THE DATA QUALITY IMPROVEMENT PROGRAM (DQIP) IS YIELDING DIVIDENDS FOR HUD

(Practice-Oriented Paper**)**

Beverly Hacker, Randy Bissett

June 2005

HUD'S EXPERIENCE WITH PROBLEM DATA

GAO reports (1994-2000): HUD at high risk of waste, fraud, abuse, and mismanagement.

FY 2000 housing subsidy overpayments from tenants misreporting their income = $978 million.

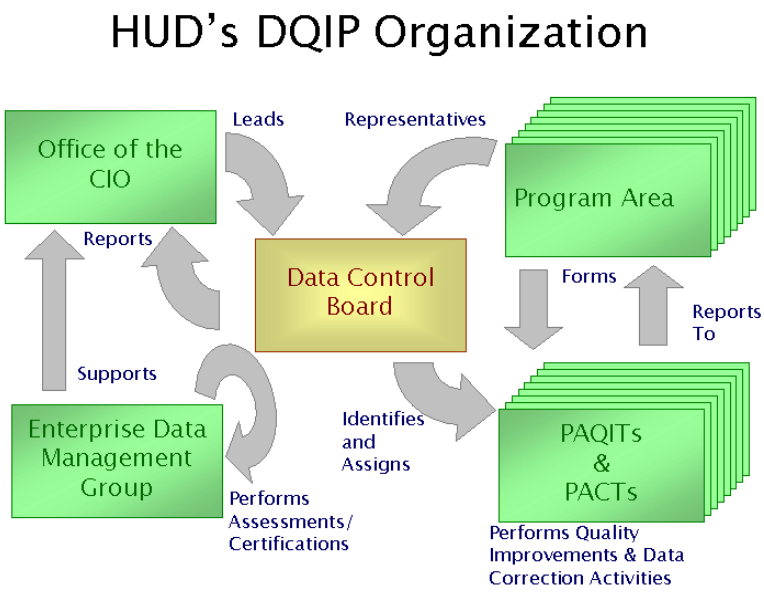
FY 2000 analysis of the database of a major financial system: half of records blank for Initial and Latest Obligation Date, resulting in unexpended appropriation of $60 billion.

A major HUD system with $5 billion in annual appropriations cannot adequately report on the accomplishments of its grantees for funded activities.

Since assuming responsibility for data quality improvement at HUD, the OCIO has been successfully applying a systematic program of data quality assessment, process improvement, data correction and certification for mission-critical data in HUD's major information systems.

DQIP OVERVIEW

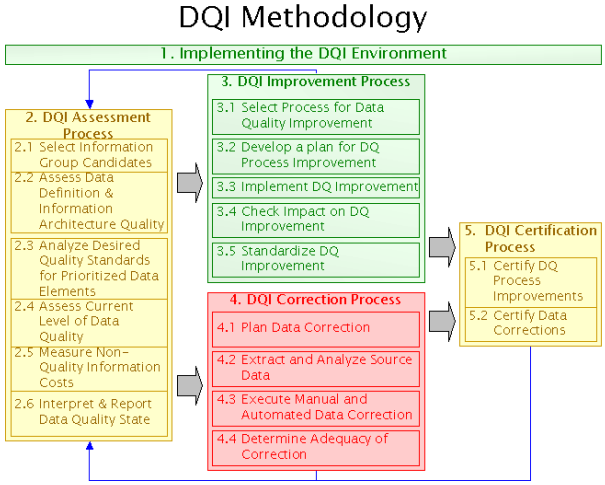
In 2001, audit recommendations of HUD's Office of the Inspector General (OIG) stressed the need for data quality improvement in HUD's mission-critical systems. The OIG report recommended that the OCIO: "(1) implement data quality standards for systems and data supporting Annual Performance Plan (APP) performance indicators, (2) require data quality plans for these systems based on a standard rationale, and (3) implement a methodology for independent verification for high priority data." At the beginning of Fiscal Year 2002 the OCIO formed an Enterprise Data Management Group (EDMG) to assess and certify the Department's mission-critical systems, ensuring that HUD's data be in compliance with Office of Management and Budget's Section 515 guidance "maximizing the quality, objectivity, utility and integrity of information … disseminated by Federal Agencies." The application of this methodology is intended to bring the Department's financial and programmatic data to a level of quality that makes the information credible and useable for all HUD's intended business purposes, while consistently meeting the quality expectations of HUD's business partners and customers. The DQIP operates as part of HUD's Enterprise Data Management (EDM) practice, with oversight provided by a Data Control Board consisting of executive- and program area-level HUD personnel. Any needed data correction activities are conducted by the program offices themselves, whose managers are accountable for the quality of the information produced or collected within the information systems operating within their area.



