TECHNICAL SPECIFICATION

ISO/TS 8000-150

First edition
2011-12-15

Data quality —

Part 150:
Master data: Quality management framework

Qualité des données —

Partie 150: Données permanentes: Cadre de management de la qualité

© ISO 2011

Contents

Page

© ISO 2011 – All rights reserved

iii

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75% of the member bodies casting a vote.

In other circumstances, particularly when there is an urgent market requirement for such documents, a technical committee may decide to publish other types of normative document:

— an ISO Publicly Available Specification (ISO/PAS) represents an agreement between technical experts in an ISO working group and is accepted for publication if it is approved by more than 50% of the members of the parent committee casting a vote;

— an ISO Technical Specification (ISO/TS) represents an agreement between the members of a technical committee and is accepted for publication if it is approved by 2/3 of the members of the committee casting a vote.

An ISO/PAS or ISO/TS is reviewed every three years with a view to deciding whether it can be transformed into an International Standard.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights.

ISO/TS 8000-150 was prepared by Technical Committee ISO/TC 184, *Automation systems and integration*, Subcommittee SC 4, *Industrial data*.

ISO 8000 is organized as a series of parts, each published separately. The structure of ISO 8000 is described in ISO 8000-1.

Each part of ISO 8000 is a member of one of the following series: general data quality, master data quality, transactional data quality and product data quality. This part of ISO 8000 is a member of the master data quality series.

A complete list of parts of ISO 8000 is available from the Internet:

        http://www.tc184-sc4.org/titles/DATA_QUALITY_Titles.htm

Introduction

The ability to create, collect, store, maintain, transfer, process and present data to support business processes in a timely and cost effective manner requires both an understanding of the characteristics of the data that determine its quality, and an ability to measure, manage and report on data quality.

ISO 8000 defines characteristics that can be tested by any organization in the data supply chain to objectively determine conformance of the data to ISO 8000.

ISO 8000 provides a framework for improving data quality that can be used independently or in conjunction with quality management systems.

There is a limit to master data quality improvement with the data-centric approach where only the data found defective is corrected. When data errors and their related data are traced and corrected, or root causes of data errors are removed through processes for data quality management, recurrence of the same data errors can be prevented. Therefore, a framework for process-centric data quality management is required to improve data quality.

For this purpose, this part of ISO 8000 specifies fundamental principles of a master data quality management, and requirements for implementation, data exchange and provenance. This standard also contains an informative framework that identifies processes for data quality management. For reader's better understanding, the framework in detail, its functional model and a business scenario with examples are provided in Annexes B, C and D, respectively. This framework can be used in conjunction with or independently of quality management systems standards, for example, ISO 9001.

This part of ISO 8000 is intended for use by organizations that have multiple systems that share master data and/or that share and exchange data with other organizations and therefore need to manage the quality of their master data.

Although the framework has been developed based on the experience of data quality management applied in industries such as finance, telecommunication, and public institutions, it is expected that this framework, with appropriate extension, can also be applied to mechanical design or manufacturing data.

Data quality —

Part 150:
Master data: Quality management framework

1 Scope

This part of ISO 8000 provides fundamental principles of a process-centred approach to master data quality management and requirements that can be used by an organization to implement master data quality management. It also contains an informative framework that identifies processes for master data quality management. This part of ISO 8000 can be used in conjunction with or independently of quality management systems standards, for example, ISO 9001.

The following are within the scope of this part of ISO 8000:

— fundamental principles of master data quality management;

— requirements

— implementation requirements;

— data exchange requirements;

— provenance requirements;

— master data quality management framework

— top-level and lower level processes;

— roles.

The following are outside the scope of this part of ISO 8000:

— data quality evaluation and certification methods;

— taxonomy of data;

— data quality maturity model.

This part of ISO 8000 is intended for use by organizations that have multiple systems that share master data and/or that share and exchange data with other organizations and therefore need to manage the quality of their master data.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 8000-2, *Data quality — Part 2: Vocabulary*

ISO 8000-110, *Data quality — Part 110: Master data: Exchange of characteristic data: Syntax, semantic encoding, and conformance to data specification*

ISO/TS 8000-120, *Data quality — Part 120: Master data: Exchange of characteristic data: Provenance*

3 Terms, definitions and abbreviated terms

3.1 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 8000-2 apply.

3.2 Abbreviated terms

For the purposes of this document, the following abbreviated terms apply.

UNSPSC   United Nations standard products and services code

GTIN   global trade item number

SQL   structured query language

4 Fundamental principles of master data quality management

To manage master data quality successfully, organizations shall keep the following fundamental principles.

— Involvement of people: people at all levels who have roles for data quality management are involved to improve data quality of an organization. Although data processing of end users with lower-level role has the most direct effect on data quality, intervention or control of data administrators with middle-level role is required to implement and settle down processes for data quality improvement in the organization. In addition, involvement of managers who are in charge of organization-wide data quality with high-level role is inevitable to change and optimize roles, authority, and processes of the organization.

— Process approach: data-centric measurement and correction is not enough to improve data quality of the whole organization. Desired data quality is achieved more efficiently when activities and related resources for data quality are managed by processes.

— Continual improvement: data quality is improved continuously through the processes of data processing, data quality measurement and data error correction. However, with these processes only, identical data errors that occur repeatedly cannot be prevented. Recurrence of data errors can be prevented when the processes to analyze, trace and improve root causes which hinder data quality goes with these processes. For this, management processes concerned with data architecture/schema, data stewardship and data flow shall also be supported. In addition, organizations shall improve not only processes for data quality management but also business processes where data are directly operated.

— Master data exchange: all processes to manage master data quality comply with requirements that can be checked by computer for the exchange, between organizations and systems, of master data that consists of characteristic data.

The framework of master data quality management, the functional model of the framework and a business scenario with examples are provided as informative in Annexes C, D and E, respectively.

5 Requirements

5.1 Implementation requirements

An organization that implements this part of ISO 8000 shall perform the following actions:

— perform processes for data quality management that include at least data processing, data quality measurement and correction, data schema design, measurement criteria setup, error cause analysis, data quality planning and data architecture/stewardship/flow management;

— assign roles for data quality management within their organization;

  NOTE 1   Each role can be assigned to multiple persons, or multiple roles assigned to one person or position. The roles assigned can be one of many other roles assigned to a person or position.

— embed processes for data quality management within the organizations business processes.

  NOTE 2   The processes defined can be embedded at multiple places within the business processes of organizations, specifically anywhere master data is created and used.

5.2 Data exchange requirements

An organization shall:

— be capable of sending and receiving master data messages that conform to ISO 8000-110;

— specify a data dictionary to be used for semantic coding of master data messages sent to and from external organizations. The data dictionary shall meet the requirements of ISO 8000-110 for use in semantic coding;

— maintain a registry of data specifications that document the organization's data requirements for master data messages.

All master data messages that the organization sends to external organizations shall conform to ISO 8000-110.

The organization shall require that all master data messages sent to it conform to ISO 8000-110.

A data supplier claiming conformance to this part of ISO 8000 shall maintain a suitable electronic means for receiving queries for master data.

EXAMPLE An email address is published on the company's website or in a registry maintained by a third party.

5.3 Provenance requirements

This clause contains requirements that are optional for data exchange in addition to those in 5.2.

