Corporate Solutions to Erroneous Dissemination in the Census Bureau

Steven S. Klement [[[1]](#footnote-1)]

U.S. Census Bureau, Research and Methodology Directorate, Quality Program Staff

Abstract: The United States Census Bureau releases thousands of data products from a multitude of statistical programs. Each one of these releases goes through numerous reviews that cover statistical, methodology, and policy aspects to ensure as much as possible that the data are correct and properly presented. Regardless of how strenuous the reviews are though, inevitably Census releases data to the public that are incorrect.

This paper describes the process the Census Bureau has developed to track and analyze these erroneous dissemination incidents. The Quality Program Staff (QPS) helps the program staff complete a standardized form on the incident that outlines what was incorrect, why it happened, and what steps they are taking to prevent similar incidents in the future. QPS then follows up to ensure the program area does correct its processes. In addition, QPS analyzes the data collected from over a hundred incidents to determine if there are systemic problems. While a program may have made a mistake once and corrected it, if multiple programs have made the same or similar errors, Census may need a corporate solution. This paper lays out this process.

**1. Background**

The Census Bureau is responsible for releasing thousands of data products every year. From the Decennial Census to the five-year Economic Census to numerous annual, quarterly, and monthly programs; each is an opportunity to supply the data-user with needed information. With each release, there is also a chance to supply numbers that are incorrect due to myriad reasons. Even with a strict review process, mistakes can occasionally be made:

*“****Census mistakes mean a recount for some.****”* June 28, 2010, Washington Post

*“****StatsCan admits five-year inflation mistake.****”* August 16, 2006, CTV

*“****Census Bureau backtracks on number of same-sex households.****”* September 27, 2011 Chicago Tribune

In times of shrinking budgets, added scrutiny, and an increasing reliance on data-driven decisions by data consumers, the statistical profession should strive to minimize data release errors.

1.1 Census Erroneous Dissemination

At the Census Bureau, reduced staffs are trying to do more with less and inevitably fail to catch some mistakes until after release to the public. In one instance, Congressional staff notified the Deputy Director about a published data error. Unaware of the issue, the Deputy Director turned to the program managers and asked what happened. The reply was not expected. It turned out that the program managers knew bad data was out in the public arena. The staff had a plan to fix the data. However, they had not implemented it yet. These mid-level managers also felt that the issue was not sufficiently important to notify their managers. This was obviously an error in judgment since members of Congress felt the issue serious enough to ask the Deputy Director of the Census Bureau about it.

