ISO/IEC JTC1/SC7
Software & Systems Engineering
Secretariat: CANADA (SCC)

ISO/IEC JTC1/SC7 3435

2006-02-05

Document Type	CD Ballot
Title	CD 25012 Software engineering: Software Quality Requirements and Evaluation (SQuaRE) – Data Quality Model
Source	JTC 1/SC7 WG06
Project	25012
Status	CD
References	N3303, N3367, N3434
Action ID	ACT
Due Date	2006-05-05
Mailing Date	2006-02-05
Distribution	SC7_AG; P, O & L Members
Medium	Acrobat
No. of Pages	29
Note	

Address reply to: ISO/IEC JTC1/SC7 Secretariat
École de technologie supérieure – Department of Software and IT Engineering
1100 Notre Dame Ouest, Montréal, Québec Canada H3C 1K3
secretariat@jtc1-sc7.org

www.jtc1-sc7.org

ISO/IEC JTC1/SC7 CD 25012	
Date 2006-02-05	Reference number ISO/JTC 1/SC 7 N3435
Supersedes document N3303	

THIS DOCUMENT IS STILL UNDER STUDY AND SUBJECT TO CHANGE. IT SHOULD NOT BE USED FOR REFERENCE PURPOSES.

ISO/JTC 1/SC 7 Committee Title **Software and Systems Engineering** Secretariat: Standards Council of Canada (SCC)	Circulated to P- and O-members, and to technical committees and organizations in liaison for: X voting by (P-members only) **2006-05-05** Please return all votes and comments in electronic form directly to the SC 7 Secretariat by the due date indicated.

ISO/IEC JTC1/SC7

Title: **CD 25012 Software engineering: Software Quality Requirements and Evaluation (SQuaRE) – Data Quality Model**

Note:

Medium: Acrobat

No. of pages: 29

Vote on CD 25012	
Date of circulation **2006-02-05** Closing date 2006-05-05	Reference number ISO/JTC 1/SC 7 N3435

ISO/JTC 1/SC 7 Committee Title **Software and Systems Engineering** Secretariat: Standards Council of Canada (SCC)	Circulated to P-members of the committee for voting Please return all votes and comments in electronic form directly to the SC 7 Secretariat by the due date indicated.

ISO/IEC JTC1/SC7

Title: **CD 25012 Software engineering: Software Quality Requirements and Evaluation (SQuaRE) – Data Quality Model**

Vote:

__ APPROVAL OF THE DRAFT AS PRESENTED

__ APPROVAL OF THE DRAFT WITH COMMENTS AS GIVEN ON THE ATTACHED

__ general:

__ technical:

__ editorial:

__ DISAPPROVAL OF THE DRAFT FOR REASONS ON THE ATTACHED

__ Acceptance of these reasons and appropriate changes in the text will change our vote to approval

__ ABSTENTION (FOR REASONS BELOW):

P-member voting:
National Body (Acronym)

Date:
YYYY-MM-DD

Submitted by:
Your Name

NOTE: do NOT submit this form for email voting. Simply type your vote into an email message (with comments where applicable in the completed attached Excel Spreadsheet) and send to Secretariat@jtc1-sc7.org.

ISO/IEC JTC1/SC7/WG6 N563
2006-01-13
ISO/IEC 25012

Title:	***Software Engineering – Software Quality Requirements and Evaluation (SQuaRE) - Data Quality Model***
Document type:	***CD***
Source:	***JTC1/SC7 WG6***
Project:	***ISO/IEC 25012***
Status:	***1st CD***
References:	***NWI N3093, WD N3303***
Distribution	***WG6***
No. of pages:	***29***
Note:	

Document Editor: ***Domenico Natale***
SOGEI
Via Mario Carucci 99
00143 Roma – Italy
dnatale@sogei.it

Co-editors: ***Michael P. Kress***
BOEING
P.O. Box 3707 MC 08-43
Seattle, WA 98124 – 2207
michael.p.kress@boeing.com

Enrico Viola
ECLAT
Via dell'aeroporto 155
00175 Roma - Italy
viola@eclat-web.com

Software engineering: Software product Quality Requirements and Evaluation (SQuaRE) – Data Quality Model

Table of content

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 by 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 such committees. International organizations, governmental and non-governmental, in liason 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 set forth in the ISO/IEC Directives, Part 2.