Any master data message that the organization sends to external organizations shall conform to ISO/TS 8000-120.

6 Conformance

An organization conforms to this part of ISO 8000 when it can present documentary evidence of the following:

— Roles for data quality management are assigned within their organization.

EXAMPLE 1   A job description is documentary evidence of a role assignment.

— Processes for data quality management are incorporated within the organizations business processes.

EXAMPLE 2   A business process model that includes the processes at appropriate places would be evidence of incorporation.

— The processes for data quality management are being executed.

EXAMPLE 3   Specifications of master data quality requirements, results of master data quality measurements, a log of defects and non-conformance, and a log of root cause analysis and corrective actions are evidence of the business processes being executed.

This part of ISO 8000 also provides for a number of options that may be supported by an implementation. These options have been grouped into the following conformance classes:

— free decoding;

— fee-based decoding;

— free decoding with provenance;

— fee-based decoding with provenance.

Conformance to the free decoding conformance class requires:

— the data dictionary, data specifications, and any incoming or outgoing master data messages conform to the free decoding conformance class of ISO 8000-110;

— all requirements of 5.2 are met.

In addition to the above, conformance to the free decoding with provenance conformance class requires that all requirements of 5.3 are met.

Conformance to the fee-based decoding conformance class requires:

— the data dictionary, data specifications, and any incoming or outgoing master data messages conform to the fee-based decoding conformance class of ISO 8000-110;

— all requirements of 5.2 are met.

In addition to the above, conformance to the fee-based decoding with provenance conformance class requires that all requirements of 5.3 are met.

Any claim of conformance to this part of ISO 8000 that does not explicitly state the conformance class shall be a claim of conformance to one of the free decoding conformance classes.

Annex A
(normative)

Document identification

To provide for unambiguous identification of an information object in an open system, the object identifier

```
{ iso standard 8000 part(150) version(1) }
```

is assigned to this part of ISO 8000. The meaning of this value is defined in ISO/IEC 8824-1, and is described in ISO 10303-1.

Annex B
(informative)

Master data quality management framework

B.1 Overview of the master data quality management framework

The structure of the framework is graphically represented in a 3x3 matrix as shown in Figure B.1. The framework consists of three top-level processes: data operations, data quality monitoring, and data quality improvement. Each top-level process is segmented into three processes by the role of the person performing the process. The processes are related to one another according to the order of processes and input/output of data.

Processes grouped under the role of data manager support those under the role of a data administrator. The process of data architecture management presents the guidelines necessary to design data structures. The process of data stewardship and flow management provides the necessary information to analyze error causes and the process of data quality plan offers objectives or guidelines for data quality that assist in setting up criteria.

Processes grouped under the role of data administrator are comprised of the followings: data design, data quality criteria setup, and data error cause analysis. These three processes control and coordinate data so as to support processes under the role of data technician. The process data design helps to ensure data operation quality by improving data schema. The process, data quality criteria setup provides with criteria and methodologies to assess data quality. The process of data error cause analysis prevents recurrence of the same data errors by analyzing root causes of the data errors.

Processes grouped under the role of data technician are split into the followings: data processing, data quality measurement, and data error correction. The three processes are performed successively: first, the process of data processing that creates, reads, modifies, transfers, and deletes data is implemented in accordance with data guidelines. In order to find unnoted data errors during the process, the process of data quality measurement is executed on a real time basis or periodically. In case that data errors are found in the process, the process of data error correction enters the execution mode.

Thus far, processes were explained in the order of data manager, data administrator and data technician. However, in certain cases results of the processes are fed back in reverse order.

Generally the data quality policy is consistent with the overall information technology (IT) policy of the organization and provides a basis for the setting of data quality objectives. Therefore, the data quality policy is considered one of control factors which affect the performance of the whole framework.
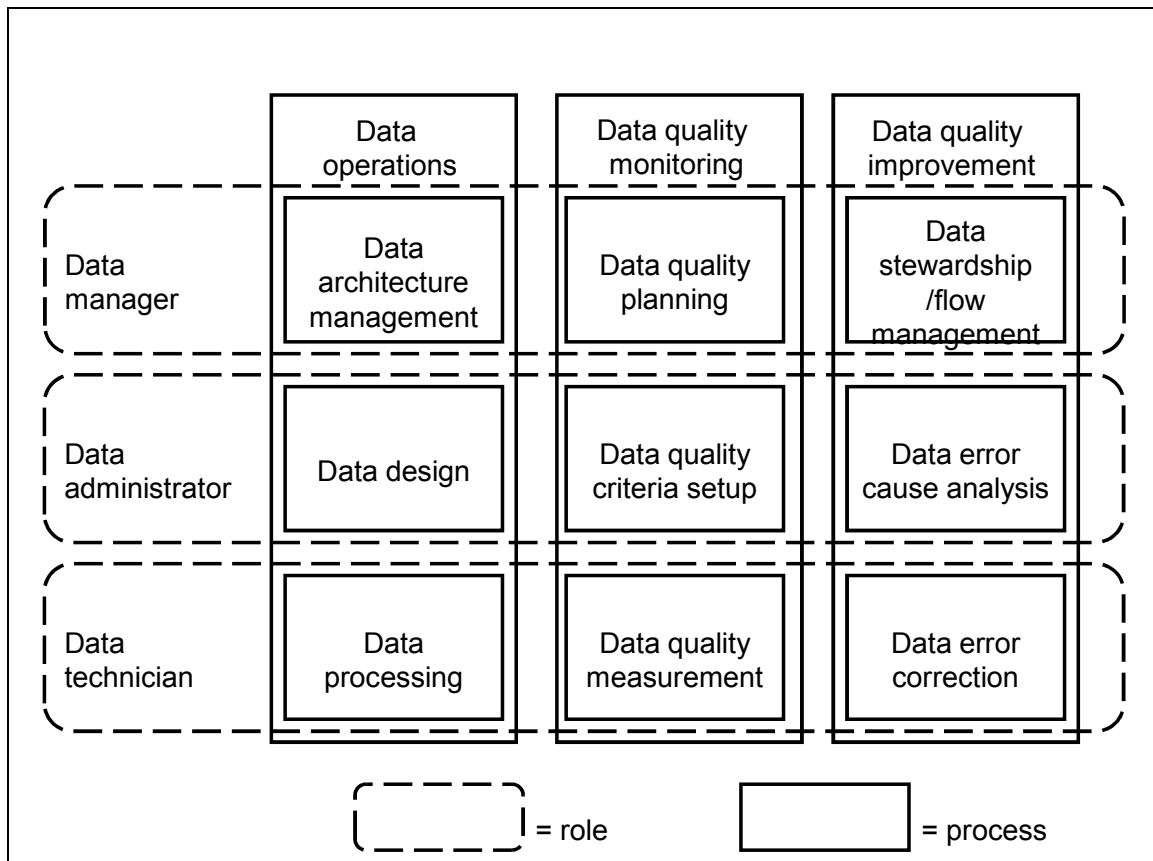
Figure B.1 — Master data quality management framework

B.2 The top-level processes of the framework

B.2.1 The three top-level processes

The three top-level processes in the framework shall be:

— data operations (see B.2.2);

— data quality monitoring (see B.2.3);

— data quality improvement (see B.2.4).

