).

Table 35: Outline of A future EU framework for data access

A future EU framework for data access
・Guidance on incentivizing businesses to share data
・Fostering the development of technical solutions for reliable identification and exchange of data (API, etc.)
・Default contract rules (including reviewing Unfair Contract Terms Order)
・Access for public interest and scientific purposes (including B2G)
・Data producer's right
・Access against remuneration (FRAND (Fair, Reasonable, And Non-Discriminatory), etc.)

Source: Prepared based on COMMISSION STAFF WORKING DOCUMENT on the free flow of data and emerging issues of the European data economy Accompanying the document Communication Building a European data economy {COM(2017) 9 final}(https: //ec.europa.eu/digital-single-market/en/news/staff-working-document-free-flow-data-and-emerging-issues-european-data-economy)

In addition, regarding personal data, "MyData Global[23]" was established on 11 October 2018, and it promotes initiatives to aim for the world where individuals fully understand their own data and utilize their personal data for themselves with independence and self-direction. This organization upholds reforms and principles through MyData, aiming to promote individual-centered visions on personal data, by recovering from unbalanced situations under the recognition of the current situation that the power balance on personal data is extremely biased on organizations who have the authority to collect and distribute personal data and make a decision on them, and individuals can't control what will happen to their personal data (Table 36).

[23] It established many bases around the world, centering on Europe. It is organized by approximately 100 organization members and 400 individual members in over 40 countries.

Table 36: MyData Global's Reforms and Principles

Reforms by MyData Global	
1. From formal to actionable rights	・We intend access and redress, portability, and the right to be forgotten, to become "one-click rights": rights that are as simple and efficient to use.
2. From data protection to data utilization	・We intend to change common practices towards a situation where individuals are both protected and empowered to use the data that organizations hold about them.
3. From closed to open ecosystems	・Today's data economy creates network effects favoring a few platforms able to collect and process the largest masses of personal data. ・By letting individuals control what happens to their data, we intend to create a truly free flow of data – freely decided by individuals, free from global choke points – and to create balance, fairness, diversity and competition in the digital economy.
MyData Global's Principles	
1. Human-centric control of personal data	・Individuals should be provided with the practical means to understand and effectively control who has access to data about them and how it is used and shared. ・Ultimately, we want the terms and conditions for using personal data to become negotiable in a fair way between individuals and organizations.
2. Individual as the point of integration	・The value of personal data grows exponentially with their diversity; however, so does the threat to privacy. ・This contradiction can be solved if individuals become the "hubs" where, or through which cross-referencing of personal data happens.
3. Individual empowerment	・We want individuals to be able to securely manage their personal data in their own preferred way. We intend to help individuals have the tools, skills and assistance to transform their personal data into useful information, knowledge and autonomous decision-making.
4. Portability: access and re-use	・The portability of personal data, that allows individuals to obtain and reuse their personal data for their own purposes and across different services, is the key to make the shift from data in closed silos to data which become reusable resources. Data portability should not be merely a legal right, but combined with practical means.
5. Transparency and accountability	・Organizations that use a person's data should say what they do with them and why, and should do what they say. ・They should take responsibility for intended, as well as unintended, consequences of holding and using personal data, including, but not limited to, security incidents, and allow individuals to call them out on this responsibility.
6. Interoperability	・In order to maximize the positive effects of open ecosystems, we will continuously work towards interoperability of data, open APIs, protocols, applications and infrastructure, so that all personal data are portable and reusable, without losing user control.

Source: Prepared based on MyData Global (https: //mydata.org/declaration/)

As for paying remuneration to users who produced data, while people in the standpoint of "Data as labor" take it

positively, people in the standpoint of "Data as capital" have pros and cons. In addition, there are expected issues in both standpoints (Table 37).

Table 37: Issues on the approach of "payment of remuneration to users who produced data"

	Data as Capital		**Data as Labor**
Payment of remuneration to users who produced data	Positive	Negative	Positive
Reasons for paying/not paying remuneration	・The importance of data will be enhanced along with the advance of AI, so incentive to provide data will be needed ・Existence of enterprises that have acquired most of the value produced from data as private profit	・Free data sets and free machine learning algorithm are available ・The value of data can be created only when data is analyzed, so data analysts should take remuneration. ・There is a possibility that the original motivation of activities might be lowered due to giving monetary consideration	・Data as Labor will give an important opportunities to complement the lack of income, and bring citizens who are suffering from the expansion of disparity the awareness of contribution to society.
Process of payment of remuneration	・Market mechanism (micro payment/information trust function (Information trust banks)) ・Regulations (competition policy/tax system/social security, etc.)	—	・(Limitations) Estimate the contribution and build an appropriate technology system to retroact and detect the value individual users produced
Expected issues	・ Market mechanism: Establishment of business models ・Regulations：objective validity of introduction (quantitative index, etc.)	There is a possibility that the uneven distribution of wealth between enterprises, data analysists and other individuals (further lowering of labor distribution rates) will arise.	・ (Limitations) Estimate the contribution and feasibility of technology ・There is a possibility that the estimation of contribution will bring surveillance society and promotion of disparity and discrimination

Source: Prepared from a published document

The conclusion of whether data as a production factor should be positioned as capital or as labor cannot be reached at this point. On the other hand, it is considered that considering the system where individuals who produce data are given some consideration (other than using the service for free) is needed so that the entire society can enjoy the benefits of active

utilization of data and effects/value of the utilization, data.

On this point, we need to consider what kind of system will function and how much consideration should be given. For example, if the estimation method of the value of data mentioned in 3.1 is established, utilizing the method will be possible. To distribute the estimated value of data to individuals in the most precise form, the degree of contribution of individuals to individual data should be estimated. However, regarding the contribution of individuals to data production, some people may contribute a little in various fields, and other people may contribute to data production greatly to one field but may not contribute to other fields very much. Therefore, grasping these situations in a precise, safe and secure manner has a technical limitation at this point. It is pointed out that if we try to estimate the degree of contribution precisely, there is a risk that it will bring surveillance society and promotion of disparity and discrimination. On the other hand, if the degrees of contribution to the value of data between data utilizers such as platformers and data scientists and individuals who are data producers in the entire value of data can be estimated, the total value of data to be distributed to individuals will be able to be estimated, so based on that, the "average" remuneration will be able to be paid to individuals[24]. However, there are some technical difficulties at this point as mentioned above.

As another approach to consider the consideration from the perspective of individuals who produce data, there is a method to check the minimum "monetary consideration" requested when they are providing their information for enterprises. When "monetary consideration" requested by individuals were checked in a questionnaire survey for individuals, the results of central values were 500 yen - 1,000 yen in many data (Table 38).

In highly confidential "Fingerprints", "Credit card numbers" and "Face photograph", the average values of data were over 10,000 yen, which shows that individuals highly hesitate to provide their data to enterprises. It is required to consider the system where monetary consideration is given, keeping in mind that average values and median values are quite different in any data, and there are a certain number of people who request high monetary consideration (don't want to provide their data without high monetary consideration).

[24] Eric A. Posner and E. Glen Weyl (2018) "Radical Markets" stated the "average value" should be paid on condition of satisfying the standard of overall quality.

Table 38: Minimum “monetary consideration” required for providing information for enterprises

	Average value (yen)	Central value (yen)
Fingerprints	10,879	5,000
Credit card number	10,641	1,000
Face photograph	10,114	3,000
Passport number	7,905	1,000
Bank account information, etc.	7,710	1,000
Mobile information	5,705	1,000
Telephone number	4,654	1,000
Information posted on SNS and blogs	4,648	500
Bio-information	4,565	1,000
Address information	4,378	1,000
Purchase information at retailer shops and restaurants	3,364	500
Workplace information	3,163	500
Electric money, usage history of IC cards, etc.	3,028	1,000
E-mail address	3,004	500
Browsing history of websites	2,915	500
Purchase information on EC sites	2,866	500
Family information, housing information	2,445	500
Sex, age, educational background	2,409	500
Insurance information	2,177	500
Annual income	1,627	100
Birthplace	1,575	100

Source: MIC “The questionnaire on awareness on data” (2,155 people responded)

3.3 The current status and issues toward marketing and visualization of data transaction

While the roles of data are enhanced in production activities of enterprises, the utilization of data has not been sufficient yet in Japan, which was mentioned in 2.1 and 3.1. Mega ICT companies have developed their business based on the data collected/accumulated from users, and the utilization of data is concentrated on large enterprises and companies in big cities in Japan. Therefore, while these enterprises enjoy benefits brought from data, other enterprises don’t because the process of utilizing data from collecting data to creating added value hasn’t prepared well in them. One of the means to lower the barrier of utilizing data at enterprises and enable many enterprises to enjoy the benefits brought from data is marketing and visualization of data transaction. Currently, data transaction is partly done, however, creating the environment where various enterprises can freely do data transaction under the condition that the value of data and contract terms are secured will enable enterprises that wish to expand their business by utilizing data to lower the cost of acquiring data. As a result, that might ease the concentration of benefits brought from data on some large enterprises.

In the following, issues to be addressed toward marketing/visualization of data transaction, after checking the current situation in Japan.

(1) The current status of data transaction in Japan

① Global trends

Here, regarding global trends on marketing/visualization of data transaction, the following 4 cases will be introduced.

The UK has a program called 'midata' run by the government[25]. This program allows individuals to receive personal information owned by private enterprises in an easy-to-use form in real time, and the objective of this program is to allow individuals to receive better services through the provided data. In this program, the targeted fields for personal information are the 4 fields of energy, banks, mobile phones, and credit card[26]. Currently, the program is operated in the 2 fields of energy and banks. At the beginning, it was an initiative with no legal ground, but in 2013, Enterprise and Regulatory Reform Act was revised, which enabled the government to enforce providing data by establishing a regulation. (Having said that, however, there is no establishment of the regulation at this point).

As an international agreement, New Zealand, Singapore, and Chile concluded Digital Economy Partnership Agreement (DEPA)[27]. This is a rule actively responding to the expansion of business transaction on the Internet by digitalizing information. More specifically, the objectives of this agreement are: ①to build a trustworthy digital transaction system, and through this system, ②to make trustworthy data distribution possible ③to promote end-to-end seamless transaction. The 3 countries aim to attract global enterprises for investment on their digital economy.

In the US, Data Transfer Project (DTP) was launched by 5 private enterprises; Apple, Facebook, Google, Microsoft and Twitter in 2018. This project has been promoted in the form of Open Source Initiative to encourage more and more providers in addition to the 5 enterprises to participate in the project. DTP has expanded data portability and provides functions to directly send data to the enterprises participating in DTP in order to save time and labor for downloading the replication of data from data owners. It is considered that providing these functions in DTP will enable both data owners and users to reduce the burden of investing in equipment, which will accelerate the utilization of data and make the rationalization of work possible as well as increase the supply of services based on the premise of mutual utilization.

One of the initiatives for utilizing data without owing it is Federated Learning[28] system whose concept Google announced. It's the construction system of learning data necessary for machine learning, etc. While the existing AI system individually collect an enormous amount of learning data and use them, this system tries to build learning data on a smartphone, etc., by decentralizing data. Through this system, the current model can be downloaded to a smartphone from a cloud service first. And the model will be partially improved on the device by learning from the data accumulated on the smartphone, and the changes focusing on the improvement will be sent to the cloud service as the update through encrypted communication, then the model will be improved in the end. Eventually, through this system, the shared model will be

---

[25] In April 2011, this program was set up by Department for Business, Innovation & Skills.

[26] These 4 fields were selected as the targets for regulation that match the concept of midata because of these: ①making a contract (use) for a long period is required ②doing transactions frequently is required, ③the payment system is complicated and it's hard to compare the payment system with that of a competitor.

[27] The websites on DEPA of the governments are as follows.

New Zealand：https://www.mfat.govt.nz/en/trade/free-trade-agreements/free-trade-agreements-concluded-but-not-in-force/digital-economy-partnership-agreement/

Singapore：https://www.mti.gov.sg/Improving-Trade/Free-Trade-Agreements/Digital-Economy-Agreements/The-Digital-Economy-Partnership-Agreement

Chile：https://www.gob.cl/noticias/chile-nueva-zelandia-y-singapur-cierran-las-negociaciones-del-primer-acuerdo-sobre-economia-digital/

[28] Federated Learning: Collaborative Machine Learning without Centralized Training Data (https://ai.googleblog.com/2017/04/federated-learning-collaborative.html)

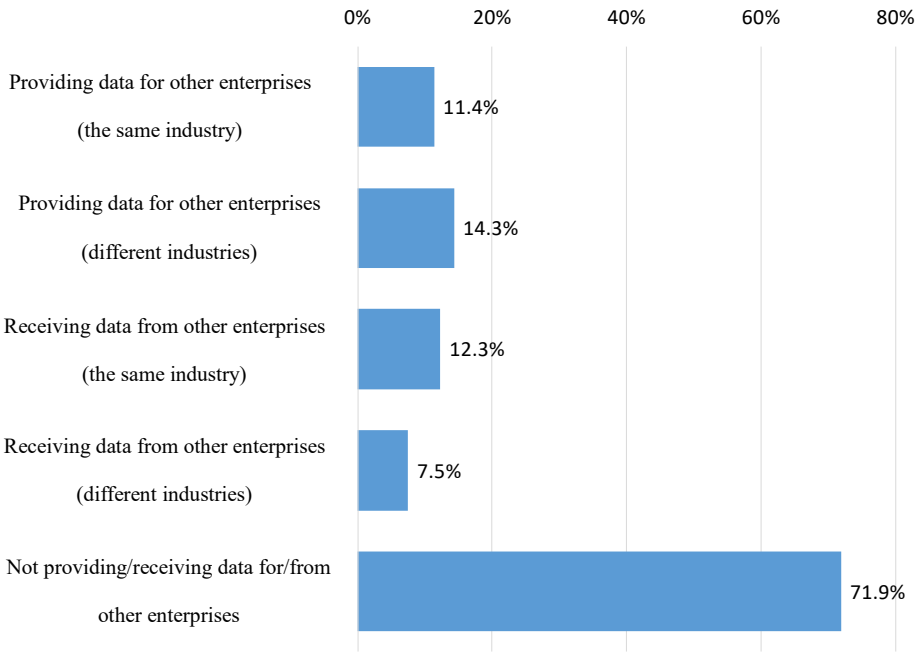
used, and the convenience for many users will be improved.

As mentioned above, initiatives for the marketing and visualization of data transaction have been implemented in the world, so these global initiatives should be noted in formulating policies, business strategies, etc. in Japan.

② Data transaction with other enterprises is inactive

In the questionnaire on the utilization of digital data for enterprises, more than 70% of the enterprises were not providing/receiving data on individuals[29] for/from other enterprises, which shows data transaction with other enterprises is inactive (Table 39). When seen by the combination of providing data or receiving data, and the same industry or different industries, the percentage of receiving data from different industries was the lowest (less than 10%), and even the highest percentage of providing data for different industries was just 14.3%. It is important to specify the factors that impede providing/receiving data in designing a system toward marketing/visualization of data transaction.

Table 39: The status of providing/receiving data on organizations and/or individuals for/from other enterprises (the same industry or different industries)



Source: MIC "The questionnaire on the utilization of digital data" (2,003 enterprises responded)

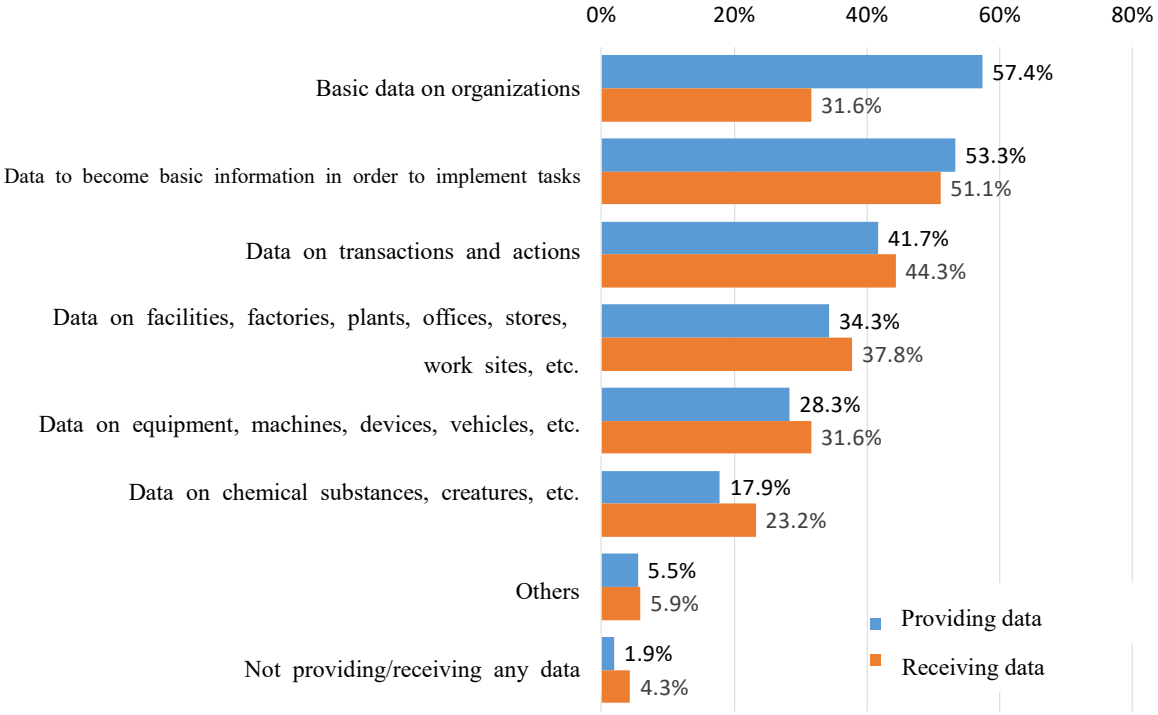
③ Organizational data have different statuses of providing/receiving data, depending on the type of data

When you see the differences in the statuses of providing/receiving data on organizations for/from other enterprises (Table 40), you will find "Basic data on organizations" and "Data to become basic information in order to implement tasks" have higher percentages in providing data, and "Data to become basic information in order to implement tasks" and "Data

[29] Please note that "data on individuals" in this report is used as a general term, and it's not used as a legal term such as "personal data" in Act on the Protection of Personal Information (Act No. 57 of May 30, 2003).

on transactions and actions" have higher percentages in receiving data. On the other hand, "Data on equipment, machines, devices, vehicles, etc." and "Data on chemical substances, creatures, etc." have lower percentages in providing/receiving data. It is important to consider the statuses of providing/receiving data are different depending on the type of data in order to smooth data transaction.

Table 40: The statuses of providing/receiving data on organizations for/from other enterprises

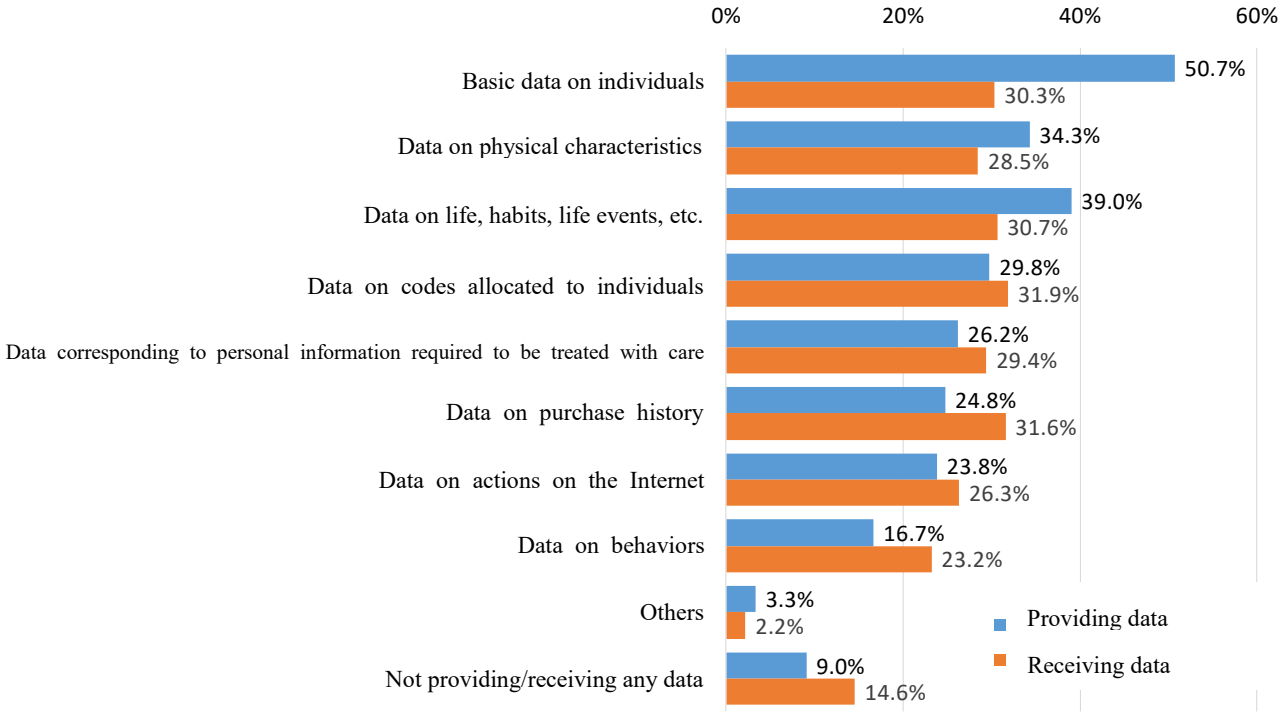


Source: MIC "The questionnaire on the utilization of digital data" (2,003 enterprises responded)

④ The use of data on individuals is asymmetric between the sides of providers and receivers

Regarding data on individuals, "Basic data on individuals" and "Data on life, habits, life events, etc." have higher percentages in providing data, and "Data on codes allocated to individuals" and "Data on purchase history" have higher percentages in receiving data, but "Data on actions on the Internet" and "Data on behaviors" have lower percentages in both providing data and receiving data. Regarding providing data and receiving data, the differences between the highest and the lowest in the percentages of providing data tend to be larger than those in the percentages of receiving data (Table 41). It is important to consider that the tendency of providing data and that of receiving data are asymmetric in order to smooth data transaction.

