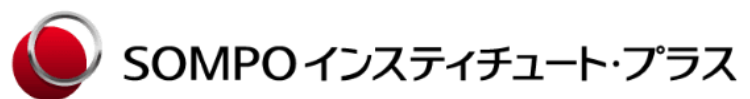# Alternative Data Factbook

January 2024

Introduction

The 2023 Alternative Data Factbook (hereinafter referred to as the Factbook) is organised by Sompo Institute Plus on behalf of the Japan Alternative Data Accelerator Association (JADAA). The Factbook was first published in 2022 to present the results of a survey of JADAA members on (i) the current use of alternative data, (ii) the future outlook of the data, and (iii) the challenges associated with the use of the data. This 2023 Factbook therefore presents the results of the second year of the survey. I would like to take this opportunity to thank the members for their cooperation and other JADAA stakeholders for their support in planning and conducting the survey.

To the best of our knowledge, this Factbook is a unique survey for alternative data in Japan, as no other surveys are regularly conducted or published. One of the features of this survey is that the results reflect the diversity of JADAA members, aggregated and sorted by each position: (i) firms that sell/provide data (sell-side), (ii) firms that buy/use data (buy-side), and (iii) fiduciaries who work between these two sides to conduct analysis. I believe that the Factbook will provide useful information not only to JADAA members, but also to a wide range of people interested in the use of alternative data. We would be grateful for continued support and honest opinions and requests regarding the survey questions and answers.

In the survey questions of the 2023 Factbook, while ensuring consistency with the 2022 Factbook, we have changed the format of some questions and added new questions. The details of these changes are presented in this Factbook and in the Digest. In this introduction, I would like to highlight two major changes relating to awareness of JADAA and Sompo Institute Plus issues.

First, for the Factbook 2023, we added questions that asked "whether the need for the use of alternative data was declining" and "whether the need of the data was growing in any fields" under the circumstances that Japan's economy and society finally began to overcome the negative impact of the COVID-19 pandemic.

In the introduction to Factbook 2022, I noted that the use of alternative data could reach a "tipping point" both domestically and overseas, pointing to a reduced need for high-frequency data after the pandemic ends. For example, the flow of people (the number of people in a given space and time), a type of high-frequency data, served as an excellent nowcasting (the prediction of the current situation) indicator of service consumption and played an important role in the real-time assessment of the economy under greater short-term fluctuations during the pandemic. However, with the shift in economic concerns from the pandemic to the impact of inflation, the benefits of fast (i.e. frequently published) data are not as great as before.

In addition, people flow data was representative of alternative data in the sense that it was easiest to understand and felt familiar to people. However, after COVID-19 was reclassified as Class V in May 2023, people flow information was much less frequently mentioned in daily news reports. In terms of promoting public awareness and understanding of alternative data, we have lost an effective communication tool with the end of the pandemic.

In Factbook 2022, on the other hand, I emphasized three points: (i) the alternative data market in Japan is still in its infancy, and its use and scale are expected to grow; (ii) high-frequency data can be used not only in pandemic situations but also for other purposes, such as quickly understanding the situation after natural disasters; and (iii) the definition and scope of alternative data are incredibly broad and diverse, such as highly granular data (e.g., massive financial data) and text data (e.g., corporate earnings disclosures). Therefore, I insisted that an inflection point is a good opportunity.

Let's take a quick look at the Factbook 2023 findings on changing needs for alternative data. The results show both negative and positive views; the decreasing need in the post-pandemic world and the increasing need for new purposes. Interestingly, the proportion of positive views was much higher than I had expected. While around a quarter of respondents said that the need to use alternative data was generally decreasing, less than 10% said that the need was decreasing in their organisation.

In the open-ended responses about areas where the use of alternative data is actually growing, respondents gave a variety of examples, ranging from highly granular data, web data, and information gathering through web scraping. Other responses pointed to financial analysis to find signs of market volatility, property rental, the use of map data and AI areas including GPT. Even if the data is not yet widely used across society, these responses suggest that new ways of using alternative data are beginning to emerge.

The second question added to Factbook 2023 is to ask for solutions and directions to the questions using alternative data.

In terms of problems in using the data, the responses are similar to those in the previous survey, such as high costs (user fees), difficulties in ensuring data reliability and continuity, lack of in-house expertise and an inadequate legal framework. Given that these issues cannot be resolved in the short term, this result was not surprising. Hence, in this survey, we added a new multiple-choice question which asks "what actions would be needed to resolve those issues and further promote the use of alternative data".

The survey results show that the most common answers to this question were "Setting voluntary rules" and "Identifying users' need" (50% each). On the other hand, only 25% selected the answer "Seeking the government's action," which was somewhat surprising. These results might imply that companies and industry associations like us still have numerous tasks to be solved before they approaching national and/or local governments to set rules for digital markets and to promote data circulation. Furthermore, a relatively large proportion (33%) of respondents selected the answer "Learning from precedents in other countries," which claims the importance of international communication, particularly with North American and European countries that are advanced in the use of alternative data.

As a user, I feel that as the usefulness of alternative data has been increasingly recognised in recent years, more

and more companies are planning not to share the data widely with the public, but to strategically monopolise it in their own businesses. There is nothing wrong with this idea or practice in a free market economy based on competition. However, if every company only pursues its own profit, it may not be possible to drive digital transformation and alternative data use. We are probably at a stage where we need to think deeply about how to continuously promote the use of alternative data across industries and the country as a whole. I hope this Factbook will help to find the answer.

Seisaku Kameda, Principal and Executive Economist,
Sompo Institute Plus Inc.

1. About This Survey and the Japan Alternative Data Accelerator Association

(1) What is alternative data?

Alternative data is a collective term used to distinguish data from traditional data, such as official government statistics. It refers to diverse data that has become available to financial institutions, private companies and economists as a result of advances in digital transformation. Typical examples of alternative data include point-of-sale data, financial news, weather information and location information. This data has been largely unavailable in the past, but is now attracting a lot of attention thanks to advances in machine learning and natural language processing technologies, as well as computing power. In the past, alternative data was mainly used by financial institutions and investors seeking to manage their assets efficiently, but in recent years its use has expanded rapidly. For example, alternative data is used by private companies for business development and by public institutions to analyse economic conditions.

Figure 1-1: Alternative Data and Traditional Data
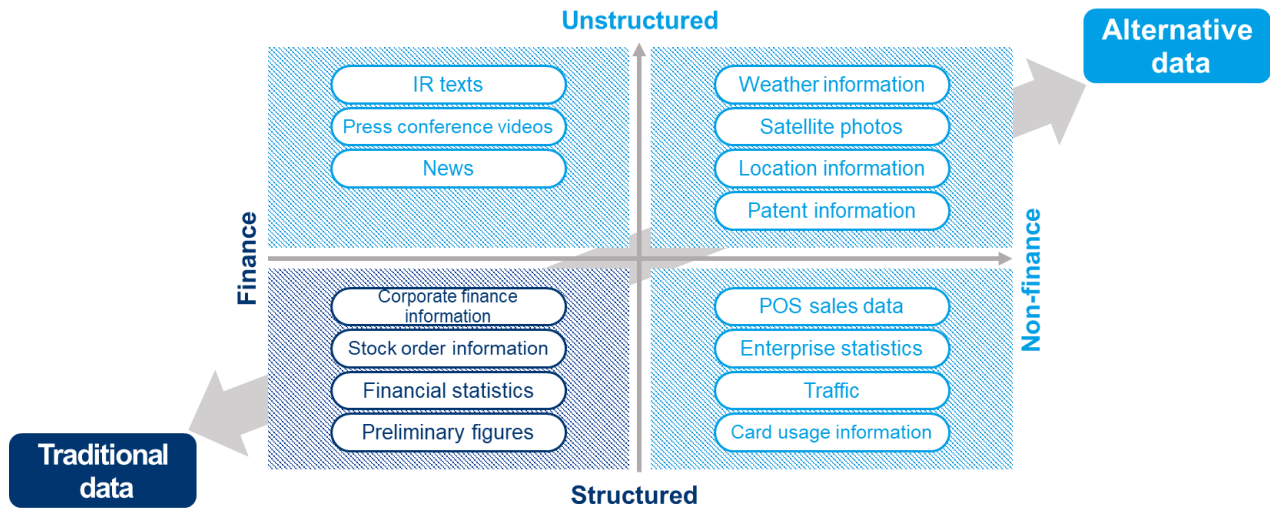
(2) About the Japan Alternative Data Accelerator Association (JADAA)

Although the use of alternative data has been expanding around the world, its use in Japan is still in the middle of the road due to several issues such as regulations, lack of experts and cost. The Japan Alternative Data Accelerator Association (JADAA) was established in 2021 to address these issues by bringing together companies from different industries. Currently, participants such as financial institutions, data providers, and data analytics companies share their activities and challenges to promote the use of alternative data.

(3) About the survey in Factbook

Overseas countries are building up knowledge and insights on alternative data, and their research and studies on data use are also progressing. On the other hand, in Japan, there is not enough research on the use of alternative data, such as "who uses alternative data", "what type of data is used", or "how often the data is used".