B.2.2 Data operations

The data operations process identifies factors that affect data quality and ensures data is available at the right place in a timely manner. This top-level process shall consist of the following processes:

— data architecture management; the process that manages organization-wide data architecture from the integrated perspective to use data in distributed information systems with consistency and therefore ensure data quality(see B.4.2).

— data design; the process that designs data schema, and implements a database to make data users apply data without mistake and ensure data quality(see B.4.3).

— data processing; the process that creates, searches, updates, deletes data in accordance with guidelines of data operations (see B.4.4).

7

B.2.3 Data quality monitoring

The data quality monitoring process identifies data errors through a systematic approach. This top-level process shall consist of the following processes:

— data quality planning; the process that sets up objectives of data quality in alignment with the strategies of an organization, identifies factors to be managed, and performs actions in order to accomplish objectives. This process also includes assurance of data quality and adjustment of objectives on the back of assurance results (see B.4.5).

— data quality criteria setup; the process that sets criteria that include characteristics of data, target data, and methods to measure(see B.4.6).

— data quality measurement; the process that measures target data with the criteria set in the process of data quality criteria setup on a real time basis or periodically (see B.4.7).

B.2.4 Data quality improvement

The data quality improvement process corrects data errors detected and eliminates root causes of the data errors by tracing and identifying them. In order to support the top-level process effectively, adjustment of data stewardship in accordance with data flows tracing is required. This process has the function of process improvement not only data quality improvement. Processes for data management are improved at the data administrator level while business processes at the data manager level. This top-level process shall consist of the following processes:

— data stewardship and flow management; the process that analyses data operations and data flows among businesses or organizations, identifies responsible parties and their data operation systems which influence data quality, and manages the stewardship of data operations(see B.4.8).

— data error cause analysis; the process that analyses root causes of data errors and prevents a recurrence of the same errors fundamentally (see B.4.9).

— data error correction; the process that corrects the data that turns out erroneous (see B.4.10).

B.3 The roles in the framework

B.3.1 The three roles

The three roles in the framework are responsible for performing the processes in the framework. These roles shall be:

— data manager (see B.3.2);

— data administrator (see B.3.3);

— data technician (see B.3.4).

B.3.2 Data manager

The data manager shall perform the following processes within the framework:

— data architecture management (see B.4.2);

— data quality planning (see B.4.5);

— data stewardship and flow management (see B.4.8).

The data manager performs the role that directs a guideline for master data quality management in compliance with objectives of an organization, manages factors that impact data quality at an organization level, and establishes the plans for performing data quality activities in the organization. Along with each major top-level process, the data manager maintains data consistency in individual information systems through the organization-wide data architecture management, and analyzes factors that affect data quality in data quality planning. In addition, the data manager takes a role of granting data administrators authority to trace and correct data over the information systems or organization.

EXAMPLE   A data manager selects one out of multiple PIC (Product Identification Code) systems used in industry. While a data administrator or a data technician operates the PIC system, a data manager oversees and manages whether they use the PIC in alignment with objectives of data quality management. A position entitled enterprise data architect (EDA) can be an example of this type of role.

B.3.3   Data administrator

The data administrator shall perform the following processes within the framework:

— data design (see B.4.3);

— data quality criteria setup (see B.4.6);

— data error cause analysis (see B.4.9).

The data administrator controls and coordinates over data technicians by defining criteria required to maintain the quality of master data, and prevents a recurrence of the same data errors by analyzing the causes of errors or designing data schema. In general, supporting resources and guidelines to data technicians, the data administrator carries the data quality plan into practice to achieve the objectives set by the data manager.

EXAMPLE   A data administrator designs data schema for data technicians to ensure data quality (for example, data consistency, standards, etc.). Additionally, the data administrator is dedicated to setting criteria to manage the data quality in alignment with organizations' objectives, and controlling data quality so that the measured data reach the level of the objectives. Furthermore, the data administrator takes an important role that searches for the root causes of data errors and removes them. In the example of PIC(Product Identification Code), if the PICs are dispersed in many individual systems, the data administrator has to identify where the representative master data is located, and re-design data schema to prevent the master data of PIC from redundancy. The program steering committee or a database administrator can take this role.

B.3.4   Data technician

The data technician shall perform the following sub-processes within the framework:

— data processing (see B.4.4);

— data quality measurement (see B.4.7);

— data error correction (see B.4.10).

The data technician creates, reads, modifies, and deletes data as per guidelines of data quality management set by the data administrator, and measures data quality and corrects erroneous data as a result of the measurement. While the data manager or administrator can handle data even across its

own business scope in accordance with data flows, the data technician handles data within its business scope.

EXAMPLE   While data users actually input and use data, data operators periodically measure whether data in use complies with business rules and correct data when errors are found. (Data technician may be called data user or data operator depending upon its detailed role.) In the example of PIC (Product Identification Code), data users apply PICs to their business directly, and data operators examine PICs with check digit checking rules periodically by an inspection tool or manually.

B.4 The lower level processes in the framework

B.4.1   The nine lower level processes

The framework shall include nine lower level processes:

— data architecture management (see B.4.2);

— data design (see B.4.3);

— data processing (see B.4.4);

— data quality planning (see B.4.5);

— data quality criteria setup (see B.4.6);

— data quality measurement (see B.4.7);

— data stewardship and flow management (see B.4.8);

— data error cause analysis (see B.4.9);

— data error correction (see B.4.10).

In this section, each process in the framework is specified in terms of its necessity, activities required at the least, responsibilities required to improve data quality from the role perspective, and relationships among processes. Processes are described in order of top-level processes, data operations, data quality monitoring and data quality improvement.

The functional model for the framework and a business scenario with examples are provided in Annexes C and D, respectively.

B.4.2   Data architecture management

B.4.2.1 Overview of data architecture management

As data is distributed in the organization, data quality cannot be ensured without systematic management. The process identifies the data commonly used throughout the organization and defines data schema that secures data quality inside and outside the organization. In addition, identifying in what type of schema the data is distributed throughout the whole information systems, the process can prevent discrepancy among the same data distributed in different information systems. The process can also manage the lifecycle of the data identified in the data architecture.

B.4.2.2 Constituent activities of data architecture management

Data architecture management shall consist of the following activities:

— Management of organization-wide conceptual data models: the activity that manages conceptual data models represented by factors which need to be shared or managed at an organization level. Data mapping or tracing can be done through the models.

— Management of organization-wide data standards: the activity that manages standards and business rules that should be observed in the organization-wide data architecture management.

B.4.2.3 Required responsibilities of data manger for data architecture management

From the role perspective, a data manager in this process shall have the following responsibilities.

— Organization-wide coordination: the responsibility that seeks convergence on the goals and plans for data quality management amongst responsible parties and assures conformance by responsible parties. Since data quality heavily depends on data users in business units, this role should have authority to control and coordinate responsible parties not only in technical units but also in business units.

— Organization-wide sharing and maintenance: the responsibility that shares conceptual data models and data standards to secure data quality, and maintain them consistently, performing activities such as modification of data mapping whenever a data schema changes.

B.4.2.4 Relationship of data architecture management to other processes

Relationships to other processes:

— Between data architecture management and data quality planning: Based on organization-wide data architecture, which shows the entire data composition of an organization, a plan for data quality can be developed. Critical findings from the plan should be reflected to conceptual data models and data standards.

