

A Data Quality Act-Based Blueprint for Correcting the Crisis in Federal Databases

Recommendations to the Administrative Conference of the United States

This document has been prepared in response to the Council's view that: "the [recommendation](#) [from the Administration and Management Committee] would benefit from additional empirical evidence concerning adverse effects associated with premature or excessive disclosure of information."

To fulfill this request we identified five databases which:

- have been reviewed by either the GAO or the respective agency Inspector General
- have demonstrated deficiencies which result in harm to the American public
- are classified as online searchable databases.

The promulgation of a final report by ACUS on this topic is of particular importance if for no other reason than to notify the American public of the existence of a statute which provides for an expeditious and science-based mechanism for the correction of errors in federal databases, the Data [Information] Quality Act (DQA).

Utilization of the DQA to correct errors in federal databases has the added advantage in that there is one agency in charge of the process—the Office of Information and Regulatory Affairs within the Office of Management and Budget.

Accolades to the Administration for the promotion of [data.gov](#). More specifically the Administration announced: "On May 9, 2013, President Obama signed an [executive order](#) that made open and machine-readable data the new default for government information. Making information about government operations more readily available and useful is also core to the promise of a more efficient and transparent government." [N. B. note the President's emphasis on [data.gov](#) and open "data" thus the emphasis on the *Data Quality Act*].

Readily available data is useless if it is not accurate; strict adherence to the DQA will ensure the accuracy of federal databases. To this end, the ACUS recommendation should be expanded to include a recommendation that the managers of [data.gov](#) announce on this [page](#) the availability of the DQA to correct errors therein. We see no possible reason for not including such a recommendation and look forward to an informed discussion of this proposal.

To this end the [data.gov](#) website has already set the stage for such a recommendation in that it states: "All information accessed through Data.gov is subject to the Information Quality Act (P.L. 106-554). For all data accessed through Data.gov, each agency has confirmed that the data being provided through this site meets the agency's Information Quality Guidelines." In addition, the aforementioned recommendation is consistent with a [number](#) the databases within [data.gov](#) which make an explicit statement regarding the necessity for a data quality program.

As of this date, January 1, 2016, the [data.gov](#) website identifies 192,410 databases. These databases contain information which affect virtually every activity that a citizen could ever undertake. We expect this number to increase continually and consequently there is an absolute necessity for ACUS to

address this matter since the resultant information is integral to virtually all federal administrative processes.

Databases constitute the infrastructure of the modern administrative state and data is its lifeblood. When the data is contaminated with errors, federal agencies have difficulty performing even the most basic administrative functions such as managing its inventory of office space and protecting the personally identifiable information (PII) of social security number holders. The dissemination of flawed data wastes money and undermines public trust in government.

Federal studies have demonstrated that at least several important online searchable databases have serious deficiencies. Discussed below are several illustrative studies from 2015 that document problems stemming from erroneous data that is disseminated by agencies via online databases. As stated above, all of the studies are conducted by the General Accountability Office and Offices of Inspectors General and should provide a basis for an informed review by the Council.

ACUS's Committee on Administration and Management has made tremendous progress in developing a practical strategy for addressing and ameliorating the financial and human costs from erroneous data. Specifically, the Information Dissemination in the Internet Era project's Draft Recommendation¹:

- 1) Began the process of updating ACUS's Recommendation 73-1, Adverse Agency Publicity.
- 2) Recognized that an updated Recommendation needs to address a broader category of data than just "adverse agency publicity" and defined the term "information dissemination" to include "agency disclosure of information to the public that may affect persons identified in the disclosure, including such information when collected by agencies and released to the public through online searchable databases."²
- 3) Recommended that Information (Data) Quality Act, which provides a framework for "ensuring that information disseminated by agencies is not erroneous, misleading. . . ." ³ be used even by agencies not subject to it and that such agencies "should adopt procedures for accepting and responding to objections to information disseminated by the agency, and for correcting and retracting materially inaccurate statements. . . ." ⁴

The case examples below demonstrate a few of the problems incurred by federal officials and the public as a result of error-laden, online searchable databases disseminated by various agencies. These examples will be followed by Conclusions and Recommendations to ACUS on how to capitalize on the Committee's excellent work to date on the database data quality issue by taking advantage of statutory law and OIRA guidance documents specifically designed to ensure federal data quality.