Table 41: The status of providing/receiving data on individuals for/from other enterprises



Source: MIC "The questionnaire on the utilization of digital data" (2,003 enterprises responded)

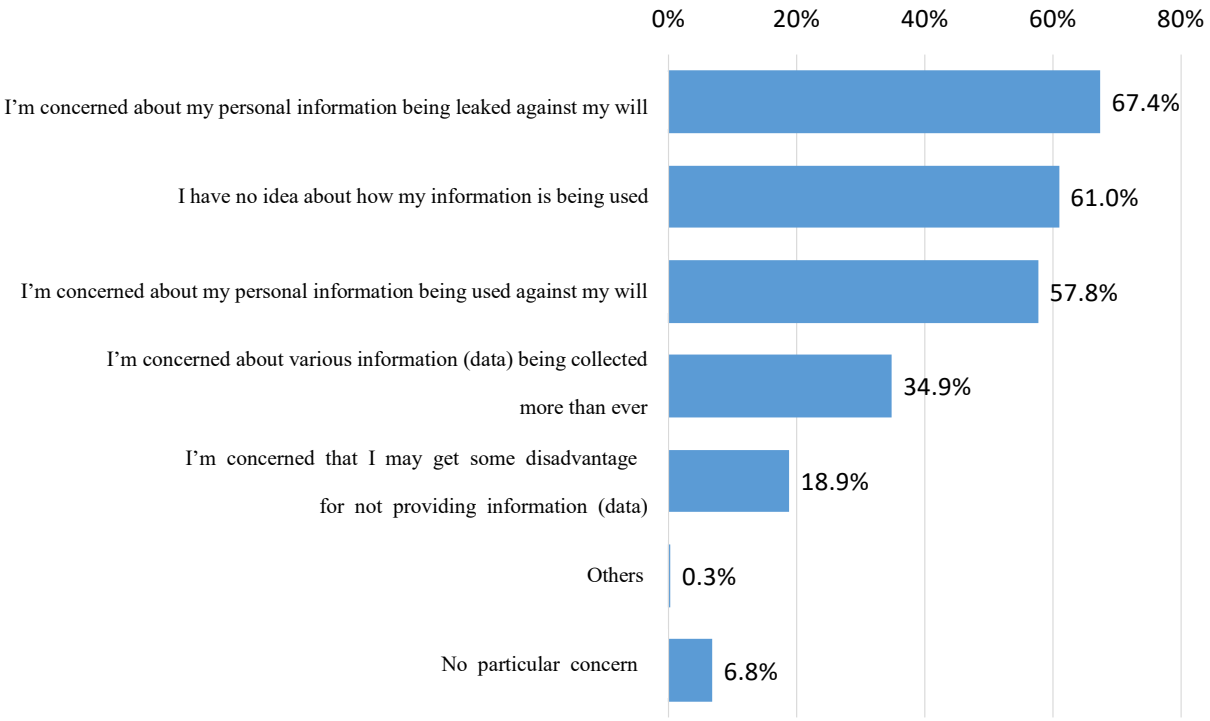
⑤ Individuals are concerned about their data being collected or used

Regarding the concerns about individuals' data being collected and used, "I'm concerned about my personal information being leaked against my will", "No idea about how my information is being used", and "I'm concerned about my personal information being used against my will" have higher percentages (more than 50%) in the responses (Table 42). It is important to consider relieving concerns of individuals in order to smooth data transaction.

⑥ Enterprises tend to be afraid of being criticized for doing data transaction by society

With regard to ⑤, it can be said that concerns about personal data being collected and used include the possibility that the enterprises that collect/use the data might be criticized by society. In particular, in the society where there is a high possibility that enterprises are criticized, the risks of collecting or using data will be excessively increased, and corporate activities on collecting/using data or data transaction might shrink. It should be noted that relieving concerns of individuals about personal data being collected or used has an impact on corporate activities on data transaction.

Table 42: Concerns about personal data being collected or utilized

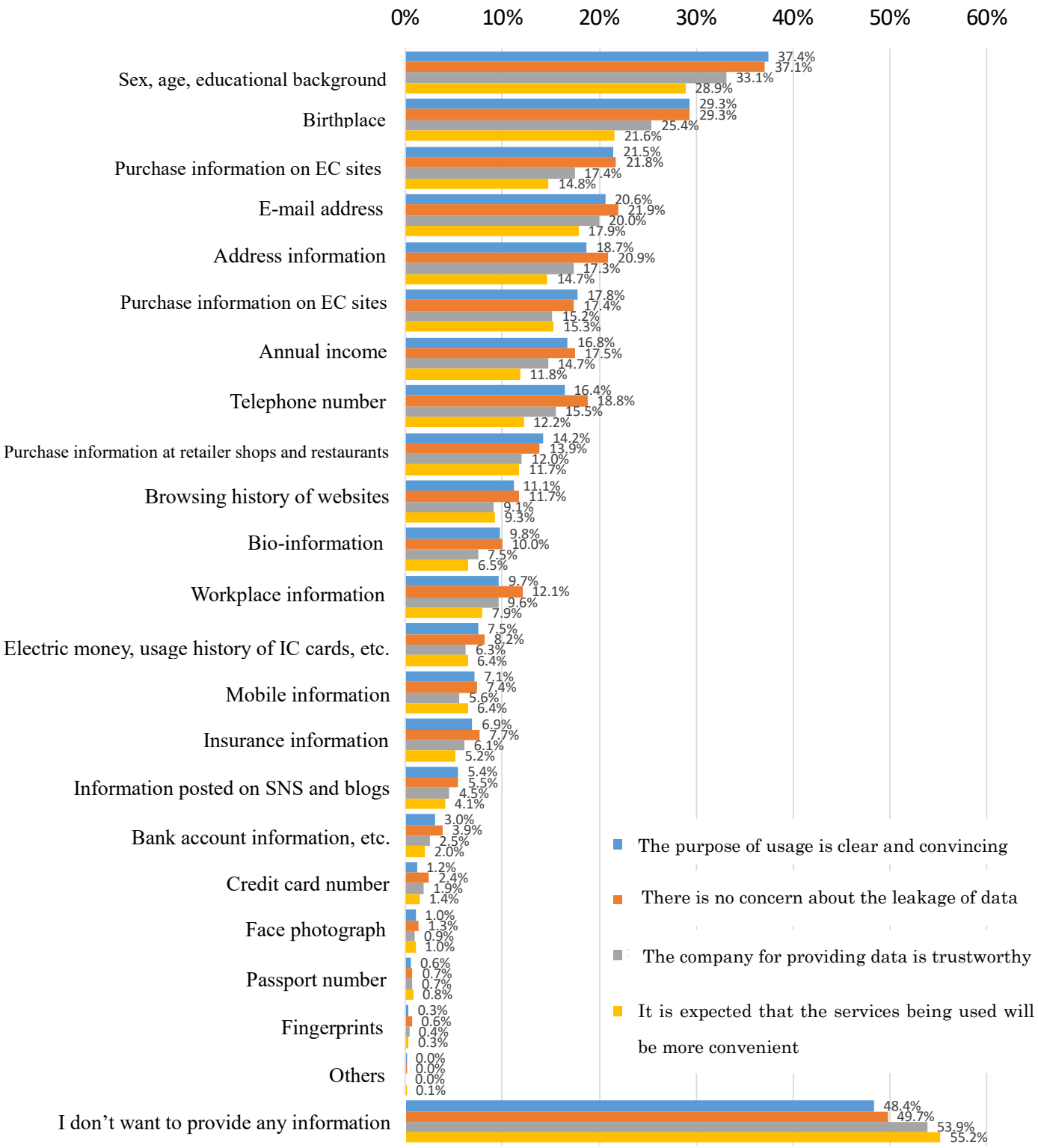


Source: MIC "The questionnaire on awareness on data" (2,155 people responded)

⑦ Individuals are expecting enterprises to ensure the purpose of using data and data security

When you see the results of "Conditions required for providing information for enterprises (Table 43), you will find "The purpose of usage is clear and convincing" and "There is no concern about the leakage of data" have higher percentages, and "The company for providing data is trustworthy" and "It is expected that the services being used will be more convenient" have lower percentages. These generally apply to the data confirmed this time. On the other hand, regarding the items of data, especially in "E-mail address", "Address information", "Telephone number", "Annual income", "Workplace information", "Insurance information", "Electric money, usage history of IC cards, etc.", the percentage of "There is no concern about the leakage of data" was highest in the responses. It is important to minimize the risks of leakage of data in order to smooth data transaction.

Table 43: Conditions required for providing information for enterprises



Source: MIC “The questionnaire on awareness on data” (2,155 people responded)

(2) Issues to be addressed toward the marketing/visualization of data transaction

① Standardization of contract formats concerning data transaction for the utilization of data

The characteristics of data utilized for AI are different from those of the existing data, so different factors are required for the contracts of data transaction. It is considered that the standardization of contract formats that incorporated necessary factors after considering the characteristics of data to be utilized for AI is required in order do appropriate transactions. The 3 important factors to be considered for the standardization will be explained below.

(a) Contractual coverage according to sales channels/price negotiation/manufacturers' responsibility

The important thing about the differences between the characteristics of existing data and those of data to be utilized for AI is that learning data to be utilized for AI might create value beyond the expected scope and keep creating value permanently by changing its form while existing data mostly end playing its role with the visualization and analysis of tendencies and characteristics. This characteristic is very close to that of proprietary technology, so the contract contents of data to be utilized for AI should be determined based on distribution channels (whether it will be distributed to a third party, etc.) as well as those of proprietary technology.

The price negotiation of data to be utilized for AI should be also determined based on distribution channels. Here, distribution channels have various factors; whether the licensee is the end user or a component user will resell the data to another manufacturer. In addition, one of the characteristics peculiar to the data to be utilized for AI is whether the quality of data is good enough for providing for business cannot be found until the data is used for trial. This characteristic should be also considered in price negotiation.

Regarding the responsibility of manufactures of products[30], it is generally considered that the contributor to the damage is liable for the damage in the Civil Code or other regulations, so the liability for damages may accrue if there is an accident in a product that incorporated the software utilizing AI (such as a learned model). However, in case of the software that utilized AI, there will be cases where it is difficult to grasp where the responsibility lies because of the distribution of responsibility (whether there is delinquency on a debt, attribution of responsibility/cause and effect relationship, etc.) and difficulties learned models have (risks of abortion of development, problems of quality and performance of learned models, integrated problems of external systems). For example, if the data to be utilized for AI have malicious data or data with large deviation, the quality of products may have a problem in quality as a result. In this case, it is considered that there might be cases where grasping the liability for damages in a learned model becomes difficult due to the following factors[31].

- It is technically difficult to assure the performance of learned models with unknown input data in advance.
- It is technically difficult to verify cause and effect relationship, etc. after an accident.
- The performance of learned models, etc. depends on the quality of learning data sets.
- The property of AI products depends on the quality of input data at the stage of utilization.

Therefore, how much responsibility manufacturers of products should take depending on distribution channels needs to be stipulated in contract forms.

---

[30] Please refer to METI "Contract Guideline on the Utilization of AI/Data－AI Edition－" (June 2018)

[31] Please refer to METI, the above mentioned, p33-p34.

(b) Necessity of grasping the distribution/quality/traceability/freshness, etc. of data

The distribution of data is important in the data to be utilized for AI. For example, when the quality of English voice recognition and that of Japanese voice recognition are compared, the quality of English voice recognition is higher than that of Japanese voice recognition. Because there are more English voice data as learning data. On the other hand, since there are little Japanese voice data of seniors and specific dialects of local people, the purchase needs for these data are considered to be larger to enhance the quality of these voice recognition. It is necessary to grasp the distribution of these languages/regions/ages, etc. in data transaction.

In addition, low-quality data (for example, data with wrong labeling such as cars labeled as persons, knives labeled as pens, railroads labeled as crosswalks, etc.) is utilized as learning data without knowing the wrong labeling, that will affect the processing of AI. Therefore, grasping the quality of data is necessary in data transaction.

To grasp the distribution and quality of data, checking where the data was produced and who gave the information such as labeling data, etc. is necessary, so grasping the traceability is necessary in data transaction.

Furthermore, when utilizing data for AI, the value of data is subject to change as time goes by depending on the field. Therefore, the freshness of data is also important. For example, if a conversation model is prepared by utilizing the conversation data of teenagers collected five years ago and the model is applied to the current teenagers, the quality will be deteriorated. However, if the model is applied to the conversation data of the youth who were teenagers 5 years ago, the deterioration of the quality will be relieved to some extent. This is because the way of speaking of the youth and terms used by the youth can greatly change in 5 years. That's why the freshness of data needs to be grasped in data transaction.

Having said that, however, it is not always necessary to grasp all the above mentioned. There are some cases where using the data won't affect and damages are indemnified or restored if some problems occur. It is desirable to determine if grasping the said distribution, etc. is necessary, considering the balance with efficiency in each data transaction.

(c) Identification and fidelity guarantee of persons in charge of data labeling/annotation

In Japan, there are a few AI vendors who collect data and conduct data labeling and annotation on their own. Therefore, there may be cases where problems occur if malicious annotation (for example, data with wrong labeling such as cars labeled as persons, knives labeled as pens, etc.) is done. In data transaction, the identification and fidelity guarantee of persons in charge of data labeling/annotation need to be stipulated when signing a contract.

② The necessity of management of data owned by enterprises by a third party

The data owned by enterprises are always exposed to the danger of being exploited in society through the enterprises or the products/services. When an enormous impact of exploited data on society is considered, it is required to prepare the system that manage data owned by enterprises if necessary, in order to smooth corporate activities in the future data economy.

The transaction of data to be utilized for AI is quite different from the existing market transaction, so many complicated factors such as contract contents following distribution channels are involved in data transaction. Therefore, if enterprises do all the procedures necessary for purchasing and utilizing data such as price negotiations, confirmation of rights, contract procedures, and security/authentication individually, they would spend much time and labor in negotiations

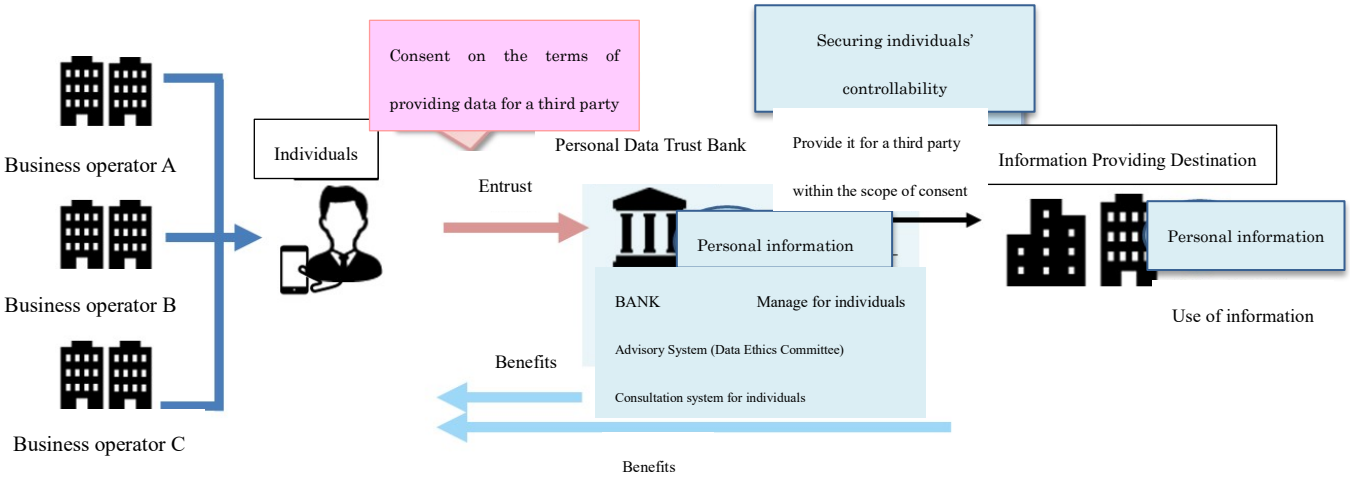
and procedures and have many risks of making mistakes, which will bring about inefficiency to the providers and users of data, or the entire society.

In case of large enterprises, the legal department can take on the role and handle confirmation of rights, contract procedures, etc. However, when small and medium-sized enterprises including start-ups don't have the system to manage data such as the legal department, they will make a mistake in various procedures to process data and data transaction itself, and they might be forced to give up doing data transaction. Therefore, to disseminate the provision/use of data especially among small and medium-sized enterprises smoothly, data management by a third party is indispensable.

③ The necessity of management of personal data by a third party (Personal Data Trust Bank's potential)

For the purpose of enhancing effective controllability and promoting the distribution and utilization of personal data, the existence of Personal Data Trust Bank who can entrust trustworthy entities to provide personal information within the scope of individuals' consent is required for the marketing of individual data transaction.

Table 44: System of Personal Data Trust Bank



Source: "Guideline of Certification Schemes Concerning Functions of Information Trust ver2.0" (https://www.soumu.go.jp/main_content/000649152.pdf)

Regarding the contents of services Personal Data Trust Bank can provide for individuals (types of data Personal Data Trust Bank can handle, conditions for choosing information providing destinations, terms of use in the information providing destinations), when the Personal Data Trust Bank appropriately makes a proposal and secure the responsibility of the Personal Data Trust Bank regarding the services by concluding contracts, etc. with the consent of individuals, screwing up important life events such as marrying and getting a job through identity theft by using personal information due to illegal acquisition of personal information or spreading/misusing of inaccurate personal information can be prevented.

There is a possibility that providing personal information may not be proceeded if this kind of environment for the utilization of data is not prepared. For example, according to the questionnaire survey conducted by NTT Data Institute of Management Consulting, Inc., consumers tend to refuse providing information that can easily identify individuals such as "location information" and "address/telephone numbers", and there are a lot of opinions that request "authentication/certification by a third party" about the information providing destination. On the other hand, regarding "hobbies/tastes", "the history of viewing TV programs", "electricity, gas, and water usage charge", "height, weight,

pedometer measurement, etc.", there are many opinions that they can provide those data on condition that they get paid[32]. Regarding the transaction of individual data, consumers' attitude is stern in terms of security and economy. And the consumers' attitude brings about withering of enterprises, so the system which can endure that is required.

The existence of the Personal Data Trust Bank is important in the transaction of individual data. The Personal Data Trust Bank business has already started, but the dissemination of this business is important in marketing and visualization of data transaction. In addition, the system of the Personal Data Trust Bank deserves to be evaluated as the market-based system that gives some consideration to individuals who produce data.

On the other hand, there is a movement that credit scores on the social evaluation of individuals are calculated and utilized for services. According to the document of the Review Meeting for the Ideal Certification Schemes Concerning Functions of Information Trust[33], the Personal Data Trust Bank and credit scores are described as follows.

- If the system of the Personal Data Trust Bank spreads, the collection of various data on individuals will also spread, and the production and distribution of "credit scores" might be promoted.
- The "credit scores" don't have a clear definition, but as for scores given to individuals, the evaluation of credit capabilities and the scores of English tests can be called scores. These scores in the broad sense are currently used widely and generally, and the convenience is expected to be improved through the distribution via the Personal Data Trust Bank.
- On the other hand, regarding the credit scores on the social evaluation of individuals beyond partial capabilities of individuals, there was an opinion that points out the risk that the utilization of credit scores might cause the discrimination or screening by impairing a diversified society due to the increase in the number of people who value scores too much or by using scores for marriage and employment.
- Considering the above, regarding how to treat credit scores that might cause discrimination through the utilization in the Personal Data Trust Bank, the treatment policy of credit scores will be shown below.

Thus, if the Personal Data Trust Bank also has the function of calculating credit scores, there might be the risks that credit scores have, but that will make the marketing/visualization of data transaction smooth and there will be potential to secure the credibility of the market transaction.

Personal Data Trust Bank business and credit scores can be provided individually. However, by utilizing both of them together, preparing credit scores based on the personal information treated at the Personal Data Trust Bank, etc. and entrusting the Personal Data Trust Bank to manage the task will be possible.

For example, while individuals can acquire credit scores in the Personal Data Trust Bank by utilizing their personal information, enterprises can do marketing, etc. based on the credit scores and scored personal information. On the other hand, individuals will have more services available, such as renting money at low rates by raising their credit scores, and

---

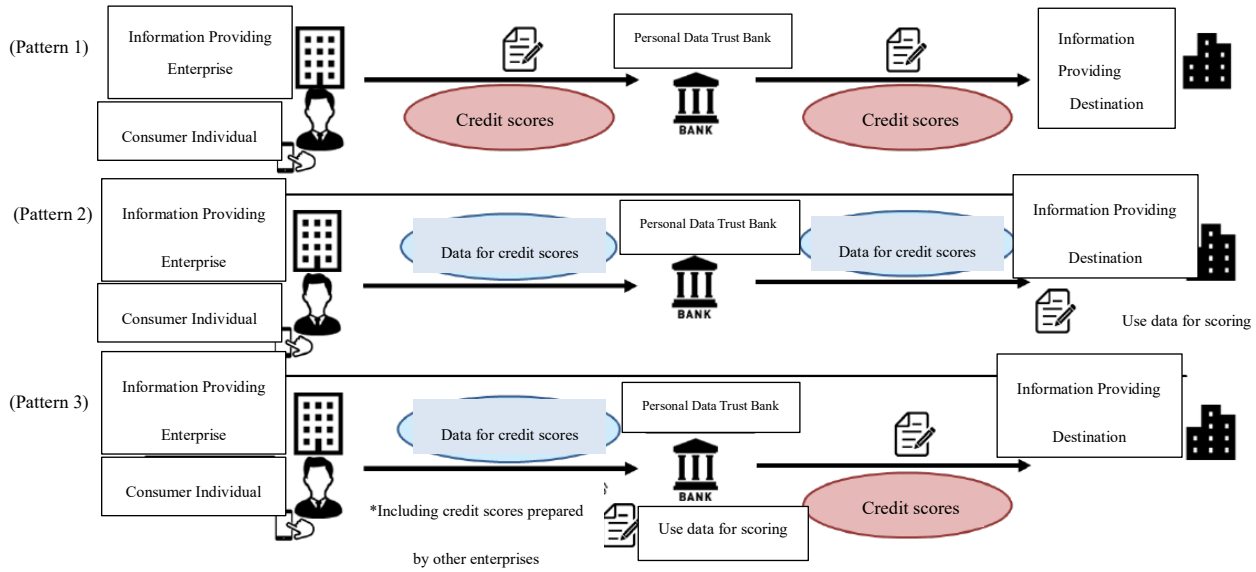
32 NTT Data Institute of Management Consulting, Inc. ""The survey of general consumers' awareness on the use of Personal Data Trust Bank"－Check the traceability of personal data / Interest in safety and the issues on the current consent process of providing personal data" (https://www.nttdata-strategy.com/newsrelease/200407.html)。

33 MIC "The outline of the review meeting on the ideal authorization scheme of Information Trust functions" (https://www.soumu.go.jp/main_content/000648745.pdf)

will enhance their life style. For enterprises, this might contribute to the expansion of good/sound customers. In addition, due to the merits of credit scores, individuals might actively provide their data for enterprises and the data enterprises can utilize will increase, which will make their marketing and product development possible, then enterprises might be able to acquire royal customers. Furthermore, it is expected that the sensitivity of providing individual data and the IT literacy will be enhanced.

