



CPS Paper

A Quality Assessment framework for Statistics based on data-science: Institutional Management Plan for Experimental Statistics of 'Statistics Korea'

Author: Bitna Kang

Submission ID: 525

Reference Number: 525

Presentation File

abstracts/ottawa-2023_620116513d082179a186b367f1da94fe.pdf

Brief Description

Due to rapid data environment change such as big data and artificial intelligence, it is more important to produce statistics in a way that is different from the survey-oriented statistics.

In order to secure the reliability and accuracy of statistics produced in a new way, it is necessary to manage them as official statistics at the national level.

In this paper, the quality management framework was designed based on the quality evaluation dimension of Statistics Korea.

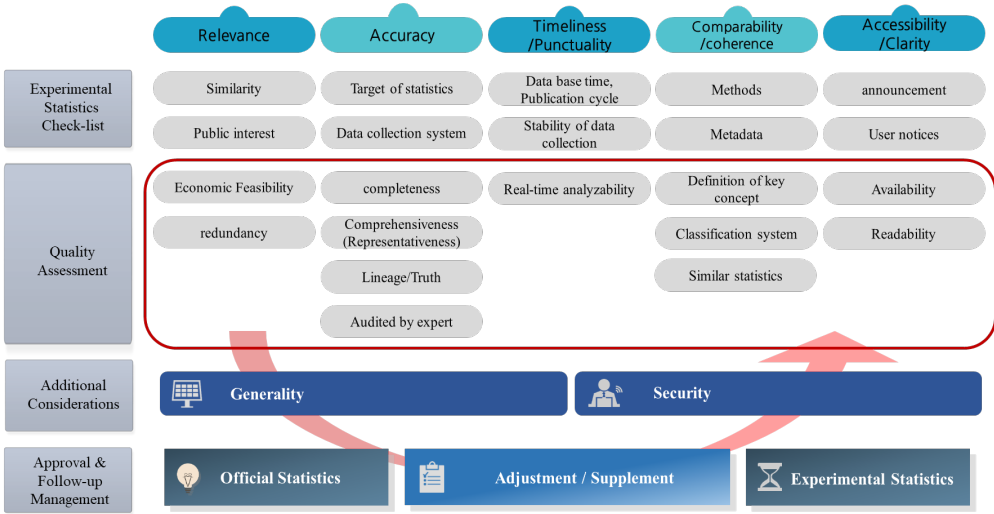
Abstract

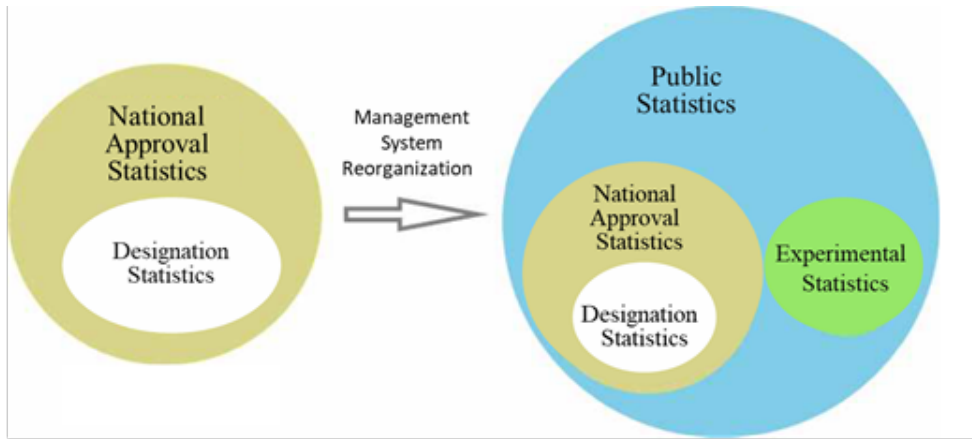
Due to rapid data environment change such as big data and artificial intelligence, it is more important to produce statistics in a way that is different from the survey-oriented statistics. In order to secure the reliability and accuracy of statistics produced in a new way, it is necessary to manage them as official statistics at the national level. For this, the Statistics Korea introduced the experimental statistics system in 2021. The goal is to discover and manage statistics that are based in various data sources and that use different method. These data science-based statistics make it difficult to manage the entire statistical production process due to the diversity and incompleteness of the data. Therefore, the same quality management method as the existing one cannot be applied. However, in order to systematically manage the quality of these statistics, there must be a quality dimensions shared with other statistics. In this paper, the quality management framework was designed based on the quality evaluation dimension of Statistics Korea. The framework includes evaluation indicators, contents and procedures for each quality dimension. In order to help understand the framework, the first experimental statistics in Korea, the case of telecommunication mobile population movement statistics('Population Mobility Statistics'), will be presented. Data science is an interdisciplinary field that uses scientific methods, process, algorithms knowledge and insights from noisy, structured and unstructure data. In this paper, data-science based statistics are defined as statistics created by utilizing various data sources(big data in a broad sense including vast public big data and administration data) or data extraction and processing using artificial intelligence. In the case of statistics based on data-science, it is difficult to establish a quality management method with a single fixed standard because the development of the underlying data related technology is very fast and the methodologies are also changing rapidly. Considering this point, the quality management(assessment) framework to be presented in this paper includes a macroscopic quality process that can encompass the overall statistics based on data-science. The framework is comprehensive and flexible taking into accounts various data sources and data complexity. This framework is only at the level of presenting the idea of a formal management method for data science-based statistics from a macro perspective. In the future, it is necessary to develop detailed inspection items and to prepare quantitative quality evaluation measures such as distribution of scores for each item and selection of weights in detail. Considering the status and influence of national statistics, as part of an effort to guarantee the quality of these statistics based on data-science, the quality evaluation framework presented in this paper should be constantly revised, supplemented, and specified.

Figures/Tables

framework

Quality Assessment Framework
for Statistics based on Data Science





Proactive Management
through
Strict Approval



Follow-up Management
through
Flexible Standards