Data Quality Management Policies at The University of Tabuk

Document Version Control

Document information	Data Quality Management Policy at the University of Tabuk
Document preparation	Data Management Office (DMO)
Version No	Version 1.0
Document status	Initial Draft
Date	11/07/2023

Document amendment/revision

Version	Date	Additions/ amendments	Added/ amended
1.0	11/07/2023	DMO	N/A

Document revisions/approvals

Date	Name	Position	Signature
11/07/2023	Dr.Mohammed Alotaibi	Director of DMO	

Table of Content

1. Document Information

1.1 Purpose of the document

The purpose of this document is to set out a clear policy framework for maintaining and increasing high levels of data quality within University of Tabuk. Good data quality is essential, and the availability of complete, accurate, consistent, unique, valid and timely data is important in supporting the University of Tabuk's operations and decision-making process. The processes associated with data quality management is defined in processes and guidelines document.

1.2 Audience

All internal stakeholders who are required to interact with data owned by University of Tabuk

1.3 Document Control

- All changes to this policy should be recorded in the document control section
- Changes to the policy should be reviewed and approved by Data Governance Steering Committee (DGSC)
- <ENTITY NAME> shall conduct regular reviews of the Data Quality Policy at defined intervals to identify initiatives for the continual improvement of the Data Quality Process

2. Introduction to Data Quality Policy

2.1 Scope

This policy is intended to cover the collection, recording, validation, further processing and reporting of all types of information generated and used within, or reported externally, by University of Tabuk

2.2 Objective

The objective of Data Quality policy is to maximize business outcomes, improve usability and reduce time and expenses associated with excessive data reconciliation.

2.3 Glossary of Terms

Terminologies used in the documentation	
CDE	Critical Data Element
DQ	Data Quality

DGSC	Data Governance Steering Committee
DMO	Data Management Office
UT	University Of Tabuk

3. Data Quality Policies

Data quality is the ability to supply accurate, timely and complete data, which can be translated into information, whenever and wherever this is required. Data quality is vital to effective decision making at all levels of the organization. Following DQ processes are covered in the policy document

- Critical Data Element (CDE) identification and update
- DQ business and technical rules development
- DQ threshold determination and attribute registration
- Data profiling and monitoring

3.1 Critical Data Element (CDE) Identification and Update

3.1.1. Critical Data Elements (CDEs) are the highest priority candidates that have an impact on a regulatory data, financial calculation, performance indicator or risk measure

3.1.2. Business Data Steward shall be responsible for CDE identification, and the selection criteria of these CDEs shall be documented appropriately

3.1.3. Potential CDEs shall be identified based on the set of criteria that can include, but not limited to:

- Key data involved in Key Business C-Level Reports
- Underlying data that is important for business continuity and business success (e.g. Revenue, Cost)
- Data for which expected Data Quality is Highest (e.g. Expenditure, Revenue, Budget Amount, Payment Amount)
- Data required to be maintained as per Saudi Government regulatory requirements
- Data involved in key risk and compliance reporting
- Data related to individual, department or organization performance
- Data related to critical employee information
- Data recognized as critical party (Beneficiary, supplier) information

3.1.4. CDEs shall be reviewed and approved by Business Data Executive and approved by the Data Management Committee to ensure the CDEs are appropriate in accordance with approved criteria

3.1.5. The Data Management Committee shall conduct a periodic review of the CDEs to determine that the CDEs continue to represent the critical data elements.

3.2 Data Quality Business and Technical Rules Development

3.2.1. Business Data Steward shall be responsible for defining business data quality rules for all the data elements in their business unit.

3.2.2. Business data quality rules shall be defined based on

- Internal data quality requirements as defined by business data consumers
- External data quality requirements as defined by applicable laws or regulations

DQ Dimension	Definition	Examples	Metric
Completeness	A measure of the absence of blank (null or empty string) values or presence of non-blank values	In order to consider the population of Zip Code in the Address table acceptable, the percentage of populated records must not be below 80%	Total number of populated records ÷ total number of records × 100
Uniqueness	A measure of the degree that define whether the data element is unique or not duplicates	In order to consider the Zip Codes master table unique, there should only be one row per zip code and the percentage of duplicate records must not exceed 0%	The number of duplicate records ÷ the total number of records × 100
Validity	A measure of the degree to which data are valid and conforms to the syntax (format, type, range) of its definition	Only invoiced orders should be dispatched and the percentage of records that meet the condition must not be below 100%	The number of records that meet the rule ÷ the total number of records × 100
Accuracy	A degree to which data values align to real values	Person's residence must correspond to the data of the Population Register the percentage of records that meet the condition must not be below 100%	The number of records that meet the rule ÷ the total number of records × 100
Consistency	A measure of the degree to which data retains consistent across different data sources	A citizen's date of birth shall be the same across multiple information systems and the percentage of inconsistent records must not exceed 0%	The number of inconsistent records ÷ the total number of records × 100