In the field of Information Technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC1. 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 require the approval of at least 75% of the member bodies' casting vote.

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

ISO/IEC 25012 is a part of the SQuaRE series of standards and was prepared by Joint Technical Committee ISO/IEC JTC 1, information technology, Subcommittee SC 7, Software and System Engineering.

SQuaRE series of standards consist of five divisions under the general title "Software product Quality Requirements and Evaluation"; these divisions are commonly represented as shown in figure 1:

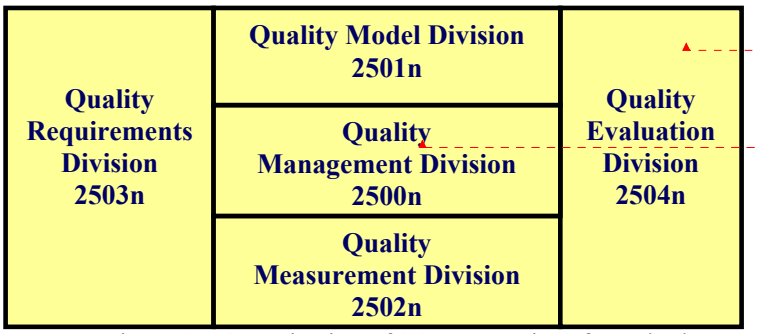

Figure 1 – Organization of SQuaRE series of standards

Formatted: Font: 12 pt

Formatted: Font: 12 pt

The Divisions within SQuaRE model are:

- **ISO/IEC 2500n - Quality Management Division**. The standards that form this division define all common models, terms and definitions referred further by all other standards from SQuaRE series. Referring paths (guidance through SQuaRE documents) and high level practical suggestions in applying proper standards to specific application cases offer help to all types of users. The division provides also requirements and guidance for a supporting function which is responsible for the management of software product requirements specification and evaluation.

- **ISO/IEC 2501n - Quality Model Division.** The standard that forms this division presents a detailed quality model including characteristics for internal, external and quality in use. Furthermore, the internal and external software quality characteristics are decomposed into subcharacteristics. Practical guidance on the use of the quality model is also provided.

- **ISO/IEC 2502n - Quality Measurement Division.** The standards that form this division include a software product quality measurement reference model, mathematical definitions of quality measures, and practical guidance for their application. Presented measures apply to internal software quality, external software quality and quality in use. Measurement primitives forming foundations for the latter measures are defined and presented.

- **ISO/IEC 2503n - Quality Requirements Division.** The standard that forms this division helps specifying quality requirements. These quality requirements can be used in the process of quality requirements elicitation for a software product to be developed or as input for an evaluation process. The requirements definition process is mapped to technical processes defined in ISO/IEC 15288 – Information Technology - Life Cycle Management - System Life Cycle Processes.

- **ISO/IEC 2504n - Quality Evaluation Division.** The standards that form this division provide requirements, recommendations and guidelines for software product evaluation, whether performed by evaluators, acquirers or developers. The support for documenting a measure as an Evaluation Module is also present.

Introduction

The quantity of data and information handled by computer systems is increasing worldwide.

Governments, industries and the private sectors are currently taking measures to ensure the enhancement of the data quality of data-bases.

The quality of software must be high to ensure the success of e-government, e-business and e-commerce. Likewise the data that is processed by the software must comply with defined data quality characteristics/subcharacteristics (e.g. consistency, semantic and syntactic accuracy, access security, etc.).

A common prerequisite to all initiatives is the quality of the data which are exchanged, processed and used in transactions between the computer systems and end-users [3].

The data quality model defined in this document has some correspondence with the principles of the Software Product Quality model described in ISO/IEC 9126-1[1] [ISO/IEC 25010].

Using the Software Product Quality Model governed by ISO/IEC 9126-1 [ISO/IEC 25010] as a reference, this new standard draws up a new framework which defines the main aspects of Data Quality and sets Data Quality standards.