— Between data architecture management and data stewardship/flow: Data stewardship is assigned based on the organization-wide data architecture. In practice, when data stewardship and flow changes in the process of data error resolution, the change should be reflected to the data architecture management.

— Between data architecture management and data design: The data architecture management provides the process of data design with information such as conceptual data models and data standards.

B.4.3   Data design

B.4.3.1 Overview of data design

Data quality errors are separated into two categories, errors by user and errors by structure. Whereas errors by user have limitations to be resolved systematically, errors by structure that mean data errors caused by wrong data schema, can be resolved by data schema redesign. On the other hand, when data is in service, correcting errors by structure is not easy. Hence, data quality should be considered from the initial stage of data design. Especially, if data is designed for a specific application system only, the quality of the data shared through the organization cannot be maintained. Therefore, data relationship with other application systems should be considered at the organization level when data is designed.

B.4.3.2 Constituent activities of data design

Data design shall consist of the following activities:

— Data design in consideration of quality: the activity that identifies data schema such as required types of data and ranges in data values. In addition, this activity should reflect requirements of data technicians sufficiently. This includes the specification of the configuration and use of data structures in software packages where the physical data structure is already defined.

— Data design in connection with the organization-wide data architecture management: the activity that implements data schema design and change to maintain the relationship with the organization-wide data architecture management, when a database is built or the need for data schema changes arise.

B.4.3.3 Required responsibilities of the data administrator for data design

From the role perspective, a data administrator in the process shall have the following responsibilities.

— An internal responsibility that reflects data quality requirements to data design through in-depth consultation with users and responsible parties of application systems.

— An external responsibility that consults with responsible parties about the relationship of data with other application systems.

B.4.3.4 Relationship of data design to other processes

Relationships to other processes:

— Between data design and data architecture management: Results of the process of data design feed back to the process of data architecture management.

— Between data design and data processing: Data processing is performed based on data schema created in the process of data design.

— Between data design and data quality criteria setup: Data quality criteria are defined based on the data schema created in the process of data design.

B.4.4   Data processing

B.4.4.1 Overview of data processing

Carelessness and lack of understanding of end users, who are the primary data technicians, impact on data quality directly, causing data errors. It is necessary to provide end users with processes and guidelines for data processing from a data quality management standpoint. In addition, when erroneous data values are input, the responsible party is identified. Data transactions not only by a person but also by application systems are managed to be able to trace the details.

B.4.4.2 Constituent activities of data processing

Data processing shall consist of the following activities:

— Data implementation: the activity that creates, reads, updates, transfers, and deletes data in accordance with data processing guidelines. End users should be educated through notification tools or training programs for better use of data guidelines. Even when data is implemented by

application systems, the data shall comply with guidelines for data processing, and especially, the implemented data should be clearly recognized.

— Data logging: the activity that records and stores data users, lapsed times, a history of data use to trace every data manipulation.

B.4.4.3 Required responsibilities of data technician for data processing

It is not easy to manage data quality systematically since the process is performed by a number of users in different business units and roles. Therefore, data processing and corresponding quality management shall be specialized by individual data technicians. In addition, for data whose quality is critical, data technicians shall confirm its quality.

B.4.4.4 Relationships of data processing to other processes

Relationships to other processes:

— Between data processing and data design: When an additional data quality issue occurs in the midst of the process of data processing, it is reflected to the process of data design.

— Between data processing and data quality measurement: Results of data processing are measured in the process of data quality measurement.

B.4.5 Data quality planning

B.4.5.1 Overview of data quality planning

If there are various requirements for data or differences in perception of data quality inside and outside an organization, it is necessary to set up objectives to be consistent with the data quality policy of an organization. To achieve the objectives, formulating the detailed data quality plan and quality assurance is also necessary.

B.4.5.2 Constituent activities of data quality planning

Data quality planning shall consist of the following activities:

— Objectives setup and management: the activity that collects data requirements of responsible parties including various customers or different perceptions of data quality inside and outside an organization, establishes unified objectives, and manages the assurance process that measures the achievement of those objectives.

— Identification of quality management items and action planning: the activity that identifies quality management items required to achieve the objectives, and establishes an action plan to improve them. A detailed action plan may include setting a scope, tasks, scheduling, securing resources, and formulating methods.

B.4.5.3 Required responsibilities of data manager for data quality planning

From the role perspective, a data manager in the process shall have the following responsibilities:

— Control of quality management factors and resources: the responsibility to control quality management factors and corresponding resources when data quality plan is carried out.

— Assurance of top executives' support: the responsibility to gain a driving force through dynamic communications with top executives. The data manager should be able to communicate to top

executives the objectives of data quality and the activities in progress, and establish governance so that the entire organization implements the plan consistently.

B.4.5.4 Relationship of data quality plan to other processes

Relationship to other processes:

— Between data quality planning and data architecture management: Results of the process of data quality planning that include a scope, tasks, scheduling, securing resources, and formulating methods, can be reflected to the process of data architecture management.

— Between data quality planning and data stewardship/flow: Results of the process of data quality planning can be reflected to the process of data stewardship and flow.

— Between data quality planning and data quality criteria setup: Data quality planning provides the process of data quality criteria setup with a detailed data quality plan.

B.4.6   Data quality criteria setup

B.4.6.1 Overview of data quality criteria setup

For the implementation of a data quality plan, it is necessary to provide the criteria which include characteristics of data (such as accuracy, provenance and completeness), target data, and methods to measure. In general, the criteria are determined by reflecting stakeholder's requirements. When there are various criteria adopted in the past by multiple users, the criteria should be set up in a standardized pattern. Through the criteria, the organization can raise data quality issues officially, allocate necessary resources, and implement the process of data quality measurement in detail. In this process, ISO 8000 master data series of standards can be used as part of criteria for data quality measurement.

B.4.6.2 Constituent activities of data quality criteria setup

Data quality criteria setup shall consist of the following activities:

— Data quality criteria determination: the activity that defines practical criteria that arouse stakeholders' empathy through the collection of their requirements for data quality. This activity includes determining characteristics of data, target data, and methods to measure in detail.

— Data quality evaluation: the activity that examines whether the target data measured satisfy given quality standards of the criteria. If necessary, results of the examination are fed back to the criteria.

B.4.6.3 Required responsibilities of data administrator for data quality criteria setup

From the role perspective, a data administrator in the process shall have the responsibility that reflects data users' requirements, and forms a consensus through full communications with data stakeholders.

B.4.6.4 Relationship of data quality criteria setup to other processes

Relationships to other processes:

— Between data quality criteria setup and data quality planning: The process of data quality criteria setup is executed as a subsequent process of data quality planning.

— Between data quality criteria setup and data quality measurement: The process of data quality measurement is implemented based on the criteria set by the process of data quality criteria setup.

B.4.7 Data quality measurement

B.4.7.1 Overview of data quality measurement

Due to difficulties in the timely finding of errors by end users, application systems, and implementation processes, a systematic quality measurement is necessary. In addition, setting up a time interval to measure data quality is necessary because certain data lose their importance as time goes by. Although it is desirable to measure data quality without delay after the process of data processing, the measurement time can be adjusted in accordance with characteristics of business tasks.

