Functional Practice Statements - Data Integration

Level 1: Initial

1.1 Data integration occurs between systems.

An example is a production database administrator (DBA) manually initiating multiple integration scripts in a predetermined order, or making manual changes to correct defective data.

Example Work Products
- Data integration scripts

Level 2: Managed

2.1 Data integration plans are documented.

Project documentation for data integration should include the following topics:
- Identification of data scope
- Identification of vendors and source systems (and SLA)
- Identification of data growth and data retentions
- Identification of applicable data standards
- Integration sequence of sources
- Decision guidance related to data trust (e.g., precedence)
- Analysis using quality dimensions
- Change-data capture rules
- Profiling guidance
- Conformance and transformation rules
- Import exception handling
- Roles and responsibilities of stakeholders and data management staff
- Validation against business requirements
- Expectations for impact analysis
- Expectations for integration testing
- Expectations for backout triggers
- Risk mitigation planning for data backup and recovery
- Data archive process
- Data privacy protection
- Data access matrix for different groups of users.

2.2. The set of data integration disciplines and tools used by the organization provides bulk transport and load, change data capture, versioning and configuration, metadata capture and management, and in-line data quality checks and controls.

Data transfer from sources to destinations typically requires complex rules, data transformations, and data standardization. To accomplish these tasks consistently and with the least amount of risk, specialized tools designed to support physical capture and movement of data should be applied, but they can also be used to provide basic analysis of the data and transformations.

These tools facilitate maintenance of the metadata associated with ETL processes and provide a means to make changes in a controlled fashion.

Refer to Metadata Management for more information on establishing the processes and infrastructure for specifying and extending clear and organized information about the data assets under management, to increase data sharing, ensure compliant use of data, increase responsiveness to business changes, and reduce data-related risks.

Refer to Architectural Standards for more information on providing an approved set of expectations for governing architectural elements regarding approved data representations, data access, and data distribution, fundamental to data asset control and the efficient use and exchange of information.

2.3 A change control process is established and followed to ensure that changes to the integration environment, including upstream sources and down-stream targets, are controlled and coordinated.

Changes to operational systems often have a direct impact on the data which are stored or managed by those systems and subsequently used by other downstream systems. Examples of downstream systems include both repositories (e.g., data warehouse) and other operational systems that rely on th data from the source system..

2.4 Remediation processes are established and followed to address selected abnormal circumstances.

It is important for the business to have confidence that the system it depends on is managed proactively—the business maintains data availability and integrity by having plans for dealing with potential interruptions and unforeseen events, including load interruptions or data validation failures.

2.5 Integration verification is performed to ensure that architecture and interface specifications are documented and will be met prior to being released into production.

Refer to Architectural Standards for more information related to interface and integration expectations.

Example Work Products
- Data integration standards
- Verification and validation plans
- Integration test environments
- Application programing interfaces (APIs)
- Data integration policy

Level 3: Defined

3.1The organization follows a standard set of practices and rules for performing data integration activities.
.

3.2 Quality checks are defined as part of the organizational integration standard and performed as part of data integration processes.

For example, the ETL process should perform a basic level of quality control checks, especially for subsequent change-data capture loads, to ensure that the data being integrated meets the quality attributes defined by the organization. These checks should define criteria by which the data are rejected, flagged, or queued as exceptions for subsequent action.

Failure to perform this check at integration may cause problems with the data that can be unknown for long periods of time and require greater cleanup efforts later. The organization may want to verify that the following tasks are performed:

- Data is profiled
- Data is complete and accurate by attribute
- Logic is instituted to handle missing data
- Provisions for alternative sources exist if data are not available from the preferred source
- Data quality and error handling reports are produced and monitored

Data Quality Strategy process area provides guidance on expectations and standards for quality checks and understanding data quality dimensions.

3.3 A standard process is established and followed to create and verify data precedence rules with business users based on use cases, requirements, and selected triggers.

Often, multiple sources contribute to data integration into a single logical destination, such as operational systems contributing to a data warehouse. In these instances, it can be expected that more than one source may contribute
data that is the same or similar to data from other sources. In this example, during the final ETL process that feeds the single repository, it is necessary to establish and apply precedence and conformance rules to guide the final integration.

The results of performing Data Lifecycle Management practices provide input to this process area by helping to understand the use and need of the organization's data.

3.4 The development and deployment of integration interfaces are specified in accordance with architectural standards supporting re-use.

Some examples include consolidated interfaces and common data services. Data integration rules and performance requirements are clearly defined. Interfaces between systems should be well documented to support future changes, as well as provide a means for accommodating integration to or from additional systems. Development of these interfaces should follow common programming practices for defining and documenting application program-ming interfaces (APIs). Organizational standards should exist to guide these activities.

Interfaces should also be included in the complete technical data package and managed following the organization's configuration management process-es and standards.

3.5 Interface and integration performance metrics are collected and analyzed to identify nonconformance with standards and criteria.

For example, tracking of defects is performed to monitor defect severity, causes, statistical occurrence, and plan resolutions.

3.6 The organization documents and manages changes to data sources and destination through the data governance process.

Example Work Products
- Verification and validation results
- Performance requirements
- Performance metrics and analysis results
- Measures and metrics for continuous improvement in data quality
- Data delivery policy and SLAs
- Integration method standards
- Integration best practices guidance
- Standard interface specifications
- Integration environment change management process

Level 4: Measured

4.1 Statistical analysis of integration metrics guides decisions on changes to
interfaces and integrations.

Variation criteria (i.e., tolerance limits) are specified and used in the analysis of interface performance. For example, analysis of data profiling results may indicate that the volume of integration-related defects is within an acceptabl
tolerance.

4.2 Selected highly shared data is fully integrated, centrally managed, and delivered as needed to integration data stores.

An organization's business processes that rely on highly shared data (e.g., reference data, master data, etc.) are greatly affected by discrepancies differ-ences in timeliness, representation, and values—in vital information about key entity types, such as customer, account, product, organization, etc.

A specific goal of data integration is to streamline, standardize, centralize, and provide trusted shared data most needed by multiple lines of business and multiple business purposes. The results of Data Lifecycle Management provide input to this practice by helping to understand the use and need of data across the organization.

Example Work Products
- Statistical analysis results
- Data profiling analyses
- Consolidated highly shared data with continuous improvement

Level 5: Optimized

5.1Performance models for data integration are periodically reviewed, and are used as input for enhancements.

5.2 The organization publishes and shares integration best practices within its industry.

Example Work Products
- Quantitative models
- Performance triggers and thresholds
- Root cause analysis results
- Presentations, white papers, or published articles.