In this survey, as in the previous year, we asked members of JADAA to complete online questionnaires to understand the use of alternative data and its issues in Japan. On behalf of JADAA, Sompo Institute Plus compiled the results into this Factbook 2023. Although this survey may not cover the overall situation of alternative data, we believe that our results capture accurate information about the current situation and issues of alternative data,

as the respondents were companies that already use or are very interested in using alternative data. We plan to continue this survey in the future to regularly monitor trends in the use of alternative data in Japan.

(4) Survey respondents

We asked JADAA members to complete questionnaires, and a total of 54 companies responded. The table below shows the respondents sorted by business category.

Figure 1-2: Survey Respondents

| Business	Number of Respondents	Ratio
Asset management	3	6%
Finance and insurance (excluding asset management)	7	13%
Manufacturing	3	6%
Think tank	8	15%
Education	1	2%
Information technology/System development	24	44%
Other	8	15%
Total	54	100%

(5) Survey data collection period

The survey collected data from respondents between 7 August and 15 September 2023.

2. Information on the use of alternative data

(1) Experience in using alternative data

Figure 2-1-1 shows that 78% answered "Yes" to the question that asked whether they have experience of using alternative data. The respondents who answered "Yes" have "Purchased" (29%), "Sold" (48%), or "Analyzed as a contractor" (24%).

Figure 2-1-1: Have you used alternative data?

n = 54   n = 42
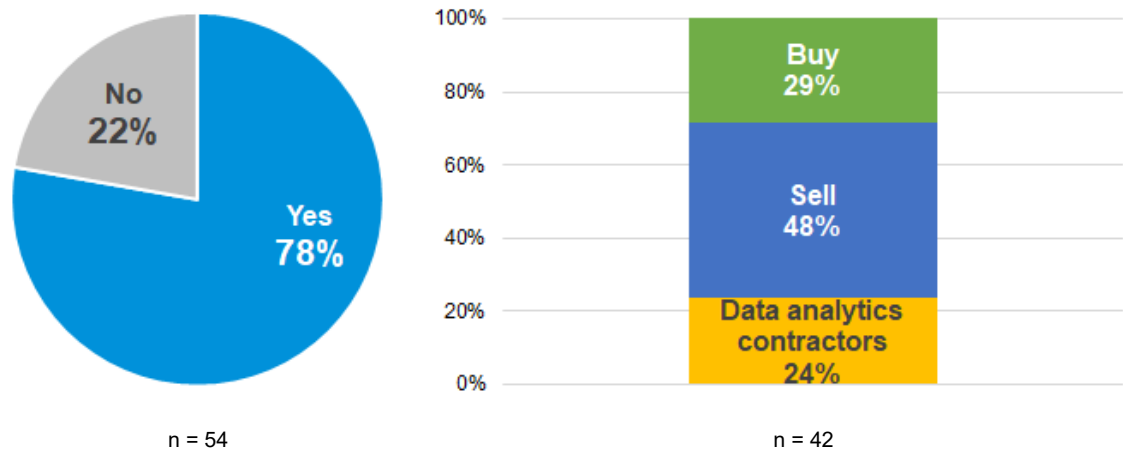
In the following section, we define respondents who have purchased alternative data as "buyers," those who have sold data as "data providers," those who have analyzed data as contractors as "data analytics contractors," and those who have never used any data as "non-users." Figure 2-1-2 shows the relationships between these players.

Figure 2-1-2: Conceptual Image of Relationships between Key Actors
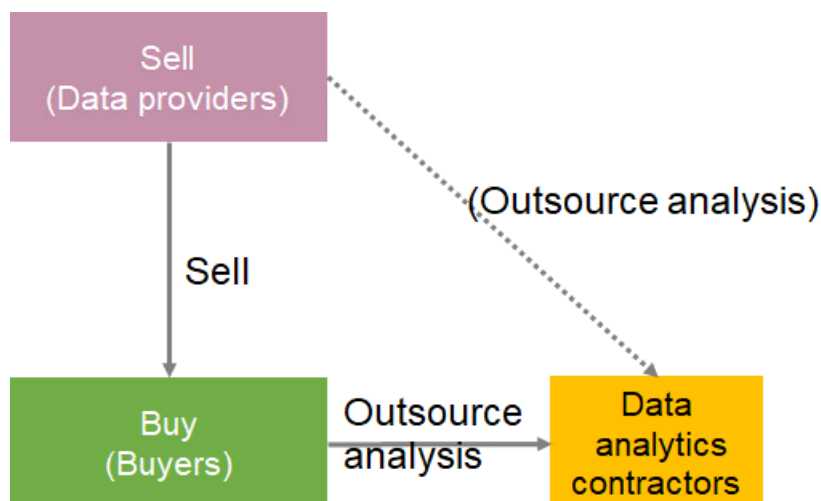
(2) Spend on alternative data compared to previous year (buyers)

Figure 2-2 shows the result to the question that asked buyers how their spending on alternative data have changed over the year. In this question, 25% answered "Significantly increased," and 33% "Increased." Compared to the survey in 2022, where all respondents answered increase ("Significantly increased (50%)" and "Increased (50%)"), although the pace of increase slowed down, increasing trend itself was continuing.

Figure 2-2: How has your spend on alternative data changed compared to previous year?

2022 survey
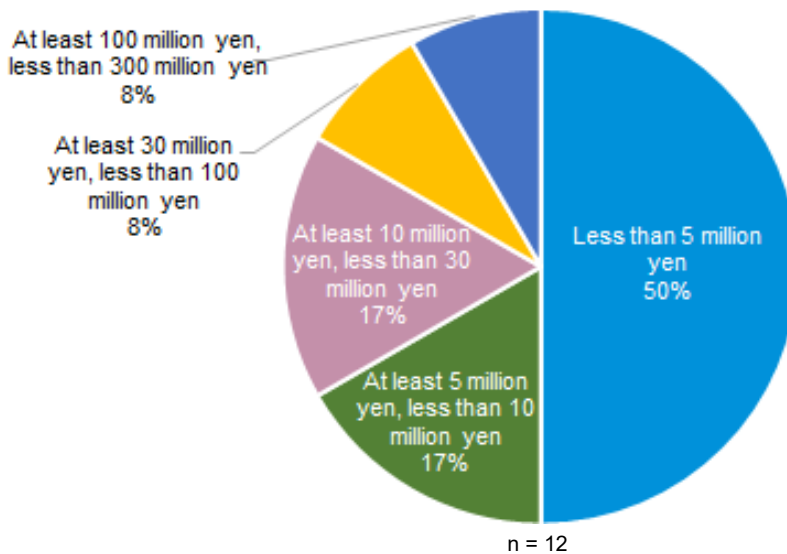
2023 survey

n = 10

n = 12
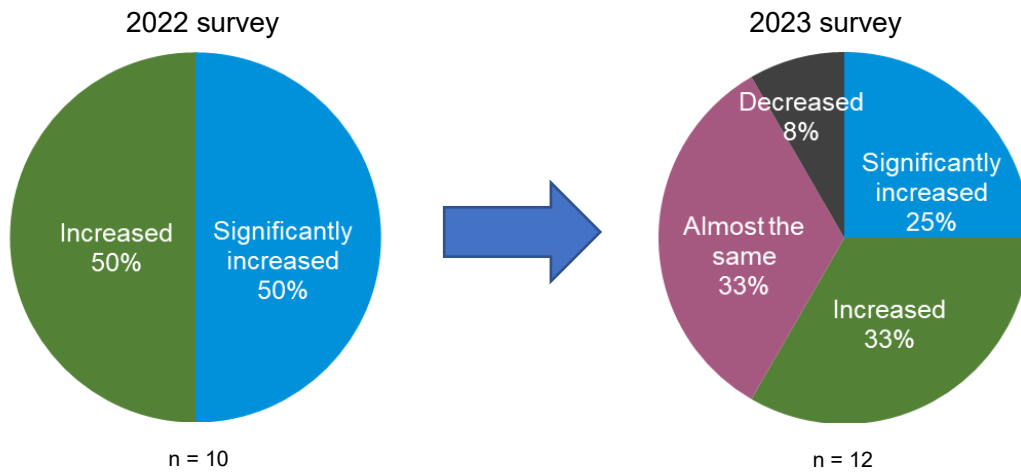
(3) Range of annual spending on alternative data

Figure 2-3 shows the result to the question that asked buyers about the range of their annual spending on alternative data. According to the figure, 50% answered "Less than 5 million yen," and 17% answered "At least 5 million yen, less than 10 million yen" and "At least 10 million yen, less than 30 million yen" each. Respondents who answered that they spent "30 million yen or more" are less than 20%.

Considering the unit prices of alternative data, this result suggests that only a limited number of companies use the data extensively and most buyers use the data to a limited extent.

Figure 2-3: How much do you spend on alternative data per year? Please select a range.

n = 12

7

(4) Start date of using alternative data

Figure 2-4 shows result to the question that asked about the timing they had started using alternative data. The figure shows that the most common answer among all respondents was "2017 or earlier" (45%). However, those who started in 2018 or later also made up large percentages, indicating that demand for alternative data grew over the pandemic years.

