Findings from Japan's Survey on Measurement of Data as an Asset

March 2026

Economic and Social Research Institute (ESRI), Cabinet Office, Government of Japan

Cabinet Office

ESRI

0. Introduction

- With the advancement of digitalization and the information society, efforts to analyze vast amounts of data to develop new goods and services are intensifying.
- For example, by analyzing customer purchase histories held by companies, it becomes possible to provide desired products and services at the optimal time.
- Reflecting this deepening and expansion of the digital economy, the 2025 SNA adopted by the UN Statistical Commission in March last year defines data included in fixed assets.

  (The definition of the data)
  Data can be defined as information content that is produced by accessing and observing phenomena, and recording and storing information elements from these phenomena in a digital format, which provide an economic benefit when used in productive activities.
- In this presentation, after briefly introducing the overview of data capital estimation at ESRI, we will explain the insights gained from the questionnaire survey conducted during the estimation process.

1. Outline of the estimation method

- How do we measure output of data?

⇒The sum of costs method is used to value data asset.

Output =Compensation of employees
+ Intermediate consumption + Consumption of fixed capital
+ Return on capital used in production.

- How do we measure intermediate consumption, consumption of fixed capital,return on capital used in production?

⇒The mark-up is applied to compensation of employees to represent intermediate consumption, consumption of fixed capital and return on capital used in production.

- How do we measure compensation of employees?

⇒$\Sigma\ W_i * (L_i * SE_i) * (H_i * ST_i)$

W_i:the wage in industry *i.*

$L_i * SE_i$: the number of workers engaged in data production in industry *i*.

$H_i * ST_i$: the time spent on data production in industry *i.*

2. Overview of the Survey

- Data plays an important role in new expanding economic activities driven by digital technologies. We have conducted surveys which aimed at gaining deeper insight into the characteristics of workers involved in the production of data assets, with the goal of measuring the nationwide scale of data asset production.
  - In November 2022, we distributed the first online survey targeting approximately 35,000 workers expected to engage in data-related occupations. These individuals were selected from a pool of 250,000 workers registered with a private market research firm. We received response from 30,295 participants.
  - In October and November 2023, we distributed a follow-up survey to 7,288 individuals from the previous year's respondents who were identified as producing data or databases. From this group, we received responses from 5,581 participants.

  **Note* Please refer to Appendices 1 and 2 for the questionnaires and aggregated results of the 2022 and 2023 surveys, respectively.*
  (https://www.esri.cao.go.jp/en/esri/prj/hou/hou020/hou014-e.html)

- Through these two-stage surveys, we found three potential insights for measuring data as an asset:
  1. A list of occupations involved in data and database production and their involvement rates
  2. Backcasting estimates based on individual retrospectives
  3. Stock estimates using the asset life survey results

3. A list of occupations involved in data and database production and their involvement rates

3-1. List of occupations and their involvement rates

- Our 2022 online survey collected responses from 30,295 workers. Participants were asked if they undertook following data-related tasks and, if so, what % of their working time was spent on each task.
- For our recent estimates, we considered Tasks 3-5 as contributing to the data production, task 6 as contributing to the production of database. For Task 3, we included responses only for a select group of occupations.

No.	Data-related task	Key statistics from survey results	
		Involvement %	If so, avg time %
1	Develop a plan to collect and utilize internal or external data	19.3% (=5,846/30,295)	24.63%
2	Prepare environment to produce data (e.g. guide and support the survey respondents)	16.5% (=5,004/30,295)	16.77%
3	Input or record data generated internally or externally (e.g. record information from business memos, surveys, customer inquiries)	28.0% (=8,494/30,295)	25.07%
4	Operate an application that collects data automatically (e.g. an app which collects search history or telephone history from smartphones)	13.4% (=4,063/30,295)	12.61%
5	Arrange and organize collected data for ease of use	10.7% (=3,326/30,295)	12.93%
6	Develop and operate a database	22.7% (=6,865/30,295)	16.93%
7	Analyze data (e.g. statistical analysis, create company management indicators)	17.3% (=5,228/30,295)	17.35%
8	Develop, manufacture, or maintain infrastructure or system to collect data automatically	9.7% (=2,925/30,295)	11.20%

(Note) The survey results went adjustments and outlier removal at the occupational and industrial level, etc. before being compiled into the involvement rate listed in the following page.

