



**Decide with Confidence**

## **D&B Webinar:**

**Ensuring Data Quality in a Transparent Government**

Dial in: 800-857-6699

Password: D&B Webinar

June 2009

**If you have problems logging in, please call 1.800.424.2495**

# Today's speaker



Decide with Confidence



## **Pelin Aylangan**

AVP, Federal Consulting  
Government Solutions

Dun & Bradstreet, Inc.

Dial in: 800-857-6699 / Password: D&B Webinar / [www.dnb.com/gov](http://www.dnb.com/gov)

# Transparency is an important building block for improving government performance and accountability – and data quality is key



Decide with Confidence

- Transparency requires that all data government agencies capture, store and maintain is accurate, timely and complete
- Recovery.gov has been created to enable the public to monitor both spending and performance associated with American Recovery and Reinvestment Act
- White House is requiring each agency establish an “information quality” point-of-contact to ensure accuracy and completeness of data reported
- The most significant benefit of both transparency and data quality is improved performance



Dial in: 800-857-6699 / Password: D&B Webinar / [www.dnb.com/gov](http://www.dnb.com/gov)

# Agencies struggle with long-standing data quality problems



Decide with Confidence

- Duplicate data within a system or across multiple systems
- Outdated or incomplete information
- Inability to consolidate information across multiple systems
- Lack of actionable information

Dial in: 800-857-6699 / Password: D&B Webinar / [www.dnb.com/gov](http://www.dnb.com/gov)

Agency officials often think of data quality as a technical issue, not a business issue and do not realize its business impact on their agencies



Decide with Confidence

Increased Risk	Increased Costs
<ul style="list-style-type: none"> <li>Supply Chain Disruption</li> <li>Loan Defaults</li> <li>Cargo Security Risk</li> <li>Public Safety Risk</li> <li>Environmental Risk</li> <li>Health Risk</li> <li>Reduced Transparency</li> </ul>	<ul style="list-style-type: none"> <li>Data Resolution</li> <li>Non-Mission Workload</li> <li>System Delays</li> <li>Increased Cycle Time</li> <li>Additional Resources</li> <li>Carrying Costs of Duplicates</li> <li>Reduced Interoperability</li> </ul>
Increased Non-Compliance	Decreased Performance
<ul style="list-style-type: none"> <li>Inaccurate Profile of Regulated Universe</li> <li>Improper Payments</li> <li>Overpayments</li> <li>Contractor Fraud Abuse</li> <li>Misuse of Public Funds</li> <li>Fraudulent Visa Applications</li> </ul>	<ul style="list-style-type: none"> <li>Inability to Create Single View of Entity</li> <li>Delayed or Lost Collections</li> <li>Unmatched Disbursements</li> <li>Limited Predictability</li> <li>Over or Under Reporting</li> <li>Poor Management Decisions</li> <li>Diminished Outreach</li> </ul>

Dial in: 800-857-6699 / Password: D&B Webinar / [www.dnb.com/gov](http://www.dnb.com/gov)

The cost of data quality is not only dependent on the information systems in place but also on when an agency tackles a data problem



Decide with Confidence

<b>Prevention</b>	<b>\$1X per Record</b>
<b>Resolution</b>	<b>\$10X per Record</b>
<b>Correction</b>	<b>\$100X per Record</b>

Dial in: 800-857-6699 / Password: D&B Webinar / [www.dnb.com/gov](http://www.dnb.com/gov)

# Proven best practices can help agencies ensure data quality, transparency and ultimately improve mission performance



Decide with Confidence

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

Dial in: 800-857-6699 / Password: D&B Webinar / [www.dnb.com/gov](http://www.dnb.com/gov)

# Define data standards, including metrics for adhering to those standards



Decide with Confidence

# 1

- The breadth of data standards depends on several factors – work with key stakeholders to determine:
  - Organization’s complexity
  - Integration needs
  - Desired level of transparency
- Identification of data quality attributes important to your organization is key to creating standards
- Standard-setting is not a one-time project

Numerous data quality attributes are available  
– it is incumbent upon your agency to identify those that are relevant to your mission