On the other hand, regarding the operation, building the system where personal data is not used by unknown enterprises without the consent of individuals and the system that prevents enterprises from screening job applicants by using credit scores is required.

Table 45: Patterns in case Personal Data Trust Bank handles "credit scores"



Source: MIC "The outline of the Review Meeting for the Ideal Certification Schemes Concerning Functions of Information Trust"
(https://www.soumu.go.jp/main_content/000648745.pdf)

④ Finding "hidden needs" in data transaction

The data transaction has been already proceeded on a bilateral basis, and the initiatives mentioned in ② and ③ have been under way. However, the data transaction is not fully active, which was explained in (1) ② (Data transaction with other enterprises is inactive).

(1) As mentioned in ⑤ (Individuals' concerns about their data being collected and used) and ⑦ (Individuals are expecting enterprises to ensure the purpose of using data and data security), concerns about individuals' data being collected and used and the needs for enterprises ensuring the purpose of using data and data security are main factors for inactive data transaction. Resolving these problems can remove the factors for impeding data transaction. However, it is considered that just doing that can't lead to the revitalization of data transaction.

It is important for enterprises that do data transaction to find out what the enterprises that utilize data and individuals who receive specific services based on the utilization potentially want about the activities and services through data transaction ("hidden needs"). Also, they need to clarify where the merits of utilizing data lie; enterprises that utilize data or individuals who use the service.

4. Future image (the image of "Inclusive AI Economy Society")

Here, the image of "Inclusive AI Economy Society" that can be seen after the issues shown in Chapter 2 and 3 are resolved will be viewed.

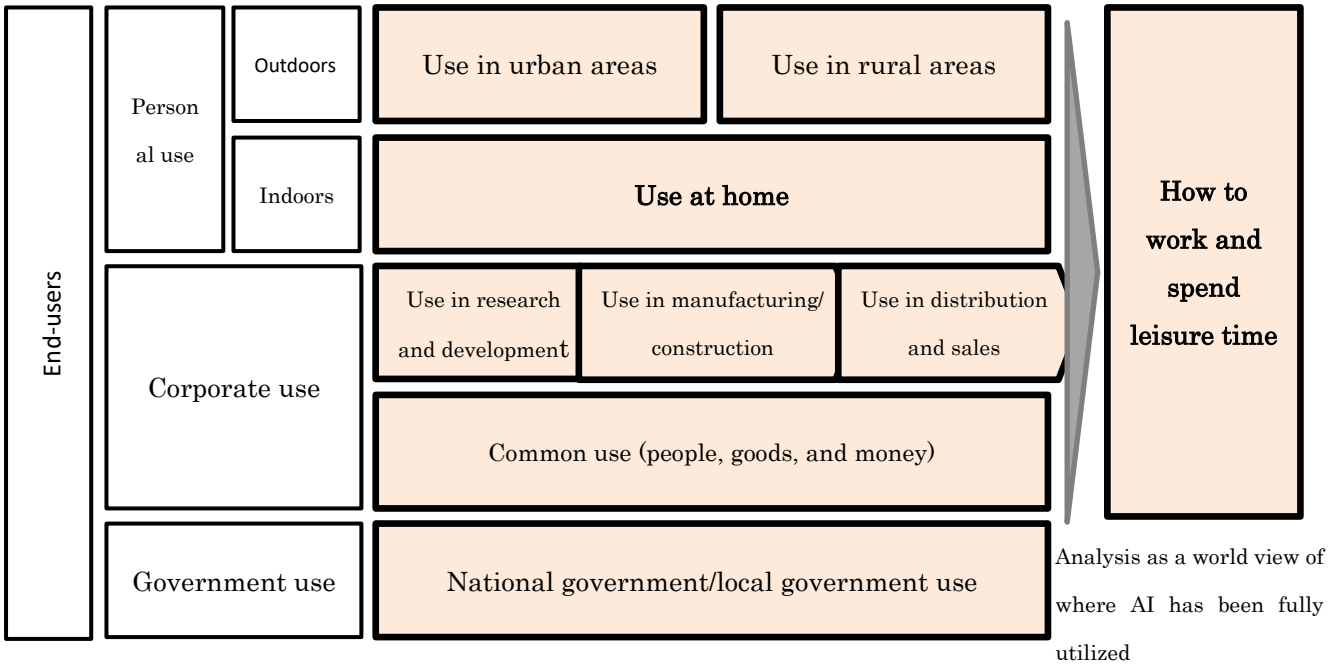
In "Inclusive AI Economy Society", the inclusive operation of AI is required. That is, it is desirable that all entities including individuals, small and medium-sized enterprises and local enterprises are actively involved in social/economic activities based on their will and decision, utilizing AI and data in an ingenious way, and contribute to the improvement of productivity, etc., acquire the distribution, fulfillment, expansion of leisure time, etc. and share the richness in the entire society rather than AI and data are utilized only by some advanced enterprises and they are used for discrimination or screening, and the richness brought about by the utilization of data is exclusively enjoyed by the enterprises.

Regarding this perspective, out of the analysis on prospect of ecosystems formed in accordance with the advancement of AI network the Conference Toward AI Network Society has considered, the part of AI utilization scenes will be referred.

■ AI utilization scenes

In the analysis on the perspective of utilization of AI, firstly the AI utilization scenes in individuals, enterprises, and governments were classified into 9 scenes (Use in urban areas, Use in rural areas, Use at home, Use in research and development, Use in manufacturing/construction, Use in distribution and sales, Common use (people, goods, and money), National government/local government use, and How to work and spend leisure time), then we picked out the 4 major AI utilization scenes in the classification as shown below and comprehensively examined the utilization of AI.

Table 46: Classification of AI utilization scenes



Source: "Sorting out of fields of AI utilization scenes from the perspective of end users"

Also, the realization timing of the AI utilization is divided into 2 stages of "already practically applied and to be

realized in the near future (until around 2025)" and "to be realized in the medium-term (until around 2035)"[34]. Then, the AI utilization scenes regarding these utilization scenes in use scenes were viewed. The details of the perspective of AI utilization scenes are described in Attachment 3.

For example, as shown in the example of AI utilization scenes: individuals (urban areas), by utilizing AI in sightseeing/traveling, tickets will be automatically arranged as soon as an optimum travel is planned, and people can comfortably enjoy sightseeing through face authentication and translation without worrying about baggage and languages. More specifically, the following practical application use cases are expected to be realized.

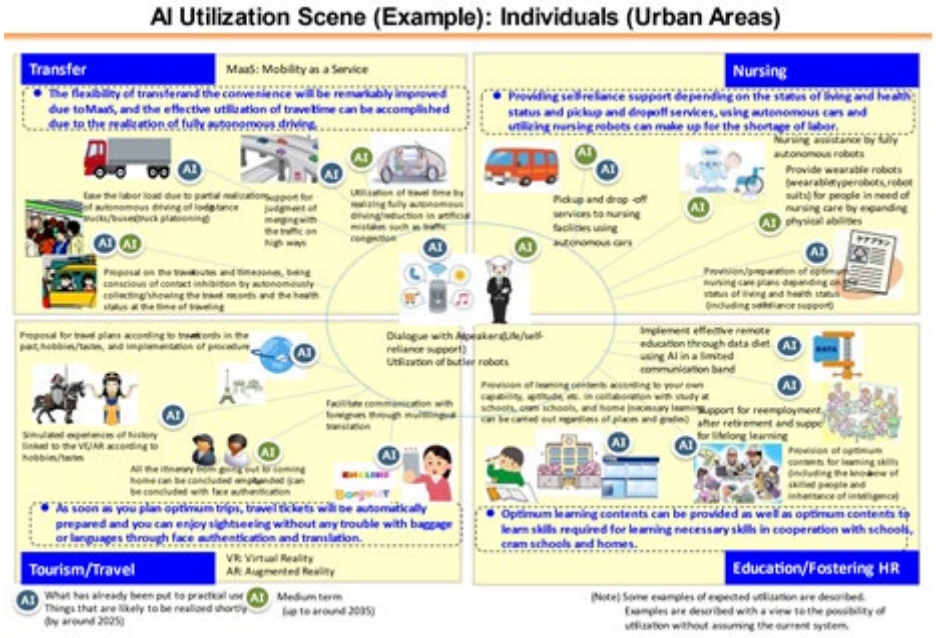
- Proposals of perfect travel plans in accordance with the past travel history, hobbies/tastes, budgets, etc. and implementation of the procedure
- Smoothing of communication with foreigners, using multi-lingual translation
- Simulation experiences of the history connected with VR/AR according to hobbies/tastes
- All the itinerary from going out to going home will be concluded with empty-handed (all concluded with face authentication, etc.)

In addition, thanks to the utilization of AI in education/fostering human resources, providing optimum learning contents through collaborating with schools, cram schools and homes and providing optimum contents for acquiring necessary skills will be possible. More specifically, the realization of practical application use cases can be expected.

- Providing optimum learning contents depending on the capability and aptitude, collaborating with learning at schools, cram schools and homes (regardless of the place and grade, necessary learning will be provided)
- Providing optimum contents for acquiring skills (including passing down the know-how and expertise of skilled persons)
- Support for reemployment after retirement, support for lifelong learning
- Effective remote education will be implemented through the AI data diet while communication zones are limited

[34] The realization timing of AI utilization scenes is supposed to be the timing when the service is placed on the market without considering the service penetration rate and functional fulfillment.

Table 47: AI Utilization Scene (Example): Individuals (Urban Areas)



Source: "Sorting out of fields of AI utilization scenes from the perspective of end users"

Also, as shown in the example of AI Utilization Scenes: Enterprises (Distribution/Sales), by utilizing AI, unattended warehouses and crewless transportation will be realized for enterprises in logistics. Management of entire distribution will be implemented by AI, so delivery services that are not possible manually will be realized. More specifically, the realization of the following practical application use cases can be expected.

- Unattended warehouses by utilizing AI robots and AGV[35], etc.
- Crewless transportation due to the realization of fully autonomous driving and delivery services that are not possible manually will be realized.
- Management of the entire distribution by AI (demand prediction, inventory adjustment, optimization of delivery roots, arrangement of transportation resources, etc.)

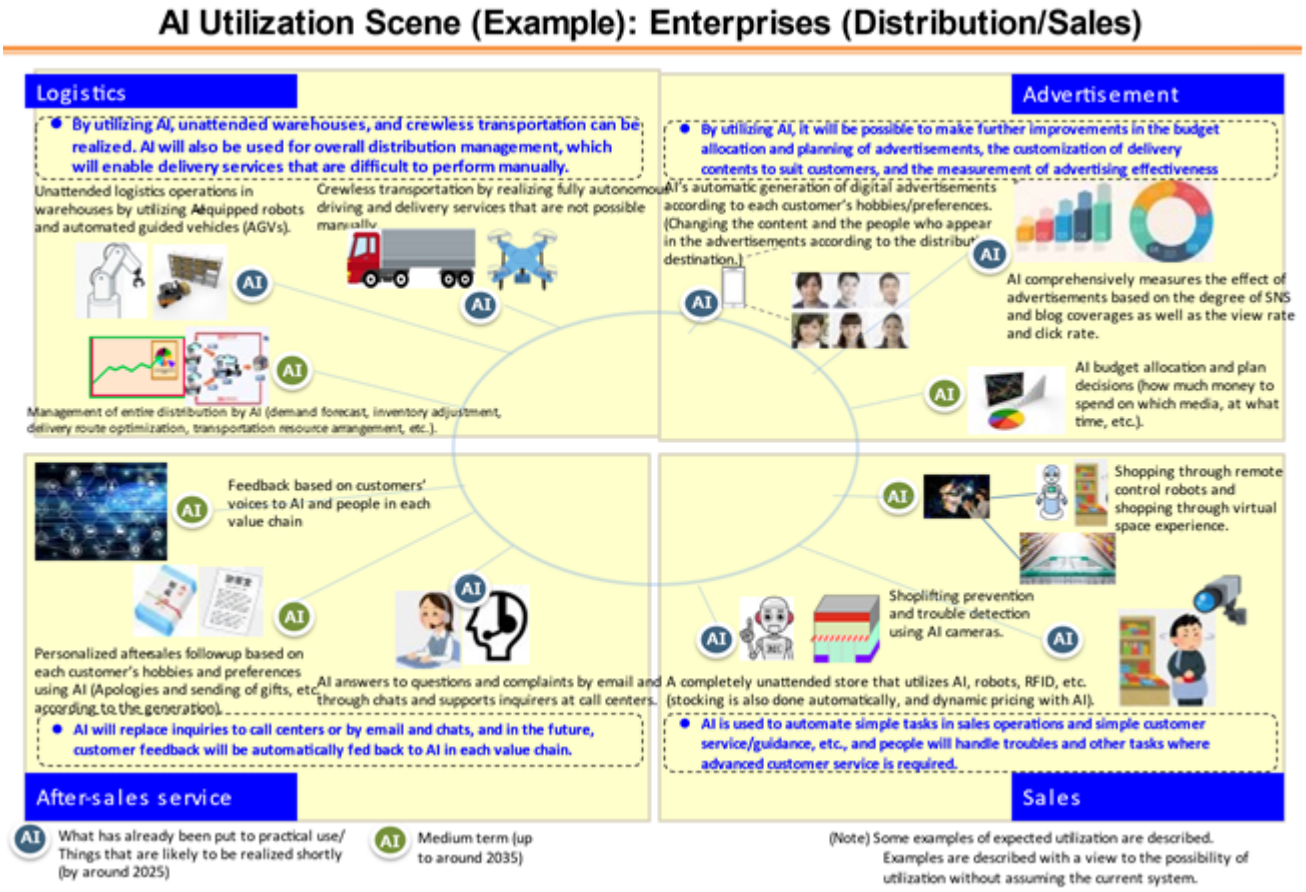
Furthermore, by utilizing AI, advertisement enterprises will be able to allocate advertisement budget, draw up plans, customize delivery contents that fit customers, and enhance the estimation of advertisement effects. More specifically, the realization of the following practical application use cases can be expected.

- Automatic production of digital advertisement suitable for the hobbies/tastes of customers (change the contents and characters, etc. depending on the delivery destination)
- Estimate/present the engagement (effects) of advertisements that integrated adoption rates on SNSs and blogs in addition to the viewing rates/click rates
- Allocation of advertisement budget and drawing up plans (how much money should be used for which

[35] AGV stands for Automated Guided Vehicle which means unmanned transport vehicle. Automated Guided Vehicles are driverless, mobile vehicles programmed to transport materials other than humans in certain areas. The vehicles cannot be used on the roads stipulated in the Road Transportation Act.
("JISD6801　Terms on Automated Guided Vehicle System"
(http://www.jiva.or.jp/pdf/Kind%20of%20AGVS.pdf))

advertisement media, when the advertisement will be done, etc.)

Table 48: AI Utilization Scenes (Example): Enterprises (Distribution/Sales)



Source: "Sorting out of fields of AI utilization scenes from the perspective of end users"

Also, as shown in the example of AI utilization scenes: Administration (National Government/Local Government), in the national government's policy making, by AI's analyzing various information and the opinions of stakeholders, and making policies, streamlining and total optimization will be done and agile type administration will be realized. More specifically, the realization of the following practical application use cases can be expected.

- Analyze policy trends of many foreign countries and make policies in an optimized form in Japan
- Do macro/micro statistics and simulate the effects of financial policies/monetary policies, using real time data, and make optimized policies
- Analyze the regulations and administrative documents related to the bills to be formulated, and analyze the consistency of the bills to be formulated and related regulations, etc. then formulate the drafts of regulations
- After making and executing policies, take the opinions of stakeholders involved in policies and optimize the adjustment instantly

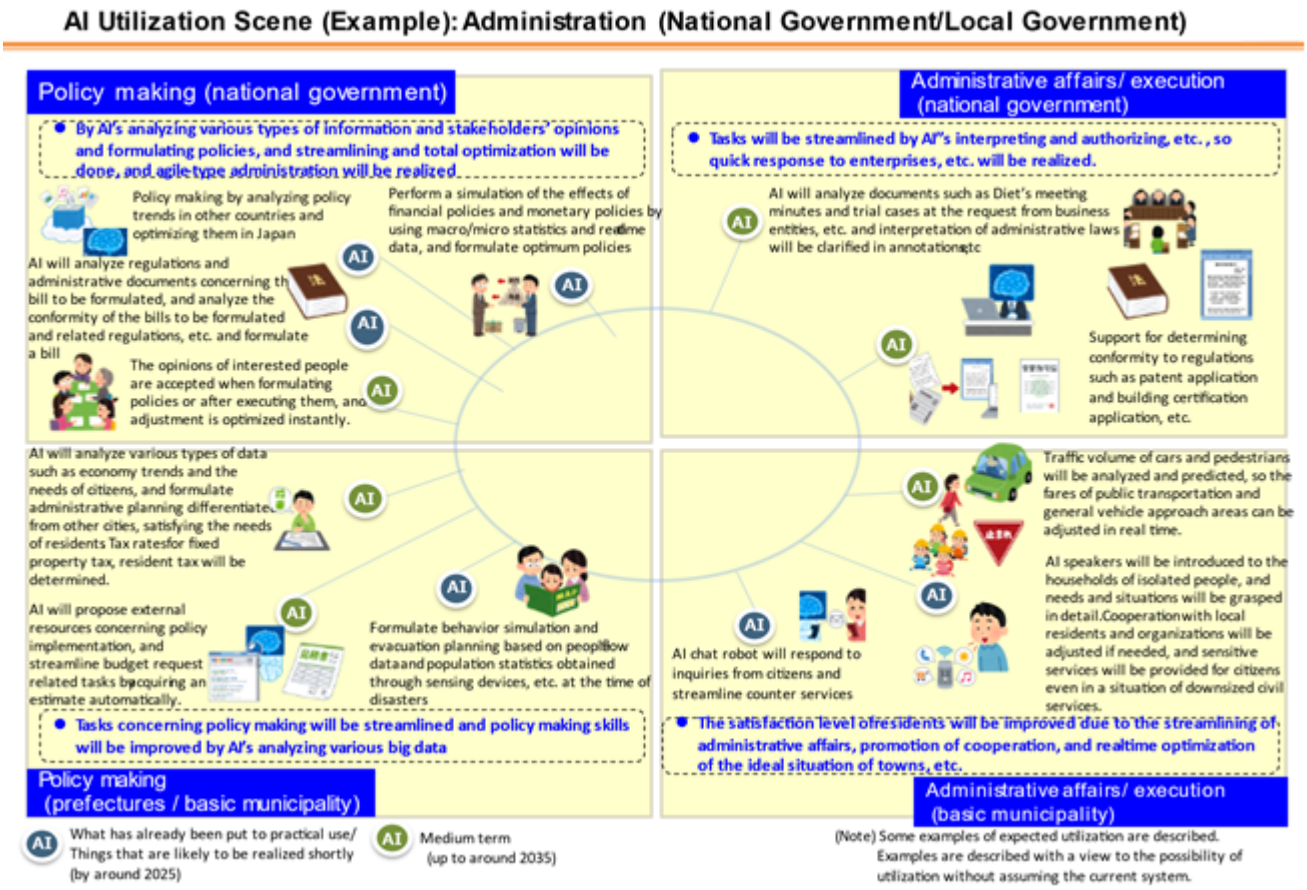
In addition, in the administrative affairs/execution in basic municipality, by AI's streamlining administrative affairs, promoting collaboration, optimizing ideal ways to run the administration in towns in real time, the satisfaction level of residents will increase. More specifically, the realization of the following practical application use cases can be expected.

- AI Chatbot will respond to inquiries from citizens and streamlining counter services
- Analyze and predict traffic volume of vehicles and volume of pedestrians, and adjust the fares of public

transportation and approach zones of private passenger vehicles in real time

- AI speakers, etc. will be introduced to the households of isolated people, and their needs and situations will be grasped in detail. Cooperation with local residents and organizations will be adjusted if needed, and sensitive services will be provided even in a situation of downsized civil services.