Managing and enhancing the quality of data is important for a number of reasons:

- The enormous quantity of information handled and the plethora of subjects which have a shared responsibility at various levels.
- The acquisition of data from computer systems of other organizations, agencies, institutions of which the quality of data production processes is unknown
- The existence of defective (incomplete, incorrect, imprecise, obsolete, inconsistent, invalid, unauthenticated, untimely) data contributing to unsatisfactory information, unusable results and dissatisfied customers.
- The dispersion of such data among various users. Data captured in accordance with the workflow needs of a single organization often lacks a coherent, integrated and transversal vision which is necessary to ensure interoperability and co-operation.
- The need of homogenous criteria for definitions, classifications, authorization and cancellation procedures.
- The existence of “special interest’ categories of data which are not immediately re-usable because of semantic ambiguity or lack of congruity between such data and other existing co-related data.
- The co-existence of legacy architecture and systems with distributed systems in the same organization often with different data bases. This inevitably leads to problems of maintainability, replication, redundancy, alignment, and coherence.

Data should also meet the predefined criteria and requirements of the computer system or systems in which they will operate.

The purpose of this data quality model, therefore, is to prompt creators of large and small scale

data bases to observe predefined criteria which will enable them to evaluate the quality of data, set up integrated and interoperable data bases, reduce ambiguity, avoid redundancy, promote ease of data maintenance and promote reliable, secure data bases. High quality data serve authorized end-users, and facilitate the migration among computer systems when different data bases are merged.

This framework can be also used to implement and re-engineer data bases and to improve the application software, processing and procedures that feed and manage these data bases.

For additional guidance, references about data are provided in:

ISO/IEC 11179 “Information Technology – Metadata registries (MDR)” [9],

ISO 19113 “Geographic Information – Quality principles” prepared by ISO/TC 211 [6].

Software engineering: Software product Quality Requirements and Evaluation (SQuaRE) – Data Quality Model

1 Scope

This document has the aim of defining a general data quality model for computer systems, to support organizations to acquire, manipulate and use data with the necessary quality characteristics to achieve its objectives.

In domains where specific data quality standards exist (e.g. aviation, pharmaceuticals, geographic information), these shall take precedence over this standard.

This model defines data quality requirements and describes data quality characteristics for any computer system application (i.e.: e-government, e-business, e-commerce).

Each characteristic/subcharacteristic of the Data Quality model may vary in importance and may have different scores depending on specific requirements of the applications of the computer system.

Data Quality determines the quality of information.

This model only refers to structured digital data; it refers particularly to data that can be shared within the same computer system or by different computer systems.

This standard takes into account all data types (i.e.: chars, date, numbers, images, sounds, etc), assigned data values and rules that impact on data structure and relationships between data (i.e.: consistency between data in the same or in different entities).

The users of this standard can be Computer System technical users (i.e.: data administrators, data base administrators, developers, data stewards) and information system end users (i.e.: managers, employees).

This standard enables data quality to be specified and evaluated from different perspectives by those involved in the data's generation, acquisition, requirements definition, implementation, use, evaluation, support, maintenance, quality assurance and audit.

Many different roles can be defined within Data Quality management processes: i.e. data quality administrator, data producer, data acquirer, simple user: the analysis of these processes and the description of these roles are ouside the scope of this document; however, an example of a data quality management approach is described in the section 6 of this document.

This standard can be used as a guideline or as a term of reference in processes that impact on Data Quality to define the characteristics to be measured to determine the necessary quality for an organization to achieve its objectives.

The quality model defined in this standard could for example be used to:

- Identify data requirements in data acquisition and evaluation processes;
- Identify data quality assurance criteria;
- Define Data Base design objectives;
- Define organization's requirements in computer systems developing or acquisition processes

The definition of metadata is addressed by ISO/IEC 11179 and is outside the scope of this document, even if it refers to metadata to evaluate data quality.

Conceptual and logical data design are also outside the scope of this document, even if the use of the model can refer to conceptual and logical design to evaluate data quality.

The geographic data domain is addressed by ISO 19113 Geographic information – Quality principles, and it is outside the scope of this document. Requirements to such specific data are characterised as data quality subcharacteristics "compliance" with this standard.