3-2. List of occupations and their involvement rates

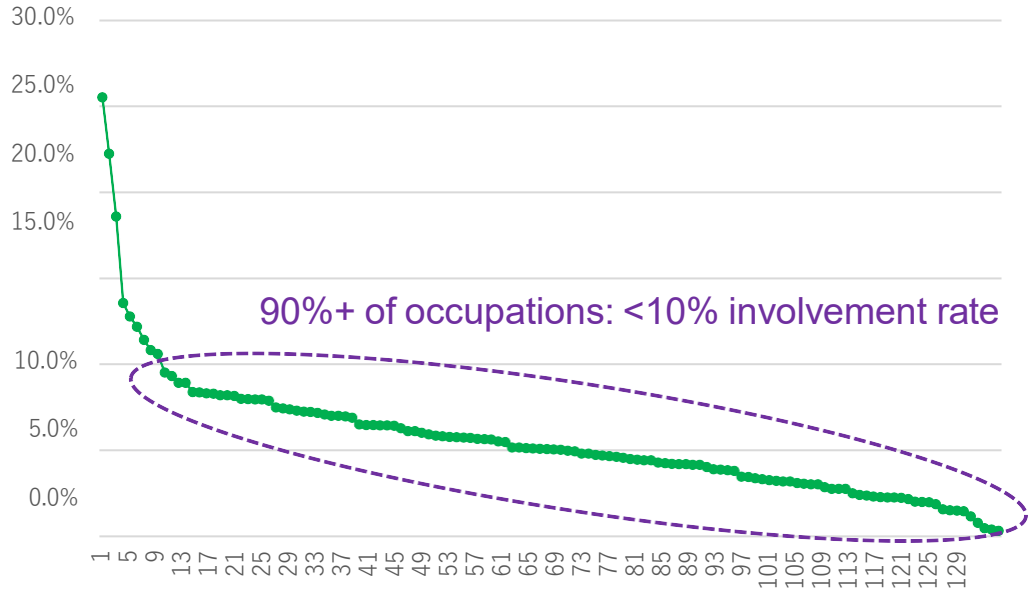
- A list of 130 data-related occupations was compiled, primarily based on the three-digit level Occupational Classifications from the Japanese 2020 Census.
- In this compilation, we avoided for potential double counting by the following methodologies:
  - Excluding two software-related occupations to avoid potential double counting with the in-house software productions
  - Adjusting involvement rates for the overlap between data assets and R&D from the survey responses

The list of 132 data-related occupations

(2 in yellow are software related occupations)

Code	Classification name	Involvement rate
312	Data entry device operators	25.5%
311	Personal computer operators	22.2%
104	Software creators	21.4%
10c	Other data processing and communication engineers	18.6%
10a	System consultants, designers	17.9%
55x	Machine maintenance and repair workers not classified elsewhere	13.6%
07e	Metal engineers	12.8%
721	Packaging workers	12.2%
332	Insurance agents and brokers	11.4%
27a	Production-related clerical workers	10.8%
18x	Other management, finance, and insurance professionals (integrated)	10.6%
07c	Machinery engineers	9.5%
34a	Machinery, communication and system sales workers	9.3%
32x	Merchandise sales workers not classified elsewhere	8.9%
06a	Agriculture, forestry, fishery, and food engineers	8.9%
343	Medicine sales workers	8.4%
03a	Corporations, organization management workers	8.4%

