Performance Improvement Council's
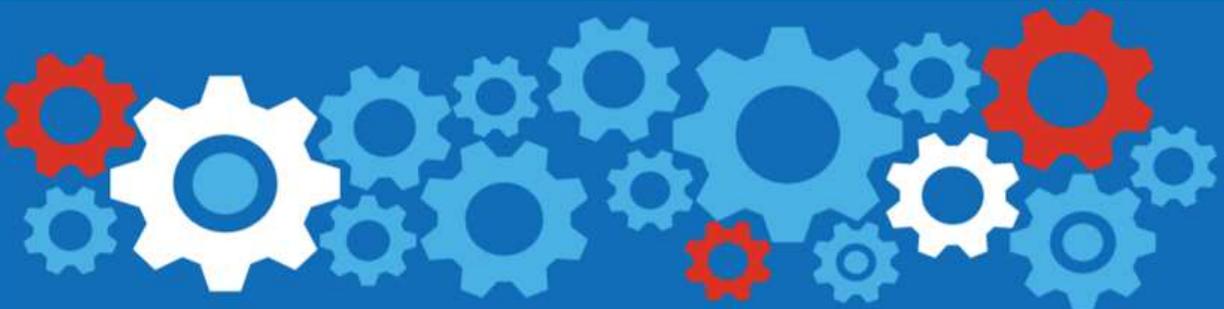# Data Quality Working Group Guide:

Data Quality Maturity Model

2016

Background

Starting in 2014, the Performance Improvement Council (PIC) has conducted annual agency- and component-level surveys to understand the scope and strength of organizational performance. The findings indicated issues in the areas of data accessibility, confidence, and accuracy. Of the 24 CFO Act Agency HQ performance offices:

- Less than one third report that performance information is frequently **accessible**
- Less than half report that they frequently have **confidence** in their data
- Less than one quarter frequently audit data for **accuracy**

In addition, the GAO has issued the following reports on data quality and asked agencies to address identified gaps.

- The Results Act: An Evaluator's Guide to Assessing Agency Annual Performance Plans, GAO/GGD-10.1.20 (1998)
- Performance Plans: Selected Approaches for Verification and Validation of Agency Performance Information, GAO-99-139 (1999)
- Performance Reporting: Few Agencies Reported on the Completeness and Reliability of Performance Data, GAO-02-372 (2002)
- Results-Oriented Government: GPRA Has Established a Solid Foundation for Achieving Greater Results, GAO-04-38 (2004)
- Managing for Results: Greater Transparency Needed in Public Reporting on the Quality of Performance Information for Selected Agencies' Priority Goals (2015)

In response, the PIC team initiated a working group to discuss data quality best practices and examine OMB validation and verification policy. The working group met nine times over the course of 2015 and 2016 to articulate the scope of the problem, share data quality practices, and examine OMB's current data quality policy language.

As a result of their early work, the working group coordinated an agency-level self-assessment of how headquarters staff implements the current OMB data quality policy language found in A-11 to better understand the state of practice. The 16 agencies who participated indicated that they actively engage in creating and/or implementing procedures to ensure high quality data, although their methods and areas of focus varied. Unsurprisingly, the A-11 factors that agencies scored the highest in their self-assessments ("fully implemented") were some of the lower effort items, while the lower scoring factors ("minority met" or "not implemented") were the more challenging policies to design and implement.

After analyzing the self-assessment results, the group agreed to develop, under the PIC's leadership, a best practices matrix and data quality maturity model to aid organizations in the creation and improvement of data quality programs. This document is the culmination of that work.

Scope

While the content in this guide may be applicable to a broad range of data sources, the working group members intend for its scope to be performance data reflected in publicly available performance measures (e.g., Annual Performance Report-level performance measures).

How to Use this Resource

This document is not official guidance or a draft of future policy. The contents are meant to be a useful guide for agencies to assess the maturity of their agency's approach to data quality and how the maturity levels correspond to OMB guidance. Agencies may also use it as a starting point to determine what improvements might be made to their program and as a resource for identifying successful practices at other similar agencies. This resource should not serve as auditing standards for data quality without first fully consulting the performance community, OMB, and other experts.