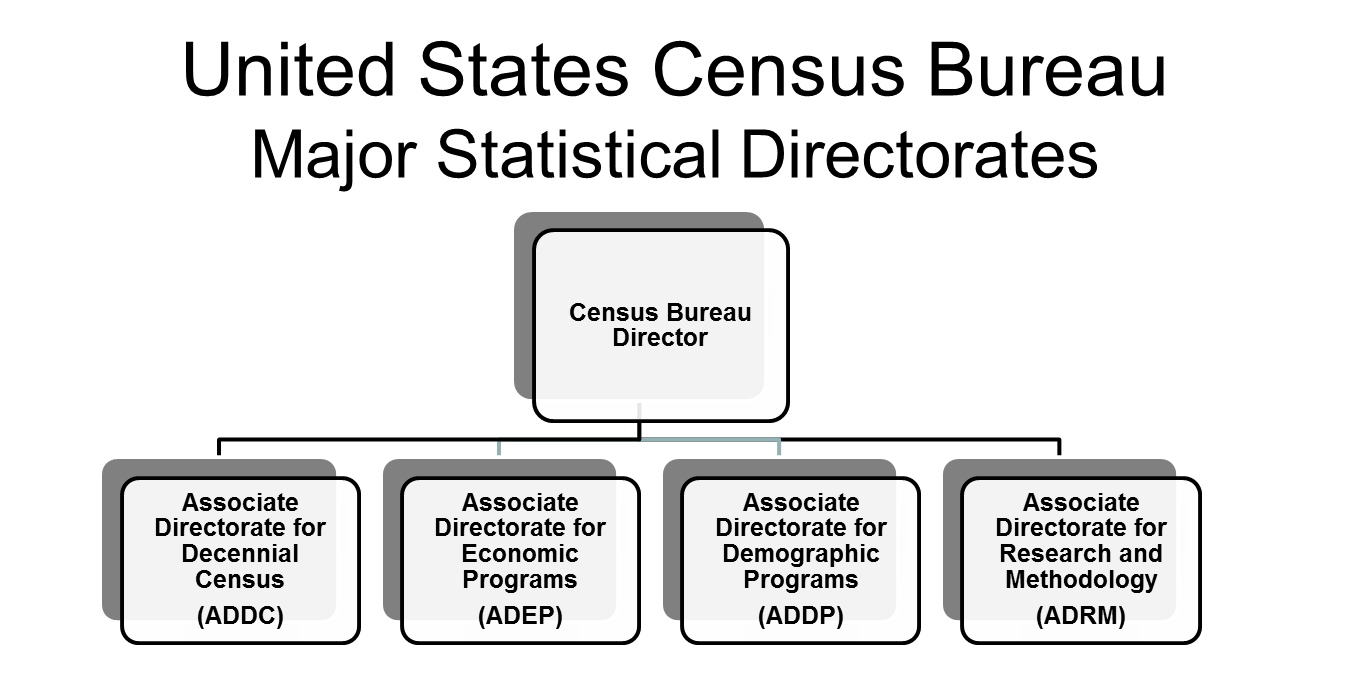
1.2 Initial Response

The Director of the Census Bureau tasked the Quality Program Staff (QPS) to develop a process to handle erroneous dissemination incidents to serve multiple purposes:

* Stress the importance of informing leadership when incorrect data is released to the public.
* Inform the public as soon as possible that the data in question has errors.
* Identify and correct the source of the error.
* Focus on finding the source of the error not affixing blame to individuals.
* Track incidents over time to determine if systemic problems needed corporate solutions.

**2. Development of the Process**

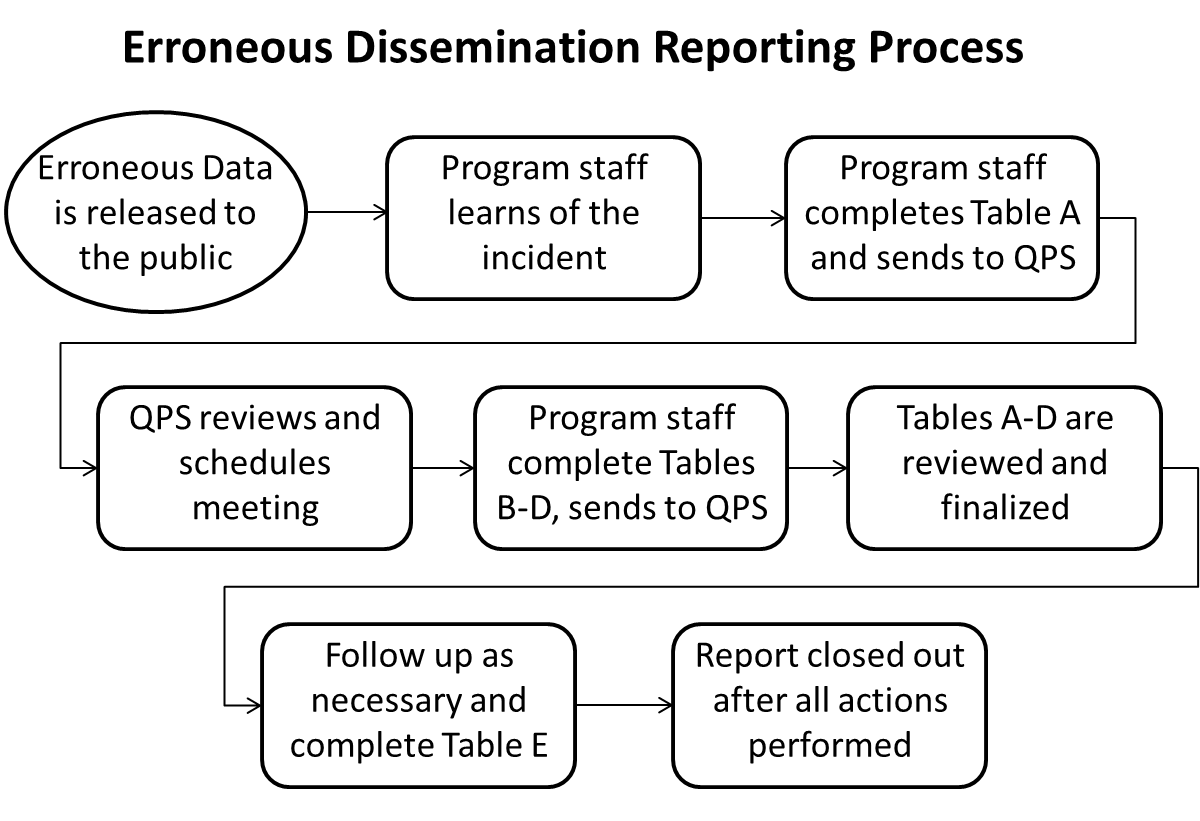
To fully understand this paper and how the process works, one needs a quick understanding of the organization of the Census Bureau. The Bureau is comprised of four major statistical Directorates outlined below:



QPS, which is part of ADRM, took this direction and developed a process to address these goals in a way that would not be onerous upon already over-taxed staff resources. First, QPS defined an erroneous dissemination as an error accessible to entities outside of the Census Bureau. This process does not cover errors found prior to leaving the Bureau at any level. The Director requested that the Associate Director for a given program area be notified of any erroneous data that was disseminated. This meant the head of the statistical directorate would be the responsible individual to determine whether the incident was serious enough to warrant informing the Director and/or Deputy Director. The reason this is important is that low- and mid-level managers may not understand the full implications, both in the public and political realm, of a particular error in an information product. Therefore, this process took the decision out of their hands.

2.1 Developing the Erroneous Dissemination Report

The QPS developed a form for program areas to use to report erroneous dissemination incidents. The form is broken into tables A through E, each table addressing different aspects of an incident and focusing on the essential parts of what happened, what caused the issue, what contributed to it, and what was done to rectify the situation. For the last table, the QPS follows up with the program area at regular intervals to document their efforts to detect and prevent similar problems from recurring. The report format is too long to include here but can be found on the Census Bureau’s website.[[[2]](#footnote-2)]



2.1.1 Table A – Documentation of the Incident

Table A contains 17 questions to document the nature of the erroneous dissemination the incident including a written description of what occurred and initial assessment of severity. These questions include:

* The description of the incident
* What type of error release occurred (inaccurate/incomplete data, wrong file, disclosure, geographic, display, or other)
* Who reported the incident when, and who was notified about the incident (at a minimum, the Associate Director of that program)
* Steps taken to notify the public
* Other immediate steps taken to correct the situation

At a minimum, the staff must place a notification on the particular webpage with the erroneous data explaining that the data has errors and by when correct data is expected.

The staff that owns the data is to complete Table A within 24 hours of notification of the error. Upon completion, the staff sends the initial report to the QPS staff. QPS immediately reviews the document and ensures the minimum actions took place; the Associate Director was notified of the incident and the public was notified that the data has errors including an estimate for when correct data is expected. If these minimum actions did not take place, QPS staff contacts the program area and reiterates these minimum requirements. Once these actions occur, QPS coordinates with the program area to set up a meeting to discuss Tables B and C.

2.1.2 Table B – Origin of the Incident

Table B serves to focus the program area staff on identifying the one failure that causes the erroneous dissemination. The program area staff may enter only one origin of the error. If there are additional entries, the QPS staff will assist in narrowing the origin of error to one. We have eight categories to classify the error that attempt to follow a production life cycle. They are:

Table B: Origin of the Incident

|  |  |
| --- | --- |
| B1. Planning/development | B5. Analyzing data/reporting results |
| B2. Collecting/acquiring data | B6. Releasing information products |
| B3. Capturing and processing data | B7. Protecting confidentiality |
| B4. Producing estimates and measures | B8. Other |

Each of these main categories is broken into sub-categories to refine the exact source of the incident. For instance, if the error occurred in B2, collecting/acquiring data, the staff can put that it was in the Interviewing, Transmitting, Quality Checks, or Other sub-category of B2. We found in developing the report format that it is essential to narrow the cause of the error to one reason. By determining a precise cause of the error, the QPS can assist the program area in developing a proper remedy to prevent future occurrences.