Distribution of 130 involvement rates



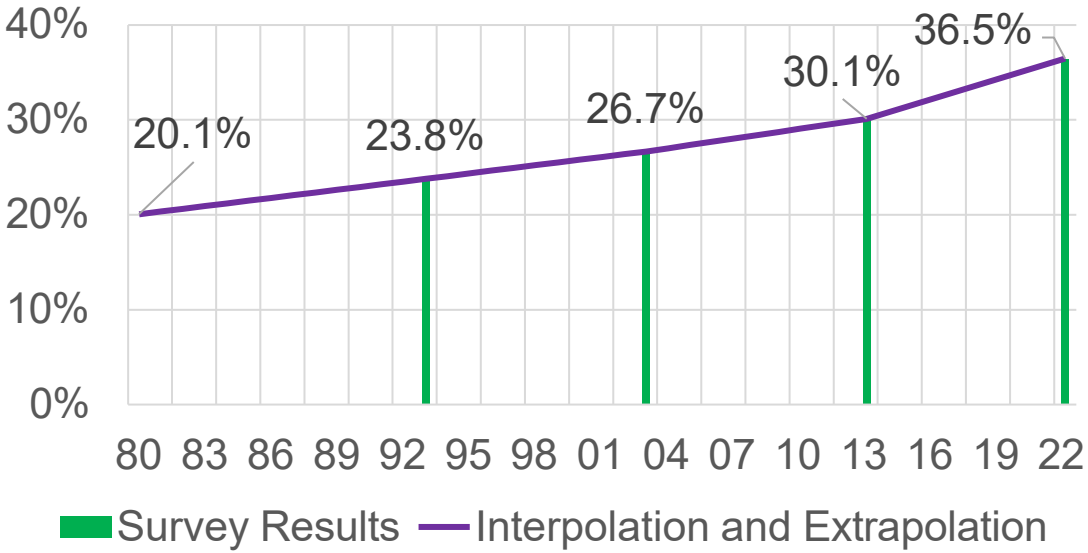
**Note* Please refer to Appendix 3 for a compiled list of data-related occupations, their involvement rates, and their overlap with computer software and research and development.*
(https://www.esri.cao.go.jp/en/esri/prj/hou/hou020/hou014-e.html)

4. Backcasting estimates based on individual retrospectives

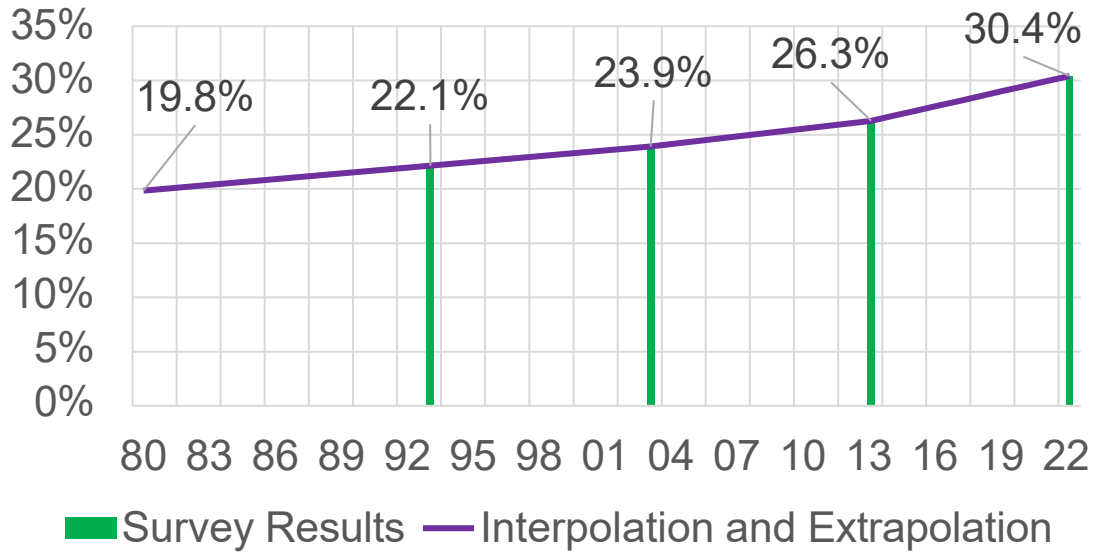
4-1. Backcasting estimates based on individual retrospectives

- Our 2023 online survey gathered responses from 5,581 workers who were identified from the previous year's survey as producing data or databases.
- These participants were asked to retrospectively provide information on the proportion of people and time engaged in data-related work in their past organizations.
  - Specifically, for four time points: 1 year ago, 10 years ago, 20 years ago, and 30 years ago.
  - For periods between these points, ratios were interpolated. For periods beyond 30 years ago, trends were extrapolated based on the ratio observed in the most recent 10-year period.

