

New Initiatives in HUD's Data Quality Improvement Program (DQIP) – Enabling Field Effectiveness

ABSTRACT-----

In 2002, HUD's Office of the Chief Information Officer (OCIO) launched an enterprise-wide initiative, the Data Quality Improvement Program (DQIP), to address HUD's information quality deficiencies. The DQIP goal was to define short- and long-term approaches to identify, assess, and continually improve the overall quality of mission-critical data used in HUD's Annual Performance Plan (APP) reporting.

Recently, the DQIP has been broadening its efforts nationally to ensure that the data quality concerns of HUD's ten major field centers and six principal American territories are also heard. Proposed DQIP initiatives are prioritized according to the severity of DQ problem reports impacting HUD field centers and the benefit a data quality solution would provide to the Department as a whole. A Root Cause Analysis is conducted early on in the process, to determine if the data quality problem can be tied to HUD training, policy, or procedure shortcomings that can be fixed more economically than an intensive data-centric approach.

This added initiative is enabling the Department, in program areas and in Lines of Business (LOBs) areas, to fully understand the accessibility, usability, and contextual clarity of mission-critical data distributed to areas outside of HUD's headquarters, and to provide guidance on improving the retrieval and management of the information used within these areas.

BIOGRAPHY-----

Shula C. Markland

Senior Data Architect/Enterprise Architecture Web Manager
Department of Housing and Urban Development

Ms. Markland is the Senior Data Architect in the Office of the Chief Information Officer (CIO) at the Department of Housing and Urban Development. She is responsible for managing the implementation of the Enterprise Information Management Practice throughout the Department. She helped establish the Data Steward Advisory Group to ensure that HUD's data assets are managed as a strategic enterprise asset. She is also responsible for managing HUD's Enterprise Architecture web site to ensure that relevant information is available on the Internet and Intranet.





Prior to Ms. Markland's current position, she worked as a Computer Specialist in the Center for IT Capital Planning, located within the Office of the CIO, at the General Services Administration. Ms. Markland was responsible for serving on several teams such as the Executive Information Services Team, the Data Warehouse Team, the Quality Assurance Team, and the Y2K Team. She designed and developed the Y2K web page. She also worked in the E-Mail Program Management Office at the GSA, where she was responsible for government-wide electronic messaging. During her tenure at GSA she installed a network for the President's Management Council (PMCNet). The database allows disseminating information across approximately 30

Federal agencies. One of her accomplishments was her 60 day developmental assignment, with the Women's Executive Leadership Program, where she developed a Briefing Memorandum on how the Department of State will implement and maintain the Internet as a communication interface.

Ms. Markland has over five years of technical experience with the private sector, and eighteen years of her career have been with the Federal Government. Ms. Markland started her career with the U.S. Office of Personnel Management, where she was responsible for the quality assurance of the data submitted to the Central Personnel Data File (CPDF). Ms. Markland received her Bachelor of Science Degree in Computer Information Science from the University of MD University College. Ms. Markland and her husband John have three daughters and live in Edgewater, MD.

Mark X. Amspoker
Citizant, Inc.



DATE	REVISION	BY	<p>New Initiatives in HUD's Data Quality Improvement Program (DQIP) – Enabling Field Effectiveness</p>
			<p>DATE: July 16, 2009 AUTHOR: Shula Markland/Mark Amspoker</p>
			<p>ENTERPRISE ARCHITECTURE PRACTICE 202-700-1021, ea_team@hud.gov</p>

