



Decide with Confidence

Transparent Government Demands Robust Data Quality

*Federal initiatives to strengthen transparency and accountability
require agencies to improve data quality practices*

WHITE PAPER



Decide with Confidence

Table of Contents

Section I: Executive Summary	2
Section II: The Impact of Poor Data Quality.....	3
Section III: Best Practices in Data Quality	5
Section IV: The D-U-N-S® Number and Transparency in the Federal Government	8
Section V: Conclusion	9



Decide with Confidence

Section I: Executive Summary

Transparency is an important building block for improving government operations and accountability. But transparency requires that the data government agencies capture, store, and maintain be accurate, timely and complete. As data quality improves, so too will transparency's benefits.

The most significant benefit of both transparency and data quality is improved performance. The Obama Administration recognizes this important link. Among its transparency initiatives, the administration has created a new website, Recovery.gov, which will enable the public to monitor both spending and performance of American Recovery and Reinvestment Act programs. As part of the Recovery.gov effort, White House officials are requiring that each agency establish an "information quality" point-of-contact to ensure the accuracy and completeness of data reported to the new website.¹

The emphasis on transparency will certainly strengthen government performance, but agencies are discovering that achieving the desired level of transparency requires them to improve considerably the way they collect and manage data. And this can be a difficult challenge. Many agencies are grappling with a number of long-standing "data quality" problems, such as duplicate data in multiple systems, outdated or incomplete information, inability to consolidate information across multiple systems and the lack of actionable information.

Agency officials often think of data quality as a technical issue, not a business issue. Surprisingly, they do not fully realize its business impact on their organizations—in terms of cost, risk and performance. For example, border inspectors using outdated information on companies shipping goods into the United States create security risks. Gaps in contractor data can lead agencies to mistakenly award contracts to businesses that are unable to perform, thus increasing costs and introducing performance delays. The rapid meltdown of the U.S. financial sector caught regulators by surprise, in

¹ Peter R. Orszag, Director Office of Management and Budget, Memorandum on Initial Implementing Guidance for the American Recovery and Reinvestment Act of 2009, February 18, 2009.



Decide with Confidence

large part because they lacked transparency into the mortgage industry. The potential cost of poor data quality is very high indeed.

Despite these challenges, agencies are solving their data problems by adopting proven best practices for improving data quality:

- Define data standards, including metrics for adhering to those standards;
- Ensure data quality at the point of origin and at key checkpoints as data flows through an agency's systems and databases;
- Adopt a unique persistent key that identifies the entity and the corresponding data that relates to that entity;
- Implement a rigorous data maintenance strategy to update constantly changing information.

By adopting these best practices, government organizations are creating genuine transparency by providing accurate, useful data to government employees, entities doing business with the government and the public. And when data quality is assured, data becomes an asset that can strengthen the public-sector capabilities, and enable agencies to achieve their strategic goals.

Section II: The Impact of Poor Data Quality

Poor data quality can undermine the everyday operations and performance of an agency in multiple ways. It can increase risk by making agencies vulnerable to defaults and fraud—for example, in loan and grant programs—because they lack accurate or up-to-date information about the entities receiving assistance. Risks to public health or safety could rise if the relevant agencies lack transparency into the entities they oversee or do business with.

Poor data quality can also increase an agency's costs by causing resources to be misdirected (e.g., an inspector visiting the wrong facility). Moreover, additional resources may be needed to resolve data inconsistencies or gaps, maintain duplicate source files, rationalize and synthesize data for insight, and address other data problems. Agencies' ability to comply with laws and



The Cost of Poor Data Quality

The cost of data quality is not only dependent on the information systems in place but is also dependent on when an agency tackles a data problem and the resources required to do so. Here's an illustrative example:

Prevention
\$1X per Record
Resolution
\$10X per Record
Correction
\$100X per Record

regulations defining their fiduciary responsibilities is also hampered because poor data quality often leads to overpayments and other improper handling of funds. And poor data quality almost certainly results in decreased mission performance because agencies lack visibility into their operations and processes, staff are reluctant to use data they do not trust, and strategic decisions may be based on inaccurate or incomplete information. For example, the Government Accountability Office (GAO) reports that federal agencies make \$55 billion annually in improper payments to contractors, Medicare recipients and their doctors, assistance programs, and others—due to incomplete record keeping.² Similarly, agencies charged with regulating specific industries, such as the Food and Drug Administration (FDA), Environmental Protection Agency (EPA), and Securities and Exchange Commission (SEC), cannot effectively enforce regulations if they do not have accurate information about the corporations and the corporate activities they oversee.