Table 49: AI utilization scenes (Example): Administration (National Government/Local Government)



Source: "Sorting out of fields of AI utilization scenes from the perspective of end users"

In addition, if AI is fully utilized in these AI utilization scenes, it is expected that the workstyle/how to spend leisure will change. For example, regarding workstyles[36], working beyond spaces will be realized due to the spread of remote working, supply and demand matching will be possible beyond the borders of companies and national borders, participating in society, providing values and experiencing that have been impossible due to temporal, spatial, regional and physical restriction become possible because of the expansion of human capabilities thanks to AI and robots, etc. As a result, it is

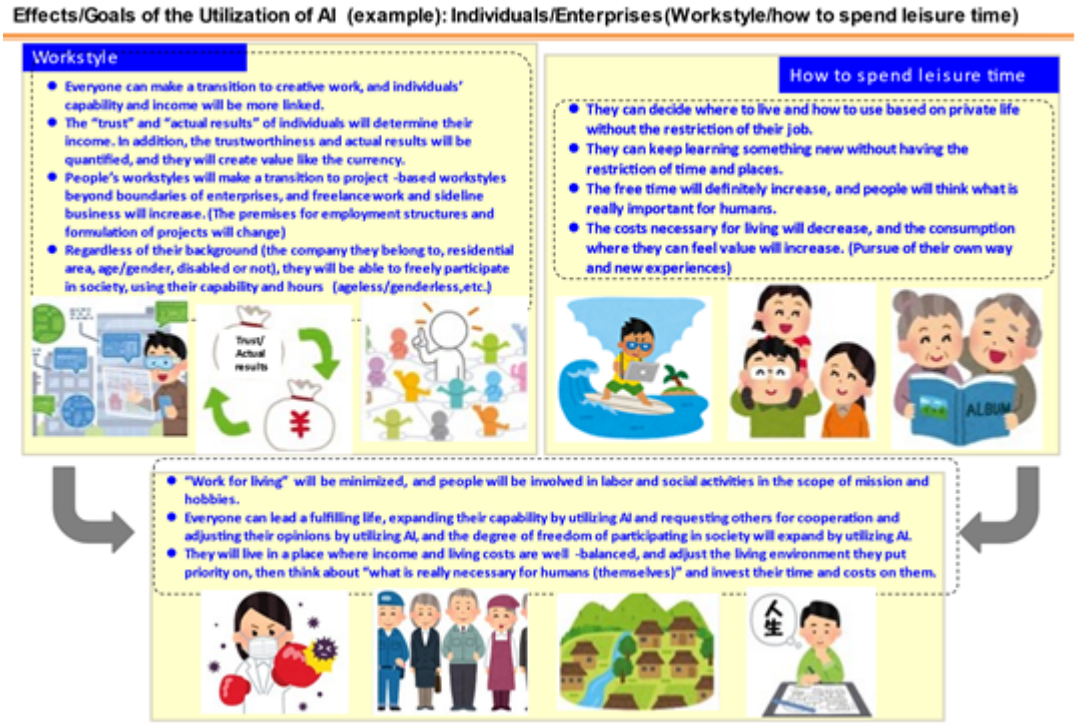
[36] The change in workstyles mentioned here is the assumed one in a society where AI is fully utilized in mid- to long-term future. There is a possibility that the COVID-19 pandemic that is ongoing in Japan and the world might accelerate this change. On May 4, 2020, in a government panel on the new coronavirus (chair: Takaji Wakita, head of the National Institute of Infectious Diseases), practical examples of "the new life style" to prevent the spread of COVID-19 were shown, and regarding workstyles, teleworking, continuing staggered commuting hours and online business meeting and online exchange of business cards were introduced as specific examples. Also, on May 14, 2020, "Guidelines for Preventing the Spread of Novel Coronavirus Disease (COVID-19) in Offices" were posted by Japan Business Federation (https://www.keidanren.or.jp/policy/2020/040.html), and as specific measures to be addressed, through a variety of forms of work programs such as remote work (work from home or from satellite offices), staggered start and finish times, rotation work (dividing workdays and working hours into multiple shifts), flexible working hour system, and four-day work weeks, to reduce the frequency of commutes and relieve congestion on public transport are proposed.

considered that “labor for living” will be minimum, and the participation in society and degree of freedom in living will expand due to the expansion of capabilities and requests for cooperation or reconciling the differences in views thanks to AI, and people will work and participate in society with a sense of mission or within the scope of hobbies. On the other hand, regarding how to spend leisure time, some companies have already adopted the four-day week system, and it is considered that human work will be streamlined by AI and our free time (leisure time) will increase in the future. Therefore, how to spend the increased leisure time will be enriched by AI, and opportunities to restudy will be provided. In addition, it is also considered that the increased time will be used for creating art works and people will pursue more creative value. As a result, it is considered that people will live in a place where income and living costs are well balanced, adjust their living environment, and spend time and costs, considering “what is really important to humans (themselves)”. On the other hand, the possibility of the expansion of the gap between workers due to the technical capabilities such as the use of AI should be noted in this social change. So far, in developed nations such as Japan and the US, the number of “medium/routine task” workers has been decreasing, and the numbers of highly skilled workers and lowly skilled workers have been increasing. There is a possibility that this trend will be spurred on due to the thorough utilization of AI and the gap in wages will be bigger (Table 52).

Table 50: AI Utilization Scenes (Example): Individuals/Enterprises (Workstyle/How to spend leisure time)

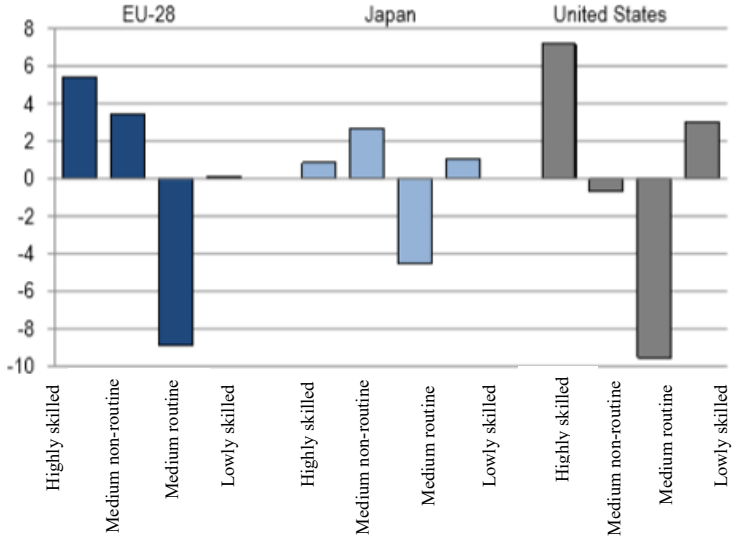
Source: "Sorting out of fields of AI utilization scenes from the "end user perspective"

Table 51: Effects/Goals of the Utilization of AI (example): Individuals/Enterprises (Workstyle/how to spend leisure time)



Source: "Sorting out of fields of AI utilization scenes from the "end user perspective"

Table 52: Percentage-point change in employment shares by occupation category (2002 - 2014)



Source: OECD "Automation and Independent Work in a Digital Economy" POLICY BRIEF ON THE FUTURE OF WORK (2016).

5. Recommendation

5.1 The Ideal Policy toward the Promotion of the Utilization of AI/Data

(1) Ideal directivity

To aim for a society where various entities are actively engaged in social/economic activities that utilize AI and data in an ingenious way based on their will or decision, the situation where Japan is positioned is described as follows. ①The global economy has changed from a finance-led economy to the one led by digital companies that utilize ICT and data, ②In the utilization of data in enterprises, there are disparities in sizes between large enterprises and small and medium-sized enterprises and disparities between enterprises in urban areas and those in rural areas, ③There are many Japanese enterprises that realize their data-related initiatives including collecting and analyzing data are lagging behind, ④Many Japanese enterprises feel the utilization of data in the office was effective, ⑤While consumers are careful about providing their personal information and behavior history and have concerns about their data being collected and used, they don't have much resistance about their data on health/medical scenes being utilized compared to those on disaster/crime prevention, transportation and education related scenes.

Based on the above, to find the ideal way to utilize data required for the implementation of AI in society, the issues to be addressed in enterprises and administration were confirmed. The issues to be addressed in enterprises are ① Positioning data and AI in management strategies and business architecture, ② Building the system to collect physical data for the utilization of data, organizational system for the utilization of data and organizational data infrastructure, ③ Relieving or reducing consumers' concerns about their data being used as well as responding to the current regulations of data for customers. The issues to be addressed in administration are ④ Accelerating initiatives of the digital government in administration, ⑤ Issues administration is facing through the outbreak of COVID-19.

The ideal directivity of AI/data will be explained in the following.

The COVID-19 brought about the economic shock that is beyond Lehman's collapse to the world when monetary easing policies were proceeding in the world. Various restrictions brought about by the COVID-19 enabled enterprises to realize the importance of communication networks and ICT tools in continuing their business, and what's more, there is a possibility that it will greatly change individuals' work styles and life styles, and the ideal way for enterprises to exist as an organization, which will create a "new normal" way to survive. For example, it is considered that many social activities will be done in virtual spaces, and the necessities of "opportunities" to create real spaces will relatively decrease by constantly continuing the utilization of teleworking and VR/AR.

There is a high possibility that the transition to the economy society based on the utilization of AI will be accelerated in the "post-coronavirus" society. WHO has built the platform to commonly utilize data and intellectual properties (COVID-19 Technology Access Pool)[37]. Voluntary sharing of COVID-19 related intellectual properties is considered to be desirable in WIPO[38]. Also, in the government of Japan, "Emergency Economic Measures to Cope with COVID-19" disclosed the initiatives of remote working promoted as the "acceleration of digital transformation due to remoteization,

[37] WHO "COVID-19 technology access pool" (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov/covid-19-technology-access-pool)

[38] WIPO "Some Considerations on Intellectual Property, Innovation, Access and COVID-19" (https://www.wipo.int/about-wipo/en/dgo/news/2020/news_0025.html)

etc." due to the spread of COVID-19 and the increased needs for remote services that utilize ICT such as remote education that can handle the emergency situation like this. Stating "we have to turn this crisis to a chance and advance digital new deal policies under the concept of wise spending and immediately accelerate the social change in order to accelerate the realization of Society 5.0"[39], the government has enforced the initiatives[40]. Furthermore, in the spread of COVID-19, the government has promoted initiatives to exert social intelligence[41], the capacity to support people, take care of people, and provide emotional support or promote people's cooperation and reconcile differences in views, using AI that is considered to lack social intelligence. Initiatives by Civic Tech can be introduced as an example. Civic Tech is a grass-roots movement in which citizens resolve social issues, etc. with other stakeholders such as administration by utilizing technology. As a specific example of Civic Tech, Taiwan's initiatives in the spread of COVID-19 can be cited. This is an example of functioning of digital technology, activities of activists and citizens, and it is attracting people's attention as a model of economy society based on the utilization of AI[42]. This kind of activities are also done in Japan, and organizations such as Code for Japan and Civic Tech Japan were established[43]. These organizations have connected activists, citizens and administration by using digital technology to tackle the spread of COVID-19, and stakeholders have voluntarily cooperated with each other and contributed to preventing the spread of COVID-19, through providing API and information disclosure utilizing API. The "social intelligence" humans should exert in the society where various entities are actively engaged in social/economic activities that utilize AI and data in an ingenious way based on their will and decision has been already required in the coronavirus crisis.

On the other hand, the struggle with the lingering COVID-19 pandemic in the world has revealed the fragility of modern economy society. For example, in the society that pursues the productivity and efficiency, the redundancy as a response to emergency was removed from the society and economy. This can be applied to the world of ICT. The utilization of ICT at home increased due to home teleworking, which caused the excessive burden of networks and servers and excessive processing capacity, and triggered the deterioration of the quality of services. It is obvious that there is a limit in responses of individual enterprises in this unprecedented crisis, and that can be true to the whole economic sectors as well as ICT, and the public roles the national government should play became clear.

To promote the utilization of AI/data, the digital transformation of the entire stakeholders in Japan will be necessary, and the changes in behaviors responding to new situation in the future will be also necessary. Also, it is necessary to fully realize that these changes are actually occurring and the utilization of ICT including AI can make these changes possible.

The ideal directivity that enterprises, individuals, and government should pursue will be explained in the following.

Specific initiatives toward the establishment of social economy based on AI has already taken in enterprises in the

---

[39] The 6th Council on Economic and Financial Policy in 2020 "Document 1, Emergency Economic Measures to Cope with COVID-19 (Cabinet Decision on April 20, 2020)」
(https://www5.cao.go.jp/keizai-shimon/kaigi/minutes/2020/0427/shiryo_01.pdf)

[40] This was featured in the joint session of the 77th IT Strategic Headquarters and the 8th Strategic Conference for the Advancement of Public and Private Sector Data Utilization. "Initiatives to fully mobilize IT and data are necessary" as a medium-to-long term-initiative, "Changes in a social structure through digital transformation". In addition, "after the COVID-19 pandemic subsides, we will turn this pinch to a chance and strongly promote digital transformation that takes digitalization as the driving force of a social reform in order to reboot the economy of Japan". (https://www.kantei.go.jp/jp/singi/it2/dai77/siryou1.pdf)

[41] Richard Baldwin, The Globotics Upheaval: Globalisation, Robotics and the Future of Work, Hachette UK, 2019.

[42] Jaron Lanier and E. Glen Weyl, "How Civic Technology Can Help Stop a Pandemic : Taiwan's Initial Success Is a Model for the Rest of the World", *Foreign Affairs*, Volume 99, Number 3, 2020.
(https://www.foreignaffairs.com/articles/asia/2020-03-20/how-civic-technology-can-help-stop-pandemic)

[43] Civic Tech Japan was established in May 2019 (https://www.civictech.jp/).
Code for Japan was established in 2013 (https://www.code4japan.org/).

world. To promote the utilization of AI/data, boosting SMEs and enterprises in rural areas that haven't started the initiatives yet to launch them is necessary.

For example, regarding digital transformation (DX), the promotion movement of boosting initiatives of enterprises by the government can be found[44], but it is important to create the system whereSMEs and enterprises in rural areas can take initiatives rather than to promote initiatives by some advanced enterprises or large enterprises. Regarding this, Digital Transformation business companies established by Fujitsu, Toshiba, and Sumitomo Corporation (Ridgelinez[45], Toshiba Digital & Consulting Corporation[46], Insight Edge[47]) and Digital Transformation promotion business of ITOCHU Corporation (business partnership with AKQA Inc. in the US)[48] are attracting attentions as advanced initiatives to promote the spread of not only AI but also the entire ICT to enterprises. It is important to realize the merits of the utilization of data to appeal for small and medium-sized enterprises that haven't started the utilization of AI/data and the basic of the utilization of data through these initiatives in the public and private sectors, and encourage SMEs to utilize data as the fourth investment factor ranking next to humans, things and money.

If individuals deepen understanding new technologies such as AI, they will recognize AI as "something to utilize" as well as computers and smartphones, not as "something brought about", and the tendency to utilize AI usefully will be promoted in the economy society. Dissemination and public relation are required for enterprises and administration to deepen the understanding. In addition, to share the image of "Inclusive AI Economy Society" in society, building the system to disseminate the ideal new AI economy society and the attitude to embody the needs of the dissemination are required for enterprises and administration.

What's more, it is necessary for administration to work on further organizing systems where people can utilize data in a way trusts of organizations and individuals are secured. Also, it is important to further accelerate initiatives of the digital government that can increase the opportunities for private sectors to use public data.

(2) Specific measures

Based on (1), specific measures to be placed as particularly important toward the promotion of the utilization of AI/data will be explained in the following.

① Support for the promotion of the utilization of AI/data in SMEs and enterprises in rural areas by the public and private sectors

For SMEs and enterprises in rural areas, the lack of management know-how necessary for fully exerting the effects of the utilization of AI/data such as positioning the utilization of data in management strategies and business architecture,

---

44 In "(4) the acceleration of Data Transformation / Promotion of the utilization of AI/data, etc." of Intellectual Property Strategy Headquarters "Promotion Program for Intellectual Property 2020－Intellectual Property Strategy toward the "New Normal" after the COVID-19" (May 27, 2020), for example, "Big change caused by Data Transformation" and the necessity of boosting by the government in "COVID-19 and Data Transformation" were pointed out, and sharing of "analysis of cases of DX in enterprise", the necessity of "preparation of rules for promotion of the utilization of real data", pointing out "various cross-cutting issues on AI/data", the necessity of "cooperation between fields", presentation of "issues in each field", "fostering human resources responding to the age of AI" or "the ideal patent system suitable for the age of AI/IoT technology" were considered.

45 Ridgelinez Limited's website(https://www.ridgelinez.com/)

46 Toshiba Digital & Consulting Corporation's website (https://toshiba-dx.com/toshiba-dx/)

47 Insight Edge, Inc.'s website (https://insightedge.jp/)

48 Through the business partnership with AKQA Inc. they review the issues of customer companies in Japan from the perspective of customers, utilizing ITOCHU Corporation's network in Japan and the world and AKQA's consulting ability in the promotion of Digital Transformation which is the strength of AKQA Inc.
(ITOCHU Corporation's website (https://www.itochu.co.jp/ja/news/press/2020/200305.html))

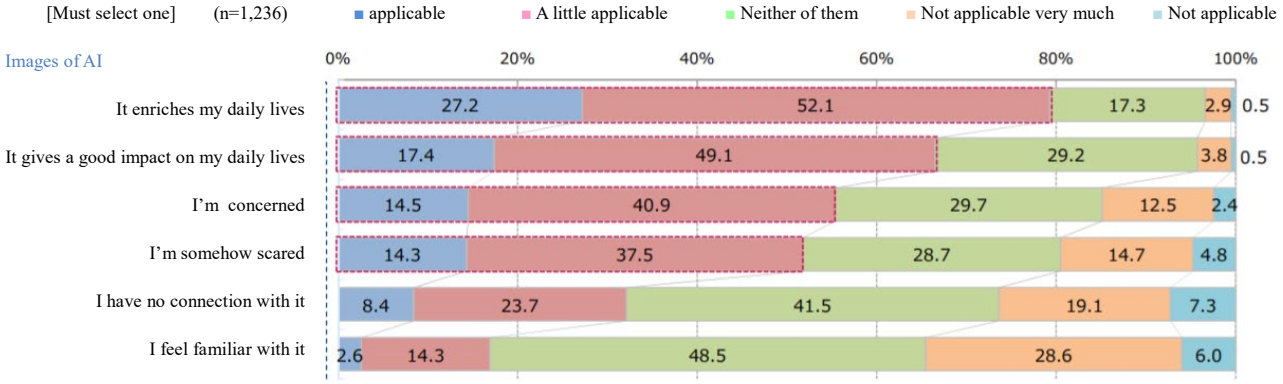
preparing the system of collecting physical data, organizational system of the utilization of data and building of data infrastructure in organizations, responding to the current regulations for data of customers, etc. and investment capital and human resources to realize the management know-how is considered to be an impediment. Therefore, it is required to clarify the issues to be addressed and prepare comprehensive consulting to realize the said management know-how by recommending necessary initiatives beyond the economic support concerning the introduction of AI, etc. Also, to penetrate the initiative, fostering human resources who can undertake the initiative is necessary.

② Enlightenment activities about understanding new technologies such as AI by public and private sectors

New technologies such as AI are just tools as well as the existing ICT devices/services and means to achieve goals. It is important to understand that using them is connected with easily achieving goals, and the system of the technologies and what can be done and can't be done with the technologies. At this point, there are quite a few people who have a negative image about AI. In the questionnaire survey conducted by Consumer Affairs Agency, more than half of the people have a negative image about AI such as "I'm concerned" or "I'm somehow scared" (Table 53). These concerns about AI can be a factor of impeding utilizing AI. Therefore, it is necessary to take initiatives to reduce the negative image about AI, and both public and private sectors need to work together for enlightenment activities to continuously encourage people to understand new technologies such as AI rightly and use AI well.

Table 53: Images of AI

Q1. Please choose an applicable one about the image of AI (Artificial Intelligence).



Source: Consumer Affairs Agency "The 1st Consumer Awareness Survey" (1,226 persons responded)[49]

③ Initiatives toward sharing the image of "Inclusive AI Economy Society"

To realize the "Inclusive AI Economy Society", it is desirable that individuals disseminate the ideal new economy society through their dissemination tools such as SNS, etc., and they are connected with making policies through big data analysis.

[49] On February 12 - February 18, this web questionnaire survey about general consumers image of AI and the status of use AI services in daily life was conducted, targeting males and females from 10s to 60s (1,226 persons) who live in Japan.
(https://www.caa.go.jp/policies/policy/consumer_policy/meeting_materials/assets/consumer_policy_cms101_20316_03.pdf)

④ Review of the system that promotes the utilization of AI/data

To promote the utilization of AI/data, it is important that systems are well-organized so that people can utilize data in a way trusts of organizations and individuals are secured. In addition, it is also important to prepare a system that is beyond the existing vertical administrative division, the nation and local districts, and public and private sectors in order to increase the opportunities to use public data in private sectors, so it is important to further accelerate initiatives of the digital government.

⑤ Promotion of the utilization of AI/data, keeping in mind the "post-COVID-19" society

Through the COVID-19 crisis, it became clear that how to settle the contradiction between the efficiency to realize inexpensive services and the redundancy to respond to this kind of pandemic is important. It is required to promote the measures to utilize AI/data, considering the balance between the efficiency and redundancy of communication networks including teleworking and the entire ICT tools.

5.2 The ideal data economy policy required for promoting the data-driven economy in the AI era

(1) Ideal directivity

It is considered that the data-driven economy will be more developed in the future. On the other hand, it can't be said that the market mechanism concerning data transaction is fully functioning now. Therefore, it is desirable that the current imperfect data transaction will be supplemented and the resource allocation based on the market mechanism will be achieved as soon as possible as an ideal form for the future.

As the mechanism of creating added value on data, the processes of producing, collection, accumulation, processing, transaction, and utilization are complicatedly interacted. And the stakeholders in the mechanism are roughly divided into 4 entities; the entity who produces resource data (net service users), the entity who analyzes data (data scientists), the entity who owns data (enterprises/platforms, etc.), data transaction mediators (Personal Data Trust Bank, etc.). It is desirable that we consider the reasonable grounds for paying remuneration to each stakeholder (visualization of digital activities) and make policies to promote the development of data economy. By visualizing digital activities by stakeholder like this, it will be possible to grasp the following;

- Reasons for paying/not paying remuneration
- How to pay remuneration (process)
- The standard of remuneration

and make policies where the market mechanism functions.