2 Normative references

ISO/IEC 9126-1: 2001 "Software Engineering – Product Quality – Part 1: Quality Model"

ISO/IEC 25000 "Software Engineering – Software Product Quality Requirements and evaluation (SQuaRE) – Guide to SQuaRE". 2005

ISO/IEC 25010 "Software Engineering – Software Product Quality Requirements and evaluation (SQuaRE) – Quality model and guide". When published

3 Conformance

This document should be used as guidance for establishing a data quality strategy or program. If used for conformance, the user shall provide evidence that each data quality characteristic and subcharacteristic in clauses 5.2 and 5.3 of this document has been addressed, giving the reason for any exclusion, or describe its own categorisation of data quality attributes and provide a mapping to the characteristics and subcharacteristics in clauses 5.2 and 5.3.

4 Terms and definitions

For the purpose of this standard, the following definitions apply.

4.1 attribute

Inherent property or characteristic of an entity that can be distinguished quantitatively or qualitatively by human or automated means [ISO/IEC 25000].

4.2 computer system

An interconnected set of hardware and software resources under the same management control that shares common functionalities.
Software includes computer programs, data and documentation.

4.3 data

A re-interpretable representation of information in formalized manner suitable for communication, interpretation, or processing.

NOTE 1 : data can be processed by humans or by automatic means [ISO/IEC 2382-1] [20]

NOTE 2 : Each data has a specific degree of information**.** Data on its own has no meaning, only when interpreted by some kind of data processing system does it take on meaning and become information. Example:

1234567.89 is data.

"Your bank balance has jumped 15% to $1234567.89", is information.

NOTE 3**:** People or computers can find patterns in data to perceive information, and information can be used to enhance knowledge. Since knowledge is prerequisite to wisdom, users always want more data and information. But, as modern societies verge on information overload, users especially need better ways to find patterns

4.4 data acquirer

The data acquirer may be an individual, an organization, an agency, an institution, which has the responsibility for data acquisition process, and determining the quality of the computer system's data.

4.5 data producer

The data producer can be an individual, an organization, an agency, an institution, who has the responsibility for data generation process. The data producer shall make data available in the format or template requested by the organization.

4.6 data quality

Capability of data to satisfy stated and implied needs when used under specified conditions.

4.7 data quality administrator

The data quality administrator can be an individual or an organization, who has the responsibility to evaluate, manage and improve data quality. The data quality administrator also determine if the data meet the relevant characteristics in clause 5.2 according to users requirements. The data quality administrator assures that data quality is compatible with the target system/s and adequate for the intended user/s.

4.8 data quality characteristics

Category of data quality attributes that have an impact on data quality.

Note : Data quality characteristics may be refined into multiple levels of subcharacteristics and finally into data quality attributes.

4.9 data quality model

A defined set of characteristics , and of relationships between them, which provides a framework for specifying data quality requirements and evaluating data quality.

4.10 data type

A named set of values; e.g.: all possible integers (integer); all possible character string (char), all possible given names (gname).

4.11 data user

The data user is the stakeholder that uses data for some specific purpose and can be affected by its quality. The data user may be the processor of the data or the recipient of the processed data. The data user collaborate to determine risks.

The standard aim to help data users in defining data quality requirements

4.12 data value

The assigned value of an attribute.

4.13 entity

A concept that can be clearly identified and defined through a predicative sentence.

An entity is to be characterised by assigning its attributes

4.14 entity occurrence

This refers to an item that satisfies the predicative sentence that defines an entity.

4.15 information

Knowledge concerning objects such as facts, events, things, processes, or ideas, including concepts, that within a certain context have a particular meaning [ISO/IEC 2382-1]

4.16 information system

An information system contains one or more computer systems and communication systems. A communication system may be empty.

4.17 metadata

Data that describe other data. [ISO/IEC 11179] [9]

4.18 structured data

Structured data refer to data that are organized to be stored in an electronic device and described through a stated structure, which is managed by a computer system.

Data are considered at different levels such as: entities, attributes, values assigned to attributes.

Note: The data description and the structure in a Data Base Management System are logically stored with the data, while in several legacy data organizations, the data description and the structure can reside separately from the data.

4.19 technological environment

The technological environment is the infrastructure used to define, manipulate and manage data. It consists of hardware components, software components of Operating Systems, Data Base Management Systems or middleware in general.

4.20 value

An individual constant with no location in time or space and that can not be changed. [29]

5 Data Quality Model

5.1 General concepts

Data Quality can be represented by the quality model shown in table 1.