2.1.3 Table C – Factors Associated With the Incident

In Table C, we determine all the associated factors that contributed to the incident. While we require that a single point of failure be identified, many other factors often go wrong that lead to the failure to detect and correct the error prior to release to the public. For example, a programmer may have typed a line of code that did not follow the specification guiding him or her. We could consider that the origin of the incident. However, a vaguely worded specification may have allowed for misinterpretation. The program may not have been properly tested or tested at all which would have caught the error. Then, the supervisor may not have checked the code as required. The program staff may not have conducted an adequate quality check on the data the program produced. Table C is where the staff captures these factors to ensure the full documentation of the scope of the incident. However, an individual’s name is not called out in the report, since one of the goals is to determine program-related issues, not issues with an individual.

2.1.4 Table D – Final Assessment of the Incident

Once the program area investigates and fully understands the incident, they fill out Table D. It includes a final assessment of the severity of the incident on the same one to five scale along with justification as to why it merits this assessment. Table D also documents the actions taken to ensure that the incident will not occur in the future. If the planned actions were incomplete at the time of the report, the program area would enter when they expect the actions to be completed. The program area uses the opportunity to assess any other vulnerabilities discovered during the course of their investigation into the incident.

2.1.5 Table E – Preventative Measures

This table is essential if Table D leaves open actions that the program area must complete to ensure the type of incident happens again. The program area states in Table D when it will complete the planned corrective actions. It is the QPS’ responsibility to follow up with the program area on or after that date to check that all planned actions are complete. The QPS fills out this table once it conducts this review. The QPS annotates whether the actions are complete or not. If they are not complete, the QPS will document when the next review will take place. The Erroneous Dissemination Report is complete once all program area actions are complete.

2.2 The Reporting Process

The QPS’ mandate was to develop a reporting process using the form that was non-adversarial and non-punitive. Senior management stressed to everyone that QPS was not to conduct this program to access blame but only to improve quality throughout the Census Bureau. In this light, the QPS works hard to understand exactly what occurred during a particular incident with less focus on who did, or did not do, a particular action. If a programmer wrote an incorrect line of code, the program area would report the coding error on the form, but not necessarily the name of the programmer who committed the error. The QPS considers that anyone could commit a particular error and the corrective action is applied to the program’s process, not its staff.

2.2.1 Initial Report

The program staff completes the initial report once the error is reported. The report may come from the public, other statistical agencies, or from within the Census Bureau. The staff completes Table A and forwards it to the QPS staff usually within twenty-four hours. Note that this is from the time the program staff learns of the erroneous data, not when the error originally occurred. There are cases where an error occurred in the past, sometimes more than a year ago, but does not become known until recently. Once the QPS receives the report, staff members ensure that the program area took proper immediate action (i.e., notifying the Associate Director, notifying the public, putting a note on the website, etc.). If any action is incomplete, the QPS contacts the program area and informs them of the unfulfilled requirement. Once the quick review is complete, the QPS contacts the program area to schedule a meeting to discuss the incident and go over the report form. This meeting normally takes place approximately two weeks after the program area makes the initial report. This is to ensure that the program area has time to research what happened to cause the erroneous dissemination.

2.2.2 Meeting to Discuss the Report

Approximately two weeks after the first report, the scheduled meeting takes place between QPS staff and the program area. The program area sends the completed Tables B-D to the QPS two to three days prior to the meeting. Again, the QPS strives to ensure this is a non-adversarial meeting. At the beginning of the meeting staff stresses that the goal is to document what occurred to cause the erroneous dissemination and prevent future occurrences. During this meeting, the QPS staff asks for a verbal explanation of the incident. The reason is to listen to how the program area describes the incident and compare it to what is in the report. What the QPS understood from reading the report prior to the meeting should match the verbal explanation. If the two explanations do not seem to match, the QPS asks clarifying questions to determine what in the report might be incorrect and need changes. The QPS corrects the report to ensure all parties agree that what is in the report clearly explains what occurred. Some incidents have relatively easy corrective actions that the program area completes prior to this meeting. If so, the QPS completes Table E and closes out the report. If there are actions that take longer to complete, the QPS will schedule the timeframe when to conduct the review.