DQ Dimension	Definition	Examples	Metric
Timeliness	A measure of the degree to which data is consistent with the most recent business event and available when required	In order to consider the arrival of Financial Market records to be timely, it should arrive on the same day of transaction to the brokerage firm and the percentage of transactions not completed on time must not exceed 1%	The number of incomplete transactions ÷ the total number of attempted transactions in a time period × 100

3.2.3. Data Quality dimensions are required to be mapped to business processes where possible. This shall provide the capability to assess the impact of both high and low quality data on business processes

3.2.4. The profiling metrics and target depth level shall be appropriately defined against each data element based on the level of complexity required for data quality rules

Dimensions	Data Profiling Depth
Completenes Uniquness Validity	Level 1 - Single Field Validation • Null/ blank values • Unique values • Valid values in range
Accuracy	Level 2 - Validation within an entry/ table • Enforced key rules • Sequencing rules • Cross validation
Consistency Timeliness	Level 3 - Validation across tables/ systems/ external data • Logical key/ date relationships • Data Obsolescene

3.2.5. Define reusable DQ business rules wherever possible that can be used across multiple data elements (e.g. Validity rule for date fields can be reused)

3.2.6. DQ technical rules shall be developed by IT Data Steward

3.2.7. IT Data Steward shall provide technical rules to support above-mentioned business rules: attribute technical table, column name, mandatory or optional, length of attribute, data values for a business attribute, normalize the data to ensure the Consistency, data dependency rules

and data relationship across entities/tables/systems. Example: Email NOT LIKE "*@@*" AND Email NOT LIKE "*..*" GENDER IN ("M" or "F")

3.2.8. Business data quality rules shall be approved by Business Data Executive, technical data quality rules shall be approved by Data Governance Manager and both shall be approved by the Data Management Committee

3.3 Data Quality Threshold Determination and Attribute Registration

3.3.1. Business Data Steward is responsible for defining the data quality thresholds for all data elements in their business unit

3.3.2. Business Data Steward shall determine minimum thresholds for meeting business expectations and the business impact of exceeding the threshold

3.3.3. Data quality thresholds shall be identified to specify the acceptable limits for the data quality metrics. A threshold is defined as upper or lower boundary of the data quality metric (degree of failure accepted) – E.g. Completeness: 2% to 4%, Uniqueness – 0%

3.3.4. Data quality thresholds shall be reviewed and approved by Business Data Executive and further approved by Data Management Committee

3.3.5. Data Governance Manager shall be responsible for centrally registering the data quality attributes that are required for each data element in the identified repository.

3.3.6. Data quality attributes shall include data quality business and technical rules, data quality thresholds and the rationale for why the data element is considered as a Critical Data Element for University of Tabuk.

3.4 Data Profiling and Monitoring

3.4.1. Data Governance Manager shall be responsible for executing data profiling activity on a defined frequency (monthly or quarterly)

3.4.2. Data profiling is usually performed at the outset of a data quality program; however, this will also be performed as an ongoing activity to monitor the data quality health in many business scenarios such as

- A data quality assessment
- Enterprise Data Warehouse Implementation
- A system upgrade or new implementation
- Any data migration
- Enhancing the National Data Maturity Index.

3.4.3. Profiling metrics shall be reviewed and confirmed by Business Data Steward for each data element based on the data quality rules and thresholds that have been defined

3.4.4. A Data Quality checklist is required to be developed for monitoring the Data Quality assessment results.

3.4.5. An agreed data profiling tool shall be used to execute the data profiling exercise across the University of Tabuk

3.4.6. Data profiling tools are required to have capabilities such as Structured data column analysis,Data structure independent integrity analysis,Pattern definition and identification,Reporting generation,Comparison of data audits over time to detect significant changes in quality

3.4.7. The result of data profiling shall be recorded in data profiling report and shall be send to Data Business Steward for appropriate evaluation

3.4.8. Data Business Steward shall evaluate the profiling results and if threshold reconfiguration or business data quality rule changes are required, then the appropriate change requests will be raised to amend the post profiling inputs

3.4.9. Data Business Steward shall identify the data quality issues and log the issue in the identified data issue management tool.