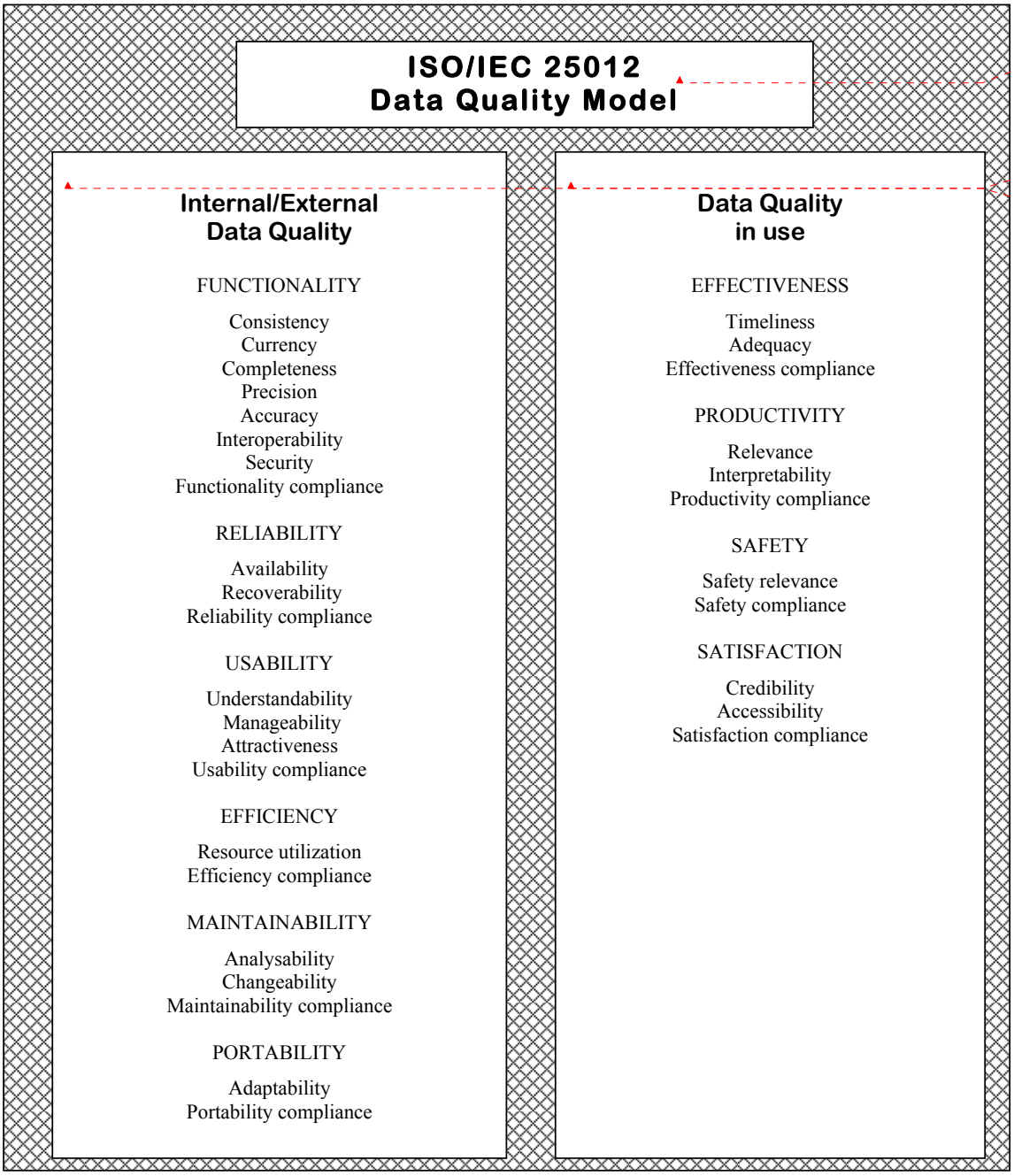

Table 1 - Data Quality Model

Table 1 outlines the Data Quality Model after having taken into account the already mentioned considerations. It categorises internal and external data quality attributes into six characteristics (functionality, reliability, usability, efficiency, maintainability and portability), which are further subdivided into subcharacteristics.