History of Data Quality Improvement at HUD

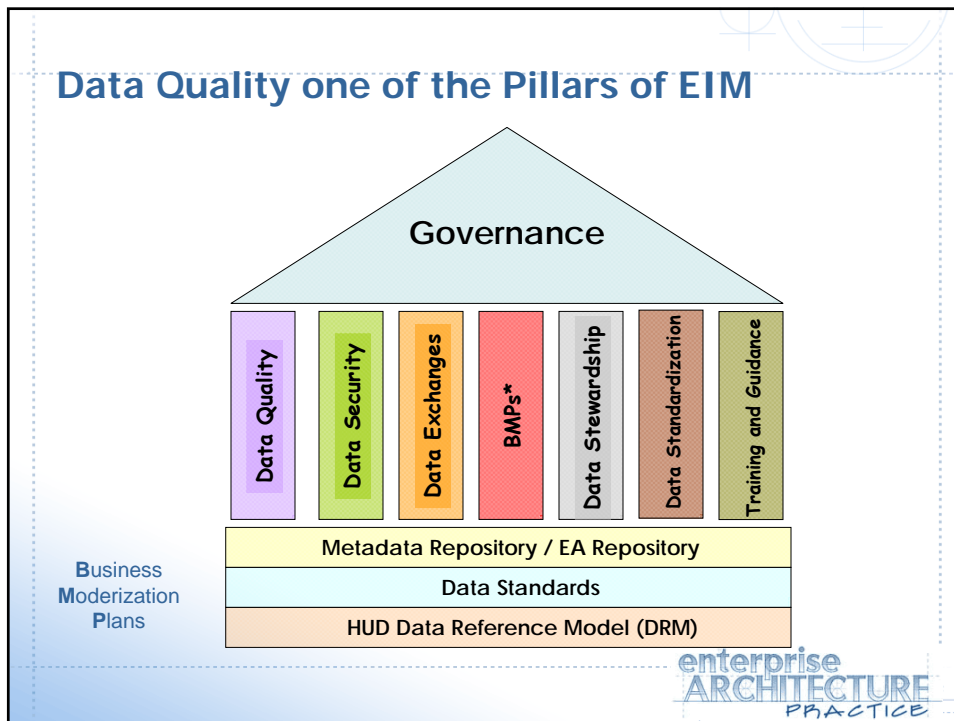
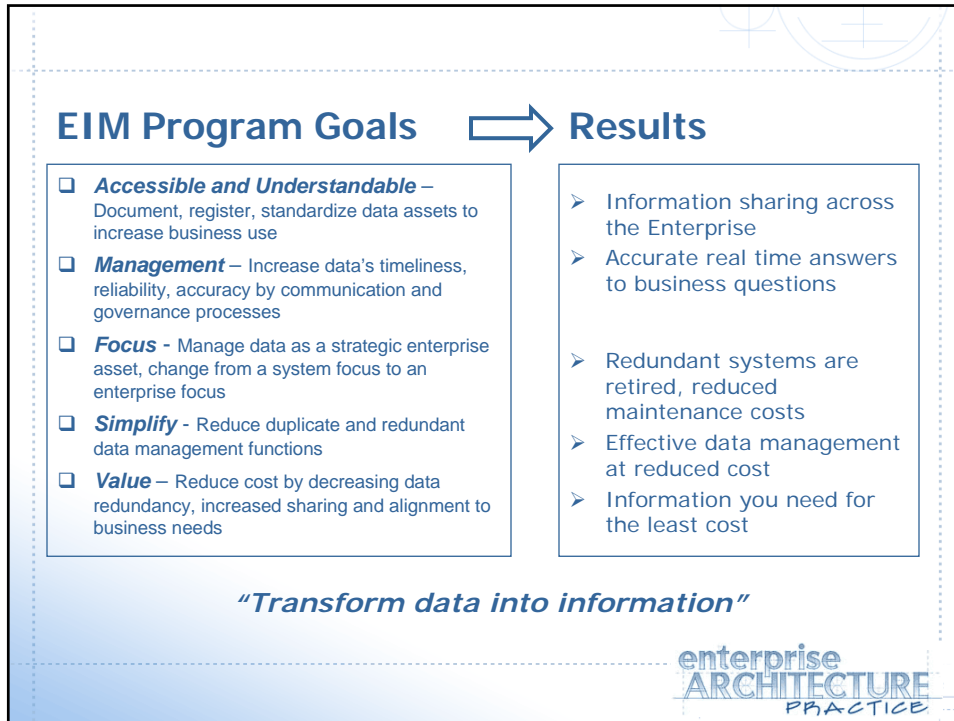
□ Where we came from...

- Standard data quality cleanup method took place across the Department led by OCFO
- An OIG Audit in FY 2001 stressed the need for continuous quality improvement in HUD's mission critical information
- In FY 2002, OCIO assumed responsibility for data quality and established the Data Quality Improvement Program (DQIP)

□ Where are we going....

