

Data Quality Standards

Purpose

The purpose of this standard document is to define the dimensions of data quality and provide best practice guidance to ensure consistent, accurate, and reliable data entry processes.

Scope

This policy applies to ODGA employees who create, manage, or use data as well as contractors and consultants.

Standards

Quality Standards

- Data quality management processes shall be incorporated to monitor, measure, and improve the quality of data exchanged between systems, ensuring that data is accurate, complete, and consistent across integrated systems.
- Data quality efforts should focus on preventing data errors and conditions that reduce the usability of data rather than correcting records.
- Data quality efforts should ensure adherence to the Six Dimensions of Data Quality ([Meet the data quality dimensions - GOV.UK \(www.gov.uk\)](https://www.gov.uk/government/publications/meet-the-data-quality-dimensions))

Completeness: Data should be complete, containing all necessary information required for its intended use. There should be no missing fields or values unless they are legitimately unavailable.

Uniqueness: Data records or entity should be distinct and not contain duplicate entries. Uniqueness measures whether each data item, such as a customer record, product ID, or transaction, appears only once in the dataset or database.

Timeliness: Data should be up-to-date and relevant for its intended use. It should be collected, processed, and made available within a reasonable timeframe to support decision-making and analysis.

Validity: Data should adhere to predefined rules, standards, and constraints. It should be valid in terms of its format, structure, and content, ensuring that it conforms to expected criteria.

Accuracy: Data should be accurate and precise, reflecting the true value or condition it represents. This means minimizing errors, discrepancies, and inconsistencies.

Consistency: Data should be consistent across different datasets, systems, and time periods. Consistency ensures that data can be reliably used and compared without unexpected variations or discrepancies.

- Data Stewards shall set specific data quality metrics and targets for COV critical assets in accordance with the tiers defined in Enterprise Architecture Standard EA-225
- Detailed documentation should accompany the data, describing its source, structure, meaning, and any transformations or processing applied to it. Documentation helps users understand the data and its context, facilitating its proper interpretation and usage.
- Data processes and changes should be auditable, allowing for tracking, monitoring, and accountability. Audit trails should be maintained to record data lineage, modifications, and access history for compliance and analysis purposes.

Data Quality Best Practices

- Define the format and structure for data entry, including the type of data (text, numeric, date, etc.), field lengths, and allowed characters. Ensure consistency in data representation to facilitate analysis and processing.
- Establish validation rules to enforce data integrity and accuracy during entry. Validate data against predefined criteria, such as range checks, format checks, and consistency checks, to prevent errors and inconsistencies.
- Identify mandatory fields that must be completed during data entry to ensure completeness and consistency. Clearly indicate required fields and provide prompts or error messages for missing information.
- Define standardized conventions for data entry, including capitalization, abbreviations, date formats, and punctuation. Consistent formatting enhances readability and ensures uniformity across datasets.
- Design user-friendly data entry interfaces with intuitive layouts, clear labels, and logical workflows. Minimize the need for manual entry wherever possible by incorporating dropdown menus, auto-population features, and data validation prompts.
- Provide comprehensive training and education programs for data entry personnel to familiarize them with data entry standards, procedures, and tools. Offer ongoing support and resources to address questions, challenges, and updates.
- Implement quality assurance procedures to review and validate entered data for accuracy, completeness, and consistency. Conduct regular audits, spot checks, and validation checks to identify and correct errors or discrepancies.

- Implement data entry controls to prevent unauthorized access, manipulation, or deletion of data. Enforce user authentication, access permissions, and logging mechanisms to track and monitor data entry activities.
- Establish version control procedures to manage updates, revisions, and changes to data entry standards and processes. Maintain a centralized repository for documenting standards, guidelines, and revisions.
- Document data entry standards, procedures, and guidelines in a comprehensive reference manual or handbook. Provide clear instructions, examples, and troubleshooting tips for data entry personnel. Generate reports and metrics to monitor data entry performance and compliance with standards.
- Enforce compliance with data entry standards through regular monitoring, evaluation, and enforcement measures. Address non-compliance issues promptly and provide feedback, training, or disciplinary action as necessary.

Data Quality Metrics

Measuring data quality is essential for ensuring that data is accurate, reliable, and fit for its intended purpose. Several metrics can be used to assess different aspects of data quality. Here are some of the best metrics for measuring data quality:

1. Completeness

Completeness Ratio: The ratio of complete records to total records in a dataset.

Missing Values Percentage: The percentage of missing values or fields in the dataset.

2. Accuracy

Error Rate: The percentage of data records containing errors or inaccuracies.

Validation Rate: The percentage of data records passing validation checks or rules.

3. Consistency

Consistency Index: A measure of the consistency of data across different datasets or systems.

Standard Deviation of Data Values: The extent to which data values deviate from the mean or expected value.

4. Timeliness

Data Age: The time elapsed between data collection or entry and its availability for analysis or decision-making.

Data Refresh Rate: The frequency at which data is updated or refreshed to reflect changes in the underlying sources.

5. Validity

Validity Rate: The percentage of data records conforming to predefined standards, formats, or rules.

Data Integrity Violations: The number or frequency of violations of data integrity constraints or rules.

6. Uniqueness

Duplicate Records Percentage: The percentage of records that are duplicates or redundant within the dataset.

Distinct Values Count: The count of unique values within specific data fields or attributes.

Standard Review

This Policy will be reviewed and updated every three (3) years from the approval date, or more frequently if appropriate. Any staff members who wish to make any comments about the Policy may forward their suggestions to the Director of Data Governance, ODGA.

Related Policies

The Office of
Governance

ODGA Policies, Standards and Procedures
Data Governance Policy
Data Protection Policy

Data
and Analytics

adheres to all Commonwealth Information Technology Resource Management (ITRM) policies and standards for security and architecture [Policies, Standards & Guidelines | Virginia IT Agency](#).

VITA Related Policies
Enterprise Architecture Information Architecture EA-225
IT Information Security Policy - SEC519
Information Security Standard (SEC501)
IT Risk Management Standard (SEC520)

Version History

Version Number	Revision Date	Description of Change	Author
V1	2/16/2024	Initial Draft	Chris Burroughs