There are three different ways of viewing Data Quality. These include:

Internal Data Quality: the capability of a set of static data attributes to satisfy stated and implied needs when the data are used under specified conditions.

Internal Data Quality characteristics refer to data itself**;** they provide the criteria to ensure and verify the quality of :

- ✓ data values
- ✓ data type and length
- ✓ data definitions (including metadata)
- ✓ data rules
- ✓ relationships between data

taking account of documentation, files, and at the real world.

An example of an internal Data Quality characteristic is Consistency : it refers to data independently of HW/SW aspects.

Internal Data Quality users are data producers and data acquirers (e.g.: Internal data quality requirements must be considered in data preparation and data acquisition processes).

External Data Quality: the capability of data to satisfy stated and implied needs when data are used under specified conditions within a computer system.

External Data Quality characteristics are "inherited" by data from computer system's capabilities, that can be implemented on such data depending on user requirements.

An example of an External Data Quality characteristic is Security since it depends on HW and SW capabilities.

External Data Quality users are mainly data acquirers (e.g.: external data quality requirements must be considered in data acquisition and software application design processes).

NOTES :

(1) Ways of viewing internal, external, in use data quality are different from the ones used to view software quality (ISO/IEC 25000 "Software Engineering – Software Product Quality Requirements and evaluation (SQuaRE) – Guide to SQuaRE". 2005)

(2) Internal Data Quality is not able to predict external data quality;

(3) metrics may be different for the same characteristic for internal/external quality.

Data Quality in Use: the capability of the data to enable specific users to achieve specific goals with effectiveness, productivity, safety and satisfaction in specific contexts of use.

Users of Data Quality in Use are software application designers and data users.

Data Quality in use aims to define the data quality characteristics that express the user's subjective point of view about data he is working with in terms of how such data satisfies his information needs.

Data quality in use is categorized into four characteristics (effectiveness, productivity, safety and satisfaction), which are refined into subcharacteristics.

5.2 Internal and External Quality

5.2.1 Overview

Internal and external quality is a set of six characteristics of data considered from an internal and external point of view. The six characteristics (functionality, reliability, usability, efficiency, maintainability and portability) are mainly related to the data value, to the format used for storing the data and the technological environment, i.e. hardware, software and middleware.

5.2.2 Functionality

5.2.2.1 Definition of functionality

Functionality is the capability of the data to enable the computer system to meet the user's functional requirements and objectives**.** Functionality requirements for data are user dependent and generally refer to the amount of functions (or capabilities) that the data can serve.

Examples: A data base of phone numbers for a country has more functionality than a data base of phone numbers for a city within a country. An index by author, date and subject matter has more functionality than a simple index by author.

Functional requirements specifically refer to requirements that ensure that the data is suitable, accurate, interoperable, secure and compliant with relevant functional standards and regulations. The suitability when referring to data, also regards: consistency, currency, completeness and precision

5.2.2.2 Consistency

Consistency refers to the absence of apparent contradictions within data [28]. Inconsistency can be verified on the same or different entities.

Example : An employee's birth date cannot be greater than his "recruitment date"

5.2.2.3 Currency

Currency is the extent to which data is up-to-date [4]. It is critical for volatile data (i.e.: frequently updated data, as wind speed or climate temperature).

Example: A flight seat cannot be available after it has been assigned to someone.

5.2.2.4 Completeness

Completeness is the extent to which all necessary values have been assigned and stored in the computer system. Completeness refers both to entity occurrences and to attributes of a single occurrence.

Example: if an entity refers to company employees, all the employees must be recorded with all attributes requested to satisfy a user's needs.

5.2.2.5 Precision

Precision is the capability of the value assigned to an attribute to provide the degree of information needed in a stated context of use.

Example: to represent the duration of a marathon race the minimum unit of time must be seconds; to represent the duration of 100 meters race the minimum unit of time must be milliseconds.

5.2.2.6 Accuracy

Accuracy is defined as the degree to which a data value conforms to its actual or specified value.

Accuracy has two main aspects:

Syntactical accuracy

Syntactical accuracy is defined as the closeness of the data values to a set of values defined in a domain considered syntactically correct.

Example: a low degree of syntactical accuracy is when the word Mary is stored as Marj.

Semantic accuracy

Semantic accuracy is defined as the closeness of the data values to a set of values defined in a domain considered semantically correct.

