Competing Effects of Scale, Scope and Complexity in the Production, Dissemination and Use of Official Statistics

Author: Dr John Lamont Eltinge

Submission ID: 616

Format: CPS Paper

Reference Number: 616

Presentation File

abstracts/ottawa-2023_72968dbf2f837793e77245287efc21dc.pdf

Files/Uploads

ISI 2023 Eltinge - Scale-Scope-Complexity 02272023 finalized

Brief Description

Competing Effects of Scale, Scope and Complexity in the Production, Dissemination and Use of Official Statistics John L.

Eltinge, United States Census Bureau John.L.Eltinge@census.gov Key Words: administrative record data; data quality; efficiency; granularity; network density and complexity; record linkage; sample surveys; trade-offs Abstract: National statistical organizations (NSOs) have encountered increasing needs to improve efficiency and data quality, while they also expand their suites of statistical information products and services.

Efforts to meet these goals often center on integration of data from multiple data sources, e.g., surveys, administrative records, sensors, and web scraping; and on expansion of platforms through which statistical information is disseminated and used.

The practical impact of these efforts can involve numerous trade-offs among competing effects of scale, scope, and complexity.

These effects may in turn require re-examination of stakeholder priorities within the space of statistical information; features of prospective data sources; the architecture and adaptability of systems for ingest and management of data, and for production and dissemination of estimates; changing requirements for personnel in high-priority technical areas; the measurability and stability of related cost structures; and multi-dimensional criteria for data quality.

This paper explores these issues, with emphasis on three points.

(1) Operational definitions of “scale,” “scope” and “complexity” within the broad context of current and prospective systems for statistical information production, dissemination, and use.

(2) Production: Application of those definitions to current practice and literature for sample surveys, administrative record systems, sensor networks, web scaping and record linkage.

These applications lead to exploration of trade-offs among scale, scope and complexity effects among multiple dimensions of efficiency and data quality in production of estimates for large-scale population aggregates and finer-scale small domains.

(3) Dissemination and use: Evaluation of scale, scope and complexity effects related to anticipated groups of data users; their patterns of data use; and related features of data dissemination systems.

User groups include segments of stakeholders who access and use statistical tables, graphs and maps provided through public-domain websites; and more specialized researchers who may integrate and analyze microdata within restricted-access environments.

These groups vary in their patterns of use of published results and related quality measures, and in the substantive context within which they interpret numerical results.

Abstract

Competing Effects of Scale, Scope and Complexity in the Production,
Dissemination and Use of Official Statistics

John L. Eltinge, United States Census Bureau John.L.Eltinge@census.gov

Key Words: administrative record data; data quality; efficiency; granularity; network density and complexity; record linkage; sample surveys; trade-offs

Abstract: National statistical organizations (NSOs) have encountered increasing needs to improve efficiency and data quality, while they also expand their suites of statistical information products and services. Efforts to meet these goals often center on integration of data from multiple

data sources, e.g., surveys, administrative records, sensors, and web scraping; and on expansion of platforms through which statistical information is disseminated and used.
The practical impact of these efforts can involve numerous trade-offs among competing effects of scale, scope, and complexity. These effects may in turn require re-examination of stakeholder priorities within the space of statistical information; features of prospective data sources; the architecture and adaptability of systems for ingest and management of data, and for production and dissemination of estimates; changing requirements for personnel in high-priority technical areas; the measurability and stability of related cost structures; and multi-dimensional criteria for data quality. This paper explores these issues, with emphasis on three points.
(1) Operational definitions of “scale,” “scope” and “complexity” within the broad context of current and prospective systems for statistical information production, dissemination, and use.

(2) Production: Application of those definitions to current practice and literature for sample surveys, administrative record systems, sensor networks, web scaping and record linkage. These applications lead to exploration of trade-offs among scale, scope and complexity effects among multiple dimensions of efficiency and data quality in production of estimates for large-scale population aggregates and finer-scale small domains.

(3) Dissemination and use: Evaluation of scale, scope and complexity effects related to anticipated groups of data users; their patterns of data use; and related features of data dissemination systems. User groups include segments of stakeholders who access and use statistical tables, graphs and maps provided through public-domain websites; and more specialized researchers who may integrate and analyze microdata within restricted-access environments. These groups vary in their patterns of use of published results and related quality measures, and in the substantive context within which they interpret numerical results.