U.S. Census Quality Standards

Paul Marck [1]

[1] *U.S. Census Bureau, 4600 Silverhill Road, Washington D.C. 20233*

Abstract

This paper reviews the challenges statistical organizations are confronting and the key changes that are occurring to keep them relevant. This paper explores possible options to revise quality standards to align with these changes, improve operations, and be more customer focused. As a result, this paper proposes to 1. Structure requirements to identify who is responsible and what they shall do. 2. Differentiate standards for products from standards for services. 3. Require product owners and service managers to establish feedback loops and metrics to monitor customer satisfaction. 4. Segment standards into management, development, and production phases. The goal of this paper is to elicit feedback from other statistical agencies in multiple countries on what legislation and standards they find effective.

Keywords: Quality, Standards, Agile, GSBPM

1. Introduction

The United States Census Bureau is starting a business transformation (U.S. Census Bureau, 2024). This is an opportune time to consider how the Census Bureau's Quality Standards may be updated to help support the transformation. This paper therefore reviews 1) the challenges we are facing, 2) the legislation and policy that we cannot dismiss, 3) the transformational efforts to which we should align, 4) some potential dilemmas that may arise requiring safeguards to be put in place, and finally 5) some proposed improvements.

2. Current Status

National Statistical Organizations are facing a few key challenges to stay relevant. Most organizations have legal frameworks to support their efforts. However, as organizations begin shifting their paradigms to improve quality a few dilemmas arise that need to be addressed.

2.1 External Challenges

The major challenges facing statistical organizations include overall declining response rates and an increasing demand for more timely, accurate, granular, and customizable information (U.S. Census Bureau, 2024). Both are indicative of unmet needs. One key to remaining

[1] Any opinions and conclusions expressed herein are those of the author and do not reflect the views of the U.S. Census Bureau.

relevant is to discover these unmet needs. Why do your respondents not want to complete your survey? How do the specific quality characteristics for new products requests differ from the existing products? Ultimately, your respondents and users will know the answers. It is critical for the statistical organization to maintain open lines of communication to learn how to tailor the statistical products and services to satisfy their respondent's and user's needs.

2.2 Internal Challenges

Statistical organizations, particularly those that have a collection of long-standing surveys coupled with newer information products may find they have several internal challenges. Given the disparate needs that existed when each information product was implemented, the data from all products may not be readily interchangeable thus hindering efforts to create new products from existing data sets. The services that support the development of each information product are also likely to be independent from one another. Such a duplication of efforts makes for a broader organizational inefficiency. A key to resolving these challenges lay in maintaining standardized information such as defined geographic boundaries or standard occupational classifications so the data is more interchangeable and utilize more centralized services to develop information products.

2.3 External Initiatives

Several laws and policies have been enacted that influence statistical quality for the United States Federal Statistical System (Office of Information and Regulatory Affairs, 2024). When considering changes to quality standards, these laws and policies are not dismissible.

The Paperwork Reduction Act of 1980 & 1995 principally established the Office of Information and Regulatory Affairs (OIRA) and its management of statistical policy and coordination, limited repeat selection of certain organizations to the Quarterly Financial Report and requires the use of sampling when selecting small businesses to participate in surveys.

The Information Quality Act of 2001 and subsequent guidelines requires agencies to 1) adopt a standard of quality. 2) review the quality of the information before it is disseminated. 3) designate a Chief Information Officer, 4) respond to complaints, and 4) every fiscal year report to the Director of the Office of Management and Budget (OMB) on complaints received and how they were handled.

The Confidential Information Protection and Statistical Efficiency Act of 2002, requires that: data acquired under a pledge of confidentiality and for exclusively statistical purposes 1) be used by the agency or agents they designate exclusively for statistical purposes and 2) not be disclosed by the agency in identifiable form except with the respondent's consent. The agency

shall also provide the public notice before it collects data for nonstatistical purpose. This Act further permits Designated Statistical Agencies (DSA) (defined as the U.S. Census Bureau, Bureau of Economic Analysis, and Bureau of Labor Statistics) to share business data in an identifiable form exclusively for statistical purposes such that any publication using the acquired data do not release respondent information in an identifiable form.