Proportion of people engaged in data-related work



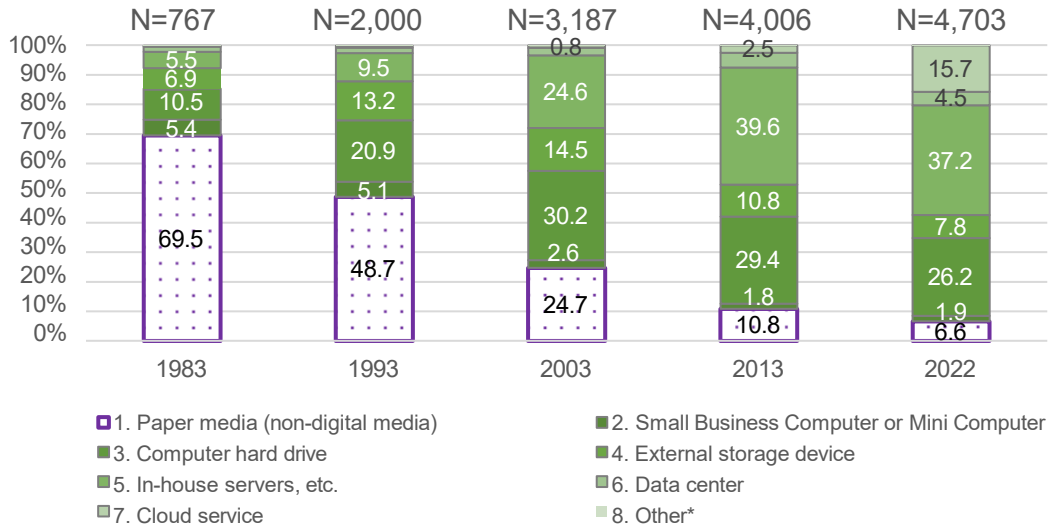
Proportion of time engaged in data-related work

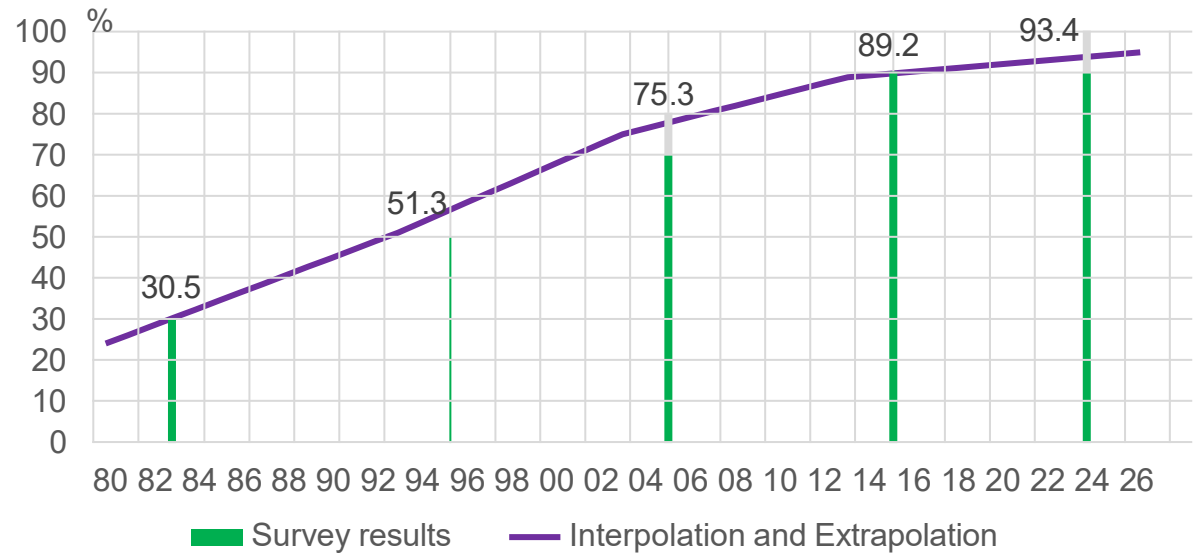


4-2. Backcasting estimates based on individual retrospectives

- Our survey also requested participants to report retrospectively on the most frequent data usage and storage methods used at five different points in time.
  - Interpolation between points and extrapolation followed the same methodology as shown in the previous slide.
- Upon further analysis, we discovered that some workers involved in data and database production reported using ‘paper’ as a storage medium. Therefore, we decided to exclude this portion of non-digital data and database production from our estimates.