- In early FY 08 decided to take the DQIP above and beyond a strictly oversight function
 - Provide a value add service to HUD Program Areas
 - Align with the goals of Enterprise Information Management (EIM): "to establish standards-based data management processes and promote data sharing across the Department"





EIM/DQIP FY 2008 & FY 2009 - Focus Areas

1. Data Stewards Advisory Group (DSAG)
2. Data Control Board (DCB)
3. Data Quality Improvement, Regions & Territories => HQ's
4. Enterprise Metadata Repository (EMR)
5. Data Dictionary Template and Data Naming Standards
6. Promotion of EIM, DQIP

enterprise
ARCHITECTURE
PRACTICE

Communication

1. Data Stewards Advisory Group

- Identifies and organizes the people who understand and manage the data that a business area needs for its mission
- Facilitates communication among peers to solve common data management/data quality problems
- Surfaces data quality issues for prompt response from the OCIO
- Assists in the management of data to improve its accessibility, reliability, accuracy, and quality
- Identifies and defines updates to HUD's EMR
- Reviews and recommends policies, standards and procedures for submission to the Data Control Board (DCB) for consideration and approval as needed to promote sharing and exchange of data

Stewardship: The process of having data stewards actively manage data and metadata to ensure they meet an organization's business needs.

Steward: Someone who *manages* or *has charge for the care* of an asset that belongs to someone else.

enterprise
ARCHITECTURE
PRACTICE

2. Data Control Board

- ❑ Reviews and approves data standardization recommendations, as provided by the DSAG
- ❑ Promotes the management of HUD's data as strategic enterprise assets consistent with Government-wide policies and standards
- ❑ Oversees DQIP initiatives and approves the selection of mission-critical data for assessment
- ❑ Reviews and recommends EIM policies, standards, and procedures for submission to the TIBEC

enterprise
ARCHITECTURE
PRACTICE

3. Data Quality Improvement Regions/Territories => HQ's FY 08 – FY 09

- ❑ Region 9 – Las Vegas Regional Office
 - Key staff members did not have access to information from HQ's, work done manually
- ❑ Region 3 – Philadelphia Regional Office
 - Used the SFH Data Warehouse for a variety of ad-hoc reporting; however, a specific query necessary for more granular calculations did not exist in SFDW metadata
- ❑ U.S. Territory of Puerto Rico (PR)
 - "Urbanization" address field, a key element of PR addresses, not captured on HUD forms, so not possible to geocode approx. 80% of addresses with precision

enterprise
ARCHITECTURE
PRACTICE

Data Quality Improvement (cont.) Las Vegas Regional Office – the Problem

- ❑ Key staff members did not have access to all the information from HUD headquarters required to timely serve their customers
- ❑ None of the information-gathering process was automated, and a manual report had to be prepared for each data-call (estimated 10% of staff time)
- ❑ May 2008 data request for “Total HUD Funding - Las Vegas Office”: of 21 major Program Area lines of business, nine categories could not be filled in for FY 2007
- ❑ Applicable IQ measures: Accessibility, Rightness, Contextual Clarity
- ❑ Applicable DQ dimensions: Accuracy, Timeliness

enterprise
ARCHITECTURE
PRACTICE

Data Quality Improvement (cont.) Las Vegas Regional Office – the Solution

- ❑ Standardize data collection procedures, using the format employed by the PIH program office in San Francisco Regional Center (a best practice)
- ❑ Distribute data meeting requirements of HUD funding categories to all regional field centers on the first day of each fiscal quarter
- ❑ Inform Region 9 of alternate data sources for acquiring the HUD funding data (five alternate sources identified)
- ❑ Modify HUD Program Office Funding Report used by Las Vegas to align funding category descriptions with actual data being presented
- ❑ Follow-up with Region 9 stakeholders to learn status of recommendations (in 6 months)
- ❑ Submitted report to HUD’s Field Advisory Council as a milestone measurement in the Secretary’s iMPACT 200 Priority Initiatives #15, *Engage Field Advisory Council and Develop Recommendations to Improve Field Effectiveness*

enterprise
ARCHITECTURE
PRACTICE

Data Quality Improvement (cont.) Philadelphia Regional Office – the Problem

- ❑ Philadelphia analysts used the Single Family Data Warehouse (SFDW) for a variety of reports and daily ad-hoc queries
- ❑ Canned queries were fine for most purposes but fell short when results needed at more granular level (i.e. # of FHA endorsements by county)
- ❑ Analysts reported that the interpretation of query results would be more precise if there were a section in the SFDW metadata to explain “official” methods for calculating commonly used data runs
- ❑ Applicable IQ measures: Usability, Contextual Clarity
- ❑ Applicable DQ dimensions: Accuracy, Consistency