Statistical Policy Directive 2 of 2006 establishes the core standards and guidelines for statistical surveys for the U.S. Federal Statistical System. It includes 20 standards plus an additional 89 guidelines organized into 7 survey lifecycle categories. In 2021, 7 standards were added for cognitive interviewing.

Statistical Policy Directive 4 of 2008, applies to the release of regular and recurring statistical products. To summarize its requirements, Agencies shall publish a) their dissemination policies, b) a schedule of release dates, and c) notices for revisions or unscheduled corrections. The Directive also permits the agency to provide pre-release access to their final statistical products under embargo or through secure pre-release access.

The Evidence Act of 2018 broadly supports developing evidence for policy making. A core aspect of the Act is the presumption of accessibility for statistical agencies to request data assets for developing evidence. Key requirements include the Director of OMB establishing a standard application process (SAP) for statistical agencies to access data assets which is now available at ResearchDataGov.org. Additional requirements include Agency heads designating Evaluation Officers – to manage the evaluation activities of the agency, Statistical Officials – to serve on the Interagency Council on Statistical Policy, and a Chief Data Officer – responsible for lifecycle data management. Agencies are also required to develop an evidence-building plan, maintain a comprehensive inventory of their data assets, establish sensitivity levels of each data asset and the corresponding level of accessibility, conduct a risk assessment for whether to release the data, and make its processes transparent.

Statistical Policy Directive 3 amended in 2024, applies to the release of Principal Federal Economic Indicators (PFEI). To summarize its requirements, Agencies shall a) issue reports promptly, b) publish a schedule of the dates and times for each release, c) announce changes that may affect the interpretation of the series, d) ensure reports are not prematurely issued but may grant prerelease access under specific conditions, e) follow certain guidelines to balance the accuracy and timeliness in releasing preliminary and revised estimates, and f) submit a performance evaluation to OIRA every 3 years.

2.4 Internal Initiatives

The U.S. Census Bureau is starting a business transformation (U.S. Census Bureau, 2024). Key initiatives for this transformation include Data Integration and Collection for the Enterprise (DICE), Enterprise Data Lake (EDL), Frames, the Census Enterprise Dissemination and Consumer Innovation (CEDSCI), and Open Census.

DICE should more efficiently manage survey collection and gather existing data from external sources. DICE should simplify survey design, integrate response data collected through multiple modes, provide rear real-time collection status, enable earlier access to data, improve the user experience, and deliver scalability and cost-effectiveness.

EDL is a central cloud-based repository for all types of Census Bureau data. The benefits of the EDL include discoverable data, streamlined security, multiple environments, quick and cost-effective, software and environments to meet your needs, and better monitoring.

Frames allows for centrally maintaining & repurposing linked datasets. The benefit of Frames includes greater alignment, advanced research, expanded analysis, increased efficiency, reduced burden, and the ability to select more targeted samples.

CEDSCI develops and implements state-of-the-art tools for discovery, visualization, and dissemination of statistical products to the public. The benefits of CEDSCI include being a trusted source, improving user experiences, providing transparency, being user driven, ensuring program efficiency and user satisfaction, and ensuring reliable operations.

Open Census allows for greater transparency and scientific reproducibility. This initiative is composed of Open Science to disseminate research publications, Open Code to disseminate source code helpful in replicating specific methods, and Open Data to disseminate disclosure approved data sets.

2.5 Potential Dilemmas

A rapid pace of change may bring about some undesired outcomes where safeguards may need to be established. The two most concerning challenges arise from an increase access to data and the production of inappropriate results by means of artificial intelligence.

2.5.1 Statical Purpose

Our first dilemma is with inappropriate access to data and the pledges given to potential survey respondents. As data products become more accessible two concerns arise. 1) Would data be used beyond the purpose for which it was originally collected? 2) Is the legal definition for what constitutes a statistical purpose still adequate?