Decide with Confidence

Data Quality Attribute	Description
Accuracy	Ensure data are the correct values, valid and attached to the right entity
Accessibility	Data should be easily accessible and legal to access with strong guidelines built in
Comprehensiveness	All required data items are included; ensure scope of data is collected and document limitations
Granularity	Attributes and values of data should be defined at the correct level of detail
Consistency	Value of data should be reliable and the same across applications
Definition	Each data element should have clear meaning so current and future users will know what the data element means
Precision	Data values should be large enough to support the application or process
Relevancy	The data are meaningful to the performance of the process for which they are collected
Timeliness	Timeliness should be determined by how the data is being used
Currency	Data should be up-to-date

# At a minimum, the metrics should encompass these four attributes of data quality



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.

# Example when records entered where no data standards exist



Decide with Confidence

**Business Name:** Gorman Manufacturing  
**Address Line 1:** 1 Main St. Suite 200  
**Address Line 2:** Attn: Paul Smith  
**City:** San Francisco  
**State:** CA  
**Zip Code:** 90210  
**Country:** USA  
**Line of Business:** Plumbing Materials

→ **Address Line 2 contains a person's name**

**Business Name:** ABC Inc  
**Address Line 1:** 500 Beverly Way  
**Address Line 2:** Suite 650  
**City:** Hollywood  
**State:** California  
**Zip Code:** 90210  
**Country:** United States  
**Line of Business:** 5645

→ **State contains spelled out name, and incorrect version**

→ **Line of Business is numeric, whereas others are text**

**Business Name:** Road Pavers Inc  
**Address Line 1:** 17 Concrete Ave  
**Address Line 2:** Finance Dept  
**City:** Sacramento  
**State:** CS  
**Zip Code:** 90431-1111  
**Country:** U.S.A  
**Line of Business:** Unknown

→ **Zip Code contains 9 characters, whereas others are 5**



# Example after data standards imposed

Data Element	Standard
Business Name	The <b>legal entity that we are issuing a license to</b> 70 characters, alphanumeric
Address Line 1	The physical location of the entity that is applying for the license 70 characters, alphanumeric
Address Line 2	The secondary address line (e.g. Suite, Office, Floor) 70 characters, alphanumeric
City	The full city name of the city where the legal entity is located
State	The <b>2-character abbreviation</b> of the state where the entity is located, Values should be table driven
Postal Code	The 5 digit code + 4 digit extension Must be a valid USPS code
Country	The 3 character ISO Code for countries Will be table driven
Line of Business	The 4 digit SIC Code that describes the primary operations of the business at the time they applied for the license <b>Can not be blank, table driven</b>

**Granularity**

**Precision**

# Example when records conform to data standards



Decide with Confidence

**Business Name:** Gorman Manufacturing  
**Address Line 1:** 1 Main St. Suite 200  
**Address Line 2:**  
**City:** San Francisco  
**State:** CA  
**Zip Code:** 90210-1234  
**Country:** USA  
**Line of Business:** 5645

**Business Name:** ABC Inc  
**Address Line 1:** 500 Beverly Way  
**Address Line 2:** Suite 650  
**City:** Hollywood  
**State:** CA  
**Zip Code:** 90210-1234  
**Country:** USA  
**Line of Business:** 5645

**Business Name:** Road Pavers Inc  
**Address Line 1:** 17 Concrete Ave  
**Address Line 2:**  
**City:** Sacramento  
**State:** CA  
**Zip Code:** 90431-1111  
**Country:** USA  
**Line of Business:** 5645

**When standards are imposed, the data enables more transparency. For example:**

- *How many licenses did we issue to businesses in CA?*
- *How many businesses with SIC 5645 do we regulate?*

# Ensure data quality at the point of origin and at key checkpoints as data flows through an agency's systems and databases



Decide with Confidence

## 2

- Put into place the data models and collection systems that support and validate information as it enters system
- Tackle data quality at the front end - or can fix data on the back end in a data warehouse environment
- Use of independent, third-party referential file on entities to help validate information