The answers sorted by player (buyers, data providers, data analytics contractors) are also in line with the overall trends, whereas data providers began to use the data earlier than buyers.

Figure 2-4: When did you start using alternative data?

(5) Cross-tabulation: Types of data and start of use

We next conducted cross-tabulation of responses from buyers, data providers, and data analytics contractors, based on the timing they began to use alternative data and types of data they use.

According to Figure 2-5, respondents who have used alternative data before the pandemic (in or before 2019) turned out to use "POS data" (40%) and "Location data" (32%) than other groups of respondents.

On the other hand, respondents who began to use alternative data after the pandemic (in or after 2020) commonly use "Web scraping data" (65%) and "Location data" (41%). These results revealed that the types of alternative data used vary according to when the use began.

Figure 2-5: Types of data used and start of use (multiple choice)

(6) Purpose of using alternative data

Figure 2-6 shows the result to the question that asked buyers the purpose of using alternative data. The result shows that the majority answered "Investment decisions/economic forecasting" (92%), followed by "Academic work" (33%), "Risk management" (17%), and "Marketing" (8%).

Figure 2-6: What do you use alternative data for? (multiple choice question, up to 3 answers)


n = 12
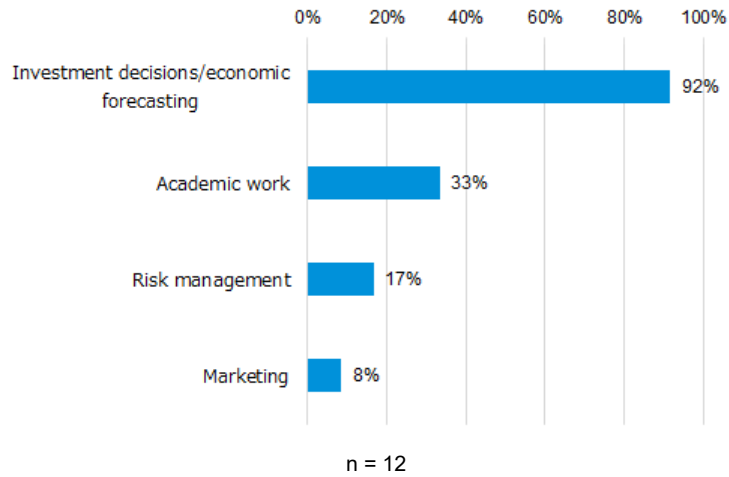
(7) Benefits of using alternative data

The benefits of using alternative data include instant availability that allows users to check real-time data, and more extensive coverage than government statistics. According to Figure 2-7, 52% of all respondents selected "Differentiation from traditional data," 48% answered "Complementarity to traditional data," and 43% "Instant availability."

The result are not quite the same when responses were sorted by players. The most common answers were "Differentiation from traditional data" (67%) and "Usability as a study subject" (50%) among buyers, whereas it is "Instant availability" (60%) among data providers.

Figure 2-7: What is/are the benefit(s) of using alternative data? (multiple choice question, up to 3 answers)

n = 42   n = 12   n = 20   n = 10

9
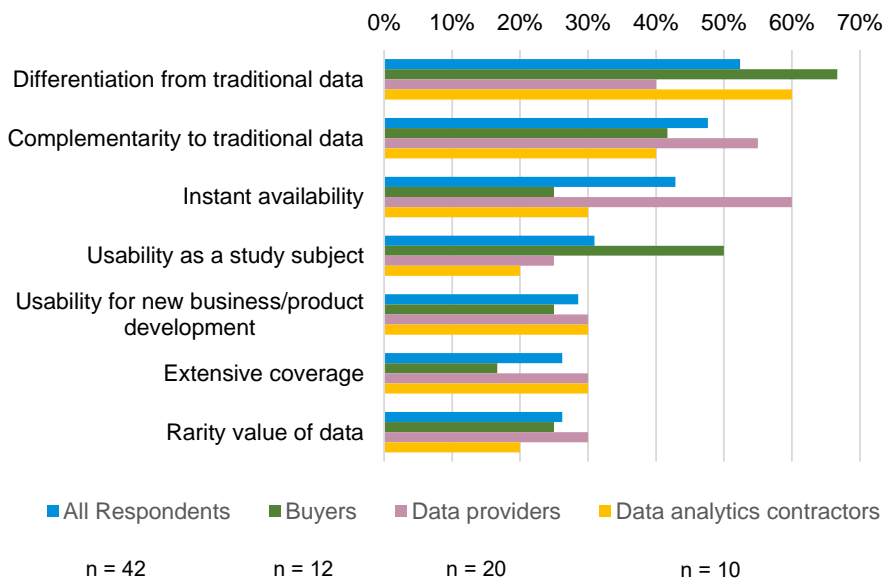
(8) Most frequently used alternative data

Figure 2-8 shows the result to the question that asked buyers the type of alternative data they most frequently use.

The most frequently used data is "Web scraping data" (33%). Some listed "POS data" (17%), "Location data" (17%). "News data" (17%), and "Credit card data" (8%).

Figure 2-8: Which type of alternative data do you use most frequently?

n = 12
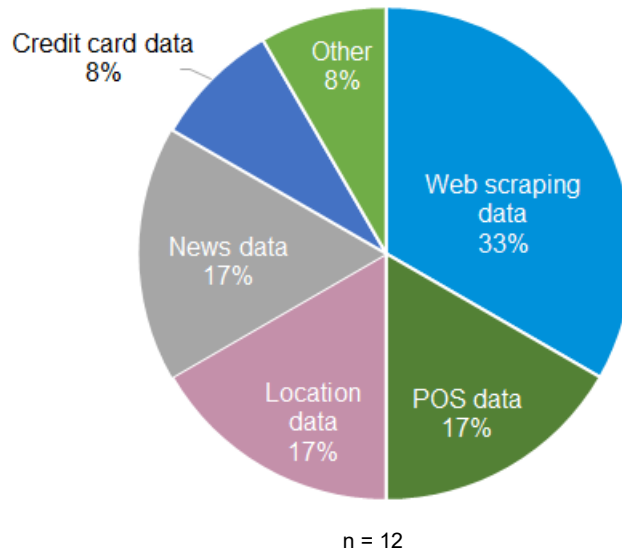
(9) Number of employees dedicated to alternative data management

We next asked buyers, data providers, and data analytics contractors about the number of employees who are dedicated to alternative data management.

Figure 2-9 shows that the most common answer was "0" (29%), followed by "1 to 2" (21%) and "3 to 5" (21%) for all respondents.

When we sort answers by key players, an interesting polarization was observed. Among buyers, 58% answered "0" and 25% "1 to 2," that is, more than 80% buyers dedicated up to only two staff members. In contrast, lots of data providers and data analytics contractors dedicated more staff members. These result suggest that adequate systems for alternative data management are not yet developed, mainly among buyers who handle the data as users.

Figure 2-9: How many of your employees are dedicated to alternative data management?
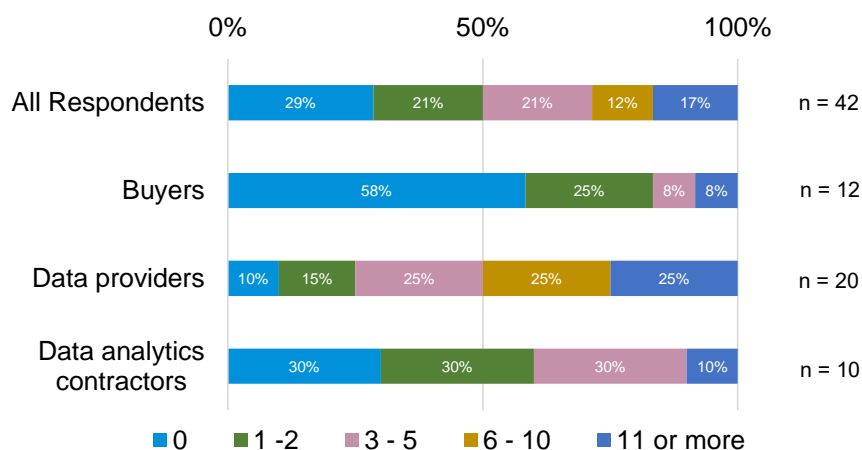
(10) Preprocessing of alternative data

We also asked buyers the preprocessing of alternative data. Figure 2-10 shows that 33% answered "We purchase raw data and preprocess internally," 17% "We purchase already preprocessed data," and 8% "We purchase raw data and outsource to an external contractor for preprocessing." On the other hand, the most common answer was "All of the above" (42%), indicating that the majority of respondents preprocess alternative data either by themselves or by outsourcing.

To the question that asked the part of preprocessing, the majority answered "Remove outliers" (89%), "Remove duplicate data" (89%), and "Convert text data to numbers" (78%).