The principles of continuous data quality improvement analyze the root causes of defective data and implement improvements that manage data for quality throughout its life cycle.

HUD'S DATA QUALITY IMPROVEMENT (DQI) HANDBOOK

Prior to the OCIO assuming responsibility for data quality at HUD, the Department's approach to quality was to stress data correction at the source rather than to discover the root causes of the problem. In early 2002, the EDMG's first task was to revise this approach and produce a Total Information Quality Management Handbook with guidance toward: (1) comprehensive standards for criteria that determine data quality; (2) a formal process for conducting data quality assessments and certifications of HUD's information systems, as well as benchmarks to determine the maturity level of the systems' data management practices; (3) continuous data quality improvement activities within the Department. The principles of continuous data quality improvement analyze the root causes of defective data and implement improvements that manage data for quality throughout its life cycle. Eliminating the causes of data defects and the production of defective data builds quality in and reduces the need to conduct data correction activities. Improvement consists of selecting the process for data quality, developing a plan for improvement, implementing the improvement in a controlled environment, checking of the impact of the improvement to make sure that results are as expected, and standardizing the improvement across the enterprise.



A study by the National Academy of Public Administration (NAPA) noted that HUD's FY 2000 APP was showing progress in the use of outcome-oriented performance measures, but the problems in documenting APP data sources still remained.

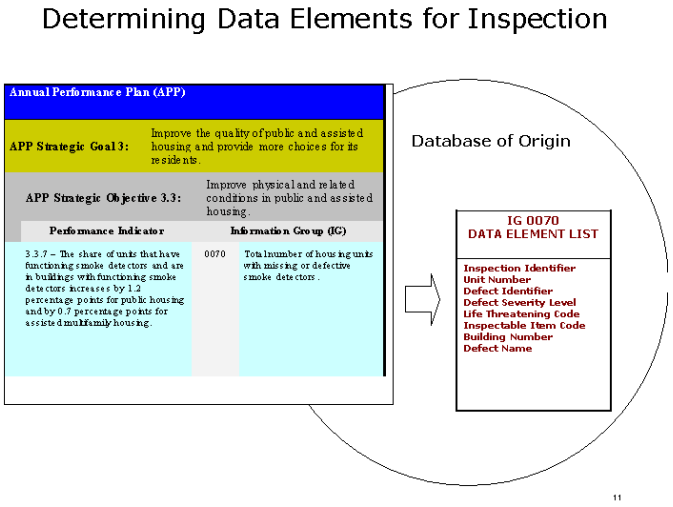
In adopting HUD's APP as its means for prioritizing information systems and source data in scope for DQIP assessments, the EDMG is detailing precisely where these points of failure exist in the performance indicator value chain for mission-critical systems across the Department.

Data quality of seven systems assessed in FY 2001 (Phase I-pre DQIP) and certified by the EDMG in FY 2002 has improved by 10-15%. Improvements are due to process and system modifications and data cleanup efforts on the part of the program areas overseeing the systems.

As the EDMG reaches the half-way point of Phase IV, 87% of all subsystems and 95% of in-scope performance indicators have been assessed. To date, 22 of 24 subsystems have been certified at 3 sigma.

THE ANNUAL PERFORMANCE PLAN (APP): HUD'S BUSINESS PERFORMANCE METRICS

In FY 2000, HUD's OIG conducted a review of the Department's initial FY 1999 APP. The OIG took a sample of data for 22 of the performance indicators and identified problems for 16 of them (72%): the data was either estimated due to lack of confidence in its accuracy, or the performance measures were immature because the data to support them was not in place yet (or, if present, was not timely). A study by the National Academy of Public Administration (NAPA) noted that HUD's FY 2000 APP was showing progress in developing outcome-oriented performance measures with a baseline of data to support them, but the problems in documenting APP data sources still remained.