enterprise
ARCHITECTURE
PRACTICE

Data Quality Improvement (cont.) Philadelphia Regional Office – the Solution

- ❑ Distribute improved formula for calculating regional default rates to all regional centers (formula provided by HUD SFH experts)
- ❑ Re-charter SFDW Users Group and/or SFH Business Analyst Working Group to collect data issues and suggest solutions
- ❑ Add section on the SFDW website as a library of commonly used data runs
- ❑ Follow-up with Region 3 stakeholders to learn status of recommendations (in 6 months)

enterprise
ARCHITECTURE
PRACTICE

Data Quality Improvement (cont.) Puerto Rico – the Problem

- ❑ HUD's Program Areas administering Puerto Rican projects often must know exactly where a specific property is located (i.e. precise geocoding requirements) to determine whether it lies in an area that has been designated for special statutory treatment
- ❑ HUD still uses stateside forms to collect PR address data, but these forms don't capture a key PR data element called "Urbanization", required for geocoding
- ❑ HUD databases and software collecting PR address data not coded to accept the Urbanization field
- ❑ Thus, only a small percentage of vouchers or public housing data can be allocated to the correct PR census geography (e.g., tract)
- ❑ Applicable IQ measures: Usability
- ❑ Applicable DQ dimensions: Accuracy, Completeness, Consistency

enterprise
ARCHITECTURE
PRACTICE

Data Quality Improvement (cont.) Puerto Rico – the Solution

- ❑ Require software vendors to add Urbanization-capture functionality
- ❑ Update HUD databases and user documentation to add Urbanization field
- ❑ Train Program Area grantees, Public Housing Authorities and building owners on Puerto Rican address data entry conventions (i.e., must document the Urbanization)
- ❑ Request change to address forms from Congress
- ❑ Follow-up with DQIP stakeholders to learn status of recommendations (in 6 months)
- ❑ Continue to promote GSC's reputation as a "Center of Excellence" for the Federal government

enterprise
ARCHITECTURE
PRACTICE

4. Enterprise Metadata Repository (EMR)

- EMR is a searchable database containing metadata
- Initially populated with DQIP artifacts
- EMR benefits:

Benefit	Value
Provide a searchable central store of HUD metadata to discover existing data definitions for reuse in new systems' databases.	Cost savings, decreased development time.
Provide a tool data users need to conduct HUD business.	Better and quicker decision making.
Increase system interoperability and data exchange by increased understanding of the current data assets at HUD.	Better interoperability, cost savings.
Provide HUD data analysis support for meeting new Government requirements.	Quicker response to new requirements.

enterprise
ARCHITECTURE
PRACTICE

5. Data Dictionary Template and Data Naming Standards

- Data Naming Standards
 - Uses the ISO 11179 three part name format: [Modifier] <Prime Word> [Modifier] [Modifier] <Class Word>
 - Promotes sharing and understanding of data assets by naming data in a way that it is readily understood
 - Has been approved by the DCB and HUD's Configuration Change Management Board as part of the System Development Methodology
- Data Dictionary Template
 - Provides a consistent level of documentation of new system's data requirements
 - Improves data quality by labeling information consistently, with agreed-upon definitions for data elements and a common set of properties for each data element, making systems and data analysis easier and business intelligence more effective

enterprise
ARCHITECTURE
PRACTICE

6. EIM Program Promotion

- ❑ To expand the awareness and knowledge of data quality/data management practices and goals using:
 - HUD Website and CIO Newsletter.
 - DM training sessions.
 - DSAG and DCB meetings.

enterprise
ARCHITECTURE
PRACTICE

Conclusion

- ❑ A more enterprise view of data and information is needed to meet HUD's mission
- ❑ Reduced resources and increased demands requires a smarter and more efficient approach
- ❑ Communication and coordination is key
 - Communicate EIM/DQIP goals and priorities
 - Accessible documentation about common information needs and current data stores
 - Coordinate EIM efforts across the enterprise to realize cost saving



Working together, WE CAN DO THIS!

enterprise
ARCHITECTURE
PRACTICE

Questions?

Shula Markland

OCIO EIM Project Manager
shula_c_markland@hud.gov

Mark Amspoker

Citizant, EIM support contractor
Mark_x.amspoker@hud.gov



enterprise
ARCHITECTURE
PRACTICE