Given these concerns, the pledge made to a respondent when the data was originally collected is still valid. To increase the utility of future data collections it would be best to have the respondents informed consent for additional planned uses. For the second concern, if some future uses are not planned, such as trying to provide information to help with a natural disaster, is generically pledging that future uses would be “for statistical purposes only” enough to encourage survey recipients to respond? Maybe, one alternative is to extend the pledge to note access is granted only for statistical purposes that are for the public benefit and reference the agency’s evidence-building plans.

2.5.2 Artificial Intelligence

Our second delimma is with potentially undesired results and the safeguards that may need to be put in place. Two concerns are of interest. 1) What steps can be taken to help ensure the validity of AI produced results and 2) what steps can be taken to reduce the risks of harm.

For the first concern, AI generated responses are only as good as the data they are is based on. To help ensure valid results, standards need to make sure the appropriate information is available and that the information is machine readable (Wise, O., Keller, S., Houed, V., 2024, January 18). For the second concern, once data is publicly released the agency cannot impose restrictions on its use. Disclosure review standards may need to be updated to account for the risk of reidentification where AI recombines data from the statistical agency with other commercially available information (CAI) (Executive Order Number 14110, 2023).

3. Proposed Improvements

To revise the existing Quality Standards, this paper proposes all requirements follow a well-formed structure. Each requirement should clearly identify who is responsible, what they are responsible for, and provide some context for when the task should be performed. Next, this paper considers the following proposed improvements when determining how to map the existing standards and business transformation efforts into a new model.

3.1 Stakeholder Focus

One cornerstone of quality is customer satisfaction. A critical component to achieve customer satisfaction is a feedback loop. Knowing your customer and their needs will help to ensure your products continue to be fit for their use. It is therefore essential to have a means to communicate with your customers, document their needs, and monitor your progress.

3.2 GSBPM

The United Nations Economic Commission for Europe (UNECE) developed the Generic Statistical Business Process Model (GSBPM) to help structure the processes used to create

official statistics (UNECE, 2019). This model gives attention to change work and ongoing work processes. This distinction gives rise to the notion of having services to produce statistics separate from the management of the information products. Adoption of this model may help to provide consistency in sharing standards and business processes to create official statistics.

3.3 Agile Quality

The agile manifesto for software development (Agile Alliance, 2001) proposes four concepts and 12 principles focused on customer satisfaction and achieving results. While established for software development we can adapt these concepts.

For the development process, customers may not know what developers can do. Developers may not know what customers want. So, work together--closely. Plans are never perfect and new ideas will emerge so, plan in shorter timescales.

For overall operations, acknowledge that procedures may not be perfect so lean towards empowerment and transparency over bureaucracy. Lastly, evaluate your effectiveness and change accordingly.

3.4 Malcolm Baldrige

In 2001 Pal's Sudden Service, a quick-service restaurant, won the Malcolm Baldrige National Quality Award (Baldrige@nist.gov, 2001). A key lesson from their experience is that excellence is a habit. Train! Repeatable processes are more effective than repetitive inspection. Given their efforts and a modest number of menu options they made several notable results. 1. Customer quality scores of 95.8 percent versus 84.1 percent for their best competitor, 2. Handout speed of 20 seconds compared to 76 seconds for competitors, 3. An error rate of 1 in every 2000 transactions, and 4. Less employee turnover.

3.5 Estonia Digital Government

Around 1990 Estonia embarked on a mission to establish a digital government. Some key design principles (Piperal, 2019) for this transformation include: 1. Guarantee Privacy & Confidentiality. This is implemented through establishing a digital identity and digital signature. 2. Once Only. This ensures data is collected once and stored in one place. This makes a distributed system that requires secure and robust access to exchange information. Data reliability and integrity are critical to the success of these operations. 3. Data Ownership. This states that the individual is the owner of the data collected about them. The system provides individuals with audit trails to see who is using their data.

Application of these principles could help statistical agencies. Imagine completing a Census in two seconds rather than two years. Alternatively, for an implementation less than a full digital

government, statistical agencies may consider developing a software application that could be included with popular accounting software programs. These applications could gather the appropriate data for a given survey and with a push of a button provide quality assured responses. An additional customer centric concept could be to provide users with digital cookies that would help find data from statistical products that is relevant to them.