The Consequences of Poor Data Quality

Increased Risk	Increased Costs
Supply Chain Disruption Loan Defaults Cargo Security Risk Public Safety Risk Environmental Risk Health Risk Reduced Transparency	Data Resolution Non-Mission Workload System Delays Increased Cycle Time Additional Resources Carrying Costs of Duplicates Reduced Interoperability
Increased Non-Compliance	Decreased Performance
Inaccurate Profile of Regulated Universe Improper Payments Overpayments Contractor Fraud Abuse Misuse of Public Funds Fraudulent Visa Applications	Inability to Create Single View of Entity Delayed or Lost Collections Unmatched Disbursements Limited Predictability Over or Under Reporting Poor Management Decisions Diminished Outreach

² "Improper Payments," (GAO-08-377R; January 23, 2008).



Even when agencies recognize that data quality issues are hindering performance, they often are uncertain about how to address the problems. Many are overwhelmed by the thought of changing longstanding processes and systems for gathering, storing and disseminating information on the entities they oversee or do business with. Some worry about how to set data management policies; others worry about the potential costs, both in terms of money and staff hours, to cleanse their data and maintain its quality. “I don’t know where to start,” is a commonly voiced concern.

Section III: Best Practices in Data Quality

Over the years, government and commercial organizations have implemented new processes and tools for ensuring that their databases store and make available accurate, up-to-date information. The specific solutions vary, depending on the specific needs of each organization. But experience shows that successful organizations typically follow a common set of best practices for achieving data quality:

Best Practices for Data Quality

- *Define data standards, including metrics for adhering to those standards;*
- *Ensure data quality at the point of origin and at key checkpoints as data flows through an agency’s systems and databases;*
- *Adopt a unique persistent key that identifies the entity and the corresponding data that relates to that entity;*
- *Implement a rigorous data maintenance strategy to update constantly changing information.*

1. Define data standards. The breath of data standards depends on several factors: an organization’s complexity; its integration needs; and the desired level of transparency. Some agencies may need transparency within individual programs, while others may need transparency across all offices within the agency. And some, such as the Integrated Acquisition Environment (IAE), may need visibility into specific information throughout the entire government.³ Understanding the level of transparency needed will help agencies define the data standards and systems needed to support those standards.

Standard setting is not a one-time project. Agencies must establish and periodically review standards for data quality, including metrics on whether the standards are being met. At a minimum, the metrics should encompass these four attributes of data quality:

³ The Integrated Acquisition Environment (IAE) is a Presidential E-Government Initiative managed by the General Services Administration (GSA). Among its goals, IAE strives to increase sharing of acquisition data among federal agencies, while creating a common, integrated business process for federal buyers and sellers that promotes competition, transparency and integrity.



Decide with Confidence

- Accuracy—the right information on the right entity;
- Completeness—information reflects all relevant attributes of the entity;
- Timeliness—the most up-to-date, current information available;
- Consistency—the information is consistent—i.e., no conflicting information—across the relevant organizational scope.

“Given the high priority placed on the accurate display of information related to Recovery.gov ...all agencies must ensure all reporting related to Recovery Act funding is complete and accurate and complies with the agency’s Information Quality Act guidelines.”

*Peter R. Orszag
Director of the Office of
Management and Budget,
OMB Memo, February 18, 2009*

2. Ensure data quality at the point of origin and at key checkpoints as data flows through an agency’s systems and databases. Agencies must put in place the data models and collection mechanisms that support and validate information when it enters their systems. Validating information can be achieved by using a “referential file”—an independent, third-party, authoritative source that can check the incoming data—thus preventing flawed data from populating databases. At the same time, many organizations also validate data as it flows downstream between key databases within their systems. Tackling data quality issues upstream ensures data integrity throughout all systems downstream.