Figure 2-10: Alternative Data Preprocessing

How do you preprocess alternative data?
Which part(s) of data do you preprocess? (multiple choice question)
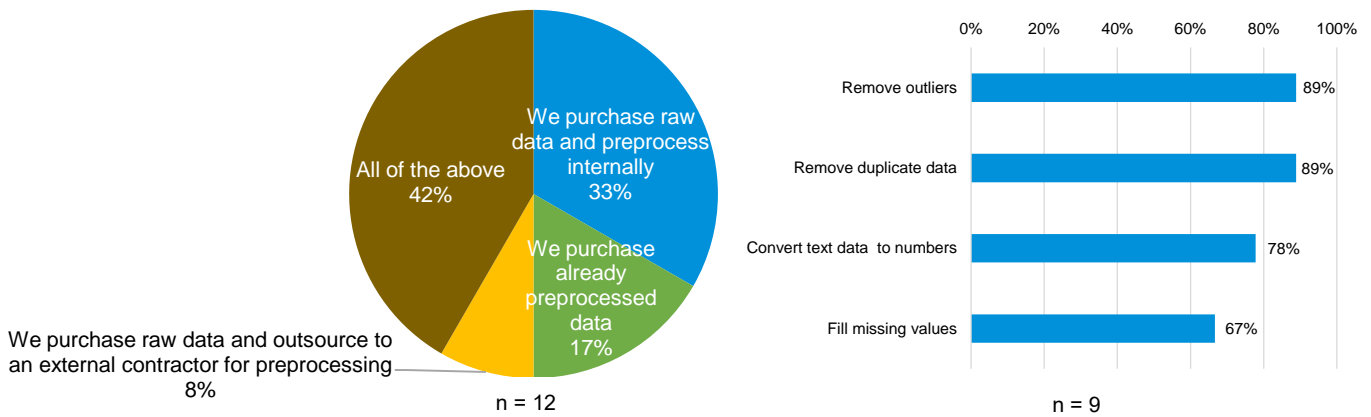
(11) Internal sharing of analyses using alternative data

We asked buyers how they share the results of analyses using alternative data internally.

Figure 2-11 shows that The most common answer in 2023 was "All the way up to the management" (75%), followed by "The team that uses the data on the ground" (17%) and "All the way up to managers of the user department and other relevant departments" (8%).

Compared to the previous survey in 2022, share of "All the way up to the management," has increased, which implies that result of analysis using alternative data have more commonly shared with top management.

Figure 2-11: Where are alternative data and results of analyses of those data shared within your company?
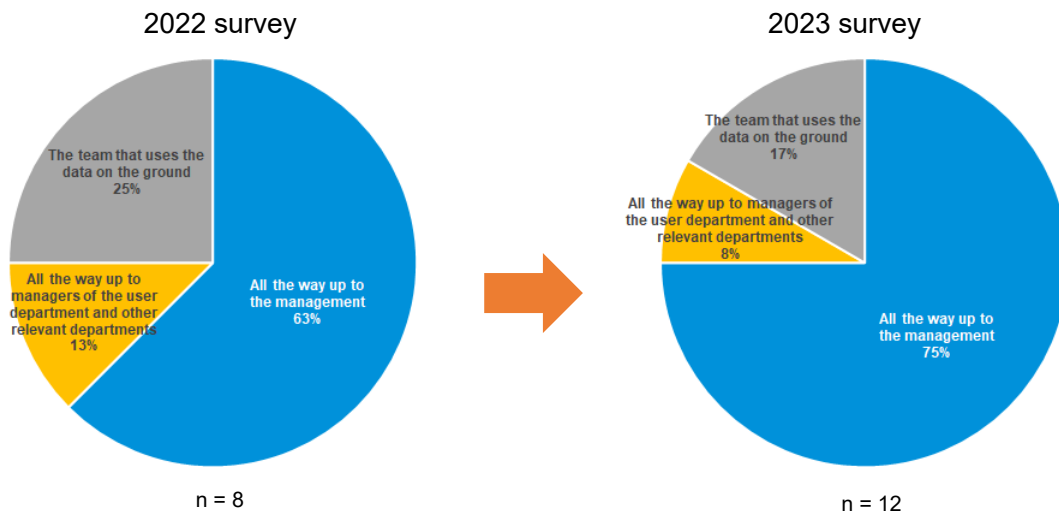
3. Information on how and what type of alternative data is traded

(1) Channels of purchasing/selling alternative data

In section 3 we first asked buyers and data providers about the channels they use to buy/sell alternative data.

Figure 3-1 indicates that the most common answer was "Direct transactions with data holders," in both buyers and data providers. Lots of buyers also answered "Purchase from vendors." For both type of respondents, only limited proportions answered that they transacted via platforms.

Figure 3-1: What purchase/sales channel do you use to buy/sell alternative data?
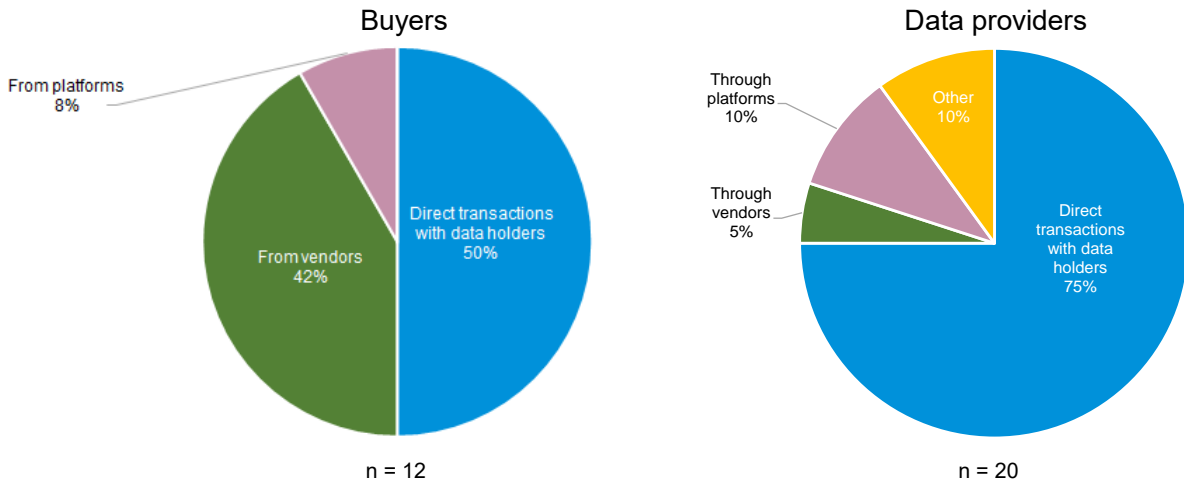
(2) Regions of alternative data

We next asked buyers the region(s) that the purchased alternative data covers. Figure 3-2 shows that the most common answer was Japan (100%). For overseas, North America (42%), Europe (33%), China (25%), and other Asian countries (25%) were listed, which implies that it is likely that purchased alternative data are mostly those on major economies.

Figure 3-2: Which of the region(s) does the alternative data you buy cover? (multiple choice question)
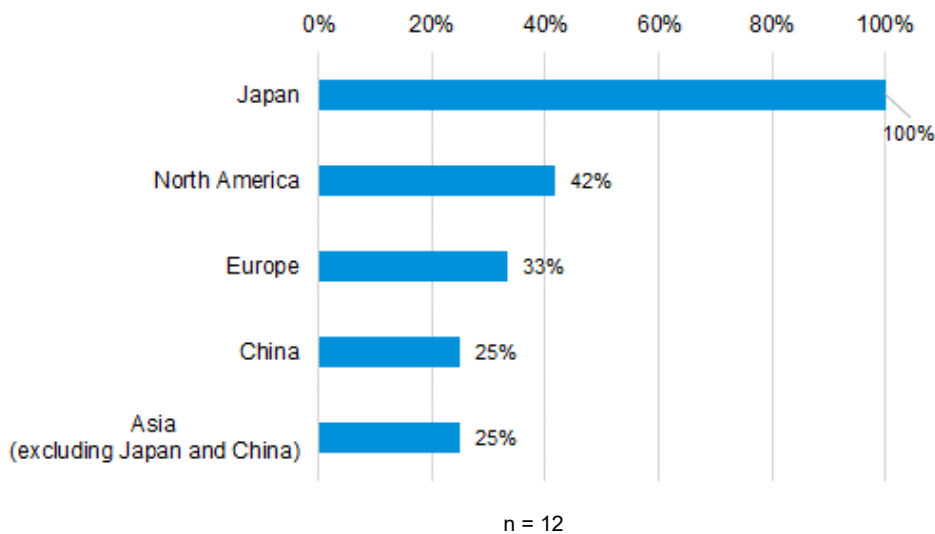
(3) Sectors of alternative data

We also asked buyers the sector(s) that the purchased alternative data covers. The most common answer turned out to be "General consumer goods" (64%), followed by "Healthcare" (36%), "Finance" (36%), "Essentials for everyday living" (36%), and "Real estate" (36%).