Connection to OMB policy

OMB circular A-11 outlines data quality validation and verification guidelines in Part 6, Section 260.9, "Assessing the completeness, reliability, and quality of performance data." OMB identifies a number of factors that agencies should consider in designing their data quality programs, but only requires that (1) agencies publish information on data reliability (including an assessment by the agency head), (2) agencies identify known data limitations, and (3) some validation and verification techniques are in place.

The Maturity Model

This model shows four levels of maturity, each distinguished by the extent to which certain characteristics and activities are present.

	CHARACTERISTICS	ACTIVITIES
Level 4	• **Cross-enterprise** set of policies and management • Validity of data is **auditable** • Data flaws recognized early in information flow • Remediation **governed** by well-defined processes • Strategic **continuous improvement**	• Governance structure to provide accountability, communication, and best practice sharing • Continuous improvement loops built into processes • Regularly-occurring third party (objective) evaluations take place • Technology enables validation and verification • Data quality expectations reflected in individual performance plans
Level 3	• **Established** set of policies • **Proactive** vs. reactive • Policies and processes are **shared** across the organization • **Framework** for responsibility and accountability • Capacity for validation of data	• Enterprise-wide policies and procedures are documented and communicated • Training around policies and procedures is developed and deployed • Framework for accountability is emerging • Leaders and managers commit to data quality principles
Level 2	• Data quality steps **documented**, repeatable • Initial policies are **defined** • Processes and policies **vary** across programs • **Limited** anticipation of data issues • **Basic** organizational management and information sharing	• Policies and procedures are documented for most programs or systems (may be different) • Simple data quality rules are in place to enable data managers to analyze data issues • Programs are sharing data quality-related information and practices • Data quality-related training exists for some systems
Level 1	• **Ad hoc** processes • Policies are **informal** and/or undocumented • **Reactive** vs. proactive • Little to no coordination across programs • Data is corrected but not in a coordinated way	• Data is available for reporting, learning, and decision-making • IT systems are in place to collect data • Programs are approaching data quality in their own way, with much variance across the enterprise • Data quality-related training may exist for some systems

Data Quality Program Elements

Agencies should consider four elements when designing, implementing, and improving data quality programs:

1. Policies and Procedures
2. Quality Control and Assurance Practices

3. Governance and Leadership (including culture)
4. Human Capital

Agencies should consider whether their data quality program maturity varies across these elements, using the detailed descriptions of each level provided below.

Level 1

Agency programs and/or components have some awareness of their data quality issues but do not have mechanisms in place across the agency to address the problems. When problems arise, they handle them in ad hoc, reactive ways. Some (not all) programs or components may have policies in place, but there is little or no coordination and communication between organizational silos.

Characteristics

Policies & Procedures	Quality Control & Assurance	Governance, Leadership, & Culture	Human Capital
Policies and procedures may be informal and/or undocumented and likely apply to specific programs or systems.	Data errors are not predictable and are corrected with no coordination with business processes. Quality issues are not connected and cannot be measured. Siloed data systems impede ability to validate and verify.	Across the organization there is varying familiarity with data quality concepts and principles and there is sporadic, ad hoc commitment from leadership to those principles. There is little to no communication from HQ or across programs regarding data quality management.	Few resources are devoted to ensuring good data quality. Data quality-related training may or may not exist for specific systems. Note: Data quality is typically a collateral duty and while more support may exist for higher maturity levels, this will likely still be the case.

Considerations for Moving to Level 2

- Programs/components could survey their systems and document the procedures they currently follow for data collection, entry, and analysis to identify existing gaps.
- Programs/components could share information and practices with others across the enterprise.
- Programs/components could develop and deliver training for data collection, entry, and analysis.
- Programs/components could require and standardize the review of data to verify accuracy.
- Data quality leaders at the program/component or HQ level could begin benchmarking component-level or other agency practices to create standard policies and procedures.
- Data quality leaders at the program/component or HQ level could publish reporting schedules.

Alignment to A-11 Factors

- Data should be available for reporting, learning, and decision-making

See Appendix 1 for descriptions of each factor and the full cross-walk of factors to the maturity model.

Level 2

Most agency programs and/or components are documenting data quality steps (thereby ensuring repeatability), but no enterprise-wide comprehensive standards or approaches exists, resulting in potentially fractured practices and variation across the agency.