# Example using front-end validation

## User 1 Input:

**Business Name:** Gorman Manufacturing  
**Address Line 1:** 1 Main St. Suite 200  
**Address Line 2:** Attn: Paul Smith  
**City:** San Francisco  
**State:** CA  
**Zip Code:** 90210  
**Country:** USA  
**Line of Business:** Plumbing Materials

## User 1 Record Stored As:

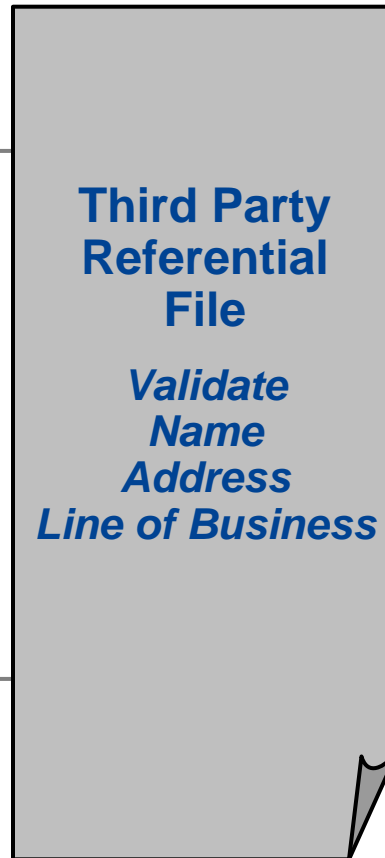
**Business Name:** Gorman Manufacturing  
**Address Line 1:** 1 Main St. Suite 200  
**Address Line 2:**  
**City:** San Francisco  
**State:** CA  
**Zip Code:** 90210-1234  
**Country:** USA  
**Line of Business:** 5645

## User 2 Input:

**Business Name:** Road Pavers Inc  
**Address Line 1:** 17 Concrete Ave  
**Address Line 2:** Finance Dept  
**City:** Sacramento  
**State:** CS  
**Zip Code:** 90431-1111  
**Country:** U.S.A  
**Line of Business:** 5643

## User 2 Record Stored As:

**Business Name:** Road Pavers Inc  
**Address Line 1:** 17 Concrete Ave  
**Address Line 2:** Suite 650  
**City:** Sacramento  
**State:** CA  
**Zip Code:** 90431-1111  
**Country:** USA  
**Line of Business:** 5645





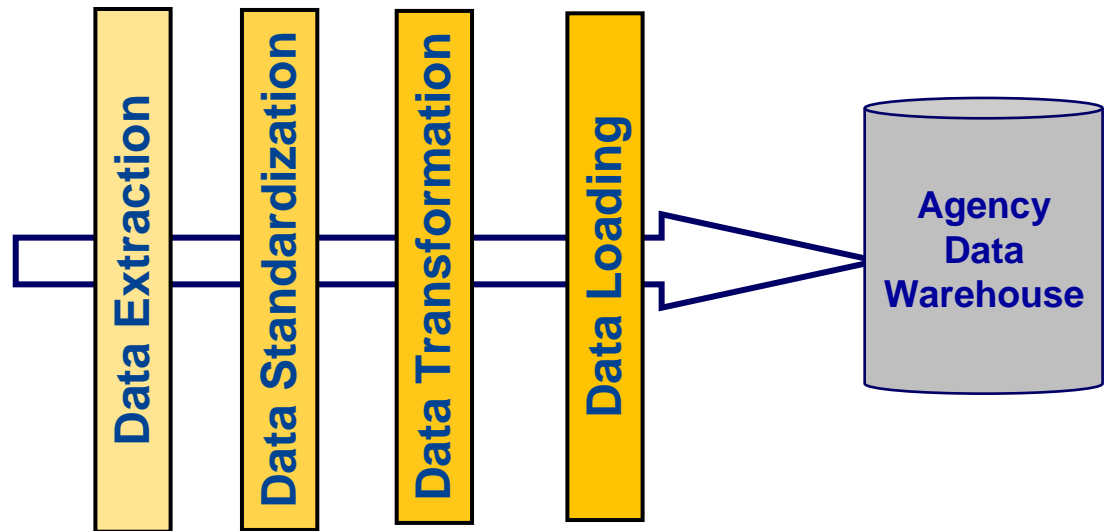
# Example using back-end rationalization