Figure 3-3: Which of the sector(s) does the alternative data you buy covers? (multiple choice question)
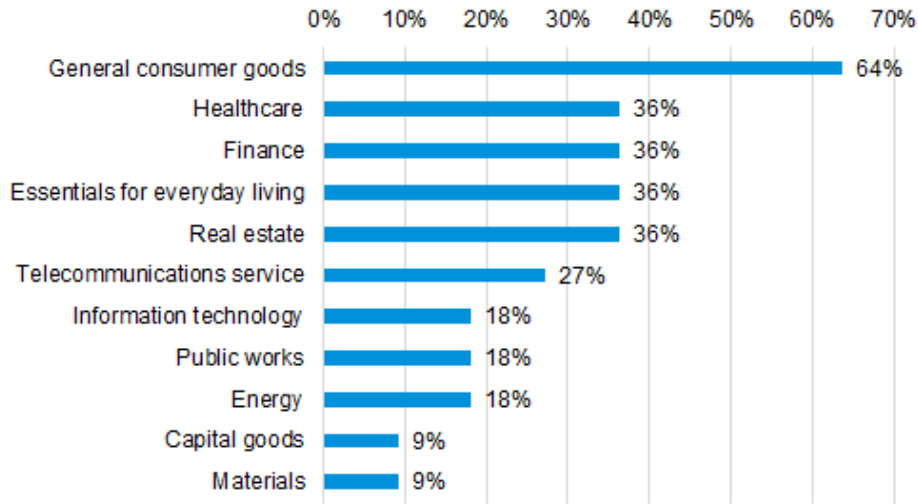
n = 12

(4) Types of alternative data handling

We next asked buyers, data providers, and data analytics contractors which type(s) of alternative data they are handling.
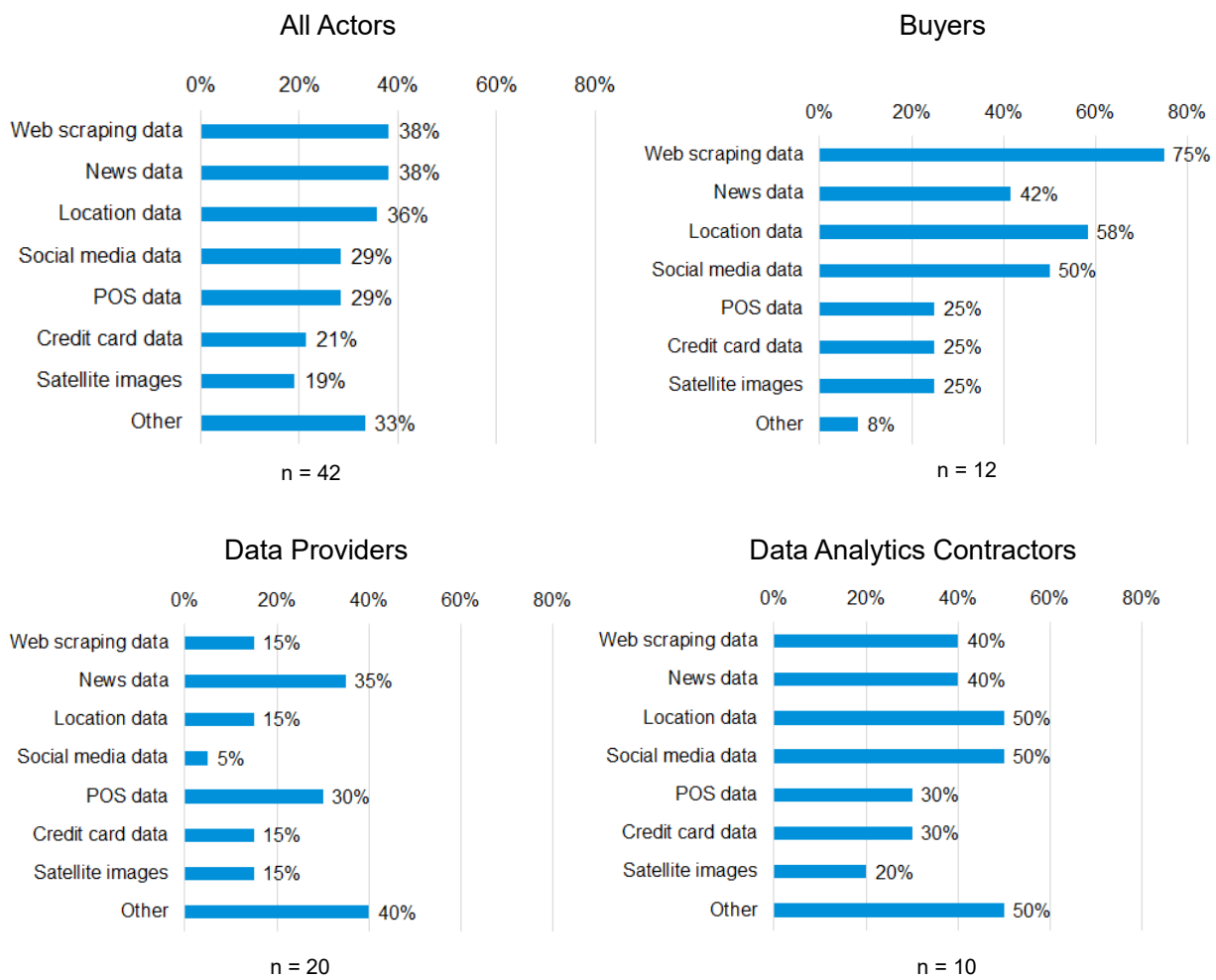
The most popular data for all respondents were "Web scraping data" (38%), followed by "News data" (38%), "Location data" (36%), "Social media data" (29%), and "POS data" (29%).

When we compared answers by players, we find that the data the three actors use are significantly different.

Buyers commonly use "Web scraping data" (75%), "Location data" (58%), and "Social media data" (50%). Data providers commonly use "News data" (35%) and "POS data" (30%). Data analytics contractors commonly use "Location data" (50%), "Social media data" (50%), "Web scraping data" (40%) and "News data" (40%).

In addition to these popular data, we find that various types of alternative data are used. Other types of data listed are finance data, maritime transport data, patent-related data, receipt data, and ESG data.

Figure 3-4: What type(s) of alternative data do you use? (multiple choice questions, up to 3 answers)

All Actors

Type	%
Web scraping data	38%
News data	38%
Location data	36%
Social media data	29%
POS data	29%
Credit card data	21%
Satellite images	19%
Other	33%

n = 42

Buyers

Type	%
Web scraping data	75%
News data	42%
Location data	58%
Social media data	50%
POS data	25%
Credit card data	25%
Satellite images	25%
Other	8%

n = 12

Data Providers

Type	%
Web scraping data	15%
News data	35%
Location data	15%
Social media data	5%
POS data	30%
Credit card data	15%
Satellite images	15%
Other	40%

n = 20

Data Analytics Contractors

Type	%
Web scraping data	40%
News data	40%
Location data	50%
Social media data	50%
POS data	30%
Credit card data	30%
Satellite images	20%
Other	50%

n = 10

(5) Customers of alternative data providers

We finally asked data providers to whom they sell alternative data. The most common answer was "Asset management companies" (80%). Other answers included "Administrative agencies" (15%) and "Government agencies" (10%).

Figure 3-5: Who buys the alternative data you offer? (multiple choice question)

Customer	%
Asset management	80%
Administrative agencies	15%
Government agencies	10%
Other	10%

n = 20

14
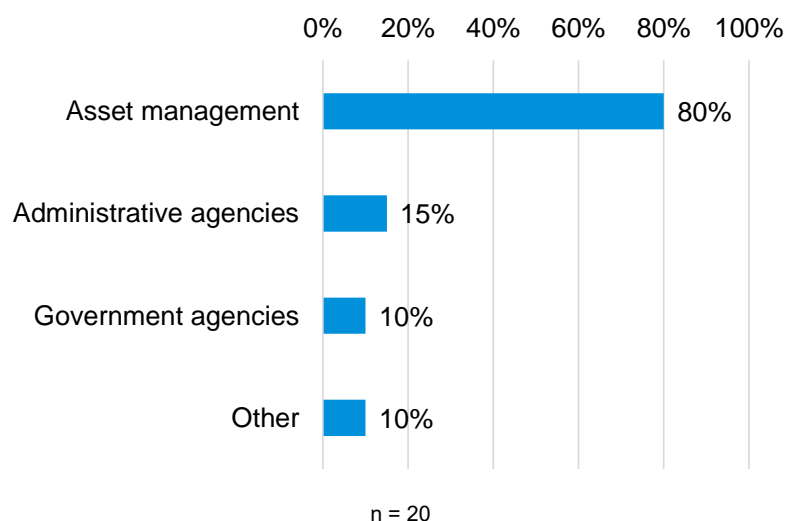
## 4. Regulations and Issues

(1) Regulations that need improvement to promote the use of alternative data

In Section 4 we asked buyers, data providers, and data analytics contractors about the regulations that need improvement to promote the use of alternative data.

Figure 4-1 shows that the common answers from all respondents were the "Act on the Protection of Personal Information" (57%), "Financial Instruments and Exchange Act" (40%), and "Copyright Act" (38%).

The results were all in all similar in all types of players. Showing detailed difference, data providers and data analytics contractors consider that the Act on the Protection of Personal Information are in particular needed to improve.