For some agencies, addressing data quality at the front end (“upstream”) poses high costs and organizational and technical challenges. In such instances, improved data quality can be achieved through data warehousing solutions, which tend to “fix” the data on the back end (“downstream”) as data from multiple sources is extracted, cleansed, transformed, and loaded into the data warehouse environment. Although data warehouses provide a “one source of truth,” operational inefficiencies and lack of standards in transactional systems can continue, thus increasing an organization’s data maintenance costs and placing more burden on the data warehouse environment to deliver high-quality data.

3. Adopt a unique persistent key that identifies the entity and the corresponding data that relates to that entity. Organizations use many different types of persistent keys, depending on their needs. One of the most common is the Social Security Number, which both government and businesses use to identify and track individuals. In the state and local government sector, the SSN is often used to identify sole proprietor businesses for tax and procurement purposes but raises increasing concerns in government about privacy and potential identity theft issues associated with the use of the SSN for business identification. There are a handful of



Decide with Confidence

unique persistent keys that are assigned to businesses and other entities, including the D-U-N-S[®] Number and Global Location Number (GLN).

The Rapid Pace of Change

A quick look at business change during normal economic times – clearly demonstrates the need for an ongoing data maintenance strategy.

Every 14 seconds – a judgment is filed against a business

Every 60 seconds – a business risk profile changes

Every 60 seconds – a business opens its doors

Every 2 minutes – a directorship (CEO, CFO, etc.) change occurs

Every 3 minutes – a business ceases operations

Every 5 minutes – a business files for bankruptcy

Every 30 minutes – a business changes their name

Every 4 hours – a business ownership changes

The importance of selecting the most appropriate or “right” unique persistent key cannot be overstated. This choice depends in large measure on the scope of its use. For example, will the agency use the unique persistent key to track business data in just a single program, across the entire organization, or throughout the entire government? Similarly, Social Security Numbers might be an appropriate unique persistent key for tracking individuals through an organization’s systems and databases, but D-U-N-S Numbers or GLN would be better suited to tracking businesses and their associated business data. Without the right unique persistent key, organizations will struggle to manage data effectively.

4. Implement a rigorous data maintenance strategy to update constantly changing information. To ensure that data remains accurate and useful, an agency must have a strategy that prescribes: a) who will update information; b) how often (or when) the information will be updated; c) how the information will be updated and validated; and d) how data quality will be measured on an on-going basis.

Whether agencies try to ensure data quality at the front end (point of origin) or the back end (data warehouse); they should conduct a data quality assessment as a key input to their data maintenance strategy. A data quality assessment identifies an agency’s data problems, such as the amount of duplicate information, number of conflicts from multiple sources, percentage of values out of range, and how outdated the data may be.

Many organizations undertake data maintenance strategies in-house, especially if their needs are relatively small and simple. However, government agencies and commercial enterprises are not typically in the business of maintaining constantly shifting business data, and so they find it valuable to outsource this important function to third-party experts. (To understand how rapidly business information changes refer to the sidebar).



Decide with Confidence

Section IV: The D-U-N-S® Number and Transparency in the Federal Government

The growing momentum for transparent government has coincided with a growing use within government of the D-U-N-S® Number as an identification code. This should not be surprising. Transparency and the D-U-N-S Number go hand in hand.

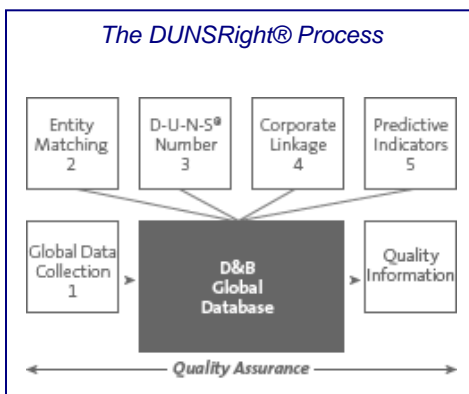
“We recommend the D-U-N-S® Number because of its universal use in industry and government, nationally and internationally, as a unique supplier identifier and because it is the only reliable mechanism for crosswalking to other numbering systems.”