4. Results and Conclusions

This paper proposes a high-level structure for a new set of standards. The new model for the quality standards makes a distinction between information products and the services used to create the products. The standards are separated into management (quality, confidentiality, data & documents), development (needs/configure, design/build), and production phases (collection, process, analyze, disseminate).

4.1.1 Management

For Quality, all staff shall check the quality of their work. The quality program staff shall ensure scientific integrity and train staff. Product owners and service managers shall prepare a quality management plan, identify their customers, establish feedback loops, and perform evaluations.

For Confidentiality, all staff shall protect confidentiality of the information entrusted to them. The Data Stewardship and Executive Policy Committee shall train staff on protecting confidentiality and document labelling. Product owners and service managers shall develop procedures for disclosure avoidance.

For Records, all staff shall retain sufficient records for the continuity of operations. The Chief Data Officer shall format data for AI ingestion, maintain a data inventory, prepare record schedules, and train staff. Product owners shall develop procedures for retention of records, metadata for data and linkage quality, and metrics. Service managers shall develop procedures for retention of records and metrics.

4.1.2 Development

For Needs/Configure, product owners shall define their products, establish product requirements, configure services, and prepare a business case. Service managers shall define their services, establish service requirements, and prepare a business case.

For Design/Build, product owners shall validate data sources, get authorization, establish contracts, register metadata in EDL, and link into frames. Service managers shall design, develop, and test processes, train staff, and get authorization to go to production.

4.1.3 Production

For collection, product owners shall manage the scope, time, and cost of their product. The cognitive manager shall pretest questions. The DICE manager shall coordinate with data suppliers and data integration manager (DIM) to collect data.

For process, product owners shall manage the scope, time, and cost of their product. The subject matter experts shall code, derive variables and units, edit, impute, calculate aggregates, and finalize data files. The DIM shall coordinate with the DICE manager to collect enough data from available sources to produce statistically valid products, perform sampling, and perform weighting. The frames manager shall link data and maintain frames.

For analysis, product owners shall manage the scope, time, and cost of their product. The subject matter experts shall prepare outputs and conduct reviews. The Disclosure Review Board shall perform disclosure reviews.

For dissemination, product owners shall compile quality and service metrics for their product and manage customer support. The CEDSCI manager shall manage release, disseminate experimental, core, and principal federal economic indicator products, and manage user support. The Open Census Manager shall disseminate research products.

References

Agile Alliance. (2001). *Manifesto for Agile Software Development*. Agile Manifesto. Retrieved April 25, 2024, from https://agilemanifesto.org/

Baldrige@nist.gov. (2001). *Pal's Sudden Service*. NIST. Retrieved April 25, 2024, from https://www.nist.gov/baldrige/pals-sudden-service

Executive Order Number 14110, (2023, October 30), 3 C.F.R. 75191 Retrieved April 25, 2024, from https://www.govinfo.gov/content/pkg/FR-2023-11-01/pdf/2023-24283.pdf

Office of Information and Regulatory Affairs. (2024, April 2). *Legislation and Related Policy Guidance*. Whitehouse.gov. Retrieved April 25, 2024, from https://www.whitehouse.gov/omb/information-regulatory-affairs/statistical-programs-standards/

Piperal, A. (2019). *What a digital government looks like*. www.ted.com. Retrieved April 25, 2024, from https://www.ted.com/talks/anna_piperal_what_a_digital_government_looks_like?language=en

U.S. Census Bureau. (2024, February 6). Building a data-centric business ecosystem. Census.gov. Retrieved April 25, 2024, from https://www.census.gov/about/what/transformation/maximizing-operational-efficiency/data-centric-business-ecosystem.Overview.html#list-tab-1020552672

UNECE. (2019). GSBPM v5.1 - Generic Statistical Business Process Model. Statswiki.unece.org. Retrieved April 28, 2024, from https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1

Wise, O., Keller, S., Houed, V. (2024, January 18). *Preparing Open Data for the Age of AI*. U.S. Department of Commerce. Retrieved April 25, 2024, from https://www.commerce.gov/news/blog/2024/01/preparing-open-data-age-ai