Characteristics

Policies & Procedures	Quality Control & Assurance	Governance, Leadership, & Culture	Human Capital
Policies and procedures exist and are documented for most programs and systems. Nascent enterprise-wide policies and procedures may be in development.	Simple errors are identified and reported. Data managers are able to analyze data issues using simple data quality rules and data validation. Efforts are made to integrate systems to allow for data V&V.	Programs are sharing data quality-related information and practices with each other and with HQ. Data quality leadership and commitment varies across the organization. Governance is limited to programs.	Data quality activities are staffed on an ad hoc basis. Roles and responsibilities are defined for some sub-organizations. Training on procedures may be system-specific and not consistently applied across the enterprise.

Considerations for Moving to Level 3

- A group of data quality leaders could be named as a central coordinators and policy owners.
- Data quality leaders could use benchmarking findings to create enterprise-wide policies and procedures.
- Data quality leaders could develop and deliver training on policies and procedures.
- A group of data quality leaders could begin developing an accountability/governance framework.
- Leaders and managers can begin regularly communicating the importance of ensuring data quality.
- Leaders and managers could designate staff resources responsible for implementation and oversight of data quality policies and procedures.
- Data quality leaders at the HQ level could establish coordination patterns with components to ensure that data is being reviewed and verified.

Alignment to A-11 Factors

In addition to the Level 1 factors:
- Externally controlled data should be documented
- Data reporting schedules should be used and distributed
- Data collection (and entry) staff should be skilled and trained in proper procedures. **(Ad hoc basis)**
- Collection standards should be documented, available, and used. **(Ad hoc basis)**
- Data should be reviewed and verified for accuracy. **(Ad hoc basis)**

Level 3

Agency-wide data quality policies are established, including a framework for responsibility and accountability. There is emerging consistency of adherence to these policies across the organization. Established processes allow for proactive quality management so that issues and errors are anticipated in many cases and/or identified early and corrected quickly.

Characteristics

Policies & Procedures	Quality Control & Assurance	Governance, Leadership, & Culture	Human Capital
Enterprise-wide policies and procedures have been established, documented, and communicated across the organization. Adoption is occurring across programs and systems. Note: Some program and system-specific policies and procedures may always exist due to unique needs and requirements.	Better defined policies, procedures, and quality expectations allow for more thorough inspection of data and determination of accuracy. More validation and information sharing enabled by technology allow for limited error prediction to emerge more consistently across the enterprise.	Enterprise-wide communication around policies, procedures, and practices is occurring. A framework for responsibility and accountability may be in development and some roles may be defined. Leaders and managers are supportive of data quality principles and set expectations.	An initial enterprise-wide management plan for data quality resources, to include roles and responsibilities, is in development. Training is developed and deployed across the organization and there is emerging consistency of use.

Considerations for Moving to Level 4

- Agency leadership could commit to:
  - Overarching principles, including a culture of data transparency and use in decision-making.
  - Resources to improve data systems, develop and deliver training, and ensure objective validation and verification by systems and third parties.
- Data quality leaders could organize into a more formal governance structure to provide accountability, streamlined communication, and best practice sharing.
- Data quality leaders could build feedback loops into data quality processes in order to provide an opportunity for continuous improvement (e.g., a survey on confidence in and use of data).
- Data quality leaders could begin incorporating expectations into individual performance plans.
- Data quality leaders could establish regularly-occurring third party evaluations of data quality.
- Responsible officials could regularly certify data accuracy and completion of procedures.

Alignment to A-11 Factors

In addition to the Level 1 and 2 factors:

- Source data must be well-defined and documented, with definitions
- Procedures for editing previously entered data should be established
- The accuracy limits of data should be acknowledged
- Data collection (and entry) staff should be skilled and trained in proper procedures. **(Enterprise)**
- Collection standards should be documented, available, and used. **(Enterprise)**
- Data should be reviewed and verified for accuracy. **(Enterprise)**

Level 4

Data quality is managed by an active and well-defined governance structure and group that employs a continuous improvement approach to agency-wide data quality policies. Agencies assess the quality of their data on an established schedule, using objective third parties (internal or external to the agency).

Characteristics

Policies & Procedures	Quality Control & Assurance	Governance, Leadership, & Culture	Human Capital
Enterprise-wide policies and procedures are adopted as standard business processes across the organization. They are continuously improved upon and communicated. Note: Some program and system-specific policies and procedures may always exist due to unique needs and requirements.	There are regularly occurring quality audits by internal or external parties. Data quality issues are recognized early in the information flow. Technology enables early detection and correction guided by defined policies and procedures.	A governance structure ensures consistent adherence to policies and procedures and enables best practice sharing/continuous improvement. Leadership is committed to data quality principles, relies on data for decision making, and creates a culture of data usage.	Data quality roles and responsibilities are documented for all staff involved in the data lifecycle. Training is provided to ensure consistent adherence to procedures. Data quality standards are incorporated into individual performance targets to enhance accountability.