Figure 4-1: Which of the established laws or regulations need(s) improvement to promote the use of alternative data? (multiple choice question)
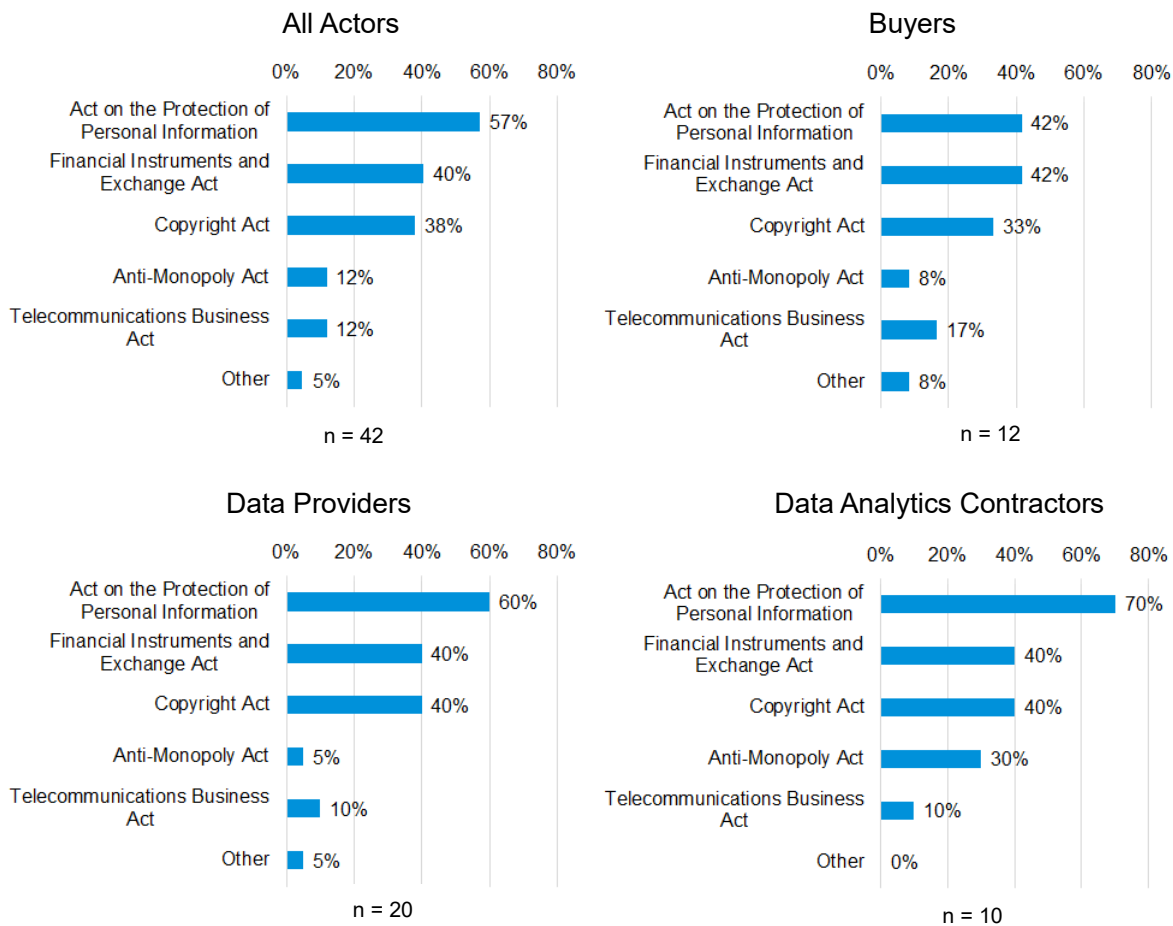
(2) Cross-tabulation: Regulations that need improvement to facilitate the use of alternative data (responses sorted by data used)

We then cross-tabulated the responses, sorting by the type of alternative data they handle and the regulations that need improvement to promote the use of alternative data.

Figure 4-2 clearly shows that the "Act on the Protection of Personal Information" is all in all believed to need improvement, regardless of the type of data handled. For the "Copyright Act", answers varies according to the type of data used. It is commonly believed to need improvement by users of "POS data" (67%), "Satellite data" (50%), and "News data" (50%), whereas the proportion were lower among users of "Web scraping data" (42%), "Social media data" (42%), and "Location data" (33%). As for the "Anti-Monopoly Act" and the "Telecommunications Business Act", smaller proportions of users answered that these laws need improvement compared to the other laws, regardless of type of data handled.

Figure 4-2: Regulations that need improvement (sorted by type of data used)

	Act on the Protection of Personal Information	Copyright Act	Financial Instruments and Exchange Act	Anti-Monopoly Act	Telecommunications Business Act	n
POS	75%	67%	33%	25%	8%	12
Credit card	89%	44%	44%	11%	11%	9
Location information	73%	33%	33%	13%	27%	15
Satellite images	88%	50%	38%	25%	0%	8
Social media	58%	42%	33%	17%	17%	12
Web scraping	58%	42%	33%	17%	17%	16
News	63%	50%	50%	6%	0%	16

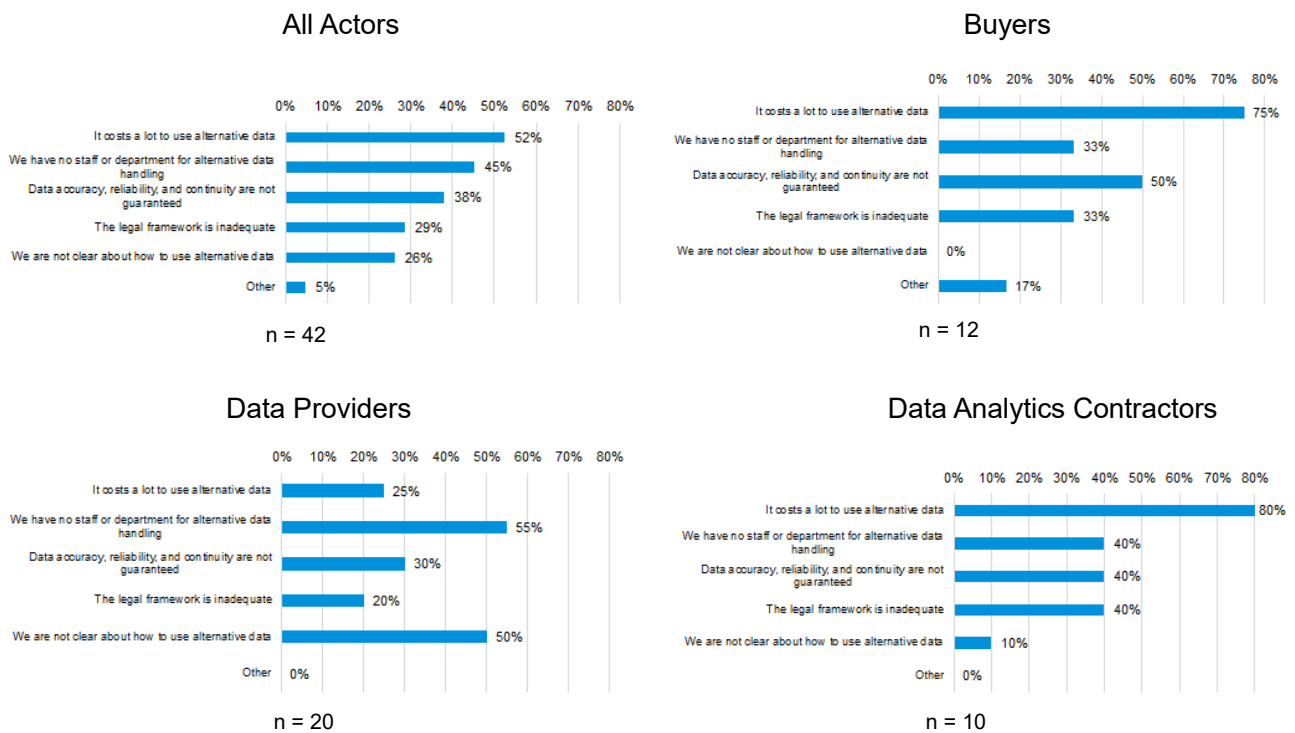## (3) Issues involved in the use of alternative data

We also asked buyers, data providers, and data analytics contractors about the issues involved in the use of alternative data. For all respondents, common answers were "It costs a lot to use alternative data" (52%), followed by "We have no staff or department for alternative data handling" (45%), "Data accuracy, reliability, and continuity are not guaranteed" (38%), and "The legal framework is inadequate" (29%).

For buyers, answers were "It costs a lot to use alternative data" (75%) and "Data accuracy, reliability, and continuity are not guaranteed" (50%).

For data providers, answers were "We have no staff or department for alternative data handling" (55%) and "We are not clear about how to use alternative data" (50%).

For data analytics contractors, the most common answer was "It costs a lot to use alternative data" (80%), followed by "We have no staff or department for alternative data handling" (40%), "Data accuracy, reliability, and continuity are not guaranteed" (40%), and "The legal framework is inadequate" (40%).

Figure 4-3: What are the issues involved in using alternative data?
(multiple choice question, up to 3 answers)

All Actors
n = 42

Buyers
n = 12

Data Providers
n = 20

Data Analytics Contractors
n = 10

In their free-form answers, respondents also mentioned the issues shown in the table below.