B.4.7.2 Constituent activities of data quality measurement

Data quality measurement shall consist of the following activities:

— Data quality measurement: the activity that measures target data in accordance with criteria by tools or manually. For repeated data, measurement can be done by tools. Yet for complicated data, measurement can be done by an expert's judgment.

— Statistical treatment of measured data: the statistical analysis of data quality measurements to support the analysis of the causes of defects and non-conformances.

B.4.7.3 Required responsibilities of data technician for data quality measurement

From the role perspective, a data technician in this process shall have the authority to access target data.

B.4.7.4 Relationship of data quality measurement to other processes

Relationships to other processes:

— Between data quality measurement and data quality criteria setup: According to the results of data quality measurement, the criteria set by the process of data quality criteria setup can be adjusted.

— Between data quality measurement and data error cause analysis: Erroneous data and their statistics found in the data quality measurement are delivered to the process of data error cause analysis.

— Between data quality measurement and data error correction: The erroneous data found in the data quality measurement are transferred to the process of data error correction.

B.4.8 Data stewardship and flow management

B.4.8.1 Overview of data stewardship and flow management

In general, data does not exist alone but is related to other applications. Hence, it is very important in data quality management to understand how data is applied. When relationships between data and data users are identified, data users' requirements for data quality become clearer and data errors can be corrected more responsibly. In addition, if data flows among applications that actually control data are identified, errors can be corrected perfectly by analyzing the impact of data errors and tracing their related data. If multiple parties use the data, a person should be assigned the responsibility of the data

quality. In this process, a data manager may be involved in business process improvement for root causes resolutions.

B.4.8.2 Constituent activities of data stewardship and flow management

Data stewardship and flow management shall consist of the following activities:

— Stewardship assignment: the activity that assigns a person the authority and responsibility to manage distributed data. When a certain data value in one application system is changed, this person should change the same types of data in all of other application systems. When the change occurs frequently, he/she may use applications specially developed to revise data automatically through the information of data flows.

— Data flow management: the activity that oversees relationships among data distributed in individual application systems so that identical types of data keep identical attributes and values. When a change or an error of data occurs in one application system, this activity identifies how the error impacts identical types of data in other application systems.

B.4.8.3   Required responsibilities of data manager for data stewardship and flow management

From the role perspective, a data manager in this process shall have the following responsibilities:

— Authorization of data processing: the responsibility that can control multiple business units that use the same data, or authorize data processing to them.

— Coordination among business units: the responsibility that controls and coordinates business units when new systems are launched or data flows are changed.

B.4.8.4 Relationship of data stewardship and flow management to other processes

Relationships to other processes:

— Between data stewardship/flow management and data architecture management: Data architecture management has to be preceded before the process of data stewardship and flow;

— Between data stewardship/flow management and data quality planning: The process of data stewardship and flow resolves conflicts over data stewardship and flows among responsible parties in accordance with the data quality plan;

— Between data stewardship/flow management and data error cause analysis: Information concerned with data stewardship and flow is used in the process of data error cause analysis.

B.4.9   Data error cause analysis

B.4.9.1 Overview of data error cause analysis

If only the data found as an error is corrected, the same error can happen again at any time. Therefore, it is necessary to analyze root causes of data errors and take actions for recurrence prevention. As there are various root causes, some may be resolved in a short term, while the others phase by phase in a long term.

B.4.9.2 Constituent activities of data error cause analysis

Data error cause analysis shall consist of the following activities:

— Cause analysis and correction: the activity that effectively eliminates data errors by finding root causes out of data schema, standards, and flows of data and by taking actions against recurrence of the same errors. Archiving data errors and their resolutions of the past can improve the performance of data error correction.

— Error recurrence prevention: the activity that prevents the recurrence of the same type of data errors from other application systems. For this activity, possible causes of errors (data itself, organizational structure, data flows, and business processes) are identified on the whole. Improving or controlling business processes and/or data management processes involves data manager's support and consultation with responsible parties. In addition, for more effective prevention, processes that describe how to correct or eliminate data errors are prepared, and responsible parties trained in their execution.

B.4.9.3 Required responsibilities of data administrator for data error cause analysis

From the role perspective, a data administrator in this process shall have the following responsibilities:

— Root cause analysis: the responsibility that can analyse causes of data errors at an organization-wide level. For the cause analysis, the data administrator should be qualified to trace data, related application systems and related users.

— Root cause elimination: the responsibility that take actions to remove root causes identified.

B.4.9.4 Relationship of data error cause analysis to other processes

Relationships to other processes:

— Between data error cause analysis and data stewardship/flow: The management of data stewardship and flow shall be secured for the process of data error cause analysis. This is because, according to the result of cause analysis, adjustment of data stewardship or data flow may be requested.

— Between data error cause analysis and data error correction: According to the result of data error cause analysis, the process of data error correction can be done to correct related data.

— Between data error cause analysis and data quality criteria setup: Data guidelines produced by the process of data error cause analysis can be fed back to the process of data quality criteria setup.

— Between data error cause analysis and data design: According to the results of the process of data error cause analysis, data schema may be changed in the process of data design.

B.4.10 Data error correction

B.4.10.1 Overview of data error correction

According to the results of two processes, data quality measurement and data error cause analysis, this process corrects data errors. In this process, not only data found in one application system but the same or related data in other application systems should be corrected. In addition, it is necessary to keep data consistency by identifying data related each other.

B.4.10.2 Constituent activities of data error correction

Data error correction shall consist of the following activities:

— Data error correction: the activity that corrects errors found through the processes of data quality measurement and data error cause analysis, and all related data that exist in other application systems. The history of correction should be accumulated for use in the process of data error cause analysis.

— Correction data sharing: the activity that notifies all responsible parties of the data before and after the correction in order to eliminate any confusion in the data operations.

B.4.10.3 Required responsibilities of data technician for data error correction

From the role perspective, a data technician in this process shall have the following responsibilities:

— Consultation with authorized persons: the responsibility that consults with the person who has authority to modify the data.

— Data correction sharing: the responsibility that should manage progressing details of data error correction, and provide an environment where data stakeholders can share details of data errors and correction.

B.4.10.4 Relationship of data error correction to other processes

Relationships to other processes:

— Between data error correction and data error cause analysis: Accumulated results of the process of data error correction can be used to analyze error causes in the process of data error cause analysis.

— Between data error correction and data processing: Refined data in the process of data error correction is used as input data to the process of data processing.

Annex C
(informative)

The functional model of the framework

This is the functional model to represent relationships and information flows among processes of the master data quality management framework in the format of IDEF0[6]. As in the framework shown in figure B.1, the functional model consists of three top-level processes and corresponding lower level processes. The indent node tree of the functional model is as follow:

A0: perform data quality management;
    A1: perform data operations;
        A11: perform data architecture management;
        A12: perform data design;
        A13: perform data processing;
    A2: perform data quality monitoring;
        A21: perform data quality planning;
        A22: perform data quality criteria setup;
        A23: perform data quality measurement;
    A3: perform data quality improvement;
        A31: perform data stewardship and flow management;
        A32: perform data error cause analysis;
        A33: perform data error correction.