Table 54: Concept of paying remuneration by stakeholder

	Resource data producers (net service users)	Data analysts (data scientists)	Data owners (enterprises/ platforms, etc.)	Data transaction mediators (processing, rating) (Personal Data Trust Bank, etc.)
Reasons for paying/not paying remuneration	Approval of the contribution by resource data producers	Contribution to analysis	Revenue from data	Commission fee from data transaction

Payment methods of remuneration (process)	Free services, information provision, points, cash, etc.	Evaluation for analysis results	Business revenue	Business revenue
Standard of remuneration	Intention of data owners/data transaction mediators	Labor market	Advertising market, etc.	Cross trading

Source: Prepared from a published document

Also, it is important to foster individuals' awareness of managing data on their own initiative as well as to promote individuals complying rules on personal information from the perspective of the revitalization of data transaction. However, it should be noted that various costs will be incurred in connection with complying with the rules.

In addition, to execute appropriate data economy policies, we have to consider the economic characteristics of data. As the economic characteristics of data, almost zero marginal costs, profits brought about from data accumulation, and the quality (contents) of information as well as nonrivalness, externality and partial exclusivity are given as examples.

(2) Specific measures

Specific measures to extract economic benefits utilizing the market mechanism as much as possible based on (1) will be explained in the following.

① Building data infrastructure

To contribute to the economic development of ICT capital centering on AI, building data infrastructure to process data at high speed is positioned as an important infrastructure. Not only speeding up of access lines including 5G/Beyond 5G/FTTH but also securing redundancy and speeding up of CPU and GPU, etc. are required from the perspective of information processing, and taking the balance between distributed processing and centralized processing should be noted.

② Realization of data cooperation

It is considered that the versatility of utilizing data will be expanded by promoting data cooperation.

In the "Digital Society Construction TF" published by the government, "by taking inventory of initiatives of data cooperation promoted by the public and private sectors in various fields such as "agriculture, disaster prevention/disaster mitigation, infrastructure, logistics, and autonomous driving, we will consider data formats, functions, and communication of rules concerning the treatment of data required for collaboration that makes the collaboration beyond fields"[50]. It is important that these initiatives will be further promoted.

③ Promotion of corporate data management by a third party/initiatives of Information Bank

In the transaction of data owned by enterprises, the relationship in transaction and the relationship between stakeholders are complicated, so it is necessary to clarify where the responsibility lies by case. It is important to clarify where the responsibility lies by case through the standardization of contract formats concerning data transaction by a third party. In addition, supporting untangling the rights of the provision/usage of data and doing the procedures of contracts

50 Council for Integrated Innovation Strategy (The 6th) "About Establishment of Digital Society Construction TF" (https://www.kantei.go.jp/jp/singi/tougou-innovation/dai6/siryo4.pdf)

by a third party is required to promote data transaction in small and medium-sized enterprises where it is hard to do data transaction systematically.

To promote distribution of data on individuals, the role of Personal Data Trust Bank as an intermediary of data transaction will be more important in the future. It is important that the Personal Data Trust Bank secures the trustworthiness of personal information, and it is promoted as a system that gives consideration to individuals who produce resource data, as well as plays an important role in enhancing individuals' awareness to manage their personal information on their own initiative.

④ Review of the legal system to promote the utilization of data

To promote ① - ③, there is a possibility that the existing legal system can be the impediment in the future. In such a case, it is required to constantly review the legal system.

(3) Issues in the estimation method of the value of data

To realize the data transaction market, the actual estimation and analysis of the reality surrounding data are important. The empirical analysis on the value of data in 3.1 is a clue to that. From the way of thinking in the production function, data is added to capital equipment to expand the capacities of workers and positioned as a provider of new capacities. Therefore, it is necessary that data is positioned as a factor of production (stock) and is examined as a clue of theoretical examination. The directivity to sort the issues of this method and further polish the method will be explained in the following.

① Building capital stock based on the economic characteristics of data

As the economic characteristics of data, the changes in value of data with the lapse of time, expansion of value due to the accumulation of data and so on are pointed out. After taking these characteristics into consideration, how to build capital stock in the year for analysis from the status of owing/utilizing data needs to be examined.

② Further analysis of the relationship between data and the value created by data

It is necessary to analyze the relationship between data quantity/quality and the value created by data by industry and production process. In addition, multidimensional analyses using many survey data are required, but to do so, it is necessary to consider building a system to collect information on a large scale and continuously.

③ Social enlightenment toward data economy

To further polish the estimation method of data value, it is necessary to improve the social recognition of the data value and expand initiatives of analysis. To do so, it is important prepare and publish data that can be used for analysis, and it is considered public statistics or reflecting public data on corporate financial statements[51] is one of the ways.

---

[51] For example, in the consideration of government's intellectual property strategy, "Discussion on the ideal way of utilizing data that can contribute to corporate value (from the perspective of creating a P&L statement, enhancing corporate value, creating a market, etc.)" are introduced as issues. Intellectual Property Strategy Headquarters Planning Committee (The 5th) "Document 4 Digital Intellectual Property" Please see (https://www.kantei.go.jp/jp/singi/titeki2/tyousakai/kousou/2020/dai5/gijisidai.html)

6. Addendum

6.1 Questionnaire

In 3.1, the questionnaire survey, “Survey on the Utilization of Data” conducted to estimate the value of data for enterprises is as follows.

Survey on the Utilization of Data

[Notes when responding to the survey]

○About the holding company system

・Regarding the companies that adopted the holding company system, please answer about the status of the entire companies in the group, not on a non-consolidated basis (the status of main business companies in the group, if it’s difficult). Furthermore, regarding listed parent and subsidiary companies, we want subsidiary companies to answer the questionnaire regardless of the answer of parent companies.

○About the timing of the status in the response

・As for the response to this survey, please answer about the status **as of the end of FY2018** if the timing is not clarified.

First, please read the following explanation about the definition of terms in this survey.

・The “data” in this survey means the one currently digitalized and processable by computers, and it is limited to the one that can be utilized for your own company for your product development, expansion of services, and streamlining of business (excluding the one that is provided for others for business or projects).

*For example, map data, meteorological data, and data sets prepared only for providing for others are excluded here.

・The data “analysis” in this survey means the act that extracts information by processing data such as characters, values, images and video. This includes the one in case you use data analysis tools/services provided by external companies, etc., but doesn’t include the one in case of browsing, automatically collecting data by computers, judging and transmitting alert (for example, transmitting alert if the temperature becomes over a certain degree).

・The data “utilization” in this survey includes browsing data, automatically collecting data using a computer, judging and transmitting alert in addition to data analysis.

1. About the utilization of data

Here, we ask you about the status of the utilization of data in your company.

<A question for everyone>

Q1: What kind of data and how processed data do you use in each field? Please choose all the corresponding ones.

[For 1 – 7, please answer in multiple responses, and if there is no corresponding one in 1 - 7, please choose 8 or 9.]

(Regarding data in A -G, please circle all the corresponding ones.)

(If there is no business area, please choose "No area". You don't have to answer in A-G regarding the area.)

(Regarding processing, please answer the following 4 questions.)

Browsing: Browsing data itself

Aggregation: Aggregating by period, and processing of aggregation, etc. by enterprise size

Analysis: Statistical analysis (analysis of correlation, variation, etc.)

AI: Prediction utilizing Artificial Intelligence such as machine learning/deep learning (including not only the cases where AI is utilized to analyze the relationship of things and optimum behaviors but also the cases where AI is utilized for the automation, judging stockout/cheating, and monitoring based on data.)

(Regarding analysis/AI, data analysis tools/services provided by external companies are included.)

	1 Management planning/ Organizational reform	2 Planning of products/services/ development	3 Marketing	4 Production/Manufacturing	5 Logistics/Inventory control	6 Maintenance/Support	7 Others (Basic research, risk management, etc.)	8 We own data, but we haven't utilized it in any field.	9 We don't own any data.
If there is no business area:	No area	No area	No area	No area	No area	No area	No area		
A. Customers' (individuals') basic data (name, address, sex, etc.)	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	8	9
B. Customers' (corporations') basic data (name, address, capital, etc.) *Including municipalities/ organizations, etc.	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	8	9
C. Data of communication with customers, etc. (daily sales reports, purchase history, inquiry history, results of market research, etc.)	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	8	9

D. Action data on the Internet (websites, SNS, mobile app, etc.) (search/access log, word of mouth/posted data, etc.)	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	8	9
E. Data on human activities (biological information, location information, camera image, etc.)	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	8	9
F. Data on the operation of machines, equipment, vehicles, etc. (operational situation, location information, speed, etc.)	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	8	9
G. Data on nature/public (map information, meteorological information, etc.)	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	a. browsing b. aggregation c. analysis d. AI	8	9

<Please answer about the data you chose 1 or more “aggregation” or “analysis” or “AI” in A-G in Q1>

Q2: When you analyze data, how old research data do you usually use?

(Please circle the corresponding one in the data of A – G.)

(If you analyze data for several purposes, please answer the status for the most frequently analyzed purpose.)

	1 Almost only the data of the day	2 Ones until around 1 week ago	3 Ones until around 1 month ago	4 Ones until around 3 month ago	5 Ones until around half a year ago	6 Ones until around one year ago	7 Including ones older than one year ago	8 No idea
A. Customers’ (individuals’) basic data (name, address, sex, etc.)	1	2	3	4	5	6	7	8
B. Customers’ (corporations’) basic data (name, address, capital, etc.) *Including municipalities/ organizations, etc.	1	2	3	4	5	6	7	8
C. Data of communication with customers, etc. (daily sales reports, purchase history, inquiry history, results of market research, etc.)	1	2	3	4	5	6	7	8

D. Action data on the Internet (websites, SNS, mobile app, etc.) (search/access log, word of mouth/posted data, etc.)	1	2	3	4	5	6	7	8
E. Data on human activities (biological information, location information, camera image, etc.)	1	2	3	4	5	6	7	8
F. Data on the operation of machines, equipment, vehicles, etc. (operational situation, location information, speed, etc.)	1	2	3	4	5	6	7	8
G. Data on nature/public (map information, meteorological information, etc.)	1	2	3	4	5	6	7	8

<Please answer about the fields where you chose “aggregation” or “analysis” or “AI” one time or more in the fields of A-F in Q1.>

Q3: How frequently do you do data analysis?

(Please circle a number in each field of A – F.)

	1 Almost every day	2 Once a week	3 Once a month	4 Once in 3 months	5 Once half a year	6 Once a year	7 Once in more than a year
A. Management planning/Organizational reform	1	2	3	4	5	6	7
B. Planning of products/services/development	1	2	3	4	5	6	7
C. Marketing	1	2	3	4	5	6	7
D. Production/Manufacturing	1	2	3	4	5	6	7
E. Logistics/Inventory control	1	2	3	4	5	6	7
F. Maintenance/Support	1	2	3	4	5	6	7

<For those who chose 1. - 7. one or more times in Q1>

Q4: What kind of initiative related to the quality of data does your company take when utilizing data? [multiple responses]

1. Check of data source
2. Check of the accuracy of data (error check, etc.)
3. Acquisition of a privacy mark
4. Acquisition of ISMS (ISO27001)
5. Other [ ]

<A question for everyone>

Q5: When your company does data analysis, what organization (departments/persons in charge) do you use for data analysis (including the use of data analysis tools/services provided by external companies, etc.)? [multiple responses]

If you have a specialized department for data analysis, please answer how many years ago the department was launched. (The specialized department for data analysis is the one that specializes in data analysis in a business unit, excluding the department that specializes in the development of hardware and software related to data analysis for external sales. Please answer about the department that is involved in data analysis.)

1. Persons in charge in a specialized department that handles data analysis --- The specialized department was launched approx. [years ago]
2. Persons in charge in each business unit who are specialized in data analysis
3. Persons in each business unit who are not specialized in data analysis
4. Entrusting an external company to do that
5. Other [ ]
6. We haven't done any data analysis (no entrusting)

<A question for everyone>

Q6: How many employees who work mainly for data analysis (including not only permanent employees but also part-timers/part-time workers/contract employees/dispatched employees) do you have? And what's the percentage of employees who graduated from graduated school in the employees who work mainly for data analysis?

(Please circle just one number corresponding to each one.)

Employees who work mainly for data analysis (as of the end of FY2018): [approx. persons]

Employees who work mainly for data analysis (as of the end of FY2015): [approx. persons]

	1 0%	2 Less than 1%	3 1% or more and less than 3%	4 3% or more and less than 5%	5 5% or more and less than 10%	6 10% or more and less than 20%	7 20% or more and less than 30%	8 30% or more and less than 40%	9 40% or more and less than 50%	10 50% or more
Employees who graduated from graduate school (as of the end of FY2018)	1	2	3	4	5	6	7	8	9	10
Employees who graduated from graduate school (as of the end of FY2015)	1	2	3	4	5	6	7	8	9	10

<For those who chose "AI" one or more times in Q1>

Q7: What kind of AI technology does your company use in the utilization of data? [multiple responses]

1. Supervised learning (regression analysis, discriminant analysis, etc.)
2. Unsupervised learning (clustering, principal component analysis, etc.)
3. Reinforcement learning (Monte Carlo method, Q learning, etc.)
4. Deep learning
5. Other [ ]
6. Utilizing AI but, but have no idea about the specific technology

<A question for those who chose 1. - 5. in Q5. >

Q8: Please answer about the effects (contribution to corporate activities) of utilizing data. How much does each process regarding the utilization of data contribute to the effects in your company?

(Please circle a corresponding number in A – E.)

	1 Contribute very much	2 Contribute to some extent	3 Neither of them	4 Not contribute very much	5 Not contribute at all
A. Data collection (Consideration of data collection, building the system to collect data, etc.)	1	2	3	4	5
B. Data processing (Error check, annotation/labeling, formatting estimating analysis/processing, etc.)	1	2	3	4	5
C. Data browsing/aggregation (Browsing and aggregation of data)	1	2	3	4	5
D. Data analysis/Utilization of AI (Statistical analysis, prediction utilizing AI, etc.)	1	2	3	4	5
E. Interpretation of analysis results (Discussion and decision-making based on analysis results)	1	2	3	4	5

2. About the accumulation (owning) of data

We ask you about the accumulation (owning) status of data in your company.

<Please answer about the data in A – G in Q1 where you chose 1. – 8. >

Q9: Regarding the data your company has accumulated (owned), where is the source of data? Also, do you provide or share the data you have owned for/with other companies, etc. after processing it if necessary? [multiple responses]

(Please circle all the corresponding numbers about each data in A-G.)

(It includes not only your own data but also the data purchased or acquired from outside, but please answer about the data limited to the data utilized for your product development/service expansion, and the efficiency in work, etc. The data temporarily kept from other companies for the delivery of analysis results should not be included here.)

	1 **Acquired** by **collecting** on our own or from the inside	2 **Purchased/acquired** the data owned by other companies (including common use by alliance, etc.	3 **Acquired** public data, etc. for free (statistical data, open data, API utilization, etc.)	4 **Provided** the owned data for other companies (including common use by alliance, etc.)	5 have not grasped that
A. Customers' (individuals') basic data (name, address, sex, etc.)	1	2	3	4	5
B. Customers' (corporations') basic data (name, address, capital, etc.) *Including municipalities/ organizations, etc.	1	2	3	4	5
C. Data of communication with customers, etc. (daily sales reports, purchase history, inquiry history, results of market research, etc.)	1	2	3	4	5
D. Action data on the Internet (websites, SNS, mobile app, etc.) (search/access log, word of mouth/posted data, etc.)	1	2	3	4	5
E. Data on human activities (biological information, location information, camera image, etc.)	1	2	3	4	5
F. Data on the operation of machines, equipment, vehicles, etc. (operational situation, location information, speed, etc.)	1	2	3	4	5
G. Data on nature/public (map information, meteorological information, etc.)	1	2	3	4	5

<A question for those who chose 1 - 8 one time or more in Q1.>

Q10: What percentage is the data purchased from outside (excluding the data published for free of charge and the data for free common use due to the alliance, etc.) in your accumulated (owned) data? Please answer **on the number of cases basis**, not on data amount basis.

(If you don't grasp that exactly, please answer within the scope of the data you have grasped.)

[Approx. %]

< Please answer about the data in A – G in Q1 where you chose 1. – 8.>

Q11: Regarding the data your company has accumulated (owned), please answer the rough makeup percentages **(on data amount basis) as of the end of FY2018**, **making the total of A-G 100%**. In addition, please answer the number of subjects (persons/companies/customers/units) of accumulated (owned) data in **round numbers with 2 significant figures** (examples: 12,000, 450,000). You don't have to answer about the items whose numbers are unclear. Besides, please answer the specific type of accumulated (owned) data within the scope of your knowledge.

(Regarding the number of subjects, please answer assuming the number of customers or units of installed equipment (sensors), etc.)

(Regarding the data in C, if both individuals and corporations are included in the subject, please answer the total number of individuals and companies/organizations.)

(It includes not only your company's data but also the data purchased or acquired from outside, but please answer about the data that can be utilized for product development/ service expansion and efficiency in work, etc. in your company. The data temporarily kept from other companies for the delivery of analysis results should not be included here.)

	Amount ratio in the accumulated (owned) data (%)	Number of subjects in accumulated (owned) data	[Arbitrary] Specific type of data you have accumulated (owned)
A. Customers' (individuals') basic data (name, address, sex, etc.)	%	persons	
B. Customers' (corporations') basic data (name, address, capital, etc.) *Including municipalities/ organizations, etc.	%	companies/organizations	
C. Data of communication with customers, etc. (daily sales reports, purchase history, inquiry history, results of market research, etc.)	%	customers (persons＋ companies/organizations)	
D. Action data on the Internet (websites, SNS, mobile app, etc.) (search/access log, word of mouth/posted data, etc.)	%	persons	
E. Data on human activities (biological information, location information, camera image, etc.)	%	persons	
F. Data on the operation of machines, equipment, vehicles, etc. (operational situation, location information, speed, etc.)	%	units	
G. Data on nature/public (map information, meteorological information, etc.)	%		
Total	100%		

About the data you answered above, please tell us about the total amount of data you owned **as of the end of FY2018**. The total amount should be rounded to **2 significant figures** (examples: 12,000, 450,000). **If you don't figure out the precise value, please answer in the intermediate value** (example: If you think it's between 1,000 and 2,000TB, please input 1,500TB).

			,			,			TB

Notes) TB (terabyte)= about 1000GB. 1PB (petabyte)= about 1000TB. 1EB (exabyte)= about 1000PB.

< Please answer about the data in A – G in Q1 where you chose 1. – 8.>

Q12: Regarding the data your company has accumulated (owned), what percentage is **the data purchased/acquired from outside as of the end of FY2018** (including the data published for free of charge and the data for free common use due to alliance, etc.) in your accumulated (owned) data? Please answer **in the number of cases basis**, not on data amount basis. (Please circle a number in each data of A – G.)

(Please answer, assuming the number of customers and units of equipment (sensors), etc.)

(Regarding the data in C, if both individuals and corporations are included in the subject, please answer the total number of individuals and companies/organizations.)

(Please answer about only the data that can be utilized for product development/service expansion and efficiency in work, etc. in your company. The data temporarily kept from other companies for the delivery of analysis results should not be included here.)

	1 0%	2 Less than 5%	3 5% or more and less than 10%	4 10% or more and less than 20%	5 20% or more and less than 30%	6 30% or more and less than 40%	7 40% or more and less than 50%	8 50% or more and less than 60%	9 60% or more and less than 70%	10 70% or more and less than 80%	11 80% or more and less than 90%	12 90% or more
A. Customers' (individuals') basic data (name, address, sex, etc.)	1	2	3	4	5	6	7	8	9	10	11	12
B. Customers' (corporations') basic data (name, address, capital, etc.) *Including municipalities/ organizations, etc.	1	2	3	4	5	6	7	8	9	10	11	12
C. Data of communication with customers, etc. (daily sales reports, purchase history, inquiry history, results of market research, etc.)	1	2	3	4	5	6	7	8	9	10	11	12
D. Action data on the Internet (websites, SNS, mobile app, etc.) (search/access log, word of mouth/posted data, etc.)	1	2	3	4	5	6	7	8	9	10	11	12
E. Data on human activities (biological information, location information, camera image, etc.)	1	2	3	4	5	6	7	8	9	10	11	12
F. Data on the operation of machines, equipment, vehicles, etc. (operational situation, location information, speed, etc.)	1	2	3	4	5	6	7	8	9	10	11	12
G. Data on nature/public (map information, meteorological information, etc.)	1	2	3	4	5	6	7	8	9	10	11	12

< Please answer about the data in A – G in Q1 where you chose 1. – 8.>

Q13: Regarding the data your company has accumulated (owned), how much the data **as of the end of FY2018** has changed compared with the one **as of the end of FY2015?** Please answer **on the number of cases basis**, not on data amount basis. (Please circle a number in each data of A – G.)

(Regarding the data in C, if both individuals and corporations are included in the subject, please answer the total number of individuals and companies/organizations.)

(It includes not only your own data but also the data purchased or acquired from outside, but please answer about the data limited to the data utilized for your product development/service expansion, and the efficiency in work, etc. The data temporarily kept from other companies for the delivery of analysis results should not be included here.)

	1 Not owned as of the end of FY2015	2 Decreased by more than 50%	3 Decreased by 10 – 40%	4 Almost no change	5 Increased by around 10%	6 Increased by 20 – 30%	7 Increased by 40 – 60%	8 Increased by 70 – 90%	9 Increased more than 2 times	10 No idea
A. Customers' (individuals') basic data (name, address, sex, etc.)	1	2	3	4	5	6	7	8	9	10
B. Customers' (corporations') basic data (name, address, capital, etc.) *Including municipalities/ organizations, etc.	1	2	3	4	5	6	7	8	9	10
C. Data of communication with customers, etc. (daily sales reports, purchase history, inquiry history, results of market research, etc.)	1	2	3	4	5	6	7	8	9	10
D. Action data on the Internet (Websites, SNS, mobile app, etc.) (search/access log, word of mouth/posted data, etc.)	1	2	3	4	5	6	7	8	9	10
E. Data on human activities (biological information, location information, camera image, etc.)	1	2	3	4	5	6	7	8	9	10
F. Data on the operation of machines, equipment, vehicles, etc. (operational situation, location information, speed, etc.)	1	2	3	4	5	6	7	8	9	10
G. Data on nature/public (map information, meteorological information, etc.)	1	2	3	4	5	6	7	8	9	10

< Please answer about the data in A – G in Q1 where you chose 1. – 8.>

Q14: Regarding the data your company has accumulated (owned), what percentage is the **non-structured data** (data in which the writing of business diaries and SNS and data in which rules (structures) such as voice/images/video are not clearly defined) in the accumulated (owned) data? Please answer **on the number of cases basis**, not on data amount basis. (For each data in A – G, please circle only one corresponding number.)