Buyers	We sometimes hesitate to start using alternative data because of uncertainty regarding regulations.
	We are unable to take enough time to analyze alternative data.
	Representativeness seems to be a major issue.
	The safety of data usage is not guaranteed.
	It is difficult to convincingly present the cost-effectiveness of alternative data to the management.
	We cannot determine if a price of alternative data is reasonable because there are no specific valuation standards for volumes and quality of alternative data.
	We do not have enough data scientists.
	We are unable to set up a team dedicated to alternative data handling.
	We have not built up enough data to identify patterns in statistics and make seasonal adjustments.
	Issues related to budgets (e.g., rising cost) are often deterrents
	It is difficult to give a convincing explanation for a considerable cost before the benefits of using alternative data become visible.
Data providers	Industries and companies need to launch their own activities to advance users' understanding.
	We need more employees on the staff for data handling.
	Natural language processing and other relevant techniques need to be more common.
	We need more employees capable of handling finance-related work.
	Japan's Copyright Act puts data vendors at a disadvantage, discouraging selling data.
	Ensuring data accuracy is a challenge, and legal risk is obscure.
	It is difficult to convincingly present to customers exactly what value the use of alternative data offers.
	Users have not acquired enough knowledge about the new type of data.
	There are no established standards or third-party institution for quality assessment.
	Reluctance to use a new type of data; cost-effectiveness of the data
	With the sensitive nature of alternative data, the availability of the data needs to be limited.
Data analytics contractors	Given the difficulty in clearly presenting the usefulness of alternative data, it is difficult to obtain approval for purchasing a new alternative data set.
	Customers sometimes abandon their plan to use alternative data because they do not see the cost-effectiveness of the data or the data does not worth the cost.
	The lead time for the use of data is long.
	There are many tasks to ensure that none of the rights involved in deliverables are violated.
	The accuracy of location information varies depending on the type of data.
	It takes time and effort to calculate cost-effectiveness, which often prevents us from making a start, even in the stage of verification.
	The data we want and laws related thereto are undeveloped.
	Expensive queries are a deterrent to analyses that use multiple data sets.
	Alternative data handled at the university level is limited in Japan compared to other countries, which means alternative data experts are hardly produced.
	The tough restrictions imposed by the Act on Protection of Personal Information remain a barrier.

(4) Actions required to resolve problems

As a special question in Factbook 2023, we asked buyers about the actions needed to address the issues in the previous question.

As actions for addressing and resolving issues regarding alternative data, the respondents selected "Setting voluntary rules" (50%) and "Identifying what users need (including potential users)" (50%), and "Learning from precedents in other countries" (33%). On the contrary, relatively fewer respondents selected "Seeking the government's action" (25%).

Figure 4-4: which of the actions do you think will be needed to address and resolve the issues? (multiple choice question, up to 3 answers)

n = 12
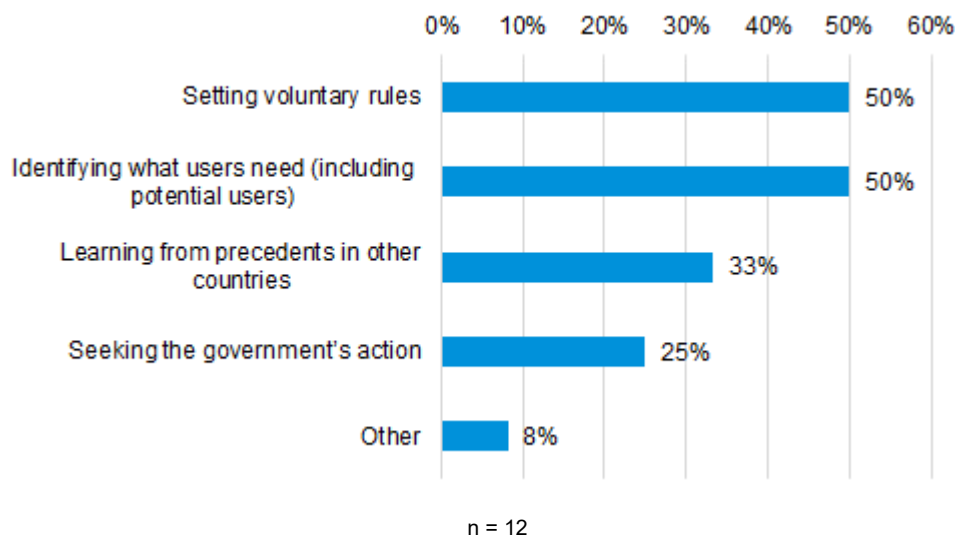
5. Need and Outlook for Alternative Data
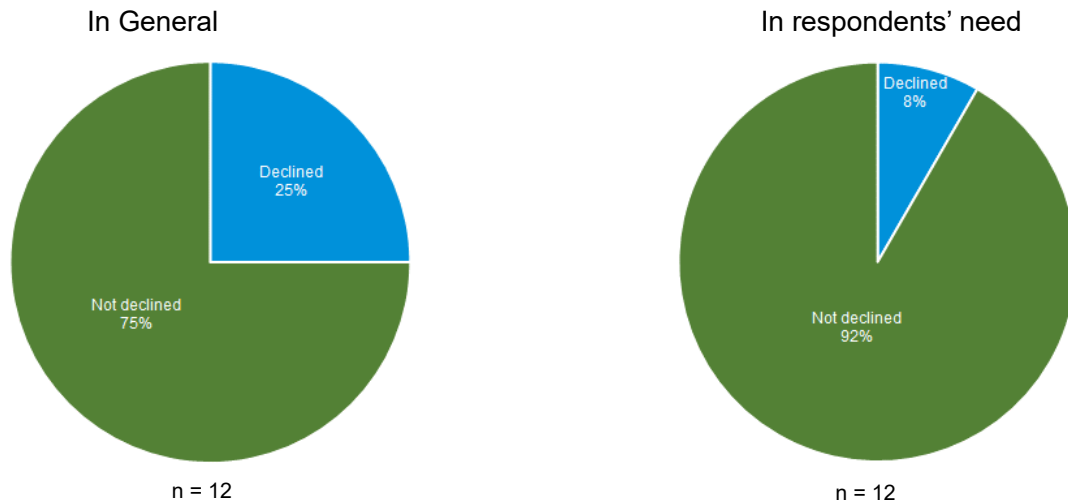
(1) Need for alternative data

In Section 5, as a special question in Factbook 2023, we first asked buyers whether the need for alternative data had changed, both in general and in their companies, following the normalisation from COVID-19.

Figure 5-1 shows that in general, 75% answered that the need of alternative data have not declined. In their company, the rate jumped up to 92%, indicating that the need for alternative data remains strong even after the decrease of emergent demand during the pandemic. The results are notable also in the sense that the need directly linked to respondents' business and activities is perceived to be much higher than the need in general.

In their free-form answers, respondents listed a variety of reasons why their need has not declined. Answers are: "There is strong demand for POS data and web data," "Alternative data is useful for quick observation of business confidence and for diversification of analytical techniques," and "Alternative data is useful for detecting signs of what may happen in financial markets." These responses serve as another corroboration for the view that in addition to the analysis during the pandemic, there are a wide range of fields where alternative data   could take advantage.

On the contrary, as a reason for the answer that the need for alternative data has declined, some respondents stated "There is less need for analysis required to be instantly available."

Figure 5-1: Has the need for alternative data declined in the post-COVID economy? normal?

In General   In respondents' need

n = 12   n = 12

In their free-form answers, respondents also provided the comments below as reasons why their need has/has not declined.

<Why the need has not declined>

| |
|---|
| There is demand for POS data and web data. |
| There is demand for data with high granularity. |
| Alternative data is useful for detecting signs of what may happen in financial markets. |
| Alternative data is useful for quick observation of business confidence and for diversification of analytical techniques. |
| The use of location data and satellite data is increasingly common in financial institutions and government agencies. |
| We use alternative data more often than before, mostly for overseas hedge funds. |
| We receive more inquiries about alternative data from customers than before. |
| There is a growing need for foot traffic data for a business. |
| There is a growing need for the use of alternative data combined with map data. |
| Alternative data is more commonly used in the asset management business and the real estate business than before. |

<Why the need has declined>

There is less need for analysis required to be instantly available.
There is less need because of budgets.
There is less need in connection with the areas of consumption we need to identify.

(2) Outlook for alternative data

In terms of outlook, we asked buyers, data providers and data analytics companies whether they expected the alternative data market to grow in the near future (over the next three years, starting this year).

Figure 5-2 shows that in each group of the players, the majority answered that the market will "grow," which indicates that respondents have high expectations for market growth of alternative data.