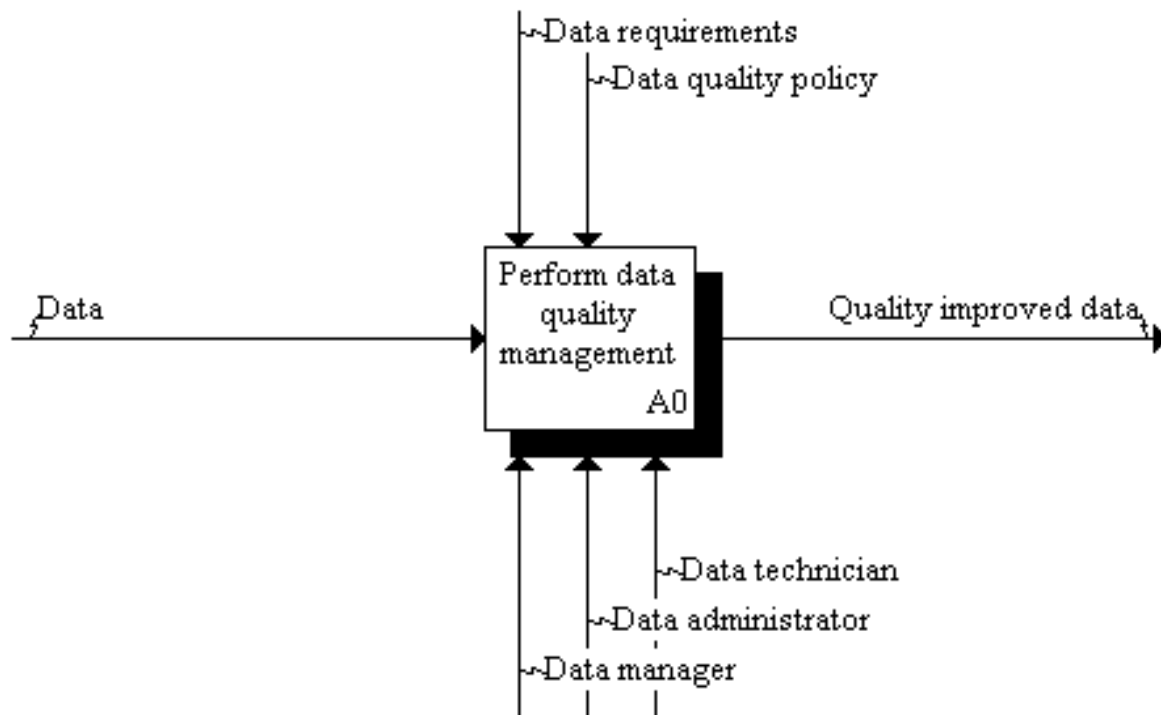Diagrams of the framework are represented in Figures C.1 to C.5 in the sequence of nodes.

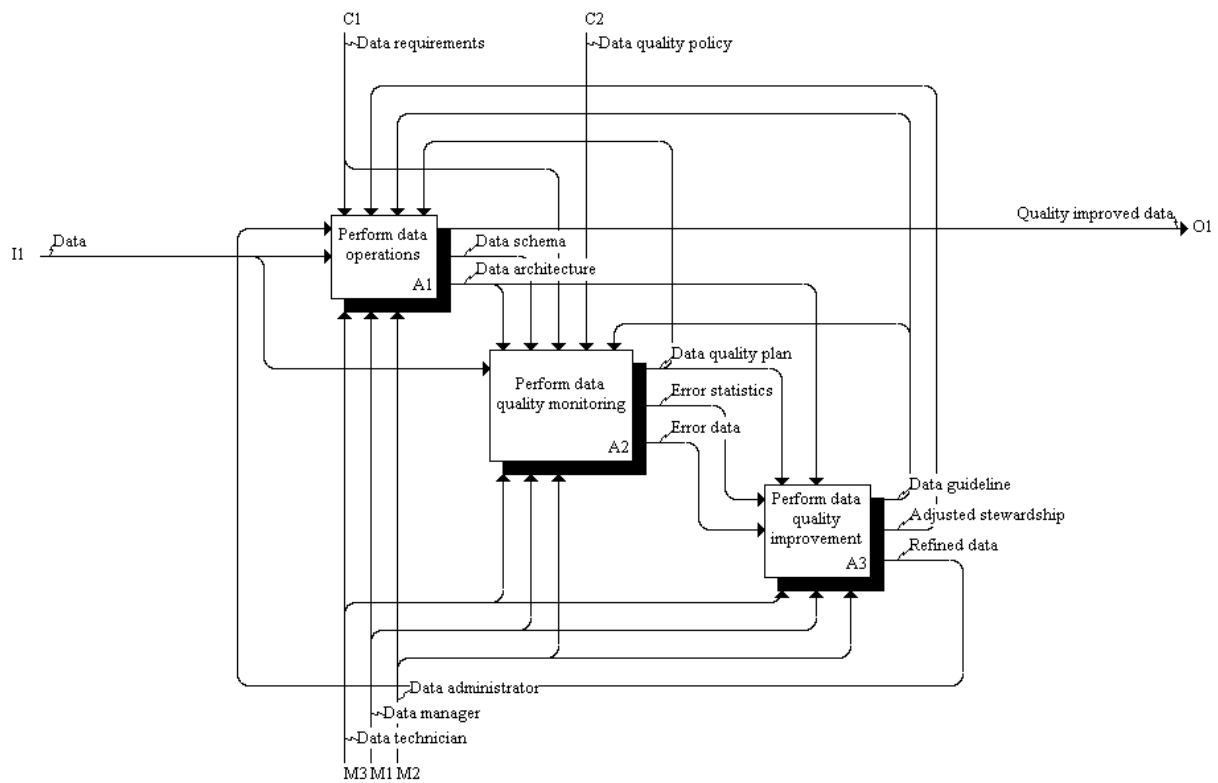Figure C.1 — The functional model of the framework at the highest level