(Regarding the data in C, if both individuals and corporations are included in the subject, please answer the total number of individuals and companies/organizations.)

(Please answer about the data limited to the data utilized for your product development/service expansion, and the efficiency in work, etc. The data temporarily kept from other companies for the delivery of analysis results should not be included here.)

	1 0%	2 Less than 5%	3 5% or more and less than 10%	4 10% or more and less than 20%	5 20% or more and less than 30%	6 30% or more and less than 40%	7 40% or more and less than 50%	8 50% or more and less than 60%	9 60% or more and less than 70%	10 70% or more and less than 80%	11 80% or more and less than 90%	12 90% or more
A. Customers' (individuals') basic data (name, address, sex, etc.)	1	2	3	4	5	6	7	8	9	10	11	12
B. Customers' (corporations') basic data (name, address, capital, etc.) *Including municipalities/ organizations, etc.	1	2	3	4	5	6	7	8	9	10	11	12
C. Data of communication with customers, etc. (daily sales reports, purchase history, inquiry history, results of market research, etc.)	1	2	3	4	5	6	7	8	9	10	11	12
D. Action data on the Internet (Websites, SNS, mobile app, etc.) (search/access log, word of mouth/posted data, etc.)	1	2	3	4	5	6	7	8	9	10	11	12
E. Data on human activities (biological information, location information, camera image, etc.)	1	2	3	4	5	6	7	8	9	10	11	12
F. Data on the operation of machines, equipment, vehicles, etc. (operational situation, location information, speed, etc.)	1	2	3	4	5	6	7	8	9	10	11	12
G. Data on nature/public (map information, meteorological information, etc.)	1	2	3	4	5	6	7	8	9	10	11	12

< A question for those who chose 1 - 7 one time or more in Q1.>

Q15: What percentage was the data actually utilized in the data accumulated (owned) in your company **in a year (FY2018)**? Please answer **on the number of cases basis**, not on data amount basis. In addition, answer about the status **as of FY2015 (3 years earlier than that).**

(Please circle a number for each fiscal year.)

(If you don't grasp that exactly, please answer within the scope of the data you have grasped.)

		1 Less than 5%	2 5% or more and less than 10%	3 10% or more and less than 20%	4 20% or more and less than 30%	5 30% or more and less than 40%	6 40% or more and less than 50%	7 50% or more and less than 60%	8 60% or more and less than 70%	9 70% or more and less than 80%	10 80% or more and less than 90%	11 90% or more
A.	FY2018	1	2	3	4	5	6	7	8	9	10	11
B.	FY2015	1	2	3	4	5	6	7	8	9	10	11

3. About market structure

We ask you about the competitive environment in the market where your company put your representative products/services.

<A question for everyone>

Q16: How much new products/new services did your company put in the market in **FY2018**?

(New products/new services include not only the ones that renewed functions/performances/designs/materials/components/purposes but also the ones that are combined with existing technologies and technically improved existing products/services. However, they don't include that ones that have the same functions of products/services and usage purposes, and only the design change or the ones just for the sales/providing for other companies' products/services,)

1. 0 case (No products/services were brought to the market)
2. 1 - 2 cases
3. 3 - 5 cases
4. 6 - 10 cases
5. 11 - 20 cases
6. 21 - 30 cases
7. 31cases or more

<A question for everyone>

Q17: In the market where your company put your main products/services, how many competitive companies did you have **at the end of FY2018** in Japan and the world in total? In addition, please answer about the changes **from the end of FY2015** to **the end of FY2018**.

(Please circle a number for each.)

(Please answer assuming the market of the same products/services as of the end of FY2015 and the end of FY2018.)

	1 0 company	2 1 - 2 companies	3 3 - 5 companies	4 6 - 10 companies	5 11 - 20 companies	6 21 companies or more
Number of competitors at the end of FY2018	1	2	3	4	5	6

	1 Decreased by 6 or more	2 Decreased by 3 - 5	3 Decreased by 1 - 2	4 No change	5 Increased by 1 - 2	6 Increased by 3 - 5	7 Increased by 6 or more
Change from the end of FY2015	1	2	3	4	5	6	7

The number of responses to the questionnaire survey (by industry and by total capital) and aggregation results are as follows.

Table 55: The number of responses to the questionnaire survey by industry

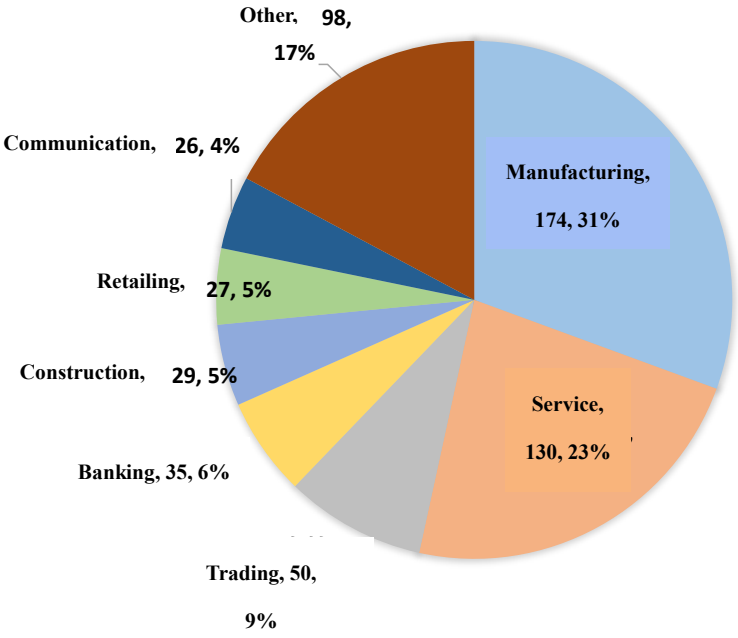


Table 56: The number of responses to the questionnaire survey by total capital

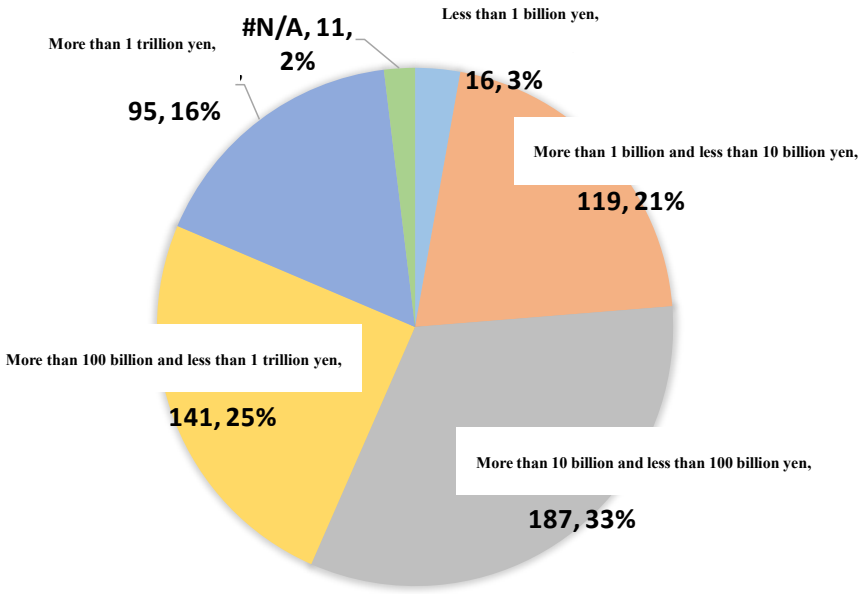
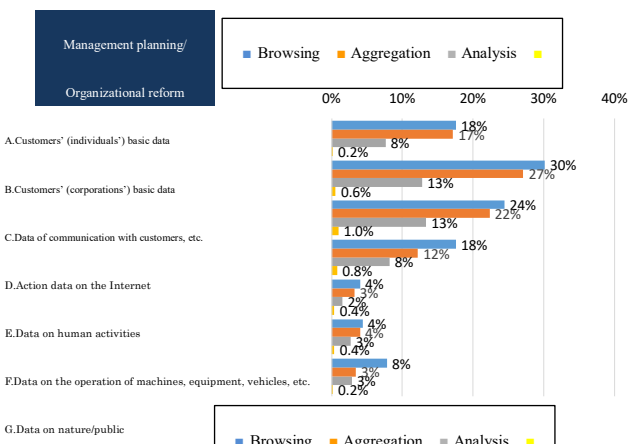
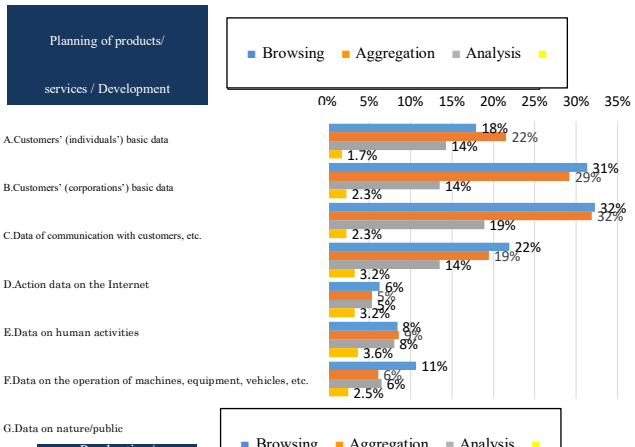
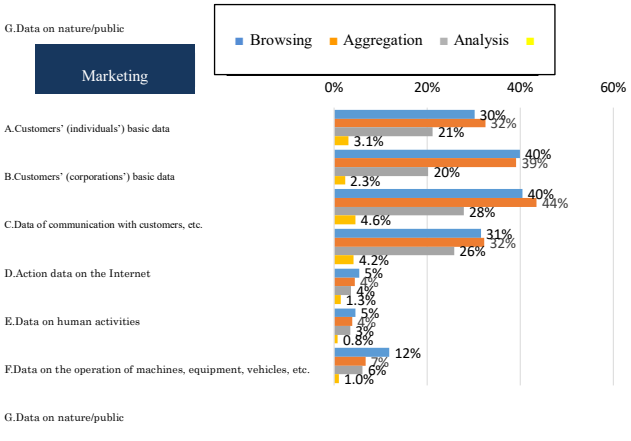


Table 57: Questionnaire results (Q1-1)(n=524)







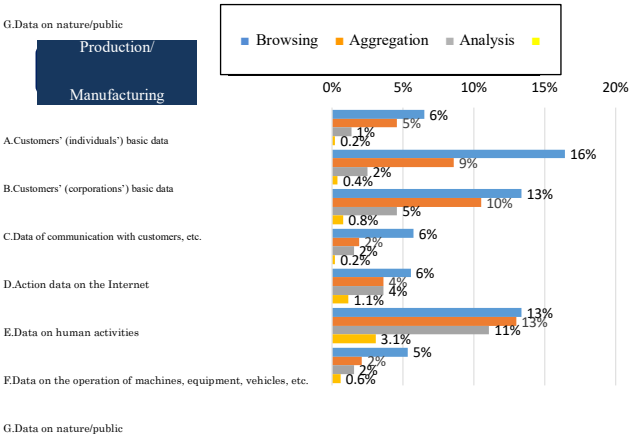
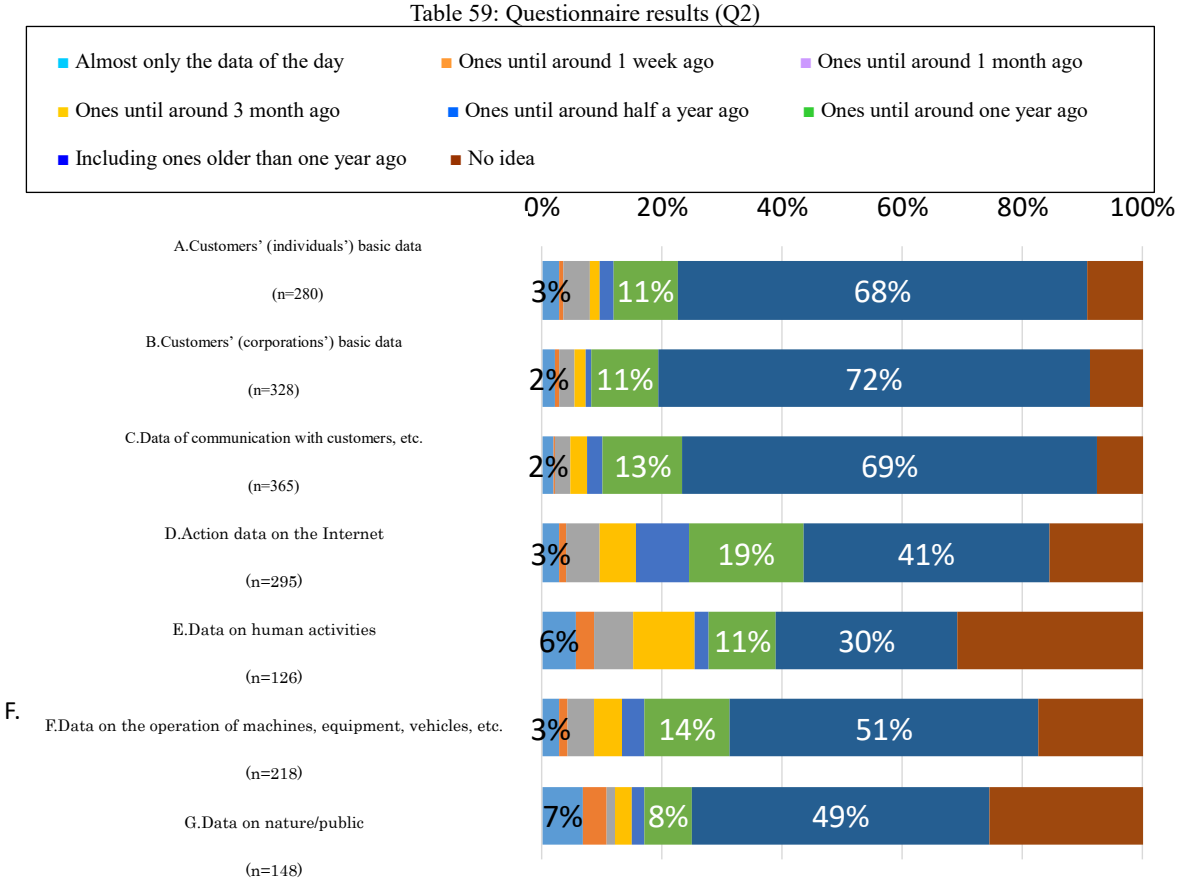


Table 58: Questionnaire results (Q1-2)(n=524)

Table 59: Questionnaire results (Q2)



F.

Table 60: Questionnaire results (Q3)

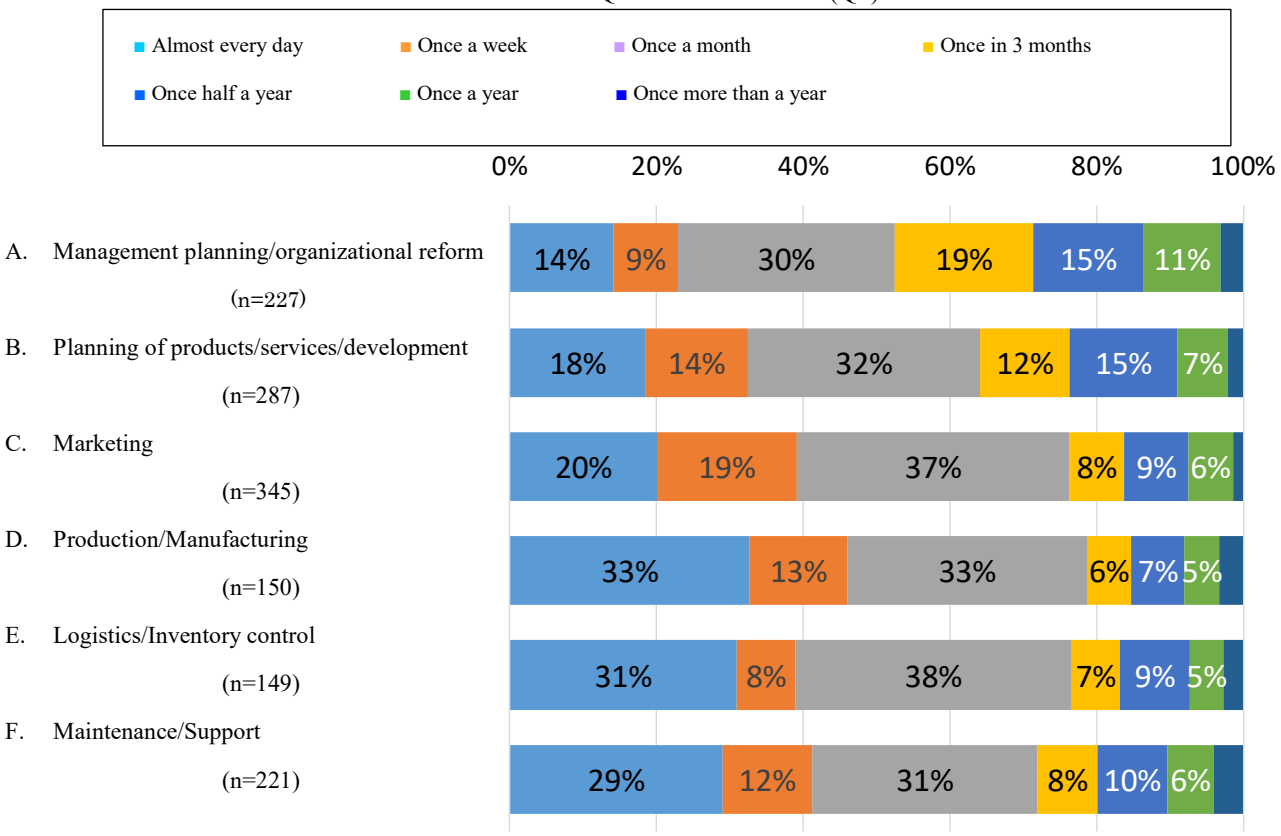


Table 61: Questionnaire results (Q4) (n=468)

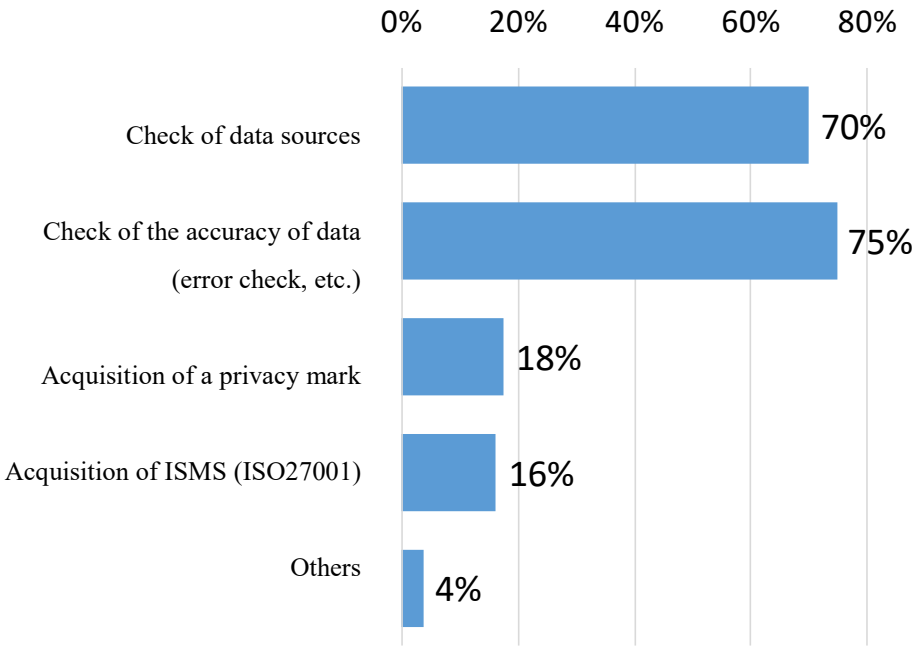


Table 62: Questionnaire results(Q5-1)(n=529)

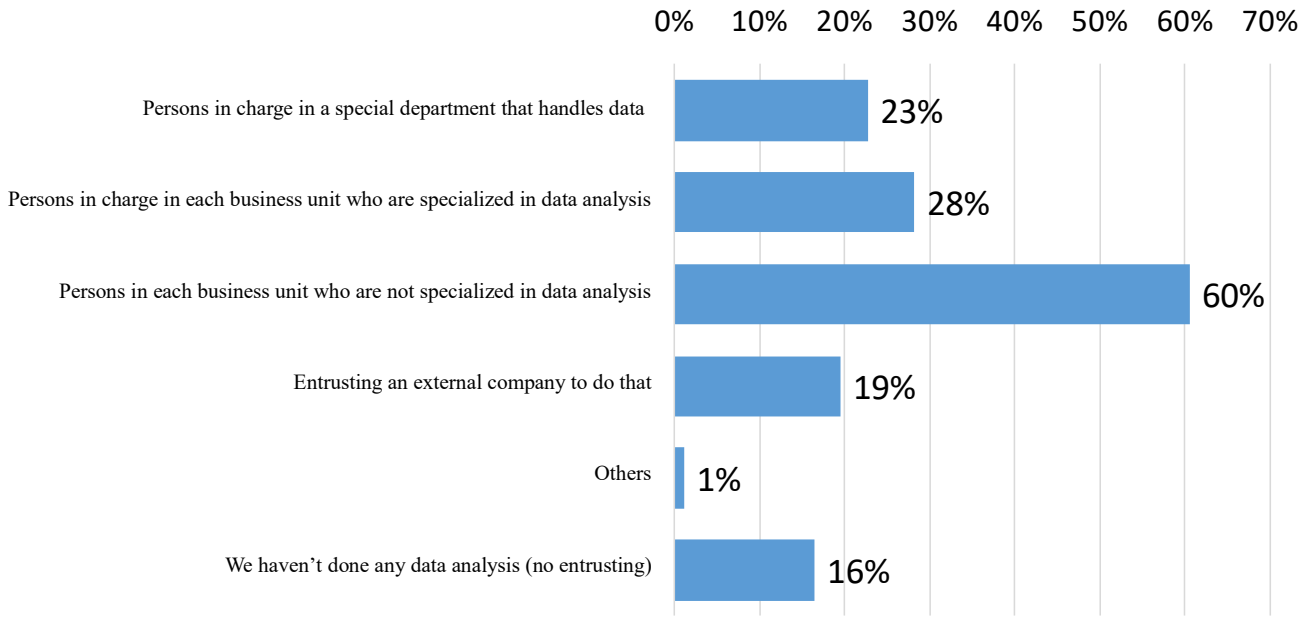


Table 63: Questionnaire results(Q5-2) (n=529) The timing when the specialized department was launched