¹ ACUS, Committee on Administration and Management, Information Dissemination in the Internet Era, "Draft Recommendation," November 4, 2015, available at <https://www.acus.gov/sites/default/files/documents/2015.11.04%20-%20Agency%20Information%20Dissemination%20Committee%20Approved.pdf>

² Id., p. 3.

³ Id., p. 6.

⁴ Id., p. 9.

Example 1: Social Security Database Errors Fuel Identify Theft

Agency/Database: Social Security Administration/Death Master File.

Public Access: <https://www.ssdmf.com/FolderID/1/SessionID/%7BA3218AFE-3B81-4F3F-9B1A-9C63D1A07514%7D/PageVars/Library/InfoManage/Guide.htm> [Limited Public Access Version]

Study: Social Security Administration, Officer of Inspector General, “Numberholders Age 112 or Older Who Did Not Have a Death Entry on the Numident,” [A-06-14-34030] March 4, 2015.

Database Function: Identify and prevent identity fraud. The public access version of the SSA’s Death Master File database was developed in response to the USA Patriot Act. [See, “SSDMF Database Uses” on the database’s public access page.]

Data Quality Findings: “The Numident includes approximately 6.5 million numberholders who, according to their dates of birth, were at least 112-years-old but did not have death information on their record. We reviewed available information in SSA records and identified further information indicating most of the numberholders were likely deceased.” [SSA OIG, p. 3]

Damage Done: “We matched the 6.5 million SSNs against SSA’s ESF and E-Verify systems and identified thousands of instances of potential identity theft or other fraud.” [SSA OIG, p. 6]

Example 2: Errors in GSA’s Database of Real Property

Agency/Database: General Services Administration/Federal Real Property Profile (FRPP)

Public Access: https://www.realpropertyprofile.gov/FRPPMS/FRPP_Login

Study: GAO, “FEDERAL REAL PROPERTY: Current Efforts, GAO Recommendations, and Proposed Legislation Could Address Challenges,” [GAO-15-688T], June 24, 2015.

Database Function: “describe[s] the nature, use, and extent of all real property under the custody and control of executive branch agencies in order to promote the efficient and economical use of the nation’s real property assets and assure management accountability for implementing reforms.” [GAO, p. 2]

Data Quality Findings: “In July 2014, the Administration released the first year results of the Freeze the Footprint policy, indicating that it exceeded its expectations for reducing the federal government’s office and warehouse space between fiscal years 2012 and 2013. However, as we reported in our 2015 High-Risk update, the data behind these results were unreliable, resulting in a potential overstatement of the progress made to date in reducing the federal government’s real property footprint. Specifically, we examined data from four of the six agencies claiming the largest reductions in the first year of implementation of the Freeze the Footprint policy and found that the actual space reductions at all four were overstated. . . .” [GAO, p. 4]

Damage Done: “Retaining unneeded real property results in significant costs to the federal government.” [GAO, p. 4]

Example 3: Army Corps of Engineers' Database of Dredging Contracts is Unreliable

Agency/Database: Army Corps of Engineers/Dredging Contracts Awarded

Public Access: <http://www.navigationdatacenter.us/dredge/dredge.htm>

Study: GAO, "ARMY CORPS OF ENGINEERS: Action Needed to Ensure the Quality of Maintenance Dredging Contract Cost Data" [GAO-15-810], September 29, 2015.

Database Function: "Information in the database is used for a variety of purposes, including tracking anticipated and actual project scheduling information, and tracking information across contracts on anticipated and actual contract costs and quantities of material dredged." [GAO, p. 8]

Data Quality Findings: "The total cost of maintenance dredging contracts during fiscal years 2004 through 2013 are unclear because data in the dredging database are unreliable." [GAO, p. 9]

Damage Done: "the Corps risks undertaking analyses and making conclusions on unreliable information, and may be missing opportunities to identify factors important to the management of maintenance dredging, such as cost elements contributing to changes in costs over time, or additional areas where it could take further actions to manage costs." [GAO, pp. 20-21]

Example 4: NRC’s Medical Events Database is Unreliable

Agency/Database: Nuclear Regulatory Commission/Medical Events Reporting Database

Public Access: <http://www.nrc.gov/site-help/search.html?site=events&q=medical&s=+#gsc.tab=0&gsc.q=medical&gsc.page=1>

Study: US Nuclear Regulatory Commission, Office of the Inspector General, “Audit of NRC’s Oversight of Medical Uses of Nuclear Material,” [OIG-16-A-02], October 8, 2015.