Figure 5-2: Do you think the alternative data market will grow in the future (over the next three years, starting from this year)?
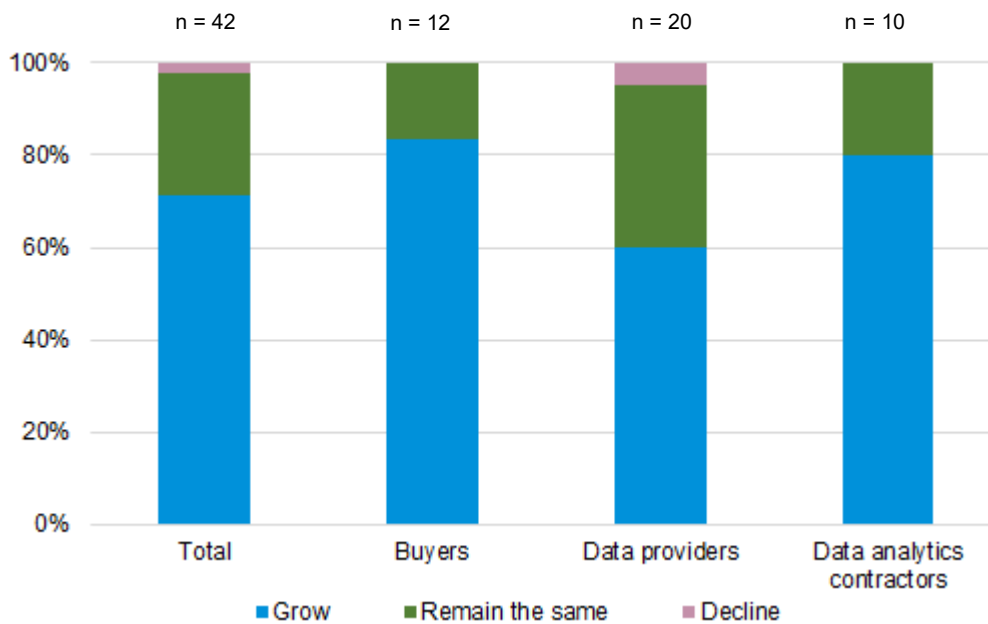
6. Non-users

(1) Reason for not using alternative data

We finally asked non-users of alternative data why they did not use it.

Figure 6-1 shows that the most selected the reason was "We have no staff or department for alternative data handling" (42%), followed by "It costs a lot to use alternative data" (33%), "We are not clear about how to use alternative data" (25%), and "Data accuracy, reliability, and continuity are not guaranteed" (17%).

Figure 6-1: Why do you not use alternative data? (multiple choice question, up to 3 answers)
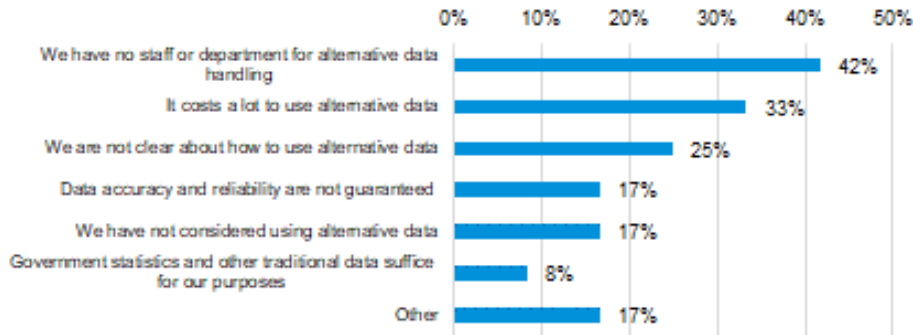
n = 12

In their open-ended responses, respondents also provided the following comments as reasons for not using alternative data.

Alternative data is not directly needed for the analyses we currently do.
We have not hired the talent needed to use alternative data.
We are considering using alternative data someday, but do not have a specific plan yet.
None of our clients have consulted us about the use of alternative data.
We cannot envision what kind of data would be available and what we could do with the data to fulfill our vision.
We do not know exactly what alternative data we should be able to effectively use.
We do not know what services we could offer by combining alternative data with the data we already have.
We have not set up any project that would need alternative data.

(2) Whether or not to use alternative data in the future

We finally asked non-users whether they have plan to use alternative data sometime in the future (the next three years starting from this year). Figure 6-2 shows that 50% of the respondents answered "We plan to start using alternative data," while the other 50% answered "We don't plan to use alternative data."

Figure 6-2: Do you plan to use alternative data sometime in the future
(the next three years, starting from this year)?

We don't plan to use alternative data
50%

We plan to start using alternative data
50%
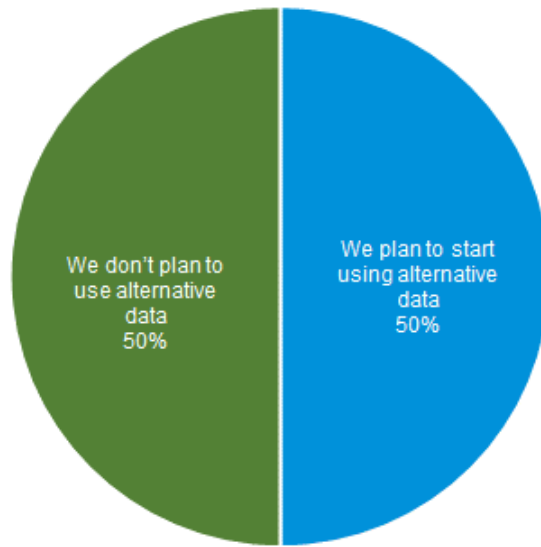
n = 12

7. Conclusion

In this 2023 Factbook, as in the previous year, we have quantitatively assessed the use of alternative data in Japan, based on respondents' answers to the various questions. The conclusion of our survey is as follows.

First of all, our survey result shows that most respondents expect that the demand for alternative data will grow, which is the opposite to the assumption that the demand will decrease after the decline of emergent demand during pandemic. While few respondents showed the view that "There is less need for analysis using instantly available data," lots of respondents answered positively for the needs of alternative data, such as: "Alternative data is useful for nowcasting economic condition and for diversifying tools of analyses," and "It is useful for detecting signs of what may happen in financial markets." These responses indicate that users have come to understand the strengths of alternative data, confirming the need for the data by many companies. Many players have started to use alternative data in recent years, and more than half of respondents said they expect the alternative data market to grow in the next three years. Taken together, we concluded that the alternative data market in Japan is likely to continue to grow.

Second, the data being used is very diverse. The survey results show that both the purpose of using alternative data and the types of data used vary considerably between different types of players, such as buyers, data providers and data analytics companies. We also find that they are using not only typical alternative data (e.g. web scraping data, location data and POS data), but also increasingly diverse types of data, including financial data and maritime data. As each type of alternative data has its own characteristics, users are still experimenting with how best to use them. We expect that as they deepen their understanding of alternative data, these actors will build their knowledge and insights to suit their purposes.

Thirdly, there still remain various unresolved issues that prevent alternative data from widespread use. In our questionnaire, the issues that a lot of respondents selected were: "It costs a lot to use alternative data," "We have no experts or department for alternative data handling," "Data accuracy, reliability, and continuity are not guaranteed," and "The legal framework is inadequate." Some respondents also offered the view that, to resolve these issues, the players should set their own rules and identify users' need, rather than expect the government's action.

In relation to the cost of using alternative data, some respondents pointed out that high cost was not the only issue. They explained that it would be difficult for their companies to make a convincing case for significant costs when the benefits of using alternative data are not yet certain. As the use of alternative data increases and its benefits become more widely known, this perception that the cost of alternative data may outweigh the benefits of using it may disappear. In particular, the cost of analysis is likely to fall gradually as the use of alternative data becomes more widespread.

In terms of staff or experts to handle alternative data, many buyers, data providers and data analytics contractors face the same problem. In particular, buyers have fewer alternative data staff than data providers and data analytics contractors. In addition to the lack of data analytics skills, buyers also face the problem that they don't see the potential benefits of using alternative data, such as the knowledge and insights that can be gained from the analytics. To address these issues, it is important to have opportunities to share use cases widely, coupled with the development of data analytics skills. The weekly workshops offered by JADAA would be useful to address these issues.

Many respondents also pointed out that the accuracy, reliability and continuity of data are not guaranteed. As alternative data are mostly a by-product of the business activities of private enterprises, they tend to contain a certain

amount of bias compared to official government statistics, which are designed to give a clear picture of the economic situation. Therefore, users should first recognise that the alternative data contain bias and then adopt an approach to minimise it. In the medium to long term, this issue will be resolved if alternative data are widely used, so that people have a deeper understanding of the bias and the knowledge is shared among all users.

Some companies were reluctant to use alternative data because they felt that the legal framework was inadequate. There is often a trade-off between the use of data and the protection of personal information. Many respondents said that there should be a better legal framework for the Personal Data Protection Act. The JADAA has a Due Diligence Questionnaire (DDQ) available to those who offer their data so that they can effectively check for potential breaches of the law and other risks. It plans to work on its proposal for a legal framework and to develop guidelines.

While it is almost universally accepted that alternative data are useful, there are still a number of issues surrounding their use. In order to promote the use of alternative data more widely, it will be essential to resolve these issues by setting voluntary rules and understanding the needs of users. We could learn from other countries where alternative data is widely used.

As mentioned above, the use of alternative data is expected to increase. While the benefits of using alternative data are becoming more widely known, we should make efforts to resolve the various issues mentioned above. We will contribute to the future publication of the Factbook to provide an accurate picture of the use of alternative data and the issues involved.