



“Build to Share”

*Federal Data Quality
Guide:
A Framework for Better
Information Sharing*

July 2008

U.S. Federal Data Architecture Subcommittee



Agenda

- ◆ Document Purpose and Intended Outcome
- ◆ Federal Data Quality Guide Overview
- ◆ Examples of Federal Agency Data Quality Practices
- ◆ About the Data Architecture Subcommittee (DAS)





Purpose

- ◆ Few agencies practice data quality at the enterprise and extended enterprise levels
- ◆ The Federal Data Quality Guide advises agencies on the key components needed for an effective enterprise-wide data quality improvement program



3



Intended Outcome

- ◆ Data quality programs among Federal agencies and Communities of Interest (COIs) align to a common description of data quality improvement practices
- ◆ Information that is shared improves in quality
- ◆ Decision support in agencies and COIs improve

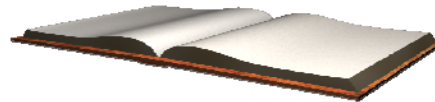


4



Federal Data Quality Guide Overview

- ◆ Build a data quality framework using EA
- ◆ The business case for data quality
- ◆ Value proposition using the reference models
- ◆ Data Quality Improvement implementation
- ◆ Advice on data quality tools
- ◆ Suggested additional reference material

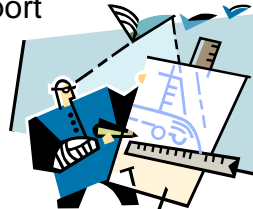


5



Key Advice Use Existing EA Program

- ◆ Establish data quality procedures and practices into existing agency and community of interest business processes that are part of their Enterprise Architecture (EA)
 - ◆ Provides a framework for improved information sharing and decision support



6



Data Quality Improvement: *The Challenge*

- ◆ Federal agencies and COIs have struggled with coordinated approaches to the quality of disseminated information due to:
 - ◆ Complexities of size and scope
 - ◆ Need to standardize and modernize technology and information technology (IT) processes
 - ◆ Internal management shortcomings



7

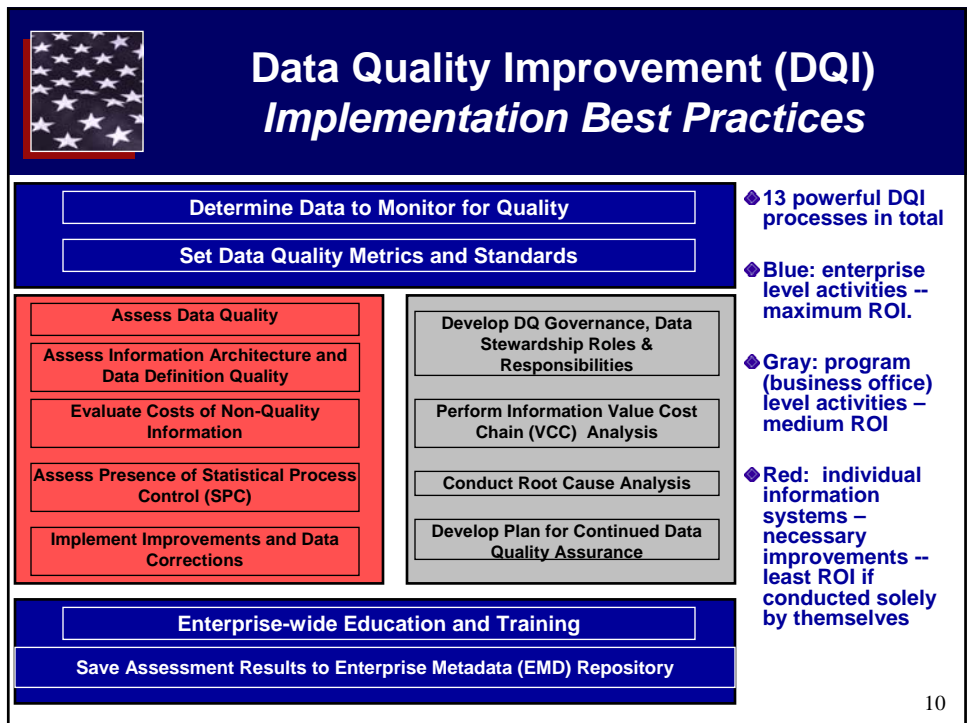
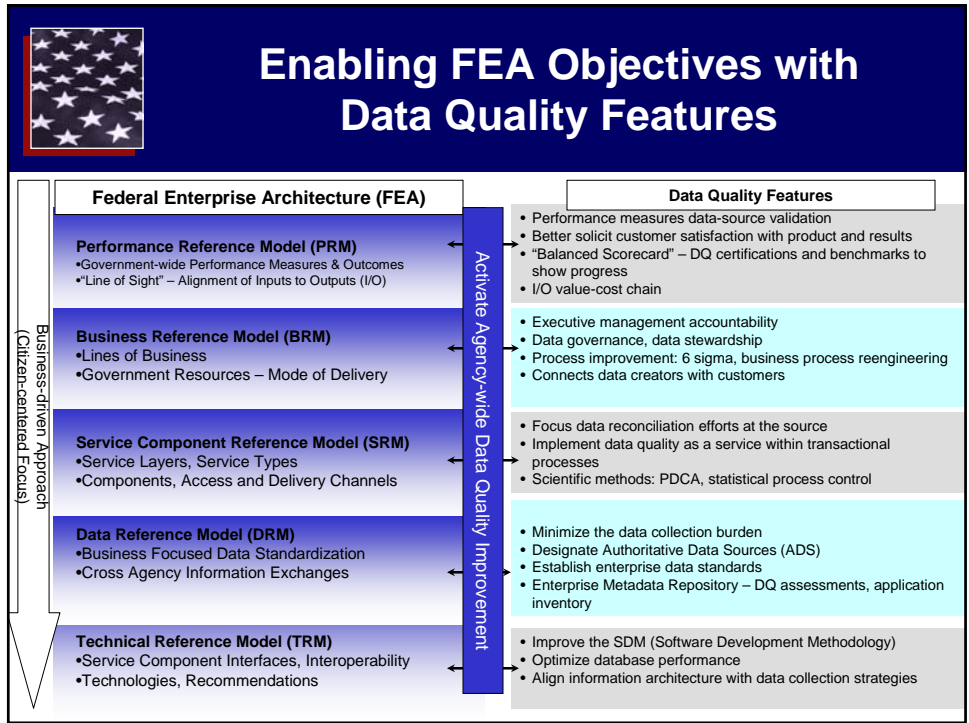