**Business Name:** Gorman Manufacturing  
**Address Line 1:** 1 Main St. Suite 200  
**Address Line 2:** Attn: Paul Smith  
**City:** San Francisco  
**State:** CA  
**Zip Code:** 90210  
**Country:** USA  
**Line of Business:** Plumbing Materials

**Business Name:** ABC Inc  
**Address Line 1:** 500 Beverly Way  
**Address Line 2:** Suite 650  
**City:** Hollywood  
**State:** California  
**Zip Code:** 90210  
**Country:** United States  
**Line of Business:** 5645

**Business Name:** Road Pavers Inc  
**Address Line 1:** 17 Concrete Ave  
**Address Line 2:** Finance Dept  
**City:** Sacramento  
**State:** CS  
**Zip Code:** 90431-1111  
**Country:** U.S.A  
**Line of Business:** Unknown

Rules must be established for dealing with each data element, as well as how errors/rejects will be resolved



Adopt a unique persistent key that identifies the entity and the corresponding data that relates to that entity



Decide with Confidence

3

- The importance of selecting the most appropriate unique persistent key cannot be overstated
- The choice depends on the scope of use – for example, tracking data...
  - In a single program
  - Across the organization
  - Across the entire government

# Reviewing key attributes of numbering schemes can help you identify the one that best meets your agency entity management needs



Decide with Confidence

- What level of granularity can be achieved, for example does the “number” represent a legal entity, a physical site, a physical structure, something else?
- Does the identifier stay with entity for their lifetime (or can it be re-assigned)?
- Will the length/type fit in your field formats, i.e. numeric vs alpha numeric?
- Does your numbering scheme need to be hierarchal?
- Does it need to handle cross border translations?
- How and how often is the numbering system maintained?
- Do you need the clout of an industry standard?

# Example of when a unique identifier is not used



Decide with Confidence

**Inspection  
System  
Record #1**

Gorman Printing Co., Inc.  
492 Koller Street  
San Francisco, CA 94110  
415-555-0000  
Leslie Smith, President

**Inspection  
System  
Record #2**

Gorman Paper Products  
492 Koller St  
Pacifica, CA 94110  
SIC 7374

**HazMat  
Permitting  
System**

Gorman Paper Products  
492 Koller St  
Pacifica, CA 94110  
SIC 7374

**Financial  
Mgmt System  
Record #1**

Gorman Paper  
P.O. Box 111  
San Francisco, CA 94110  
Fine to be collected: \$5,000

**Financial  
Mgmt System  
Record #2**

GPC  
492 Kooler St  
San Francisco, CA 94110  
415-555-0000  
Fine to be Collected: \$3,000

**Can the agency correctly answer these questions:**

What is the total amount of fines to be collected from Gorman?

How many times have we inspected Gorman in the last 2 years?

How many inspectors will we need to cover San Francisco?



# Example of when a unique identifier is used

## Inspection System Record #1

Gorman Printing Co., Inc.  
492 Koller Street  
San Francisco, CA 94110  
415-555-0000  
Leslie Smith, President

## Inspection System Record #2

Gorman Paper Products  
492 Koller St  
Pacifica, CA 94110  
SIC 7374

## HazMat Permitting System

Gorman Paper Products  
492 Koller St  
Pacifica, CA 94110  
SIC 7374

## Financial Mgmt System Record #1

Gorman Paper  
P.O. Box 111  
San Francisco, CA 94110  
Fine to be collected: \$5,000

## Financial Mgmt System Record #2

GPC  
492 Kooler St  
San Francisco, CA 94110  
415-555-0000  
Fine to be Collected: \$3,000

## Standardized Record with Unique ID

D-U-N-S #:80-480-0217  
Gorman Printing Co., Inc.  
+ Gorman Paper Products  
+ GPC  
P.O. Box 111  
San Francisco, CA 94110  
492 Koller Street  
San Francisco, CA 94110  
415-555-0000  
Leslie Smith, President  
  
HazMat Permit: Yes  
Inspected: Yes  
Fine to be Collected: \$8,000

### How do the agency's answers to the questions change?

- What is the total amount of fines to be collected from Gorman?
- How many times have we inspected Gorman in the last 2 years?
- How many inspectors will we need in California?