In adopting HUD's APP as its means for prioritizing information systems and source data in scope for DQIP assessments, the EDMG is detailing precisely where these points of failure exist in the performance indicator value chain for mission-critical systems across the Department. Since FY 2002, the EDMG has been assessing eight systems per year of the 24 subsystems currently supporting the 2005 APP. Systems have top priority for assessment if they support multiple performance indicators and have been the subject of previous audits by the GAO or OIG. The EDMG assigns an "information group" to each performance indicator and learns from system representatives the data elements comprising each information group, which become the targets for data quality inspection against the quality criteria used in the DQI methodology. Data that is found to be external to HUD is declared out of scope. If the EDMG discovers that a performance indicator is not actually supported by the system under assessment (i.e., the "database of origin" for this indicator exists in another system), then these facts are documented in the DQIP Assessment Certified Repository.

RESULTS TO DATE

DQIP analysis of eight systems in FY 2002 revealed widely different levels of data quality/data management practices. Generally, systems with documented, repeatable data management practices scored higher in data content quality than systems with undocumented, informal practices. In order to be certified, systems now must not only

As part of the OCIO corrective action plan (CA), the EDMG plans to complete data quality assessments for all 24 HUD subsystems supporting APP performance indicators by the end of calendar year 2005. Thereafter, the DQIP will transition from an assessment-focused program to a certification-focused one.

score above 3-sigma (93.319% correct) in eleven content quality characteristics, but also have documented processes by which they extract their data to calculate APP results. For example, a system can fail certification if it exhibits a significant level of "work-around" internal data manipulations in order to achieve acceptable levels of quality. Likewise, a system can be found to be noncompliant, even if the data is of 3-sigma quality, if its information architecture prevents the accurate aggregation of data to adequately validate the performance indicator metric.

DQIP Progress Report

Year	No. In-scope APP PIs/IGs	% PIs Assessed by Phase End	No. of Systems/ Subsystems	2005 APP Subsystems-Completed Assessments	Subsystems Certified at 3-sigma
Phase II (includes Ph I pre-DQIP)	87/92	49%	33/38	11 (40%)	10 (one system not certified to date)
Phase III (FY 2004)	73/76	76%	33/38	19 (70%)	8
Phase IV (FY 2005) - projected	60/62	100%	19/24	24 (100%)	2 (three systems not yet assessed)

Yet holding program areas accountable for the quality of their data is certainly producing dividends for HUD. Data quality of seven systems assessed in FY 2001 as part of Phase I (pre-DQIP) and certified by the EDMG in FY 2002 has improved by 10-15%. Improvements are due to process and system modifications and data cleanup efforts on the part of the program areas overseeing the systems. There are currently 19 systems comprising 24 subsystems supporting 60 APP performance indicators whose data is in scope for a DQIP assessment. As the EDMG reaches the halfway point of Phase IV, 87% of all subsystems and 95% of in-scope performance indicators have been assessed. To date, 22 of 24 subsystems have been certified at 3 sigma.

THE FUTURE OF DQIP

In its most recent financial statements audit (2005-FO-0003), OIG named performance measures data reliability a reportable condition, with CFO and CIO sharing lead responsibility for corrective action, and program areas responsible for participation. As part of the OCIO corrective action plan (CA), the EDMG plans to complete data quality assessments for all 24 HUD subsystems supporting APP performance indicators by the end of calendar year 2005. Thereafter, the DQIP will transition from an assessment-focused program to a certification-focused one. In that role, the EDMG will evaluate newly-introduced performance indicators and ensure that the supporting systems and data meet and are certified at HUD's quality standard. The OCIO is currently proposing that HUD's target data quality standard be revised to 4 sigma (99.379% correct). Based on lessons learned during Phase I through Phase IV assessments, OCIO believes that the 4 sigma standard is more appropriate, since it will not be feasible in all cases to design the system edits that would enable quickly achieving 6 sigma data quality.

The EDMG is alleviating the problem of poorly documented APP data sources by creating a new policy with guidance for developing APP performance indicators. The policy creates a structured process for program offices to document the indicator "Background" and "Data Source" sections for information systems under their charge. Guidance focuses on the creation of:

- Data flow "value chain diagram" of program office source systems, showing the data's end-to-end transformation process
- Subject matter experts' roles and responsibilities in the APP performance indicator process, including review of APP prior to publication
- Means for determining which system or systems manage the complete data for calculating a performance indicator, and whether a sufficient baseline of APP data is present for the calculation