Future Direction of Statistics Capacity Building In Response to New Data Source

8th December, 2017

Masato AIDA
Statistical Research and Training Institute,
Ministry of Internal Affairs and Communications

For production of official statistics, it is extremely important to use Big Data, administrative record information and other New Data Sources. Production of official statistics based on administrative record information is widespread, especially in Northern Europe.

In modern society, composite analysis of data of various types has major potential to solve problems and create new values.

EBPM also suggests to "establish a system for researchers to use and analyze in-depth data held by governments."

Because Big Data is generally created and collected in a context different from that of statistics survey data, there are many challenges including:

- There are not enough items for analysis because Big Data is not collected with any defined analysis purpose.
- Huge amount of data
- Insufficient standardization, etc. for the classifications used
- Sometimes data is not cleaned.

Systems have not been developed to make personal information including micro data of official statistics held by administrative organs available for various analyses after taking security measures.

The government has not developed or secured sufficient data science human resources who conduct analyses using New Data Sources.

The Statistical Research and Training Institute has been offering the "R Language Course." In 2017, the institute started "Introduction to Big Data," "Statistical Techniques Useful for Practice – Missing Value Complementation" and other courses. "Introduction to Data Science" is offered jointly with Shiga University.

The Introduction to Big Data introduces legal issues related to Big Data and examples of its utilization in the public sector.

Statistical Techniques Useful for Practice also introduces various techniques including data cleaning and examination of result tables.

Provide training courses for cross-departmental sharing of information on best practices that will be made by government departments in the future.

Ensure capacity building of highly skilled professionals in analysis and information processing in cooperation and exchange with universities and other institutions of higher education and research as well as private enterprises.

Offer courses to teach techniques to provide and use official statistics and other micro data. Implement training on sophisticated programming application techniques for analysis including R language.

End