Proportion of digital storage in data-related work

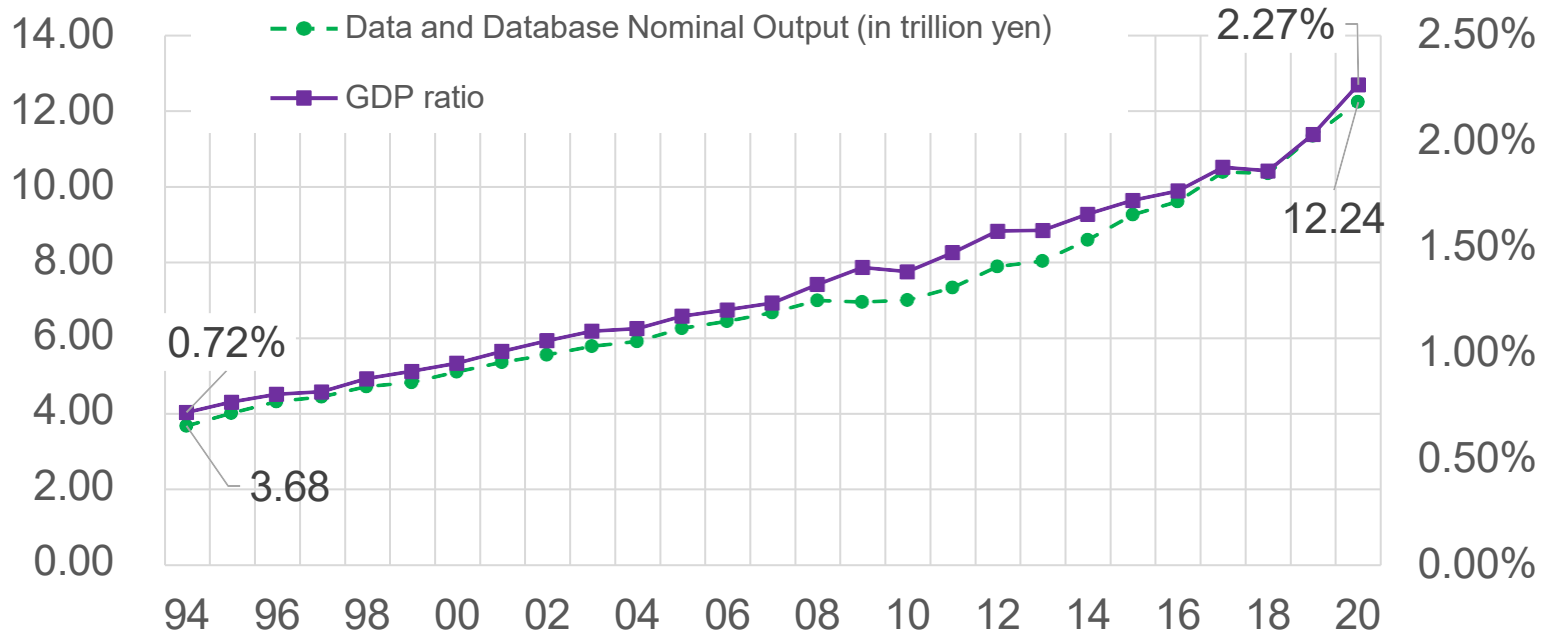




4-3. Backcasting estimates based on individual retrospectives

- We calculated a time series of labor costs by combining three explained components: proportion of people involved and time spent on data-related tasks, and the ratio of digital storage usage. Subsequently, we applied markup ratio to generate backcasting estimates.