Database Function: “Medical events refer to a potential problem with how a medical facility uses radioactive material. These events may involve doses to a patient of the wrong amount, the wrong radioactive drug, incorrect administration of a drug, or dose to the wrong patient or wrong part of the body. . . . NRC analyzes each event to see if further action is needed.” [NRC OIG, pp. 5-6]

Data Quality Findings: “medical event reporting is unreliable due to the underreporting and retracting of events by licensees who do not consistently understand the requirements.” [NRC OIG, p. 13]

Damage Done: “NRC is not effectively achieving all the possible benefits of medical event reporting.” [NRC OIG, p. 13]

Example 5: Federal Procurement Data System Misrepresents Contracts to Small Businesses

Agency/Database: GSA/Federal Procurement Data System (FPDS)

Public Access: <https://www.fpds.gov/fpdsng/cms/index.php/en/>

Study: US Small Business Administration, Office of the Inspector General, “Report on the Most Serious Management and Performance Challenges in Fiscal Year 2016,” [Report Number 16-01], October 15, 2015.

Database Function: “FPDS provides a comprehensive web-based tool for agencies to report their contract actions. ... The Small Business Goaling Report (SBGR) is used by agencies to measure their respective accomplishments in relation to their annual small business program goals. The SBGR provides summary information about small business purchasing for all the departments in the Federal government.” [GSA, “Manipulation and Fraud in Reporting of VA Small Business Goals”]⁵

Data Quality Findings: “Congress has expressed concerns about the accuracy of the *Small Business Goaling Report*; this has been substantiated by OIG audits and other Government studies, which have shown widespread misreporting by procuring agencies, since many contract awards that were reported as having gone to small firms have actually been substantially performed by larger companies.” [SBA OIG, p. 1]

Damage Done: “Awards made to ineligible firms impact procurement opportunities for small businesses and damage SBA’s credibility in reporting accurate small business contracting goals achievements. . . . Without reliable data, SBA cannot accurately measure the Federal Government’s small business procurement goals achievements, which in turn weakens the ability of Congress and other Federal policy makers to determine whether the Government is maximizing contracting opportunities for small businesses.” [SBA OIG, p. 1]

⁵ Written Statement of Kevin Youel Page, Deputy Commissioner Federal Acquisition Service, General Services Administration Before the Subcommittee on Investigations, Oversight and Regulation on Manipulation and Fraud in Reporting of VA Small Business Goals, June 23, 2015, available at <http://www.gsa.gov/portal/content/234623>.

Conclusions

1. Unreliable data disseminated via agency online searchable databases harms the ability of federal agencies to efficiently manage the modern administrative state.
2. Unreliable data disseminated via agency online searchable databases originates from diverse sources including from every federal agency and from numerous outside parties.
3. The Data (Information) Quality Act was passed by Congress for the purpose of ensuring that “information disseminated by agencies is not erroneous, misleading. . . .”
4. OIRA’s government-wide “Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies,” were developed through a *Federal Register* notice-and-comment process in response to a statutory mandate to ensure and maximize the quality of federally disseminated data.
5. The DQA, which is applicable to federally-disseminated data irrespective of source and which includes an established administrative data correction process, should be the cornerstone of ACUS’s updated recommendation on agency adverse publicity.

Recommendations

1. The Data (Information) Quality Act and its implementing guidance should form the basis of an updated ACUS Recommendation on Adverse Agency Publicity.
2. The updated Recommendation should advise agencies to subject each database they disseminate to a DQA pre-dissemination review process to assess its reliability, the review process should include the opportunity for public comment.
3. The updated Recommendation should advise agencies to inform the public that the DQA is the appropriate mechanism for anyone seeking corrections to an agency-disseminated database.