Federal Electronic Commerce Acquisition Team, 1994, recommending the D-U-N-S® Number be adopted government-wide as a contractor identifier within the Central Contractor Registration (CCR) system.

The D-U-N-S Number, originally the Data Universal Numbering System, is Dun & Bradstreet’s unique means of identifying and tracking an entity globally throughout all phases of its life. Because the D-U-N-S Number is an industry standard, universally recognized and used throughout the globe, the U.S. government in 1994 adopted the D-U-N-S Number as a unique identifier for organizations doing business with the government. Consequently, federal contractors must now provide a D-U-N-S Number when registering in the Central Contractor Registration (CCR) system to provide products or services, so that agencies can use the number to track contractors in their acquisitions systems. Grants also require a DUNS Number which led to the D-U-N-S Number being used as the unique identifier for usaspending.gov.⁴

Over time, as the D-U-N-S® Number became more widely adopted within government, agencies discovered that it can be used for more than just tracking customers. Many agencies are using the D-U-N-S Number to uncover redundancies, eliminate inconsistencies and update incomplete information, thus significantly improving data quality and enabling transparency.

In addition, the D-U-N-S Number enables agencies to tap into D&B’s DUNSRight® process to further strengthen data quality. The DUNSRight process contains five main components or drivers that work sequentially to



⁴ The Federal Funding Accountability and Transparency Act (FFATA) requires information on federal awards (federal financial assistance and expenditures) be made available to the public via a single, searchable website, now commonly known as usaspending.gov.



Decide with Confidence

aggregate, standardize, enrich and, ultimately, create the highest quality business information:

1. **Global Data Collection**—gathers detailed information about more than 146 million global entities from 20,000 worldwide data sources.
2. **Entity Matching**—integrates data into the D&B global database through a process that cleanses and validates information to create a single, more accurate picture of each business entity.
3. **D-U-N-S Number**—tracks entities through every business activity through the lifetime of a business with a unique persistent key.
4. **Corporate Linkage**—reveals the entire universe of an entity's connections, such as parent companies, branches and subsidiaries so agencies can understand the total risk associated with that entity.
5. **Predictive Indicators**—uses statistical analysis to provide risk information about entities, such as their likelihood of failure or ability to make payments.

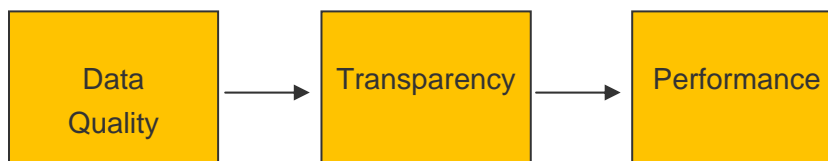
The D-U-N-S Number is the lynchpin of the DUNSRight process for collecting, cleansing and enhancing information in Dun & Bradstreet's global database of public and private entities. Government organizations, of course, tailor solutions to fit their unique systems and requirements. But federal, state and local governments can use the D-U-N-S Number on its own or with the more comprehensive DUNSRight process as a foundation for implementing best practices in data quality.

Section V: Conclusion

Transparency is a process, not an endpoint. And data quality is key to that process. When the data that agencies collect and store is accurate, timely, consistent and complete, the barriers to transparency will be removed. Increased transparency will, in turn, enable improved performance and all its intended benefits.



Decide with Confidence



- Complete names
- Complete addresses
- Complete profiles

- Who are we regulating/overseeing?
- Where are they located?
- Who owns them?
- What is the total business relationship?
- How risky are they?

- Accountability
- Fraud Prevention
- Oversight
- Public Safety
- Streamlined Procurement
- Vendor Management

Many agency officials tend to dismiss concerns about data quality as technical or process issues unrelated to performance. But as agencies know only too well, inaccurate or poor data can hamper their ability to carry out mission activities. And conversely, strengthening data quality is a key step for improving operations and achieving an agency's goals.

For more information on D&B's Data Quality approach, call 1.800.424.2495 or email government@dnb.com

When information is reliable—and agency staff have confidence in that information—then data becomes a strategic asset with multiplying benefits for transparency, accountability and performance.