Data and Databases Outputs and Their GDP Ratios in Nominal Terms



5. Stock estimates using the asset life of data assets

5-1. Stock estimates using the asset life of data assets

- In our 2022 survey, we inquired about the average usage period of produced data, offering three options: "Less than 1 month," "1 month to less than 1 year," or "1 year or more.“ In our 2023 survey, we expanded our inquiry by asking more specific questions about usage periods of 1 year or more, differentiating between data and databases.
- Combining the results from both surveys, we calculated the average service life:
  - 3.63 years for all data (6.61 years for data used for 1 year or more)
  - 7.05 years for databases

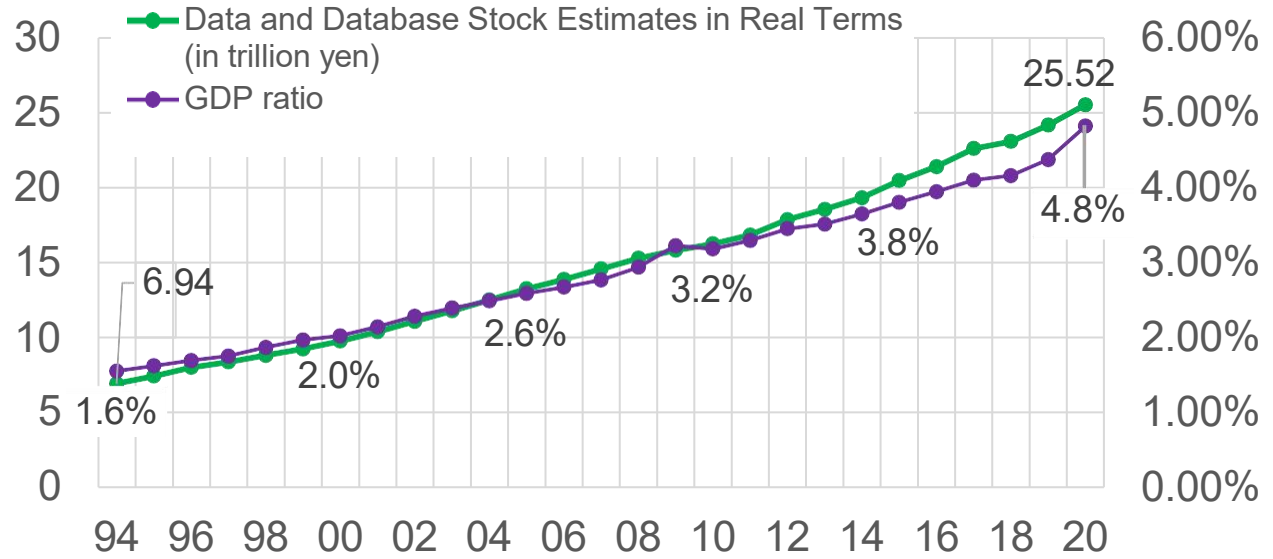
Calculated Average Usage Periods of Data and Database

	Data				Database		
Response	Comp %	Class Val	Product		Comp %	Class Val	Product
<1 month	15%	0.042	0.01		14%	0.5	0.07
1<= months <12	32%	0.542	0.18				
1<=years<3	21%	2	0.41		24%	2	0.48
3<=years<5	12%	4	0.47		20%	4	0.80
5<=years<7	5%	6	0.30		10%	6	0.62
7<=years<10	4%	8.5	0.35		7%	8.5	0.59
10<=years<15	3%	12.5	0.36		7%	12.5	0.83
15<=years	8%	20	1.56		18%	20	3.68
Total	100%		**3.63**		100%		**7.05**

5-2. Stock estimates using the asset life of data assets

- We conducted stock estimates using the PIM, incorporating long-term time series of flow estimates, depreciation rates corresponding to usage periods, and calculated chain index cost-input deflator.
- The depreciation rate was calculated using the formula δ = DBR / T, where we set the DBR to 1.65 and average service life (T) from our survey results.