Average value	9.8 years
Median value	5.0 years
Minimum value	1.0 year
Maximum value	60.0 years
Mode	10.0 years

Table 64: Questionnaire results (Q6-1)

(Unit: persons)

	FY2018	FY2015
Average	48	40
Average excluding 0	80	76
Minimum excluding 0	1	1
Maximum	9,999	9,999

Table 65: Questionnaire results (Q6-2) Percentage of employees who graduated from graduate school

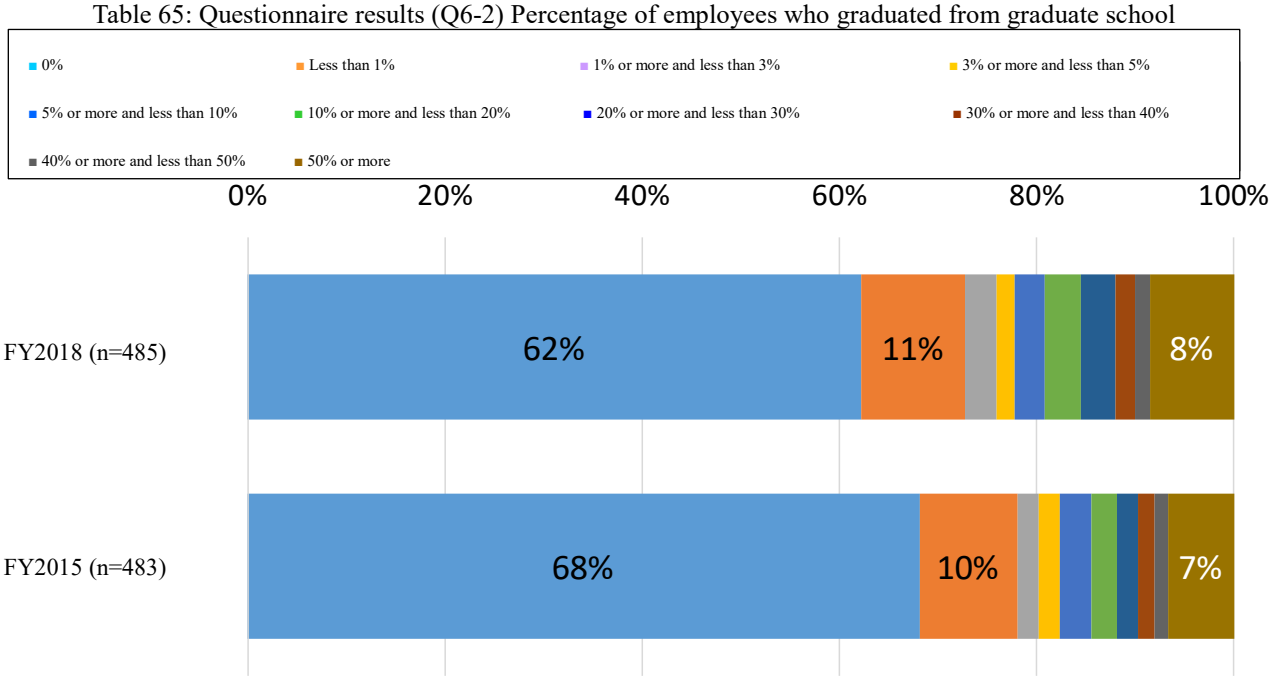


Table 66: Questionnaire results (Q7) (n=104)

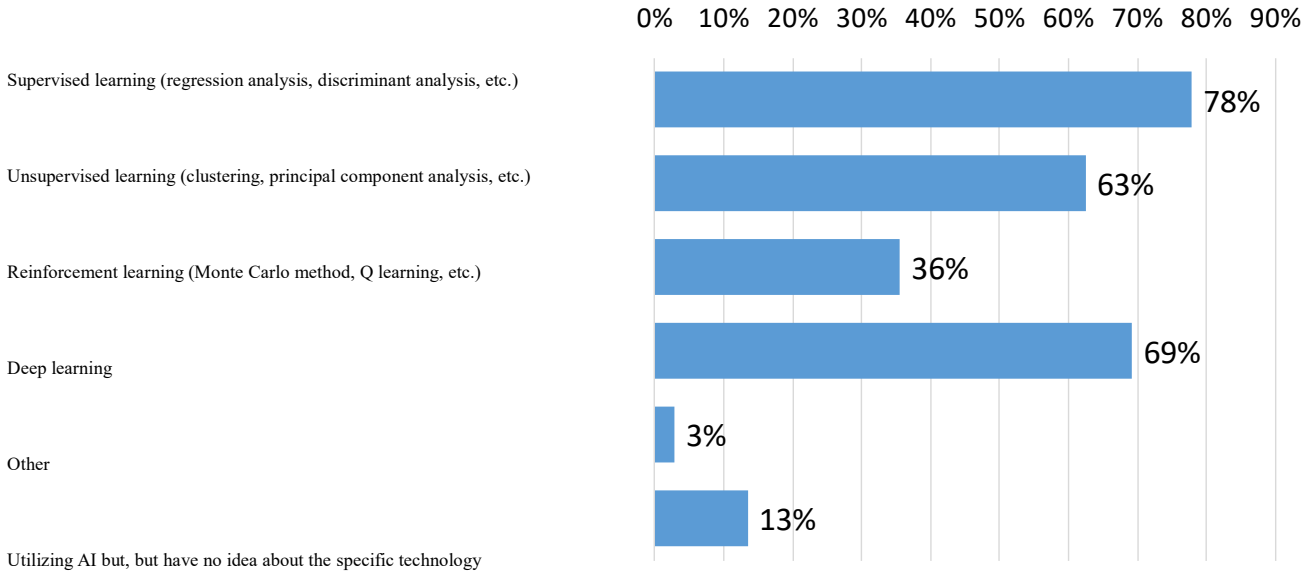
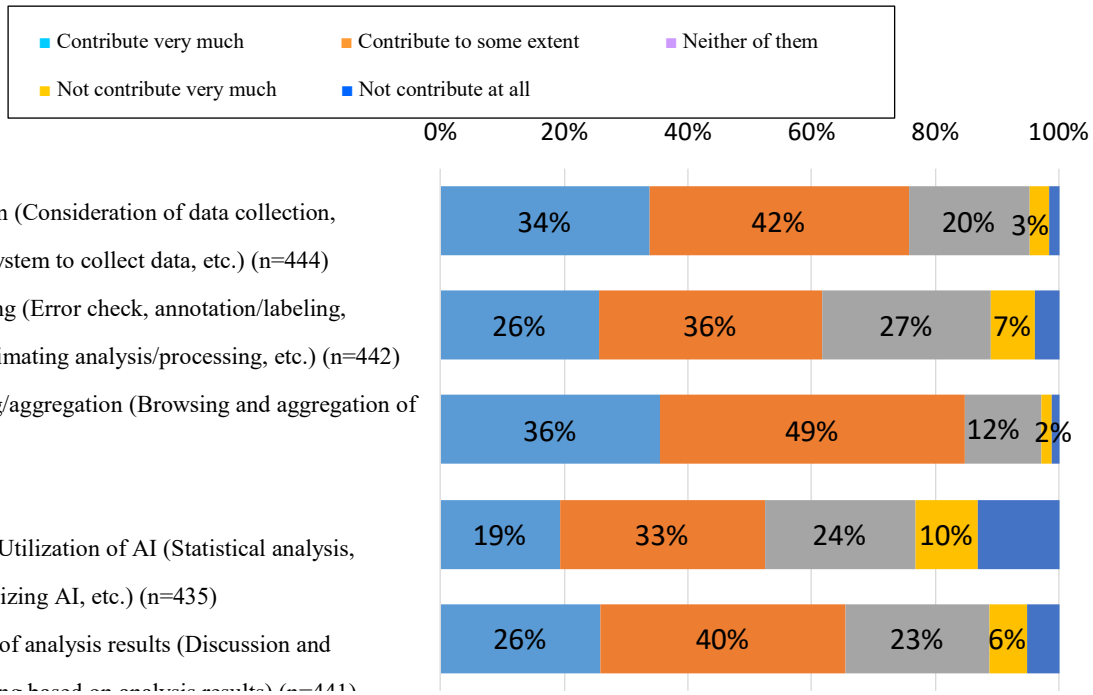


Table 67: Questionnaire results (Q8)



A. Data collection (Consideration of data collection, building the system to collect data, etc.) (n=444)

B. Data processing (Error check, annotation/labeling, formatting estimating analysis/processing, etc.) (n=442)

C. Data browsing/aggregation (Browsing and aggregation of data)
(n=444)

D. Data analysis/Utilization of AI (Statistical analysis, estimation utilizing AI, etc.) (n=435)

E. Interpretation of analysis results (Discussion and decision-making based on analysis results) (n=441)

Table 68: Questionnaire results (Q9) (n=508)

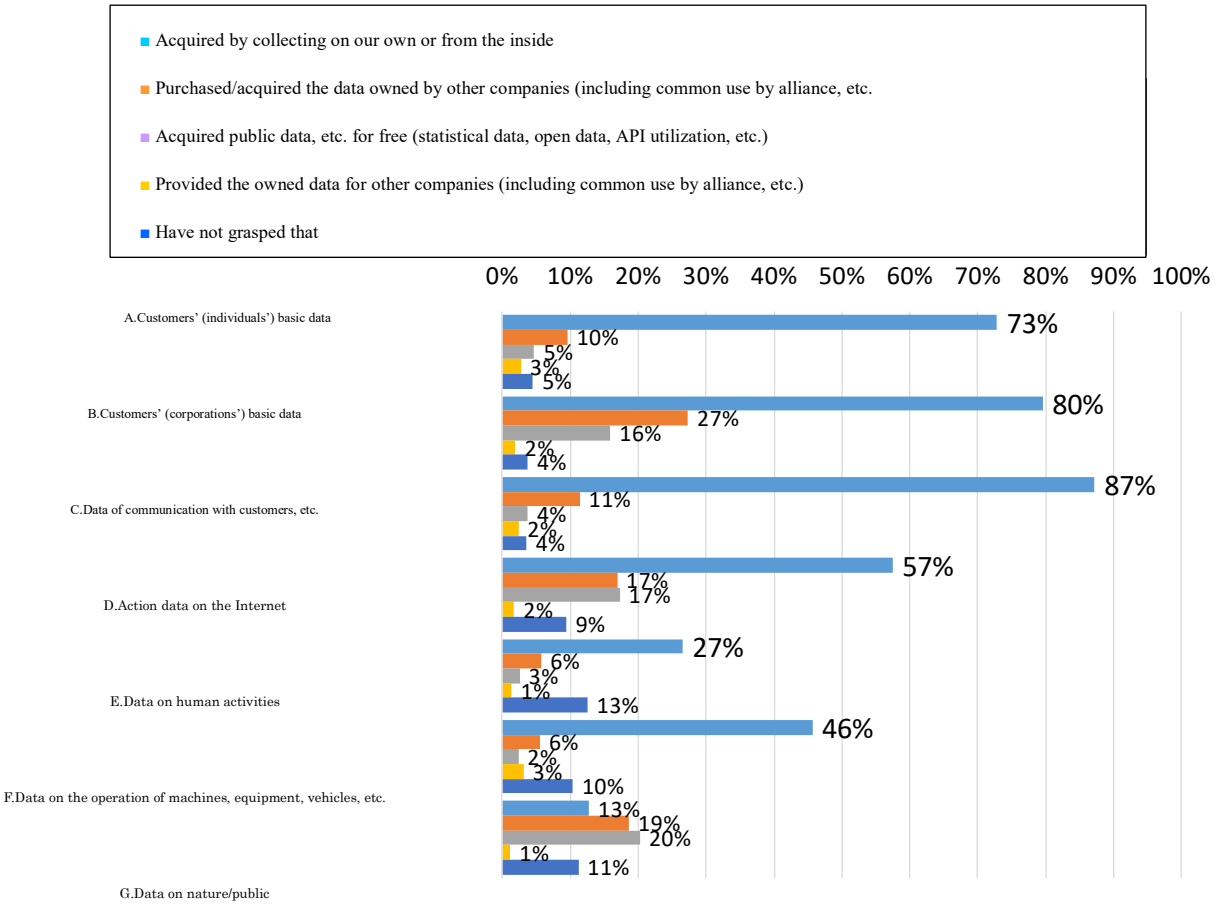


Table 69: Questionnaire results (Q10) (n=461)

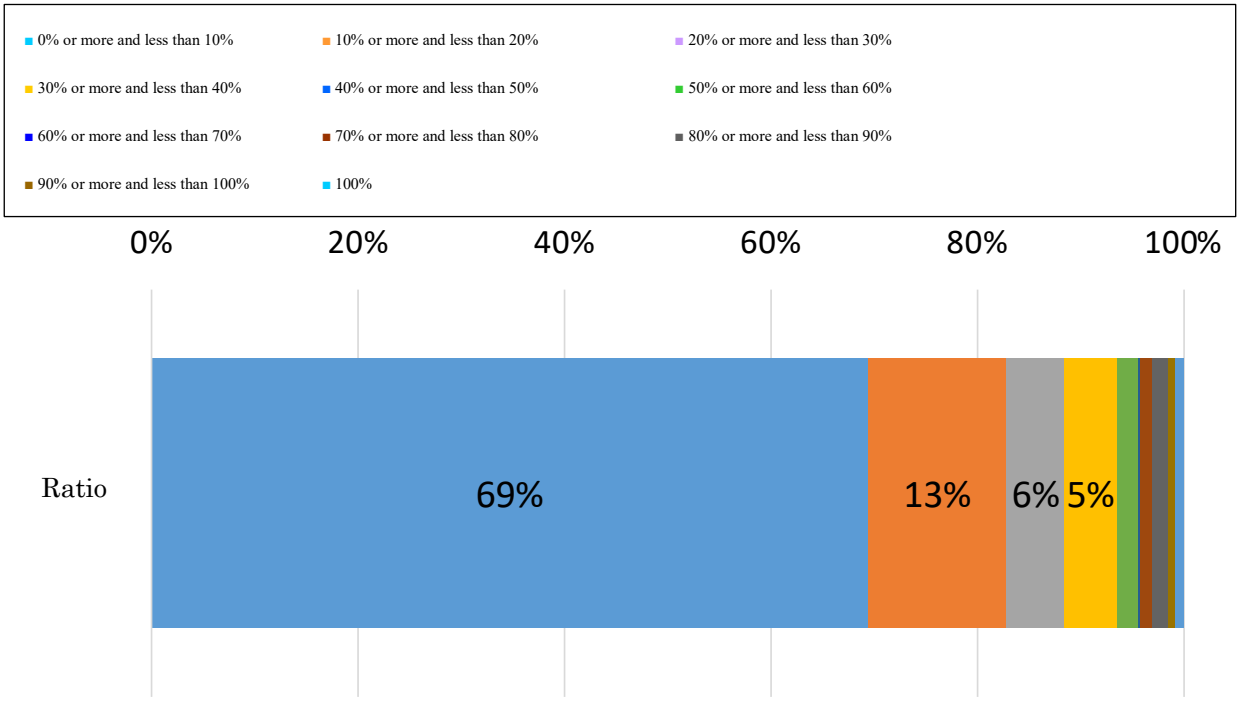


Table 70: Questionnaire results (Q11-1) (n=426)

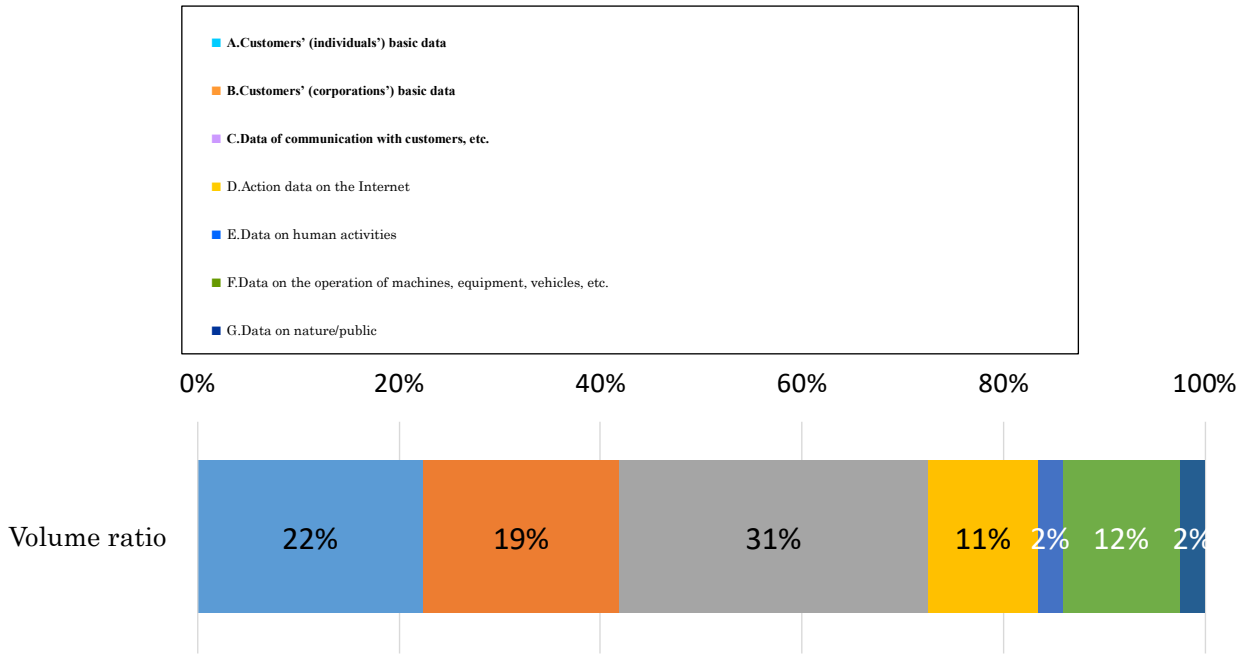


Table 71: Questionnaire results (Q11-2) Number of subjects of data

	A. Customers' (individuals') basic data (name, address, sex, etc.) [persons] (n=141)	B. Customers' (corporations') basic data (name, address, capital, etc.) [companies/ organizations] (n=169)	C. Data of communication with customers, etc. (daily sales reports, purchase history, inquiry history, results of market research, etc.) customers [persons + companies/ organizations] (n=143)	D. Action data on the Internet (Websites, SNS, mobile app, etc.) (search/access log, word of mouth/posted data, etc.) [persons] (n=87)	E. Data on human activities (biological information, location information, camera image, etc.) [persons] (n=33)	F. Data on the operation of machines, equipment, vehicles, etc. (operational situation, location information, speed, etc.) [units] (n=76)
Average value	9,299,791	100,287	12,905,002	1,302,429	246,115	32,212
Median value	60,000	2,000	10,000	1,000	300	200
Average value excluding 0	9,959,351	103,254	13,356,226	1,526,985	365,443	36,451
Median value excluding 0	20	5	2	1	10	4
Maximum value	1,000,000,000	4,500,000	1,000,000,000	35,000,000	10,000,000	700,000

Table 72: Questionnaire results(Q11-3) (n=380) Total of data amount

Average value	6,373 TB
Median value	10 TB
Minimum value	0 TB
Maximum value	1,000,000 TB
Mode	1 TB

Table 73: Questionnaire results (Q12)

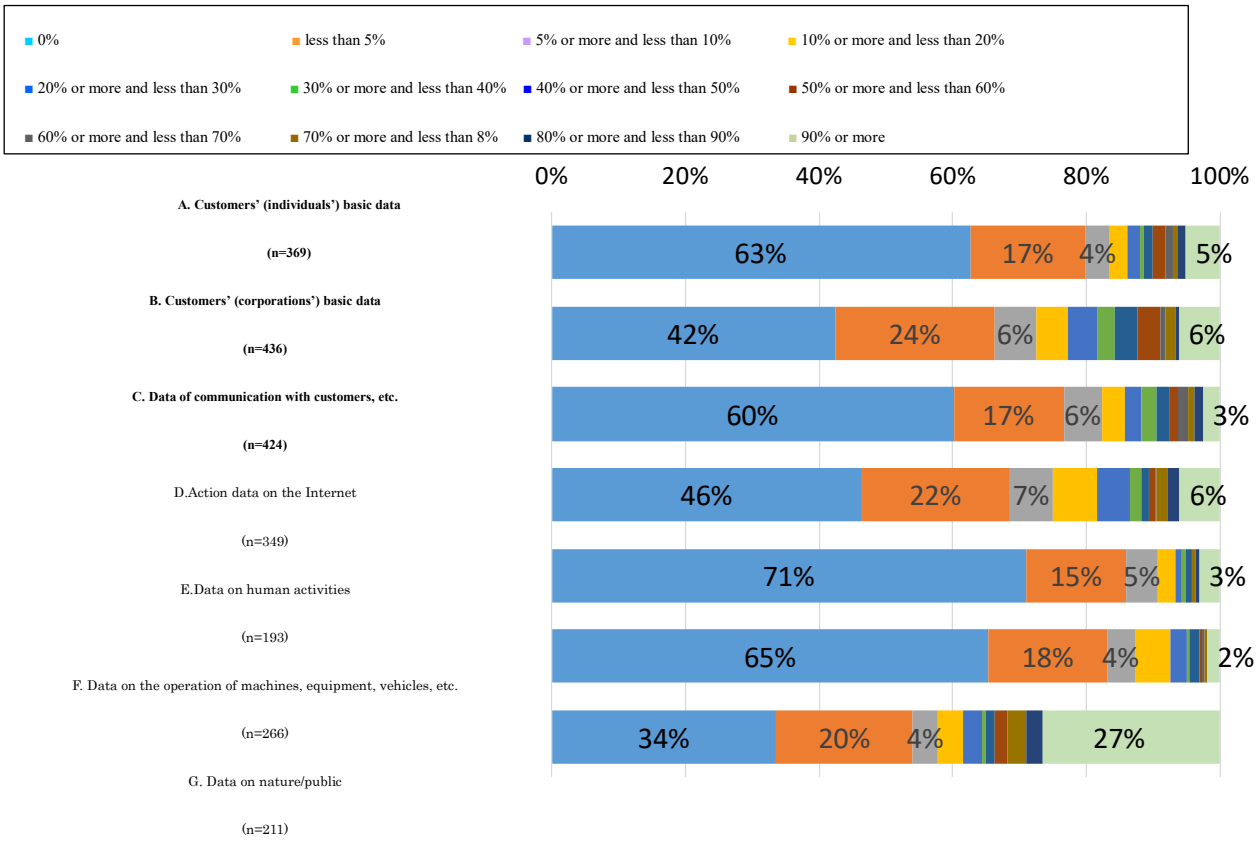


Table 74: Questionnaire results (Q13)