# Implement a rigorous data maintenance strategy to update constantly changing information



Decide with Confidence

# 4

- Perform a data quality assessment as a benchmark that will identify:
  - Amount of duplicate information
  - Number of conflicts from multiple sources
  - Percentage of values out of range
  - How outdated the data may be
- Develop a strategy to ensure data remains accurate and identifies:
  - who will update information
  - how often (or when) information will be updated
  - how the information will be updated and validated
  - how data quality will be measured on an ongoing basis
- Create a dashboard or scorecard for easy reporting on metrics

# The rapid pace of business change clearly demonstrates the need for ongoing data maintenance strategy



Decide with Confidence

## **EVERY...**



**4 hours**

a business' ownership changes

**14 seconds**

a judgment is filed against a business

**8 minutes**

a business files bankruptcy

**60 seconds**

a business' risk profile changes

**3 minutes**

a business ceases operations

**2 minutes**

a directorship (CEO, CFO, etc.) change occurs

**2 minutes**

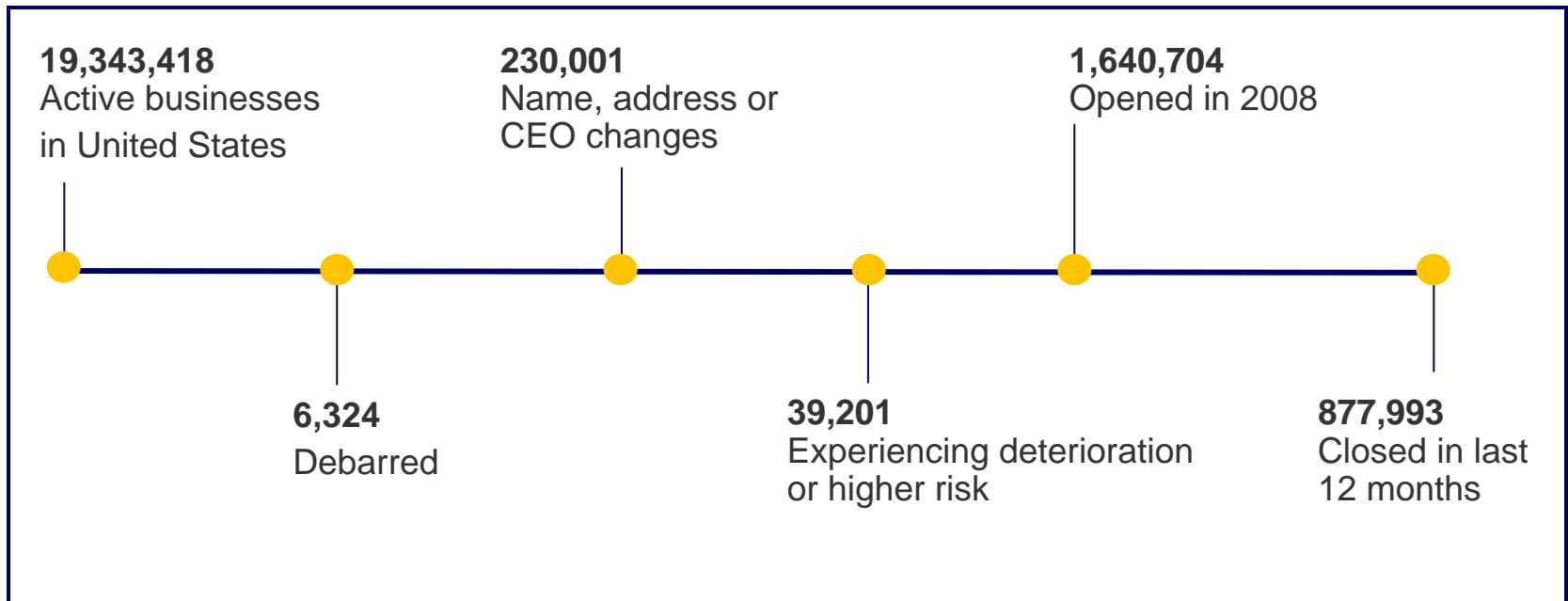
a new business opens

...and even more so today with the current economic conditions



Decide with Confidence

### Business activity over the past year nationwide



# Proper data maintenance enables transparency and performance



Decide with Confidence

**Are the businesses who filed taxes still in business today?**

**Will we need to inspect more or fewer facilities this year, and do we have enough inspectors to handle the load?**

**How many entities in the Oil Refinery industry have undergone an acquisition or merger in the last year?**

**Do we have valid names/addresses and industry classification for outreach to life sciences businesses in neighboring states?**

**How many new businesses applied for housing construction permits in the last 9 months?**

- How often is business status updated in systems?
- Can we accurately determine the number of active facilities, and how many will require inspection?