Business Case for Enterprise-wide Data Quality Improvement

- ◆ Data Quality Improvement (DQI) provides agencies and COIs with repeatable processes for:
 - ◆ detecting faulty data,
 - ◆ establishing data quality benchmarks,
 - ◆ certifying (statistically measuring) their quality, and
 - ◆ continuously monitoring their quality compliance



8





Some Agency Examples

- ◆ Agencies that have strong data quality programs at the enterprise level



- ◆ Defense Logistics Agency



- ◆ Housing and Urban Development

11




Defense Logistics Agency (DLA) *Data Quality Challenges*




- ◆ Building understanding of data and functional process flows of four feeder data systems into a DLA portal
- ◆ Analyzing multiple data entry points of the same classes of mission-critical data
- ◆ Determining authoritative source for multiple data “instances”
- ◆ Determining data stewardship responsibilities

12



Defense Logistics Agency (DLA) DQI Implementation



Identified 4-5 key business processes impacting agency performance

DQ Manual set thresholds for compliance with the dimensions of Completeness, Uniqueness, Timeliness and Currency

Sampled data at key feeder system points and compared with legacy instances, documenting the results according to required DQ dimensions

Reengineered some business processes at the source to align feeder data with legacy requirements


Enforced information stewardship by holding feeder systems' business process owners accountable for their quality

Identified and designated official record-of-origin, record-of-reference, and Authoritative Data Source


Developed ongoing Data Quality Monitoring & Trend Analysis

Educate the Enterprise

13




Defense Logistics Agency (DLA) Internal DQI Scorecard

	 Enterprise Level (minimal DQI impact felt here)	Program Level (most DQI impact felt here)	System Level (modest DQI impact here)
Successes	1. Some key business processes and their sequencing (operational "racetrack") developed for first time 2. DQ Manual developed with metrics and standards	1. Data Integrity Branch (DIB), program area stewardship defined 2. Data Quality Monitoring & Trend Analysis program taken up by DIB	1. Assessment points for sampling feeder data developed strategically 2. Reengineered some business processes to decrease data redundancy
Challenges remaining	1. EMD Repository solution required 2. Training required across the enterprise	1. Authoritative Data Source (ADS) analysis completed, but full information Value Cost Chain from feeders to legacy not understood	1. Refining Statistical Process Control methodology 2. Determining ROI for DQ improvement 3. Defining investment threshold for reaching point of diminishing return

