Deciphering & Interpreting "For Statistical Purposes"

Sallie Ann Keller & Michael B. Hawes
U.S. Census Bureau

October 24, 2023

Origins

The idea of “functional separation”

Privacy Protection Study Commission (1977)

“…separating the use of information about an individual for a research or statistical purpose from its use in arriving at an administrative or other decision about that individual.”

Current Federal Statistical System Definition

Legal Definition

CIPSEA (2019)

Statistical purpose primarily means the description, estimation, or analysis of the characteristics of groups, without identifying the individuals or organizations that comprise such groups.

Nonstatistical purpose primarily means the use of data in identifiable form for any purpose that is not a statistical purpose, including any administrative, regulatory, law enforcement, adjudicatory, or other purpose that affects the rights, privileges, or benefits of a particular identifiable respondent.

Application to the Census Bureau

Key reference to “Statistical Purposes” in Title 13

”[The Census Bureau may not] use the information furnished under the provisions of this title for any purpose other than the **statistical purposes** for which it is supplied” [13 USC §9(a)]

Why Does This Matter?

Will our current concepts be practical in the future?

- Privacy concerns
- Potential for public harm
- Data quality concerns.
- Scientific integrity and trust
- Enhance relevance

Discussion Points

Why does it matter?

What has informed statistical purpose?

What Are Future Challenges?

Use case – For Statistical Purpose

Statistical Agencies are established to use data for the creation of statistical products that adhere to codified statistical purposes that are in public's interest and for the public good.

Examples for statistical purpose include:

- Decennial Census – apportionment and representation
- American Community Survey – to understand the changing shape of the nation and its communities
- Principal Economic Indicators – to understand and improve the economy
- Sector-specific statistics (health/education/etc) – to assess important aspects of society and support policymaking and research

Use case – Not For Statistical Purpose

As noted, data have been used for non-statistical purposes, the most widely known example was:

- Second War Powers Act (1942) rescinded Presidential and Census Bureau promises regarding the confidentiality of the census.
- Provided the Secret Service with lists of every individual of Japanese descent in Washington DC (1943)

Use case – Grey Areas

As noted, data have been used for non-statistical purposes, the most widely known examples include:

- The U.S. Census Bureau lending staff and expertise to INS (1940)
- Provision of tract and block-level tabulations of Japanese-, German-, and Italian-Americans (1941-1942)
- Custom tabulations for DHS of Arab-American households by country of origin at the zip code level (2002-2003)

Interpretation and understanding of “Statistical Purposes” varies by stakeholder group

Messaging about data protections may hurt rather than help

- "…respondents will extract information about the nature of a survey from the confidentiality assurances given to them. If these assurances are very elaborate, respondents are likely to infer that the survey is sensitive and that they will be asked a number of questions that may be unpleasant, embarrassing, or incriminating." -Singer, Hippler, and Schwarz (1992)

Landreth, Gerber, and DeMaio (2008) also note:

- Messaging about "confidentiality," "statistical purposes," and linkage to administrative records is often misunderstood, or seen as vague or open-ended
- Messaging about legal protections and penalties for unauthorized disclosure was variously interpreted as "legalese," the "fine print," or "over the top"

Need to Clarify Statistical Purpose

Given efforts to improve the usability of statistics making finer grain details available and with the accessibility of more data sources some concerns arise regarding:

- The risk of reidentification
- Risk of group harm
- Private (or partisan) capture of public goods

How do definitions compare?

United States – ‘statistical purpose’ primarily means the description, estimation, or analysis of the characteristics of groups, without identifying the individuals or organizations that comprise such groups (44 U.S.C.A. 3561)

Eurostat – 'use for statistical purposes’ means the exclusive use for the development and production of statistical results and analyses; and ‘statistics’ means quantitative and qualitative, aggregated and representative information characterizing a collective phenomenon in a considered population.

Statistics Canada – “Information collected by Statistics Canada is for statistical purposes, defined as description or analysis of characteristics of a population to which the individual belongs.” (Statistics Canada, 2017, p.18)

New Zealand – Official statistics may be produced on any matter relating to New Zealand including, its economic, social, cultural, and environmental situation (Section 43). Statistician must be satisfied research is in the public interest and account for the public benefit, contribution to the well-being of the Maori, and any risks of harm (Section 49). (Statistics New Zealand, 2022)

Concerns

- Privacy concerns/Harm to the individual or oganization
- Confidentiality / Disclosure risks
- Risk of Group Harm
- Private capture of public goods, partisan rather than public interest

Challenges

Maintaining functional separation with diverse data **sources** including public and private administrative data, designed data collection, and opportunity (open) data

Statistical **output** and tiered access controls **or restrictions** need to be linked to approved purposes and uses

Proliferation of AI

Moving forward…What should a future definition include?

Balancing trade-offs for "Statistical Purpose"

Implementing controls/restrictions in a dynamically changing environment

Managing the statistics for their purpose and use

Functional separation vis-à-vis measurement priorities