3.4.10. Chief Data Officer is required to ensure the Data Quality assessment helps in enhancing the National Data Maturity Index

3.4.11. Data Governance Manager shall summarize the data profiling results and present it to Data Governance Head for post profiling discussions. Data quality profiling metrics include, but are not limited to:

- Data Quality Profiling Index: representing the defect count across DQ dimensions
- Data Quality Profiling Coverage: Index representing percentage of data elements that are currently being profiled

3.4.12. Data Quality of third-party data suppliers is required to be assessed before incorporating into the Entity's data repository. Assessment shall be in the form of performing spot checks to ensure conformance to Data Quality rules

3.4.13. The results of the Data Quality assessment is required to be audited. The audited results are required to be gathered and updated as part of the metadata which is registered within the Data Catalog automated tool.

3.5 Data Quality Audits

3.5.1. University of Tabuk is required to perform Data Quality Audit to verify the accuracy of DQ assesment. The following Data Quality audits shall be adopted based on the <Entity Name> requirements:

3.5.2. One-off data quality audit to ensure full coverage

3.5.3. Incremental data quality audit as part of checkpoint processes

3.5.4. Focused data quality audit categorized by data and systems (e.g., strategic or critical systems first)

3.5.5. Data Quality audits for all the systems undergoing change

3.6 Data Quality Continuous Improvement

3.6.1. The Entity is required to set goals to continuously improve the quality of their data.

3.6.2. The Entity is required to identify the gaps between the stated data quality definitions and the monitored data quality measures and execute an iterative data cleansing initiative to improve data quality and ensure continuous improvement in line with changes in business and technical requirements

3.6.3. Data quality improvement initiatives shall be determined, and may be carried out:

- On a system-by-system basis
- By master profile or other data type, across multiple systems
- According to business benefit

3.6.4. Entities are required to ensure the overall health of their data by regularly reviewing and improving how data is modelled.

3.6.5. To ensure continuous improvement, the Entity is required to take the following actions.

- *Ensuring the adherence to the periodic updates to the Data Quality standards.*
- *Introducing data quality checks for data entry, such as data validation*
- *Improving data storage and data architecture*
- *Improving training and guidance for those involved in data entry*
- *Introducing automation, such as validation on data entry, automated quality checks or using specialist coding tools rather than spreadsheets*

3.7 Data Quality communication

3.7.1. The Entity is required to enhance their Data Quality communication process by following the below activities

3.7.2. Proactively engage with users to understand their Data Quality priorities

3.7.3. Effectively communicate for understanding the Data Quality issues and provide guidance for resolving any internal conflicts on Data Quality

3.7.4. Regularly communicate with users to understand any changes in their requirements.

3.7.5. Provide clear data quality information and describe its impact on use of the data.

3.7.6. Communicate trade-offs in data quality clearly to aid understanding of the data's strengths and weaknesses.

3.7.7. Be transparent about the quality assurance approach taken and communicate data quality issues clearly to users

3.7.8. Build strong relationships by continuously communicating with suppliers of external data to identify data quality problems at source.

3.7.9. Inform users in advance about changes made to data processes which could impact on quality.

3.7.10. Provide clear definitions of terminology used and not presume a high level of user understanding of data quality

4. Exceptions

Compliance with the policies of this document is mandatory. All exceptions to this policy shall be explicitly reviewed by respective Heads and approved by Data Governance Steering Committee. The exception shall be reassessed and re-approved if necessary

5. Enforcement

Violations of this policy and supporting policies shall result in corrective action and reviewed by respective Heads. Disciplinary action will be consistent with the severity of the incident as determined by an investigation and may include, but may not be limited to:

- Loss of access privileges to information assets.
- Other actions as deemed appropriate by management

An internal audit shall be carried out based on audit annual plan and a report shall be submitted to the Data Governance Steering Committee.

6. References

- <ENTITY NAME>-DG-PO-DG Data Governance Management Policy
- <ENTITY NAME>-DG-PR-DQ Data Quality Management Procedure

End of Policy