

Functional Practice Statements - Data Cleansing
<div><div>Level 1: Initial</div><div><p>1.1 Data cleansing requirements are defined and performed.</p><p>Refer to Data Requirements Definition for more information related to development and management of requirements.</p><p>Example Work Products</p><ul style="list-style-type: none">• Data cleansing requirements• Data cleansing guidelines</div></div>
<div><div>Level 2: Managed</div><div><p>2.1 Data cleansing activities adhere to data cleansing requirements, which are linked to process improvements to achieve business objectives.</p><p>Data cleansing requirements are often derived from data profiling, data quality assessments, change requests, identification of report discrepancies, and regulatory audits. Requirements typically identify the following topics:</p><ul style="list-style-type: none">• What problem to solve (i.e., business objective not being met as a result)• What data set to clean (data or metadata)• How the target state of the data will solve the problem<p>Describing the target state of the data involves an understanding of the business rules, data definitions, current and future usage, quality criteria, business objectives, cost and benefit analysis, data, software and system architecture, etc.</p><p>2.2 Data cleansing activities conform with data quality requirements (e.g., quality dimensions such as conformity, accuracy, uniqueness) and quality criteria.</p><p>Data cleansing processes and rules should be documented and linked to business objectives and requirements. The resulting cleansed data should be evaluated for consistency with the data definitions. The data dictionary, in turn, is updated as a result of deficiencies discovered during data cleansing efforts.</p><ul style="list-style-type: none">• Data cleansing rules are used to support the achievement of data cleansing objectives and processes.<p>The development of rules typically involves:</p><ul style="list-style-type: none">• Validating data cleansing rules against business definitions• Linking data cleansing rules to business rules• Conformance with quality criteria• Use of business SMEs to create rules• Documentation and approval of data cleansing rules and their validation• A process for adding new rules• Measures to assure consistent application of cleansing rules across repositories<p>2.3 The scope of data cleansing is defined.</p><p>The scope includes the identification of the data set, data elements and rules that will define the cleansing activity, as well as the criteria for prioritizing data cleansing activities. The scope needs to be balanced with any time-bounds or resource constraints that may exist. The impact analyses produced by the data quality assessment process are input to estimating and setting the boundaries of data cleansing efforts.</p><p>Data Quality Assessment and Data Profiling provide complementary information that can be leveraged for this practice.</p><p>2.4 The process for performing data cleansing is defined by a plan.</p><p>Typical planning steps include the following:</p><ul style="list-style-type: none">• Evaluate the downstream impacts of the change• Determine where the changes needs to be made• Reconcile any changes with respect to the data's original intent, versus avoids unintentionally changing the meaning of the data, which results in inefficiencies and confusion<p>Data cleansing plans typically include:</p><ul style="list-style-type: none">• The tools and methods that will be used for correcting the data (i.e., multiple repository comparison, verification against valid source, logic checks, etc.)• Defined process for resolving differences of opinion on data validity• Logging of corrections, audit trails, version control (configuration management)• Feedback on needed adjustments to the data store maintenance plans• Definition of how referential integrity will be maintained after data changes are implemented<p>2.5 A data cleansing policy is established and maintained.</p><p>During development, it is important to ensure that appropriate stakeholders (governance through execution) are involved. If prescribed methods and tools will be used to support data cleansing activities, the policy should identify them.</p><p>The policy should include a mechanism for communication to affected stakeholders, as well as a means to manage their feedback. In addition, the policy should include any specific criteria or thresholds expected to drive data cleansing activities.</p><p>The policy should clearly articulate expectations for root cause analysis, business, and technical impact. It should also specify the criteria for prioritizing and scheduling data cleansing projects, as well as guidelines for resource allocation.</p><p>2.6 Methods for correcting the data have been established and are defined within a plan.</p><p>Methods may include multiple repository comparison, verification against valid source, logic checks, referential integrity, or range tolerance.</p><p>Methods typically involve functions to ensure the following:</p><ul style="list-style-type: none">• Content is standardized, validated, and maintained with the use of external sources as needed• There is a process for resolving differences of opinion on the validity of the data• Log corrections, audit trails, and version control (configuration management) have been defined• Feedback to data store maintenance plans has been delivered• Referential integrity is maintained after data changes are implemented.<p>2.7 Data cleansing issues are communicated and resolved, when possible, in the internal or external source.</p><p>The best defense against spending the time and effort to cleanse the same data set for multiple instances on multiple occasions is prevention at the point of origin.</p><p>Example Work Products</p><ul style="list-style-type: none">• Data cleansing policy• Data cleansing processing and rules• Data cleansing metrics• Data cleansing plans• Data correction methodologies• Data cleansing issues</div></div>
<div><div>Level 3: Defined</div><div><p>3.1 Data change history is maintained through cleansing activities.</p><p>Data change history should be supported by traceability of cleansing rules from their requirements to where the changes are being made and down through their interconnected processes.</p><p>See the Metadata Management for guidance on ensuring that necessary information is managed to support this activity.</p><p>3.2 Policies, processes, and procedures exist to ensure that data cleansing activities are applied at the point of origination in accordance with published rules.</p><p>3.3 Data cleansing rules are applied consistently across the organization.</p><p>Quality rules and business rules for data elements that are shared among multiple business lines should not diverge unless a considered exception has been approved.</p><p>3.4 A governance group establishes, maintains, and ensures adherence to data cleansing rules.</p><p>3.5 Standard data cleansing results report templates, at the detail and summary level, are employed.</p><p>The organization can monitor these reports to develop a picture of duplicative efforts (cleansing the same data sets in multiple data stores). In circumstances where the same shared data is used (e.g., customer data), rather than spend the time and allow separate efforts to fix several data repositories, resources may be better spent to modify systems and processes to consume from a single standard data set that can meet the quality requirements of all users.</p><p>Example Work Products</p><ul style="list-style-type: none">• Data change history log• Traceability matrix• Data cleansing feedback• RACI matrix for data cleansing governance, activities, and rule development• Data cleansing results report templates</div></div>
<div><div>Level 4: Measured</div><div><p>4.1 Service level agreements include data quality criteria to hold data providers accountable for cleansed data.</p><p>The results of data validations should be provided to internal and external suppliers to enable the improvement of their cleansing processes.</p><p>Refer to Provider Management for more information related to these expectations.</p><p>Example Work Products</p><ul style="list-style-type: none">• Service level agreements• Feedback documentation</div></div>
<div><div>Level 5: Optimized</div><div><p>5.1 The organization is involved in the establishment and maintenance of external or industry standards for improving the quality of data content.</p><p>This is especially applicable when the organization is involved in industry initiatives to standardize data produced or consumed by multiple organizations.</p><p>5.2 Data cleansing requirements for data providers are managed in accordance with standardized processes.</p><p>Example Work Products</p><ul style="list-style-type: none">• Meeting minutes showing involvement in standards• SLAs include cleansing processes and expectations for data providers</div></div>