- Who is responsible for maintaining this information, the agency or the business?
- How do our systems know about name and address changes, how is SIC/NAICS Code assigned?
- Can your data maintenance strategy detect duplicates?



**Decide with Confidence**

The D-U-N-S<sup>®</sup> Number and Transparency  
in the Federal Government

# D&B's D-U-N-S Number was adopted by the U.S. Government as the unique identifier in the Central Contractor Registration (CCR)



Decide with Confidence

*“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**

In order to improve the statistical reporting of **federal grants and cooperative agreements**, the Office of Management and Budget (OMB) has directed federal agencies to require all applicants to provide a Dun and Bradstreet DUNS Number when applying for Federal grants or cooperative agreements on or after October 1, 2003.

Federal Acquisition Regulation (FAR) – October 2003

CCR portal is a registration portal for all government contracts that validates key business information at the point of origin



Decide with Confidence



Home | Contractors | Grantees | International Registrants | Small Businesses

Start New  
Registration

Update or  
Renew  
Registration

Search CCR

Dynamic  
Small  
Business  
Search

Request  
Access to  
CCR Data

## Welcome to Central Contractor Registration (CCR)

Central Contractor Registration (CCR) is the primary registrant database for the U.S. Federal Government. CCR collects, validates, stores and disseminates data in support of agency acquisition missions. Click [here](#) to learn more about CCR Policy and Background.

**ATTENTION EXISTING USERS:** CCR login procedures have changed. If you currently have a CCR record, but do not have a user ID and password, you will not be able to access your record unless you **extended the deadline to make this change to December 21, 2011** prior to your renewal date.

Please take a moment to review the [User Account Guide](#) before getting you through the process.

Related

[O](#)