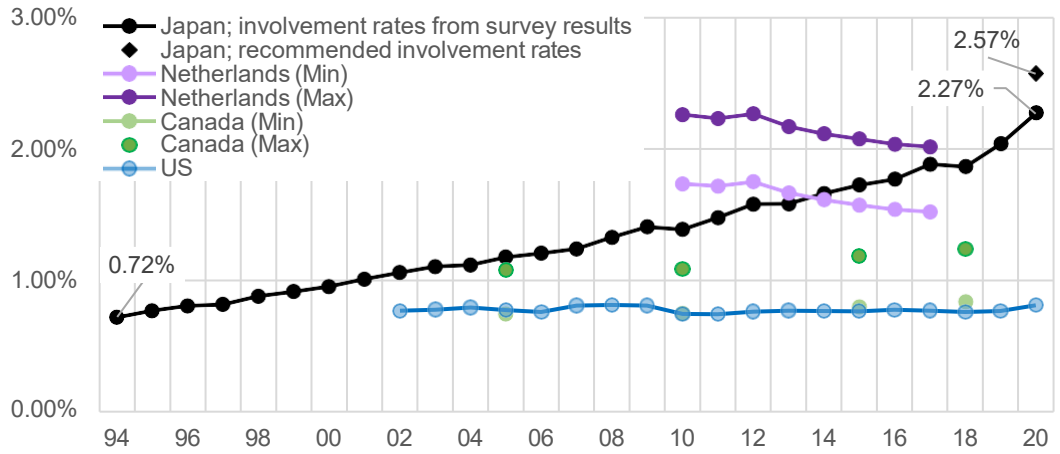
Data and Databases Stock Estimates and Their GDP Ratios in Real Terms



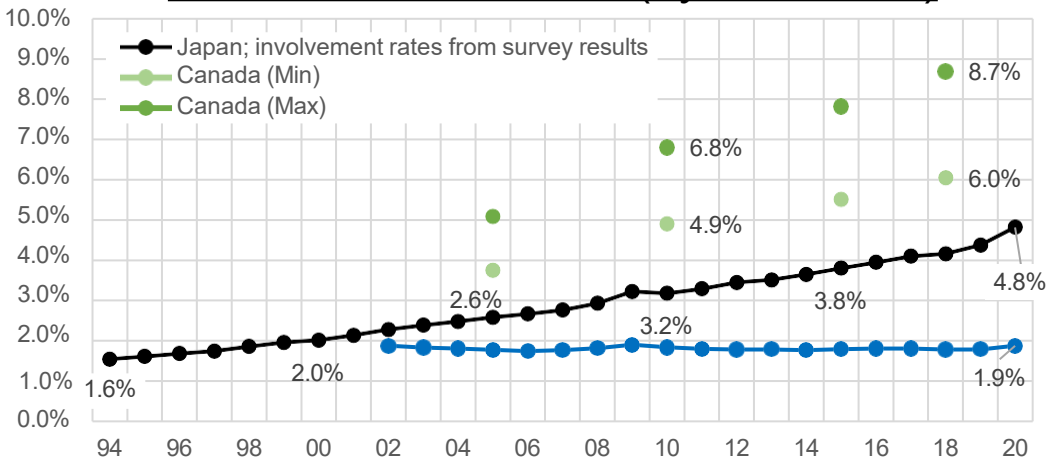
Comparison with previous studies

- It should be noted that this comparison, while informative, may lack accuracy due to two main factors: firstly, the latest estimates may not be reflected for all countries; secondly, estimation methods vary across countries.
- Despite these limitations, the results reveal a consistent pattern: almost all countries estimated data and database output as a ratio to nominal GDP falling within a range of 1% to slightly over 2%.
- For Japan, the increasing proportion of data and database output over time can be attributed to our adjustment of the proportion of people, time, and digital data storage based on retrospective information from our online surveys. Additionally, our estimation using the list of occupations and their involvement ratio recommended by the Task Team yields a similar level to our estimates derived from the internet survey results for the nominal output of 2020.

Nominal Output (by GDP ratio)



Real Stock Estimates (by GDP ratio)



References

- Calderón, J.B.S., & Rassier, D. (2022). Valuing the U.S. Data Economy Using Machine Learning and Online Job Postings, *BEA Working Paper Series, WP2022-132022*
- De Bondt, H., & Mushkudiani, N. (2021). Estimating the Value of Data in the Netherlands, *paper presented at the IARIW-ESCoE Conference*
- Statistics Canada (2019). "The value of data in Canada: Experimental estimates." *Latest Developments in the Canadian Economic Accounts*