14


352



Housing & Urban Development (HUD) Data Quality Challenges

- ◆ Information Architecture required redesign to better support accuracy and quality of information exchange
- ◆ Legacy Grants Monitoring System
 - ◆ Business Goal:
 - Support job creation in underprivileged areas
 - ◆ Reporting Method:
 - Data from multiple collection points aggregated to report on job creation statistics in HUD's Annual Performance Plan
 - ◆ Challenge:
 - Allowable data entry points did not use common method to convert jobs data

15



Housing & Urban Development (HUD) DQI Implementation


"Number of jobs created" performance measurement from Annual Performance Plan identified as key business process

DQ Handbook set thresholds for compliance with the dimensions of Validity, Uniqueness and Completeness


<div style="border: 1px solid red; padding: 2px; margin-bottom: 2px;"> <p>Assessment gave excellent results, but issue was in enforcing uniform business rules at the entry points</p> </div> <div style="border: 1px solid red; padding: 2px; margin-bottom: 2px;"> <p>Recommended Database Design and Data Definition improvements</p> </div> <div style="border: 1px solid red; padding: 2px; margin-bottom: 2px;"> <p>Estimated costs of non-quality information only</p> </div> <div style="border: 1px solid red; padding: 2px;"> <p>Program area completed necessary reengineering of system to enforce FTE job data entry on a single screen, and business rules across the database were made uniform</p> </div>	<div style="border: 1px solid gray; padding: 2px; margin-bottom: 2px;"> <p>Identified database of origin, mapped data entry fields to database locations, & identified business rules (allowable values) for each</p> </div> <div style="border: 1px solid gray; padding: 2px;"> <p>"Jobs created" can now be reported to management with 6 sigma accuracy, and steps are being made for improvements in other key business processes</p> </div>
---	--

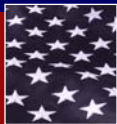
Assessment results saved to EDM staging area

16



Housing & Urban Development Internal DQI Scorecard

	Enterprise Level (some DQI impact felt here)	Program Level (modest DQI impact felt here)	System Level (most DQI impact here)
Successes	<ol style="list-style-type: none"> 1. Annual Performance Plan effective blueprint for identifying key business processes/data sources 2. Development of DQ Handbook with consistent standards and DQI procedures 3. Data Control Board created for DQ governance 	<ol style="list-style-type: none"> 1. Reengineered system to 6 sigma for this metric 2. Information Value Cost Chain completed for in-scope data showing transformations, data classes, and system interfaces 	<ol style="list-style-type: none"> 1. Costs of non-quality information estimated 2. Information Architecture alignment with database improved 3. System functionality improved 4. New Data Dictionary developed
Challenges remaining	<ol style="list-style-type: none"> 1. EDM staging area not secure, robust enterprise solution required 2. Training required across the enterprise 	<ol style="list-style-type: none"> 1. Data Quality Assurance plan not formalized 2. Root Cause Analysis not undertaken – errors may return and impact other business processes 3. DQ stewardship lacking at program level 	<ol style="list-style-type: none"> 1. Lack of Statistical Process Control 2. Database partitioned between grants programs, resulting in data overlap and lack of visibility



Data Quality Tools Advice

Enabling tools for data quality at minimum:

- ◆ Data Profiling (Business Rule Discovery)
- ◆ Data Defect Prevention
- ◆ Metadata Management
- ◆ Data Re-engineering and Correction





Current Status

- ◆ The Federal Data Quality Guide is in draft form undergoing review.
- ◆ A copy of the draft is available on the Data Architecture Subcommittee collab site on Core.gov.
- ◆ A copy of the draft can also be obtained via e-mail request: suzanne_acar@ios.doi.gov

19



About the Federal Data Quality Guide Authors:

**Federal Data Architecture Subcommittee
(DAS)**

20



Data Architecture Subcommittee

◆ Federal Data Architecture Subcommittee (DAS) Facts

- Chartered by the Federal CIO Council
- 2 appointed Co-chairs
 - Suzanne Acar, DOI
 - Adrian Gardner, NWS
- Membership Federal CIO representation + contributors (135)
- Eight work groups



◆ Key FY08/09 Activities/Deliverables

1. Federal Data Quality Guide
2. Final Draft Person Framework Standard
3. DRM Implementation Guide

21



Summary

22



Summary

- ◆ The Federal Data Quality Guide informs agencies on features of an enterprise-wide data quality program.
- ◆ The key advice is to leverage existing EA programs.
- ◆ The outcome is improved information sharing, interoperability, and decision support.
- ◆ Supports key principle to manage information as a national asset.

23



Questions



Contact info:

Mark Amspoker
Lead, Federal DQ Guide Working Group
Citizant, Inc.
mampoker@citizant.com

24