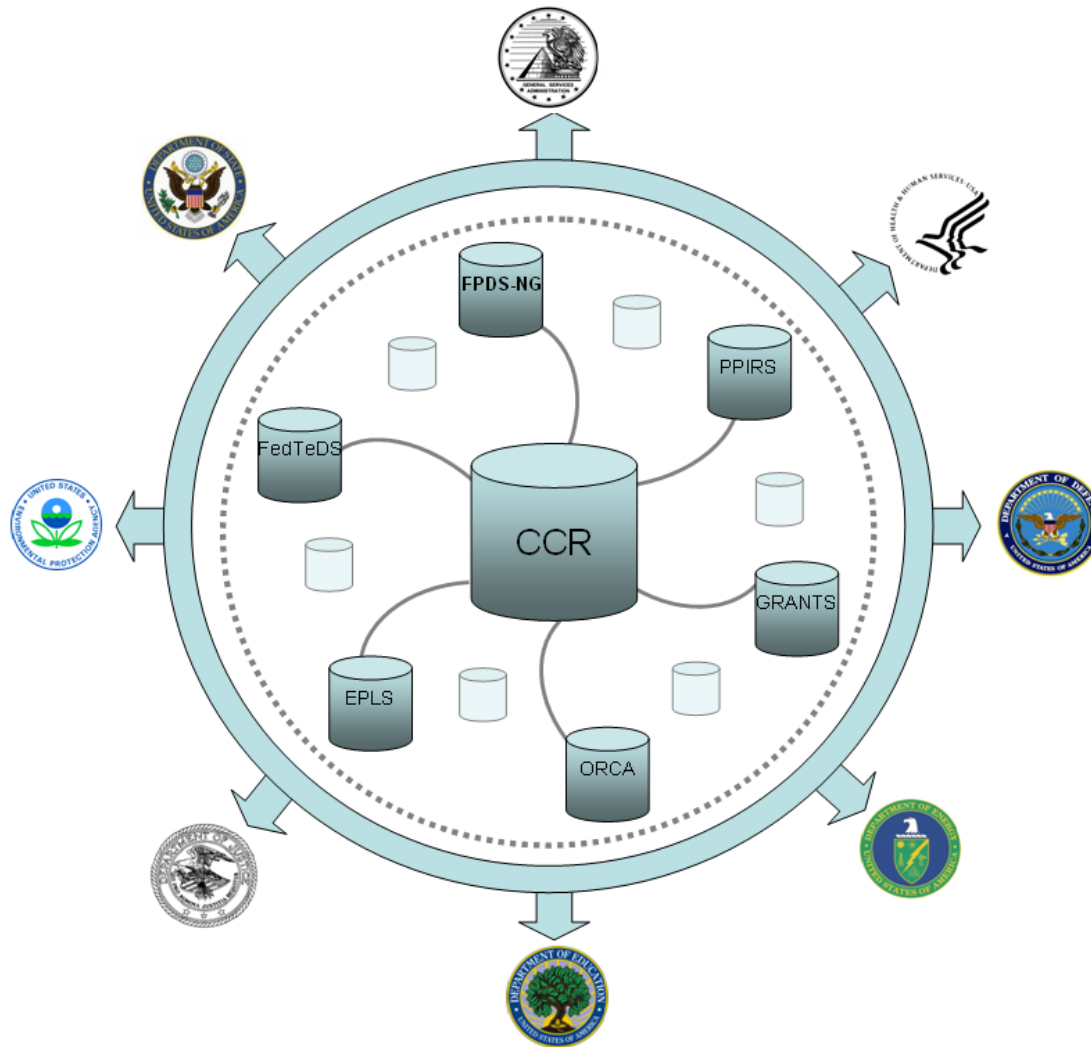
[S](#)

[D](#)

