Tentative Translation

Data Value Measurement

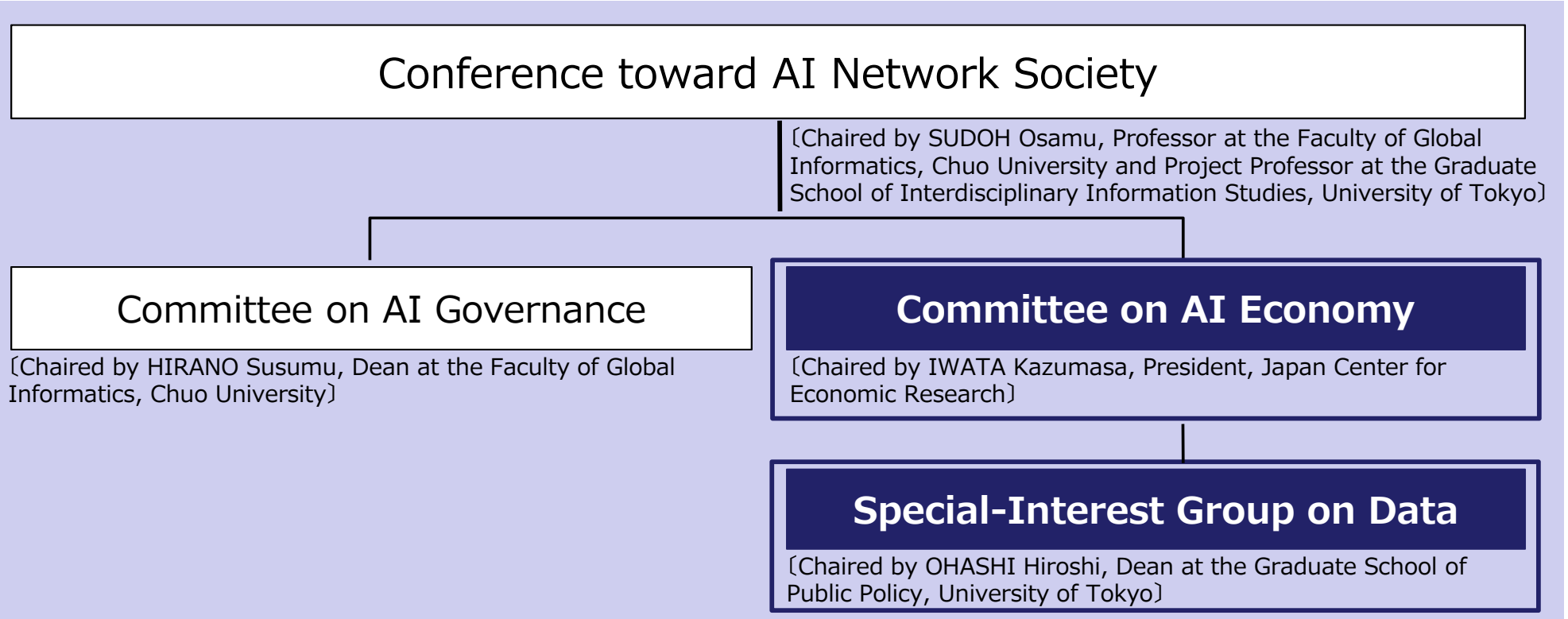
-Findings through Estimates of Data Contribution to Added Value of Japanese Enterprises-

Institute for Information and Communications Policy (IICP),
Ministry of Internal Affairs and Communications (MIC), Japan

Issues Considered at the Committee on AI Economy

- The MIC's Institute for Information and Communications Policy (IICP) established 'the Committee on AI Economy' in January 2019 to examine what kind of socio-economics should be pursued in promoting the social implementation of AI.
- The Committee has been examining specialized and technical issues on data economic policies in the AI era. Data was regarded as a factor of production and its effect on productivity improvement was estimated.

*The Committee on AI Economy 2020 Report was published in July 2020.
https://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/pressrelease/2020/7/21_2.html



Data Value Measurement Methods

- The roles and characteristics of data are diverse. It is difficult to uniquely identify the value of data.
- Approaches for measuring the value of data are classified in the following three ways:
  - ✓ Cost-based Approach - Calculation method which pays attention to costs related to the creation, management, and utilization of data.
  - ✓ Market-based Approach - Calculation method which is based on the market prices of similar products or of users' willingness to pay for those products.
  - ✓ Impact-based Approach - Calculation method where the effects (productivity, and revenue, etc.) of utilizing data is estimated.
- From the point of view of versatility and reproducibility of method application to different industries and countries, the **impact-based approach was adopted** and **the effect of data on the added value of companies was estimated according to production function**.

Analysis Model

- 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**.

$$V = A_o K^{\alpha} L^{\beta} Data^{\gamma}$$

※Estimate is based on the results of a survey with Japanese enterprises.

(Survey target) 4,286 Japanese enterprises (569 responses)

(Survey period) From February 2020 to March 2020

(Details of Survey) Amount/number of data utilized by enterprises, and method used for data analysis, etc.

Analysis Results

- Results from an empirical analysis highlighted the following:

  - ✓ The utilized data amount/number has a positive relationship with the added value as well as other production factors (capital and labor).
  - ✓ The added value increased by 0.05% when the utilized data amount/number increased by 1%.

- However, it should be noted:

  - ✓ That this reflects the current status of (not the change in) data utilization efforts, and that increasing the amount/number of data used does not necessarily lead to an increase in added value, and
  - ✓ That the results show the average between enterprises with more gains by data utilization and otherwise.

Remaining Challenges

- Necessity of further study on the relationship between data and value generated by data

  - ✓ Studies to refine the model for analysis
  - ✓ Precise analysis that considers the difference by industry sector and production process
  - ✓ Continuous research to collect more data for diversified analysis

⇒ These issues are to be discussed at the Committee on AI Economy from December 2020. The outcome will be published as a report next summer.

(Reference) Method of Data Value Measurement

Approaches	Methods	Tools	Limitations	Issues
Cost-based Approach	• Calculation 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	• Calculation 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 measured (not individual data)? • Which point of view is more appropriate for data measurement, the business side or the consumer side?
Impact-based Approach	• Calculation 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 objects for which data is to be measured (the definition and the 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?

(Reference) Analysis Results

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

※Manufacturing/Non-manufacturing

Data variable	Sample size	Adj R^2	Capital	Labor	Data
Utilized data amount (= total data amount x ratio of data used for analysis)	258	0.8343	0.44 ◎	0.50 ◎	0.05 ◎
Utilized data number (= total number of data x 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 x ratio of data acquired from outside)	267	0.8401	0.47 ◎	0.47 ◎	0.05 ◎
Externally acquired data number (= total number of data x 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 x data utilization rate (use area/type/processing method)	261	0.8334	0.44 ◎	0.49 ◎	0.05 ◎
Total number of data x data utilization rate	137	0.8150	0.32 ◎	0.58 ◎	0.05 ○
Total data amount x data diversity (diversity of data acquisition and provision)	173	0.8537	0.53 ◎	0.42 ◎	0.04 ○
Total number of data x data diversity	85	0.8576	0.41 ◎	0.49 ◎	0.06 ◎

※Variables not described as FY2015 are values for FY2018.
The ratio of externally obtained data was corrected based on the situation of external data acquisition.

(Note) Significance level ◎:1%、○:5%