GEODIS DATA QUALITY

Functional Model

for data quality governance and management

GEODIS

Table of Contents

Document – Version follow-up

VERSION	DATE	
V1	10/10/2023	Chief Data Officer (Pierre Lenclud)

DOCUMENT RESPONSIBILITIES	
Owner	Chief Data Officer (P. Lenclud)
Editor	Data Quality & Governance Manager (M. Wolski)
Reviewer	Chief Data Officer (P. Lenclud)
Approver	EVP, Chief Digital & Technology Officer (F. Bottin)

1. INTRODUCTION

The model described herein and adopted by GEODIS constitutes the functional reference of the target Data Quality framework. The adoption of the present model must be framed in the perspective of a path of progressive acquisition of the maturity and skills needed to implement – and improve, in the long term – the key aspects of the model itself.

1.1. Objectives of the document

The objective of this document is to describe the **methodology to define and manage Data Quality controls and quality indicators**, accurately illustrating all the functional elements involved in the exercise.

The scope of this model includes all areas whose data is subject to be put under surveillance. The elements described in the model should therefore be detailed and contextualised for each new area to be included in the scope, in accordance with local principles and guidelines.

This document is complementary to the Data Quality Policy – please refer to the DQ Policy for background information.

1.2. Description of the document content

This document is divided into four main parts:

Part 1 (this part) expresses the context of the document, its objectives, and guidelines

Part 2 provides the Data Quality components

Part 3 gives a detailed view of the Data Quality functional model

Part 4 deep dives into the evolution model

1.3. Distribution methods

Once the cycle of verification, validation and approval of the document has been completed, the Chief Data Officer publishes and makes the functional model available on the MYGEODIS intranet to all stakeholders in the Group, for contextualisation and implementation.

1.4. Methods of approval and review

The Data Governance & Quality Manager (DGQM) is responsible for preparing and regularly updating this document to ensure its relevance and accuracy.

The Data Governance and Quality Manager submits the document to the Chief Data Officer. The Chief Data Officer verifies its **consistency with GEODIS' IT process model**, as well as its **compliance with the GEODIS data quality policies**.

The CDO provides the initial validation of the document, and the Chief Digital & Technology Officer (CDTO) grants the final approval.

The document may be subject to updates in several circumstances, including the following cases: the introduction of new significant regulatory provisions, changes in the scope of

management of the Information Systems Service, changes in the organizational structure, modifications in the IT governance system. The DGQM needs to conduct a comprehensive review of the policy at least once a year to identify any necessary updates.

In case information needs to be updated or modified, contact the DGQM.

Any amendments or additions to the document follow the same approval process outlined above.

2. KEY COMPONENTS

The functional model defines the elements and rules necessary for the effective implementation of quality controls and quality indicators. This model will:

- Enable a quantitative approach for an objective and coherent assessment of the quality of data
- Provide methodology to synthetise the results of quality controls and the preparation of adequate reports for functions / Lines of Business / Regions concerned
- Be applicable to all perimeters of GEODIS group

2.1. Functional objects

There are two key elements to the functional model: the quality control and the Key Quality Indicator (KQI).

2.1.1 Quality control

Controls ensures the respect of a given rule for one individual data element and help prevent anomalies. They answer the question: “Is the rule respected?”. Hence, the answer to this question can only be “Yes” or “No”.

Controls can be blocking or non-blocking, depending on whether the data flow is released to the user in case the control fails.

Controls are usually implemented at the data creation or acquisition step in operational systems to prevent anomalies to occur. However, a business case is to be carried out and studied to decide where controls are to be positioned. In fact, depending on the number of source systems, the further upstream, the better the quality throughout, but the more expensive the control. The further downstream, the less expensive the control but the quality of the data may be poor throughout the chain before the control.

2.1.2 Key Quality Indicator (KQI)

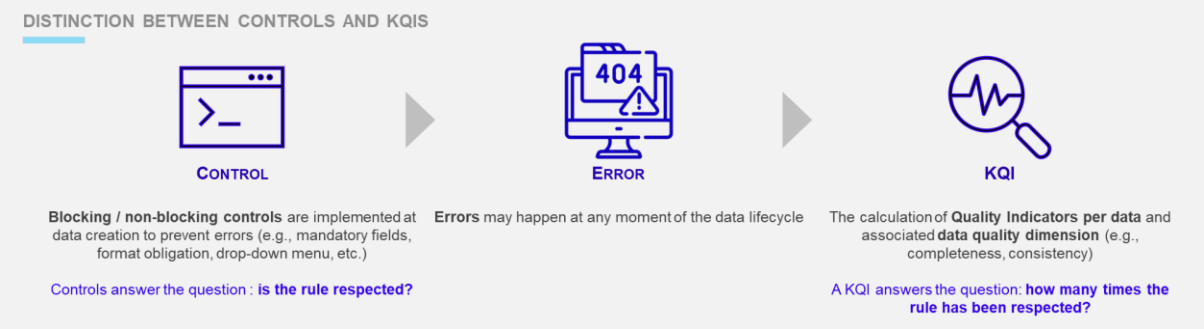
A KQI is the result of a comparison between a given measure and a threshold. It is a metric which shows to what extent a given rule is respected on a certain perimeter of data or dataset, and it gives an idea of the quality of data at a given point in time and along the transformation journey. Therefore, it also provides a measure of the effectiveness of eventual non-blocking[1] controls implemented on the same path.

[1] KQIs apply only to non-blocking controls – blocking controls do not necessitate any monitoring, as only correct values can be input.

A KQI answer the question: “How many times has the rule been respected?”. Hence, the answer can acquire multiple forms, depending on how the final user desires to represent the information.