# CCR has both complex organization and integration needs as it supports all facets of government acquisition



Decide with Confidence

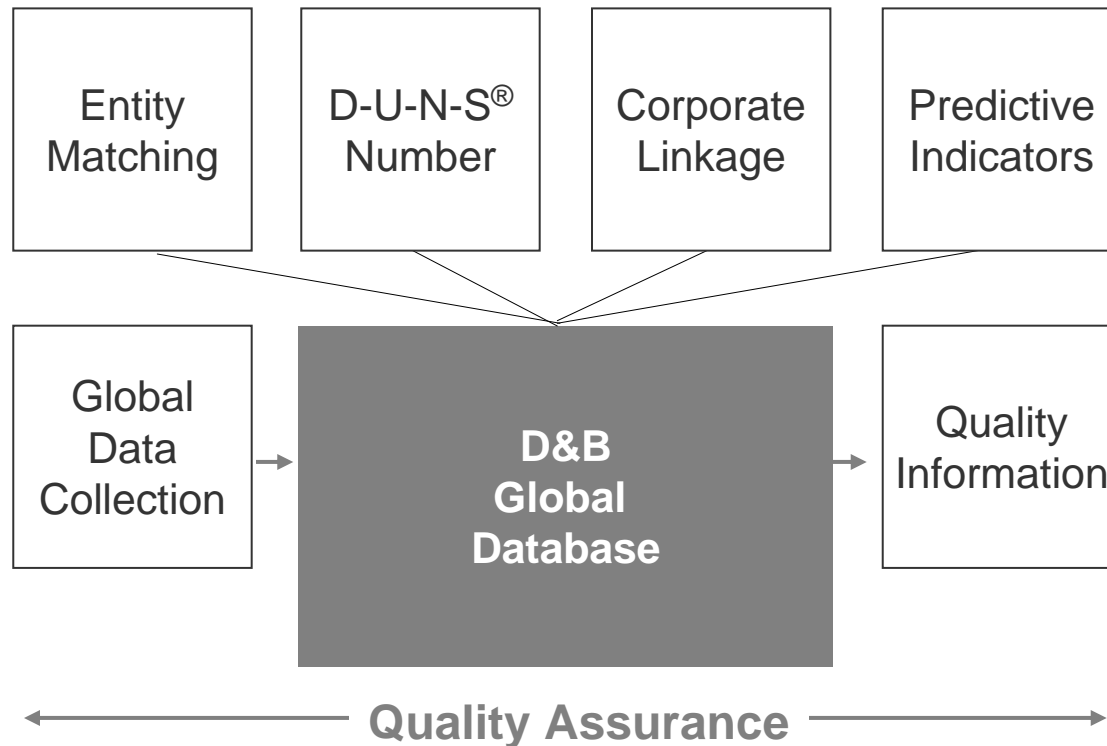


In addition to the D-U-N-S Number, some agencies take full advantage of our DUNSRight process to help solve agency objectives



Decide with Confidence

## D&B's DUNSRight® Process



Remember, data quality is an ongoing process, which when addressed through best practices can enable increased transparency and lead to better government performance



Decide with Confidence



- Complete Names
- Complete Addresses
- Complete Profiles

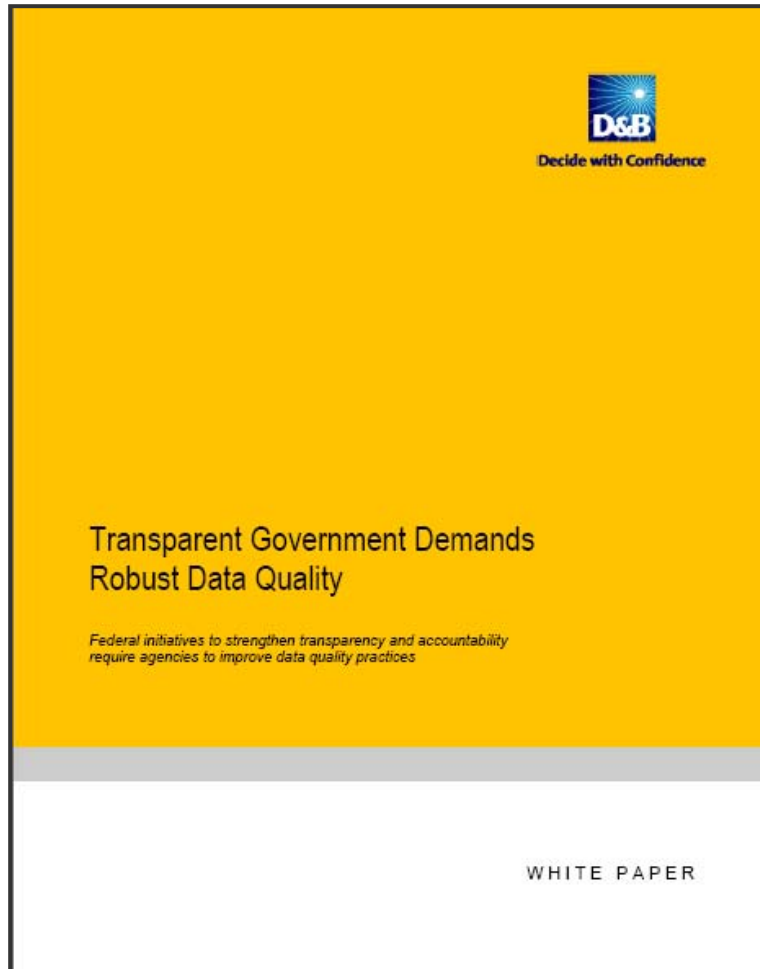
- Who Are We Regulating/Overseeing?
- Where Are They Located?
- Who Owns Them?
- What is the Total Business Relationship?
- How Risky are They?

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

Questions? Or to request the whitepaper...  
[government@dnb.com](mailto:government@dnb.com)  
1-800-424-2495



Decide with Confidence



[www.dnb.com/gov](http://www.dnb.com/gov)