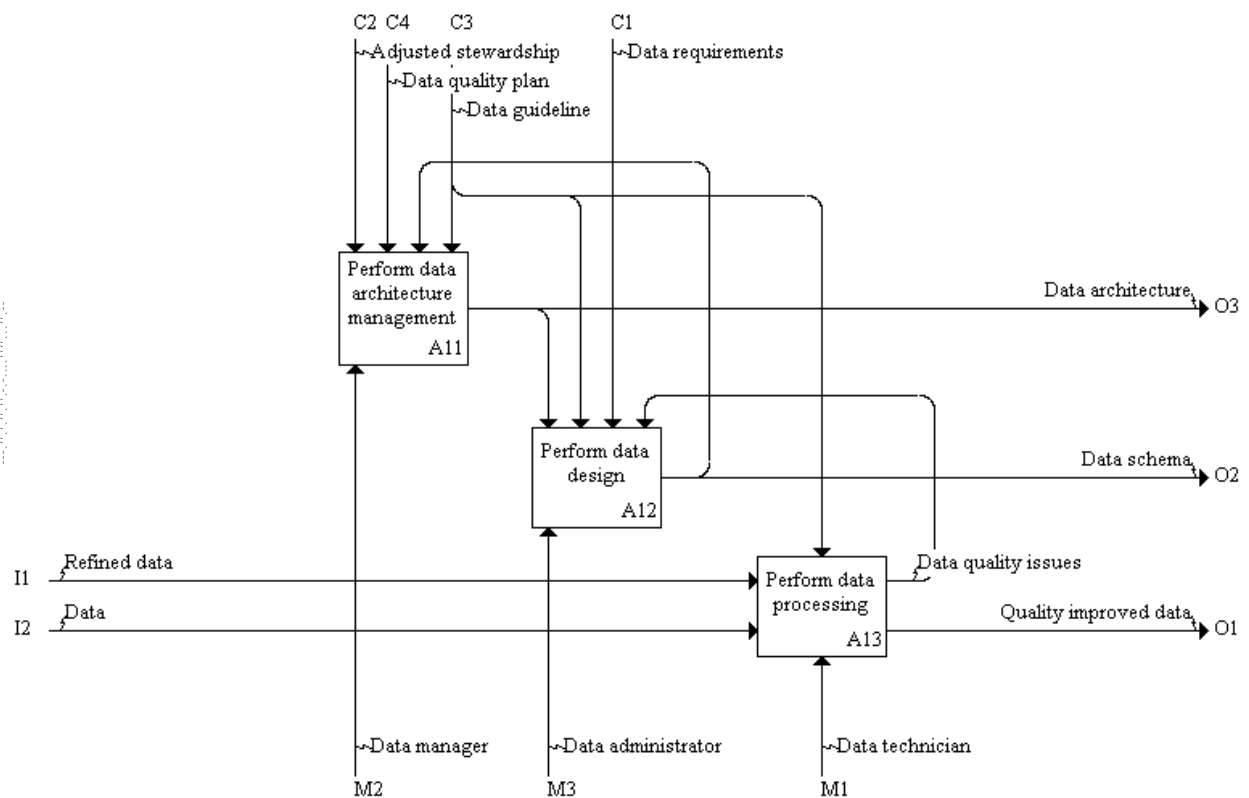Figure C.2 — Top-level processes at the second level

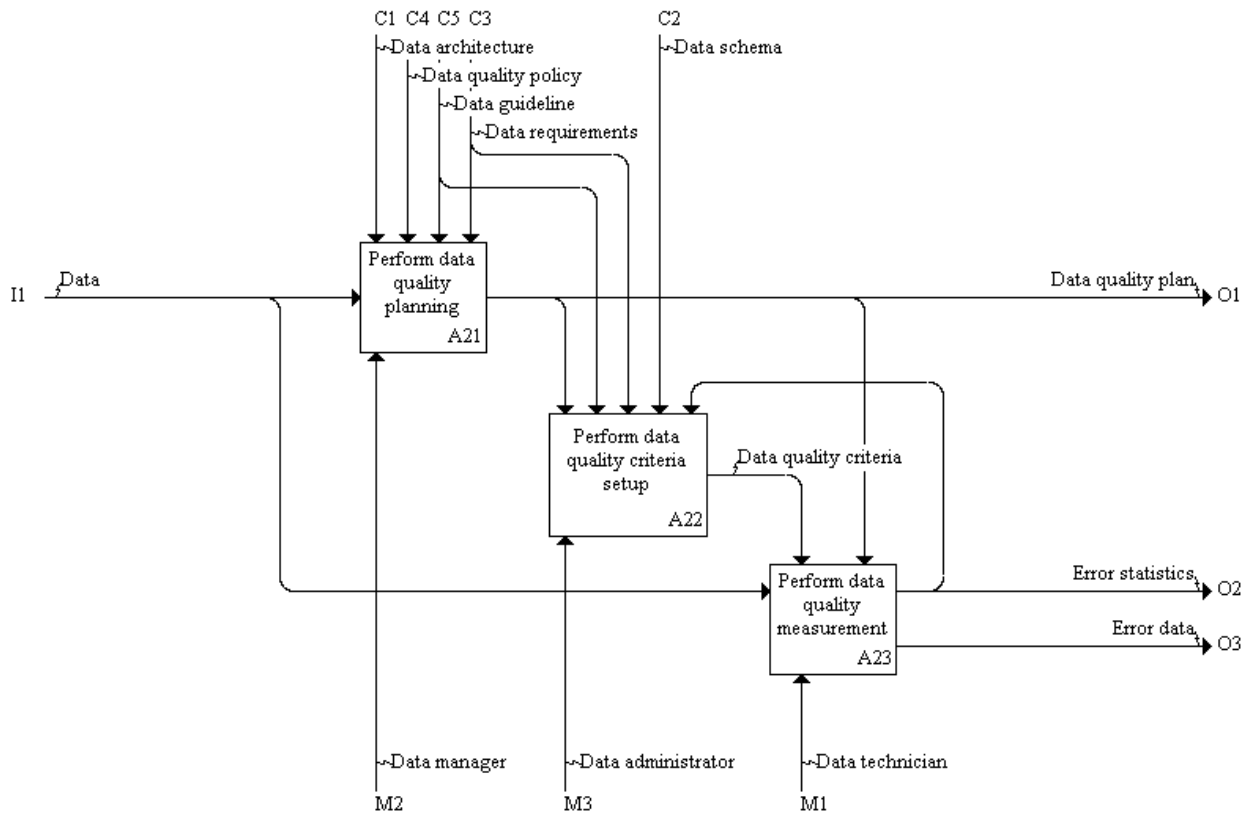Figure C.3 — Top-level process data operations at the third level

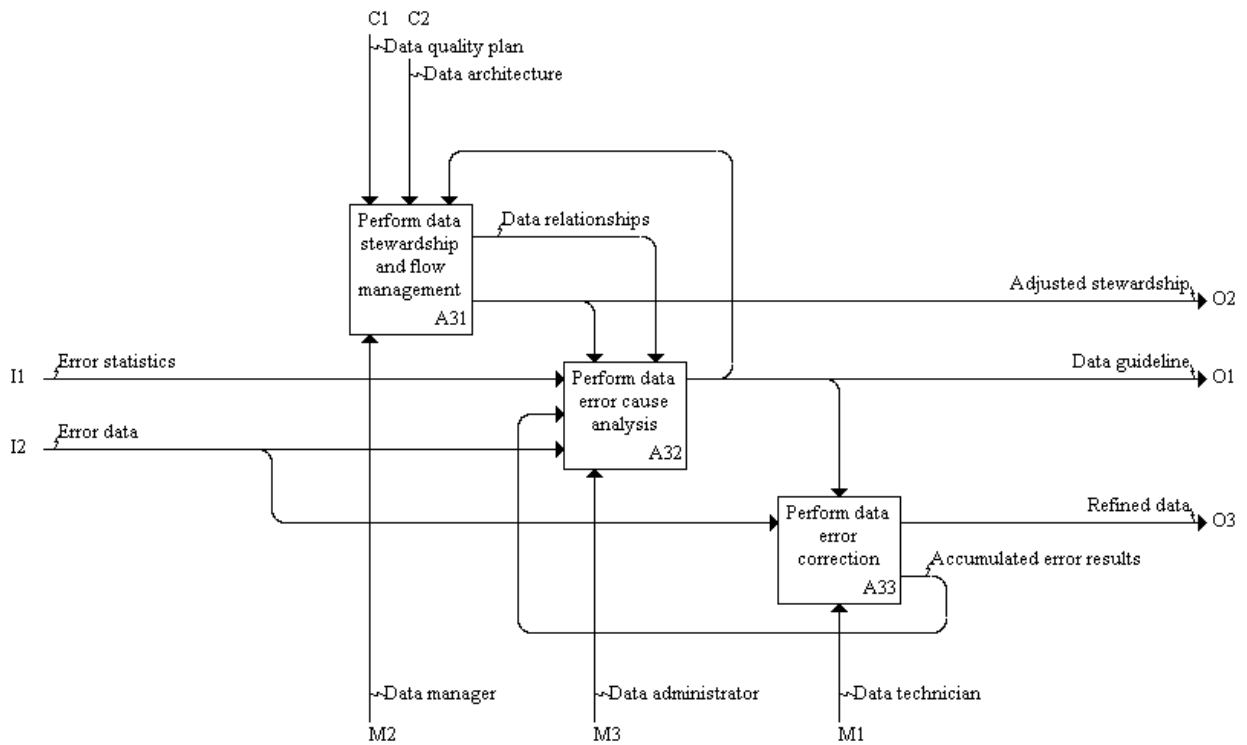Figure C.4 — Top-level process data quality monitoring at the third level