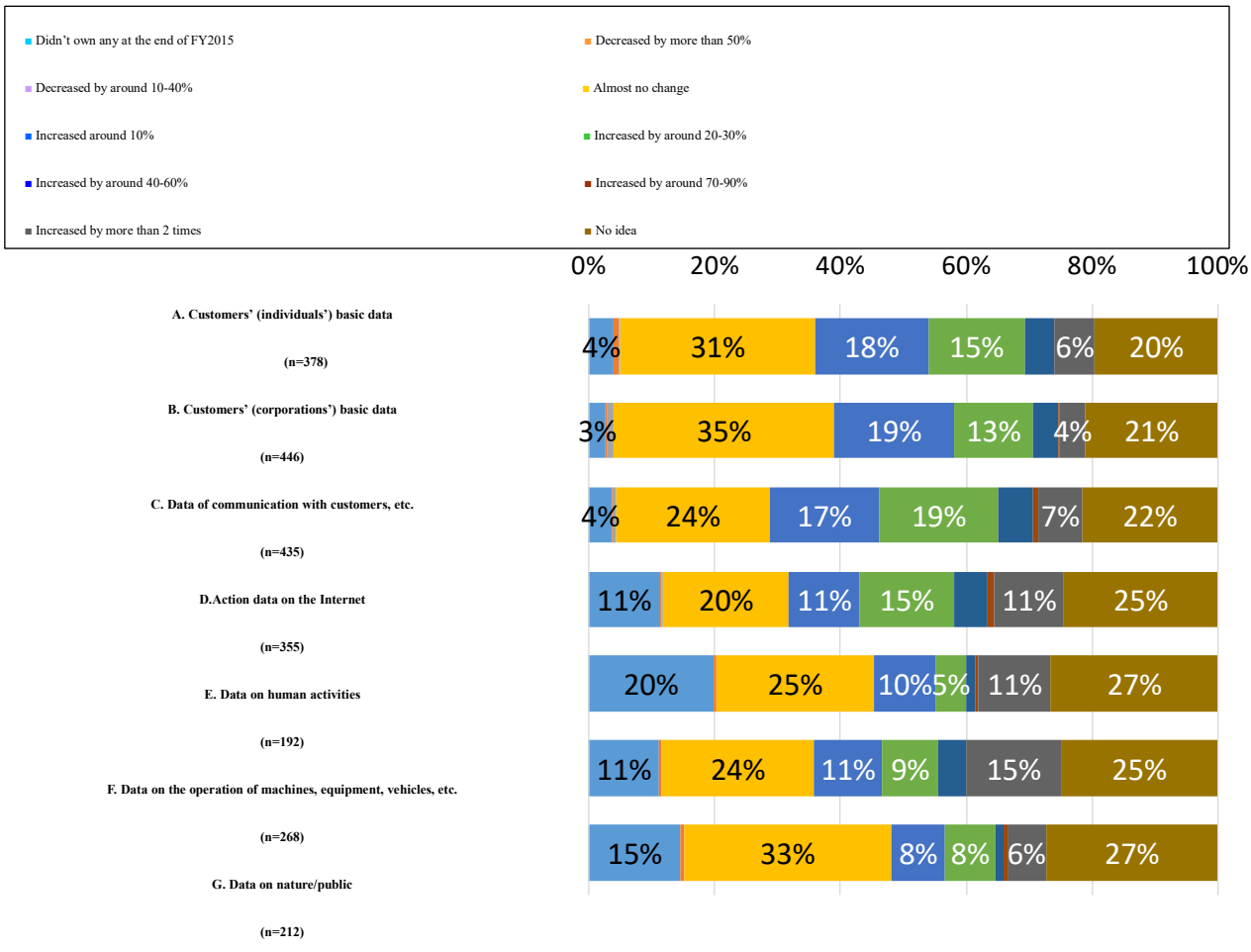


Table 75: Questionnaire results(Q14)

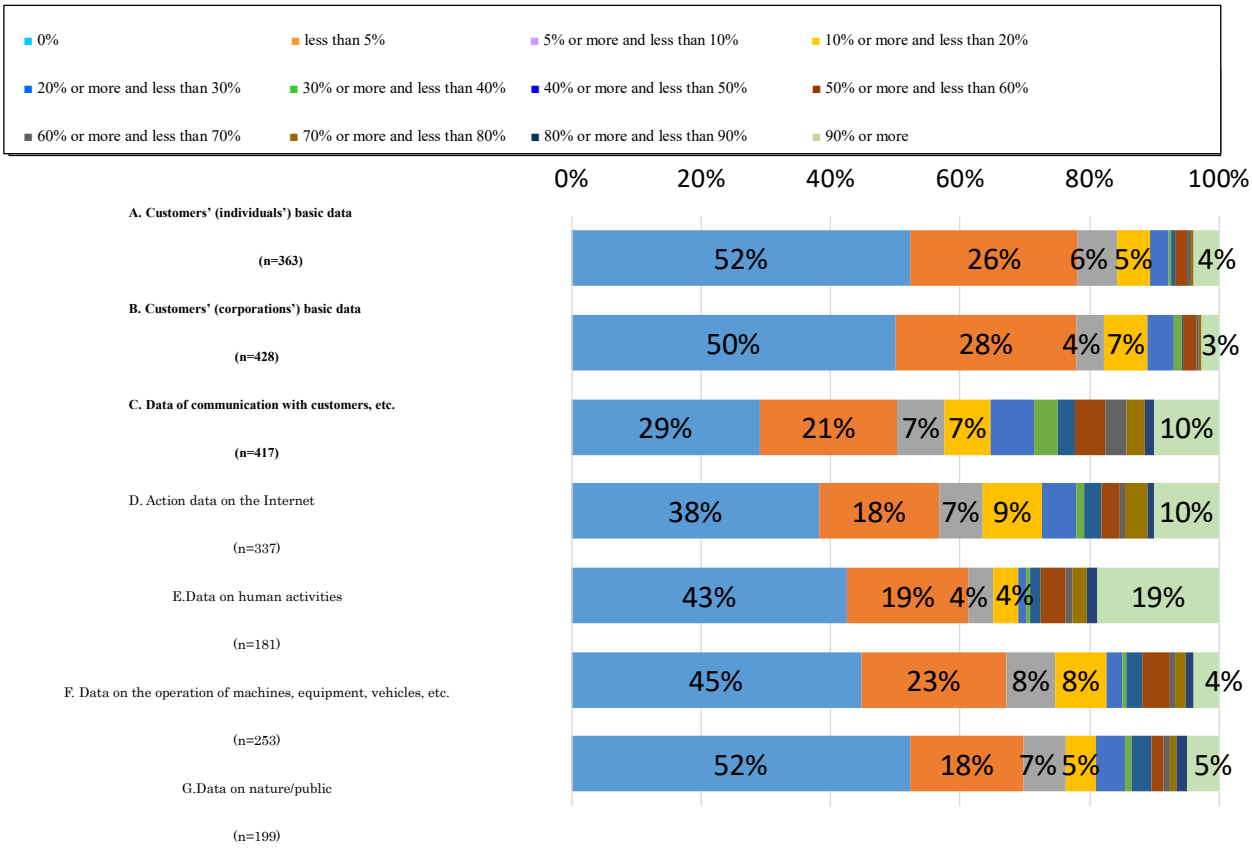


Table 76: Questionnaire results (Q15)

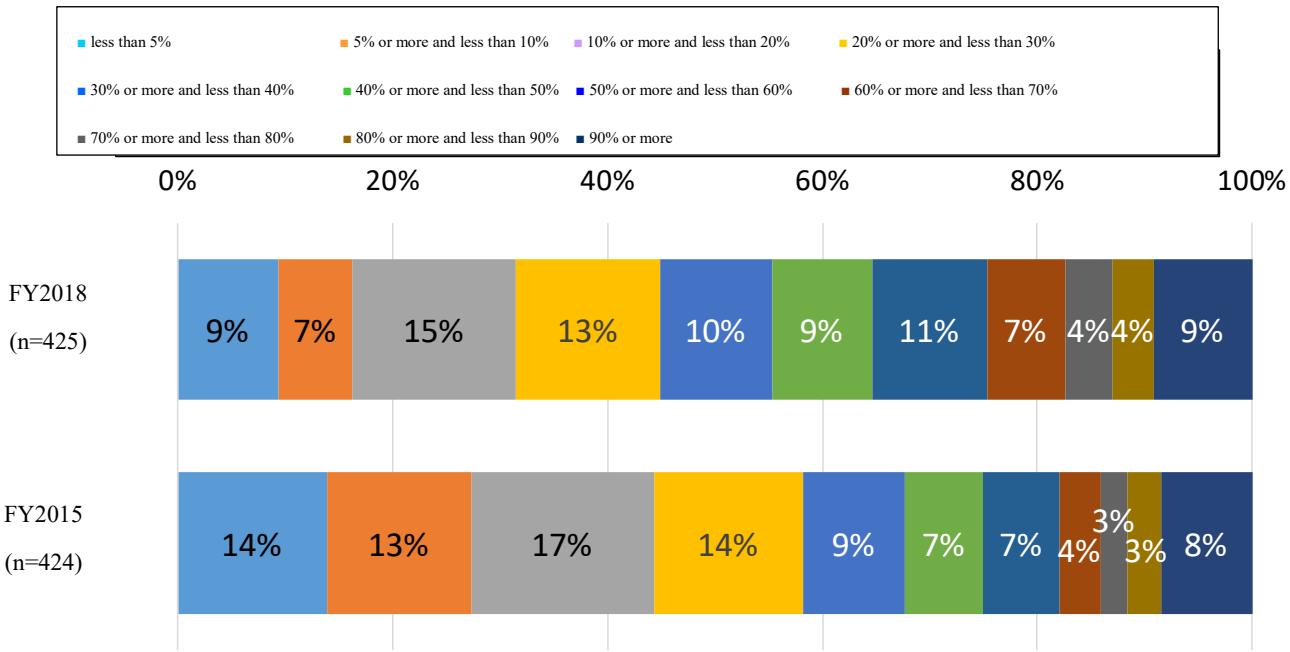


Table 77: Questionnaire results (Q16) (n=492)

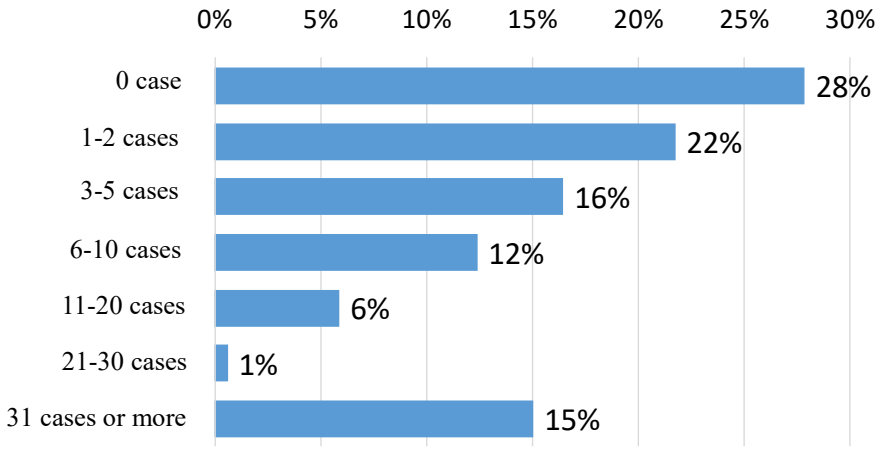


Table 78: Questionnaire results (Q17-1) (n=484)

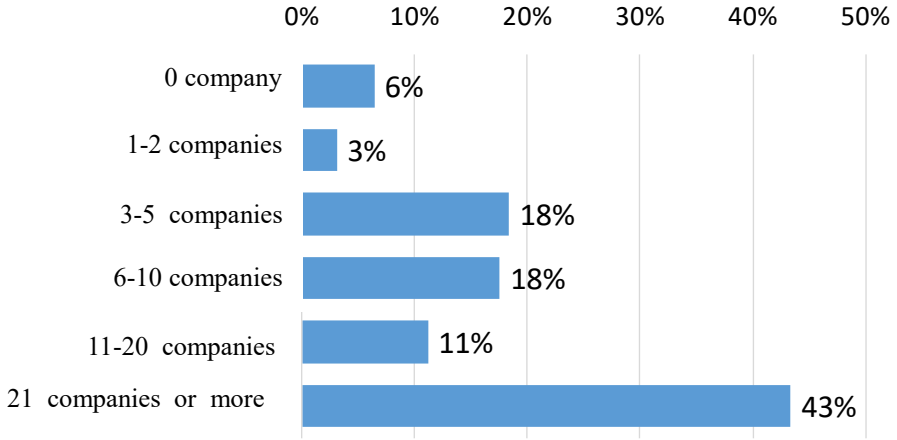
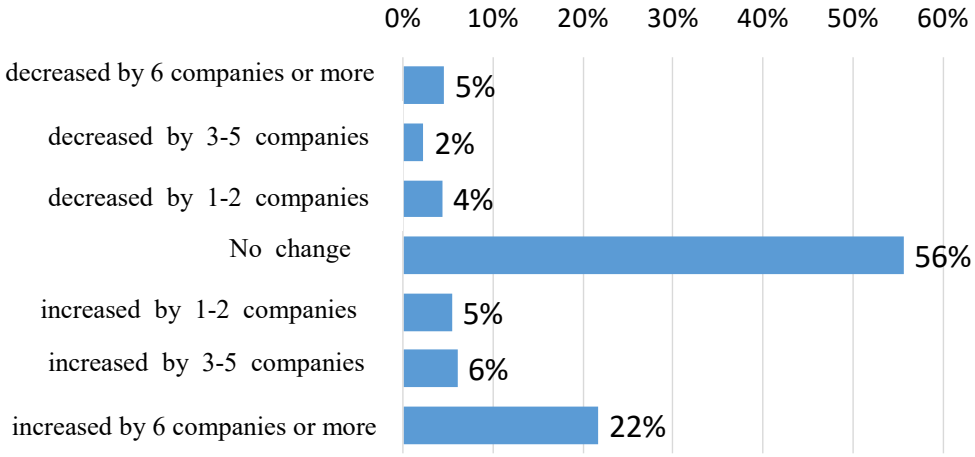


Table 79: Questionnaire results (Q17-2) (n=480)



6.2 Empirical analysis

The detail of empirical analysis to conduct the trial estimation of the value of data in 3.1 is as follows. First of all, the estimated formula is as mentioned in 3.1, and it's the Cobb-Douglas Production Function modified to include data as a factor (linear homogeneity is not assumed). In estimation, it was estimated by putting logarithms in both sides and adding Industry Dummy (1=manufacturing、0=non-manufacturing).

$$V = A_o K^{\alpha} L^{\beta} Data^{\gamma} \qquad \log(V) = \log A_o + \alpha \log(K) + \beta \log(L) + \gamma \log(Data) + Industry\ Dummy^{*}$$

*Classified into manufacturing and non-manufacturing

In addition, fundamental statistics of variables are as shown in Table 80. Because logarithmic transformation is done in the estimation, the enterprises who have the value zero in either variable are excluded from the targets of estimation. Added value, tangible fixed assets＋intangible fixed assets, and total number of employees were acquired from Nikkei NEEDS-FinancialQUEST, and the results of the questionnaire survey in 6.1 was utilized for variables related to data.

Table 80: Fundamental statistics

Variable	Unit	N-Number	Average value	Standard deviation
Added value	100 million yen	509	428	5,781
Tangible fixed assets ＋ Intangible fixed assets	100 million yen	478	2,112	7,734
Total number of employees	persons	493	12,049	37,157
Utilized data amount	TB	349	3,622	49,968
Number of utilized data cases	cases	189	8,271,450	55,300,000
Utilized data amount (FY2015)	TB	349	3,053	48,340
Number of utilized data cases (FY2015)	cases	189	5,312,146	31,000,000
External data amount	TB	373	2,100	37,360
Number of external data cases	cases	196	7,076,037	68,300,000
Internally owned data amount	TB	373	4,385	53,316
Number of internally owned data cases	cases	196	10,600,000	57,300,000
Total data amount x data utilization degree	TB	352	2,475	33,960
Total number of data cases x data utilization degree	cases	192	2,388,490	13,900,000
Total data amount x data diversity	TB	380	2,223	34,627
Total number of data cases x data diversity	cases	201	2,691,044	18,400,000
Total data amount	TB	380	6,373	72,761
Total number of data cases	cases	201	17,300,000	104,000,000
Data utilization degree	-	473	0.10	0.09
Data diversity	-	508	0.08	0.11
Industry dummy (manufacturing = 1)	-	569	0.31	0.46

The details of estimated results are shown in Table 81 - 83. The variables related to data are significant in any model, and coefficients are around 0.05. Also, the Industry Dummy are not significant in any model, so it is considered that the differences between industries are relatively small when they are seen in added values and the entire production factors.

Table 81: Estimated results①

Explanatory variable \ Unexplanatory variable	log (added value)	log (added value)	log (added value)	log (added value)
log (Tangible fixed assets＋Intangible fixed assets)	0.44 (10.44) ***	0.338 (5.69) ***	0.436 (10.01) ***	0.337 (5.64) ***
log (Total number of employees)	0.50 (8.76) ***	0.55 (7.78) ***	0.51 (8.79) ***	0.56 (7.86) ***
log (Utilized data amount)	0.05 (2.96) ***			
log (Number of utilized data cases)		0.07 (2.69) ***		
log (Utilized data amount, FY2015)			0.05 (2.64) ***	
log (Number of utilized data cases, FY2015)				0.06 (2.38) **
Industry Dummy (Manufacturing)	-0.05 (-0.40)	0.10 (0.54)	-0.05 (-0.44)	0.09 (0.46)
Constant	-2.11 (-7.78) ***	-2.71 (-7.55) ***	-2.11 (-7.74) ***	-2.67 (-7.39) ***
Number of samples	258	135	258	135
Adj R-squared	0.8343	0.8157	0.8332	0.8136

From left, coefficient, t value, p value

(Note) *Significant level 10%, ** Significant level 5%, *** Significant level 1%

Table 82: Estimated results②

Explanatory variable \ Unexplanatory variable	log (added value)	log (added value)	log (added value)	log (added value)
log (Tangible fixed assets＋Intangible fixed assets)	0.47 (11.24) ***	0.35 (6.43) ***	0.47 (11.14) ***	0.34 (6.16) ***
log (Total number of employees)	0.47 (8.48) ***	0.51 (7.67) ***	0.48 (8.60) ***	0.55 (8.25) ***
log (External data amount	0.04 (3.11) ***			
log (Number of external data cases)		0.06 (3.28) ***		
log (Internally owned data amount)			0.05 (2.44) **	
log (Number of internally owned data cases)				0.06 (2.48) **
Industry Dummy (Manufacturing)	-0.02 (-0.16)	0.09 (0.49)	-0.04 (-0.31)	0.07 (0.41)
Constant	-1.96 (-7.25) ***	-2.34 (-7.71) ***	-2.17 (-8.24) ***	-2.71 (-7.89) ***
Number of samples	267	140	267	140
Adj R-squared	0.8401	0.8193	0.8379	0.8134

From left, coefficient, t value, p value

(Note) *Significant level 10%, ** Significant level 5%, *** Significant level 1%

Table 83: Estimated results③

Explanatory variable \ Unexplanatory variable	log (added value)	log (added value)	log (added value)	log (added value)
log (Tangible fixed assets＋Intangible fixed assets)	0.44 (10.18) ***	0.32 (5.42) ***	0.53 (9.36) ***	0.41 (5.17) ***
log (Total number of employees)	0.49 (8.68) ***	0.58 (8.11) ***	0.42 (5.61) ***	0.49 (4.96) ***
log (Total data amount x data utilization degree)	0.05 (2.77) ***			
log (Total number of data cases x data utilization degree)		0.06 (2.38) **		
log (Total data amount x data diversity)			0.04 (2.06) **	
log (Total number of data cases x data diversity)				0.06 (2.67) ***
Industry Dummy (Manufacturing)	-0.07 (-0.60)	0.07 (0.38)	-0.04 (-0.29)	0.08 (0.40)
Constant	-1.96 (-7.15) ***	-2.64 (-7.93) ***	-1.92 (-5.51) ***	-2.52 (-5.47) ***
Number of samples	261	137	173	85
Adj R-squared	0.8334	0.815	0.8537	0.8576

From left, coefficient, t value, p value

(Note) *Significant level 10%, ** Significant level 5%, *** Significant level 1%