At the conclusion of the meeting, the QPS staff amends the report as needed based on the discussion and forwards to all parties for review. The QPS finalizes the report once all parties agree to the verbiage.

2.2.3 Follow Up

If the program area completed all actions prior to the meeting, there is no follow up. If not all the proposed corrective actions were complete, the QPS schedules meetings to review progress periodically until all actions are complete. This is to ensure that current program work does not supersede taking the actions necessary to prevent future erroneous dissemination incidents. By conducting these reviews, the action plans do not collect dust on the shelf due to other work taking precedence.

**3 Lessons Learned from the Process and Some Results**

3.1 Lessons Learned

The goal of the program in addition to ensuring that the erroneous dissemination is not hidden within the Bureau is to improve processes to reduce future incidents in the future. To this end, the QPS reviewed the incidents to date and made the following observations.

3.1.1 The Reporting Process

The process itself got off to a slow start since staff members were unsure of the true intentions of the process. At first, there were only a few reports and most of those were simply because the error was too big and too well known to not report. The QPS has the authority to ask whether an incident that they hear about could be a dissemination incident but rarely exercises that option to avoid the perception of being the “Dissemination Police.” Once some reports were completed and employees realized that management’s goal was to correct problems and not punish staff more reports started coming in to the QPS. We found there was a direct correlation with reporting prevalence and Directorates which have their own internal audit programs.[[[3]](#footnote-3)]

3.1.2 Corporate Lessons Learned

Many errors causing erroneous dissemination are small and unique to a particular program. Solutions to these normally would not be useful for other Census programs to incorporate into their processes. However, we have found some types of error are systemic across the programs and even Directorates.

The source of a about half of errors occur when a long-standing program makes major changes to its methodology. The changes to programming, imputation, editing, etc. introduce opportunities for error by staff co-mingling the old and new. Old code is recycled and not always properly updated. Processes pull from the old parameter file instead of the new. The data analyst who performed editing for the last 20 years on the same program forgets to use the new process. The erroneous dissemination reporting process is bringing these issues to light and causing the Directorates to focus more attention to ensuring methodological changes do not cause erroneous dissemination.

Many errors occur when staff members manually transfer data files from one software program to another. For example, when copying an Excel spreadsheet table to PDF it is very easy to copy and paste the wrong column or row. Since many organizations name files sequentially by date, it is easy to accidently open FILE2013 instead of FILE2014. Since the files are formatted the same, once open it is hard to realize that one is looking at last year’s data. We have found that program areas committing this type of error, especially if it has led to a high-severity incident, are more likely to ensure they automate this type of process to minimize chance for human error.

Transfer of institutional knowledge was an issue in many incidents. When a staff member with 20 years experience retired and did not document his knowledge prior to leaving, inexperienced staff members were left to implement an unfamiliar process allowing the introduction for error.

We have noted that notifying the Associate Director, one of the main goals of the program, happens only 56% with the initial Table A report. Further analysis shows that the probability of notifying the Associate Director goes up as the severity of the incident increases. Notifying data users of the error immediately occurs about 60% of the time. Completing the correction of the data by the time of the meeting also occurs about 60% of the time. These numbers show the importance and need of the Erroneous Dissemination Reporting Program. If it were not required, these numbers would probably be even lower.

**4. Future**

The QPS is working with the Directorates to share lessons learned from the collection of these reports. Annual reporting to share what was learned in the last year is planned to ensure Directorates and program areas can learn from the experience of others. We are also developing a database to store the information gathered by the reports. The QPS is planning to analyze the data as we eventually collect enough reports to make statistical testing a plausible possibility. The long-term goal is that Census will have fewer of these incidents and when they do occur all actions are done in a timely manner and reports are properly completed.

1. [] Any views expressed are those of the author and not necessarily those of the U.S. Census Bureau. [↑](#footnote-ref-1)
2. [] http://www.census.gov/quality/standards/Appendix\_F1.doc [↑](#footnote-ref-2)
3. [] Steven S. Klement and Joel Fowler. *Implementing OMB’s Standards and Guidelines for Statistical Surveys in the Census Bureau’s Economic Directorate and Some Results*. Paper presented at the Q2010 European Conference on Quality in Official Statistics, Helsinki, Finland, 2010 [↑](#footnote-ref-3)