Figure C.5 —Top-level process data quality improvement at the third level

Annex D
(informative)

Business scenario with examples for the framework

This annex contains a business scenario with examples to help readers understand better the concept of the master data quality management framework. In this scenario, cases for product identification codes or product classification codes are provided as examples. The business scenario and its examples are given in the sequence of processes described in Annex B.

D.1 Business environment and situation

With the expansion of information usage in business, the number of information systems is gradually increasing within the organization, and data is accordingly distributed over these information systems. In this distributed environment, some of these data are used commonly throughout internal information systems, while others are shared with external information systems. Accordingly, data quality issues could not only create confusion internally because of data discrepancy, but also make a serious impact on external relationship. For these reasons, people responsible for data quality should pay careful attention to distributed data to keep data consistency.

D.2 Data architecture management

A data manager manages the organization-wide data architecture to identify target data for quality control distributed in information systems and its data schema in an easily accessible manner. Organization-wide data architecture management is used not only in identifying locations and relationships of existing data, but also in ensuring that new data designed keeps consistency with existing data schema.

EXAMPLE   When Factory A uses custom-developed product identification codes and Factory B uses GTIN identification codes, these two different codes systems need to be consolidated for management at the organization level. For consolidated management, a conceptual data model that maps custom-developed product identification codes and GTIN identification codes should be developed.

D.3 Data design

Data errors arising from a structural problem tend to be repeated. For the resolution of the problem, data schema should be re-designed at an organization level.

EXAMPLE   If there is no 'check digit' feature in a product identification code system used in Factory A, the structure and corresponding DB schema should be redesigned by adding a 'check digit' to prevent data entry errors.

D.4 Data processing

Even if data schema is well designed, when users enter data without care or all of related data errors are not changed, the data values that should be the same end up with different ones. To prevent such errors, data users should observe the guidelines or rules of data processing.

EXAMPLE   When entering GTIN codes, information systems or users should keep the rule of examining the check digit.

D.5 Data quality plan

It is impossible to check data quality by applying all standards to all data as targets or standards for data quality management are relative - not absolute. In other words, according to the policy of an

organization, priority, standard, or quality level may be different by data. Therefore, data manager should develop a data quality management plan aligned with the policy.

EXAMPLE   Factory A which uses eCl@ss codes for product classification supplies products to Factory B which uses UNSPSC codes. Disparate codes between the two companies resulted in wrong product supply as well as data errors. Thus, Factory B requested Factory A to supply products based on UNSPSC codes. Factory A decided to use two classification codes after all and established a plan to adopt a consolidated code management system which supports mapping between two codes and prevent any entry error in advance.

D.6 Data quality criteria setup

If criteria such as characteristics of data, target data, methods to measure are not pre-defined, the process of data quality measurement cannot be carried out in a systematic way. Data administrators in the organization should set up criteria to be used systematically and officially through consultation with responsible parties.

EXAMPLE   For the measurement of the quality of product identification codes, domain integrity and reference integrity can be selected as characteristics of data. All data of information systems used in Factory A can be target data to measure, and a commercially available profiling tool or a custom-developed SQL could be used as a method to measure.

D.7 Data quality measurement

Data errors are typically discovered by chance in the process of data processing, and the data errors tend to be corrected within user's capability or business scope. Therefore, if data error correction is carried out depending on users only, the number of unidentified data errors will gradually increase. For this reason, it is necessary to measure and inspect data errors continuously and systematically. This process can be performed with SQL programs or data quality profiling tools by data operators.

EXAMPLE   In Factory B, a data operator measures and inspects whether errors exist in any GTIN code entered in the past or in any attribute where GTIN codes are used. At this time, a data operator uses a self-developed SQL program to find errors in databases.

D.8 Data stewardship and flow

As there are a variety of causes for data errors, data flows should be traced from data creation to final use to reveal exact causes. Stewardship and responsibility in data use should also be clearly understood and recognized. A data manager should analyze data stewardship and flow to identify the person or process related to data errors. In addition, he/she should prepare resolutions to root causes through consultation with other stakeholders.

EXAMPLE   Factory A uses UNSPSC version 10 for product classification while factory B uses the latest version 11. Factory A therefore should update to the latest available version. However, factory A cannot simply opt for update since it shares UNSPSC version 10 with suppliers and outsourcing partners. Responsible parties of the code system (Factory A, Factory B and partner companies) should consult with each other to determine the party in charge, scope and time concerned with revision.

D.9 Data error cause analysis

Data error correction is carried out for visible data. However, unless root causes of data error is identified and eliminated, similar types of error will occur repeatedly as time goes by. Recurrence of similar errors can be prevented when a data administrator traces and eliminates root causes.

EXAMPLE   Different classification codes for product X have been repeatedly discovered at Factory A and B. Through the cause analysis, it was found that Factory A was using UNSPSC version 10 while Factory B was using version 11. To root out the problem, Factory A's version should be updated to version 11.

D.10 Data error correction

Data technicians should correct any data error found during the process of data quality measurement. However, one data could exist in several different information systems, or its related data could exist in multiple. Data technicians should, therefore, correct all the data which has such relationships.

EXAMPLE   The classification code of product X in UNSPSC version 10 was 12345671, but in version 11, the code has changed to 12345680. Then, all of the classification code of product X distributed in information systems should be changed together.

Bibliography

[1] ISO/TS 8000-100, *Data quality — Part 100: Master data: Overview*

[2] ISO 9000, *Quality management systems — Fundamentals and vocabulary*

[3] ISO 9001, *Quality management systems — Requirements*

[4] ISO 10303-1, *Industrial automation systems and integration — Product data representation and exchange — Part 1: Overview and fundamental principles*

[5] ISO/IEC 8824-1, *Information technology — Abstract Syntax Notation One (ASN.1) — Part 1: Specification of basic notation*

[6] IEE Std 1320.1-1998, *Standard for functional modelling language — Syntax and semantics for IDEF0.*

[7] UNSPSC v10.0501, v11.0501 http://www.unspsc.org/download.aspx.

Index

ICS 25.040.40

Price based on 26 pages