Considerations for Optimization/Continuous Improvement

- Data procedures should be thoroughly documented and made accessible to all staff responsible for reporting data. Data quality leaders should annually train staff in data reporting procedures and on-demand training should be made available.
- Data quality leaders could review the effectiveness of the existing governance structure to improve accountability and communication.
- Data quality leaders could identify and document data limitations (including the impact on the usefulness of performance data), and implement action plans to improve where necessary.
- Data quality leaders could incorporate appropriate responsibilities for data quality and use in individual performance plans across the enterprise, using it as a standard criteria for evaluation.
- Data quality leaders could implement annual third party evaluations of data quality based on GAO best practices or OMB guidelines.
- Certification becomes an automatic process with standardized procedures and business rules.

Alignment to A-11 Factors

In addition to the Level 1, 2, and 3 factors:
- Methodology for data entry should be documented and followed
- Data limitations should be explained and documented
- Method for handling anomalous/irregular data should be established
- Third party evaluations should be conducted
- Officials should certify that procedures were followed/data accuracy checked each reporting period
- Employee's performance standards should include accountability for data accuracy

Appendix 2: Agency Examples by A-11 Factor

OMB Factors	Description of Recommended Factors	Maturity Level
Standards and Procedures	Source data must be well-defined and documented, with definitions. Data definitions are well documented and distributed to all responsible for specific data collection, responsible offices can document adherence to data definitions, definitions and standards are used in a consistent manner for all parties involved in specific data collection.	**3**
	Collection standards should be documented, available, and used. Protocols and methodology for data collection are documented, distributed to all responsible for data collection, and adherence to the protocols is required and verified, data sources are documented.	**2 (ad hoc) 3 (enterprise)**
	Data reporting schedules should be used and distributed. GPRA and other data reporting schedules linked to decision-making are issued to all parties responsible for data collection; timely data collection and reporting is routinely practiced.	**2**
	Data collection staffs should be skilled and trained in proper procedures. Those responsible for collecting/assembling data are trained.	**2 (ad hoc) 3 (enterprise)**
Data Entry and Transfer	Methodology for data entry should be documented and followed. Documentation of data entry procedures/protocols is available, understood, used by data entry personnel. Network of data sources is identified/methods used are comparable for all entry locations.	**4**
	Data should be reviewed and verified for accuracy. Calculations checked, data consistency checks employed (e.g. electronic editing).	**2 (ad hoc) 3 (enterprise)**
	Procedures for editing previously entered data should be established. Procedures for making changes are documented and followed.	**3**
	Data should be available for reporting, learning, and decision-making. Data available for GPRA reporting/critical decision making cycles.	**1**
	Data entry staff should be appropriately trained and skilled. Data entry staff are skilled and trained in proper procedures.	**2 (ad hoc) 3 (enterprise)**
Data Quality and Limitations	The accuracy limits of data should be acknowledged. Estimated data are identified; methodology for estimation is documented and is supportable; use of estimates are minimized; data with margins of error due to accuracy of instrumentation or interpretive leeway, are identified and margin of error is reported; incomplete data are identified and extent of missing data is reported; preliminary data are identified and qualifications on data are described.	**3**
	Data limitations should be explained and documented. Any other data limitations explained/documented. Mitigation plans described.	**4**
	Method for handling anomalous/irregular data should be established. Data that appears to be incongruous compared to most other data obtained is re-evaluated and handled appropriately.	**4**
	Third party evaluations should be conducted. Objective internal/external parties are periodically used to verify accuracy/quality of data. Use of other crosschecks on data quality such as comparison to similar databases are employed and documented.	**4**
	Externally controlled data should be documented. Need to use external data is established; external is identified as such.	**2**
	Employee's performance standards should include accountability for data accuracy. Accountability exists in performance standards.	**4**
	Officials should certify that procedures were followed/data accuracy has been checked each reporting period. Responsible officials certify that procedures were followed each reporting period/data are accurate by signing/filing certifications.	**4**