Depending on the type of information they represent, the quality indicators can be classified as:

- Indices, which express the ratio between the number of cases with a favourable outcome compared to the total number of observations. Indices can vary between 0 and 1.
- Percentage deviations, which express the distance, in percentage terms, between the value of the index and one of the following:
    - A reference value deriving, for example, from other information flows
    - The historical average or trend of the values of the quality indicator in the last N periods



Please note: Quality controls and KQIs are defined and implemented independently of each other. This means that controls can be implemented independently of quality indicators.

2.1.3 Surveillance scope

A surveillance scope is any functional, organisational or technical perimeter where either controls are implemented or KQIs provide information on the quality of the data included in the perimeter itself. The surveillance scope corresponds to only one scope of activity because, since the scope and the use of the data changes, the needs regarding quality change too.

Within a surveillance scope, controls and KQIs can be implemented and carried out using automatic tools or manual activities. In the first case, the organisational units responsible for the check take note of the results produced by the system and decide on the actions resulting from them, while in the second the checking activity also is carried out directly by the operators.

In both cases, the results of the check performed should always be recorded in an administration tool / repository.

The GEODIS group’s position is to automate as much as possible controls and KQIs.

2.2 Functional model building blocks

The main components of the functional model are described below, providing at the end of each paragraph the building block – control or KQI – to which it applies.

2.2.1 Quality Dimensions

The measurement of quality will be carried out according to a multidimensional functional model, which allows to consider in a structured way aspects related to data quality, in accordance with current regulations, if any.

Six dimensions constitute the basis of data quality management[2]:

❖ Completeness: data has all required values, filled with all the necessary information.

It refers to whether all required data is present. It can be measured at the data set, record, or column level.

Example of incomplete data: the client's address is missing.

❖ Validity: data contains values that are compliant to defined business rules.

It refers to whether data values are consistent with a defined domain of values.

Example of invalid data: a user enters a date in the wrong format (American format instead of French one).

❖ Uniqueness: data is unique, i.e., it has no duplicates.

It states that there are no two identical data objects within the data set.

Example of non-unique data: two same products exist in the product referential.

❖ Consistency: data in different systems / datasets and in different moments in time conveys plausible and non-contradictory information.

It evaluates whether same data in different systems / datasets conveys pieces of information that can be valid at the same time; moreover, no inexplicable deviations are identified over time, according to a predefined behaviour.

Example 1 of inconsistent data (non-contradiction across different systems): a GEODIS employee has different referred managers across systems.
Example 2 of inconsistent data (plausibility over time): the FTE count triples in a day, without major organisational changes.

❖ Timeliness: data is provided within the expected timeframes.

It refers to the punctuality with which the data is provided to the data user and to its level of freshness compared to a predefined update frequency. This dimension must be managed by monitoring the data made available and comparing the moment of release with the expected one (in time for the user). These controls are usually referred to the whole information flow and rarely to the single information.

Example of untimely data: transportation milestones are provided a month after the event, not providing sufficient lead time, creating delays, or wrong expectations.

❖ Accuracy: data conveys the right information, according to a predefined source of truth.

It indicates the level of precision of data in terms of deviation from a value that is deemed true / correct. The source can be an officially recognised document / dataset (a KBIS, the SIREN / SIRET national base, etc.) or a business rule.

[2] Source of definitions based on the DM-BOK v2 by the Data Management Association (DAMA)

Example of inaccurate data: inputting the wrong date of shipping from a bill of lading.

The use of different evaluation dimensions is a key aspect of the model; in fact, depending on the purpose for which the data under surveillance is intended, the relevance of each dimension changes.

Dimensions apply to both controls and KQIs.

2.2.2 Calculation logic

As seen in section 2.1.2, a KQI is the result of a comparison between a given measure and a threshold. Such a measure is the result of a calculation logic, which is the set of mathematical or logical operations providing the ratio of data respecting a certain quality control and the overall data set / data flow, within a given perimeter.

The measure is then compared with an identified threshold or target value, in order to provide insightful information.

2.2.3 Thresholds and Target Values

The threshold is the minimum value that the quality indicator must assume for the data to have a certain, desired level of quality. There is an eligibility threshold and a compliance threshold, detailed in section 3.2.1. The outcome of the operation comparing the quality indicator with the threshold determines the level of quality of the data.

The target value represents the value that a quality indicator should assume, in relation to the current level of maturity of the company environment. Therefore, the target values are actively included in the activities of monitoring and continuous improvement of the quality level of the scope.

For any quality check, it is necessary to determine the desired target value and threshold in a given period. In fact, the target value for year N+1 can be 85% and N+2 97% with a threshold of 75% for N+1 and 85% for N+2.

The threshold and target value must be determined by the Data Owners and Data Stewards, preferably with the data consumers, based on their needs and constraints. Several methods exist for setting thresholds:

- Based on the history of the result of a control
- By iteration: first defining an alert threshold a priori and correcting this threshold according to the results of an a posteriori control
- By using existing benchmarks
- By using an impact analysis based on simulations or projections

The Data Stewards can suggest realistic alert thresholds and its periodic updates to improve the data quality over time.

Threshold and target value components are applied on Key Quality Indicators.

2.2.4 Weights

Weights indicate, within a surveillance scope, the relevance of a single quality check compared to another. Weights can be used during aggregation operations to weight both the outcome of the quality check and the quality dimensions (described below).

The weight component is applied to KQIs.

2.2.5 Synthetic overview

Overview of key components and their application to controls and KQIs

Component	Quality Control	Key Quality Indicator
Quality dimensions	X	X
Calculation logic		X
Thresholds and target values		X
Weights		X

Summary table of functional model components