Example: a low degree of semantic accuracy is when the name John is stored as George. Both names are syntactically accurate, because of the domain of reference in which they reside, but George is a different name.

5.2.2.7 Interoperability

The capability of data to be accessed, extracted and exchanged among different platforms and systems.

5.2.2.8 Security

The capability of the data to be accessed only by authorized users.

5.2.2.9 Functionality compliance

The capability of the data to adhere to standards, conventions or regulations in force and similar rules relating to functionality.

5.2.3 Reliability

5.2.3.1 Definition of reliability

The capability of the data to maintain a specified level of operations in a specific technological environment.

5.2.3.2 Availability

The capability of data to be always retrievable.

Note*:* a particular case of availability is concurrent access (both to read or to update data) by more than one user and/or application.

5.2.3.3 Recoverability

The capability of the data to maintain and preserve its physical and logical integrity, even in the event of failure.

This standard refers to the quality of data, not the system: the characteristic of recoverability is related to the data that must be recoverable.

Note : Recoverability can be provided by features like commit/synchpoint, rollback (fault-tolerance capability) or by backup-recovery mechanisms.

5.2.3.4 Reliability compliance

The capability of the data to adhere to standards, conventions or regulations in force and similar rules relating to reliability.

5.2.4 Usability

5.2.4.1 Definition of usability

The capability of the data to be understood, managed and used to catch the user's attention, when used under specified conditions.

5.2.4.2 Understandability

The extent to which data is in appropriate languages, symbols and units, and the degree to which its definitions are clear [5], [11].

Note : some information about data understandability are provided by metadata.

Example : To represent a State, the standard acronym is more understandable than a numeric code

5.2.4.3 Manageability

The capability of data to be stored appropriately from a functional point of view.

Example : Data representing costs should be stored like numeric and not like string, to allow users to carry out algebraic operations.

5.2.4.4 Attractiveness

The capability of data to be attractive to users.

Example : to describe a holiday resort a data stored as an image is more attractive than its description like a character string.

5.2.4.5 Usability compliance

The capability of data to adhere to standards, conventions or regulations in force and similar rules relating to usability.

5.2.5 Efficiency

5.2.5.1 Definition of efficiency

The capability of data to be processed (accessed, acquired, updated, etc) and to provide appropriate levels of performance, relative to the amount of resources used, under stated conditions.

5.2.5.2 Resource utilization

The capability of data to be stored or accessed using the appropriate amounts and types of resources under stated conditions.

Example : Using more space than necessary to store data cause waste of storage, memory and CPU.

5.2.5.3 Efficiency compliance

The capability of data to adhere to standards, conventions or regulations in force and similar rules relating to efficiency.

5.2.6 Maintainability

5.2.6.1 Definition of maintainability

The capability of data to be modified for changes in technological environment, in requirements or in functional specifications

5.2.6.2 Analysability

The capability of data to be inspected for checking.

Note: it includes the capability of data to be tested.

5.2.6.3 Changeability

The capability of data to be changed in its type, length or assigned value

5.2.6.4 Maintainability compliance

The capability of data to adhere to standards, conventions or regulations in force and similar rules relating to maintainability.

5.2.7 Portability

5.2.7.1 Portability definition

The capability of data to be transferred from one technological environment to another.

5.2.7.2 Adaptability

The capability of data to be moved from one platform to another; this includes the capability of data to be also installed and replaced in its destination platform. It refers particularly to homogeneous and coherent set of data (i.e.: data that are requested by a software application).

5.2.7.3 Portability compliance

The capability of data to adhere to standards, conventions or regulations in force and similar rules relating to maintainability

5.3 Data Quality in Use

5.3.1 Data Quality in use overview

The capability of data to enable users to achieve specific goals with effectiveness, productivity, safety and satisfaction in specified contexts of use.

5.3.2 Effectiveness

5.3.2.1 Definition of effectiveness

The capability of data to enable users to achieve specific goals from a quantitative and qualitative point of view in a specified context of use.

5.3.2.2 Timeliness

The extent to which data is sufficiently up-to-date for the task at hand [5].

Example : When attending a course a student needs to know the right timetable before its starting date.

5.3.2.3 Adequacy

The extent to which data are sufficient to satisfy user's stated needs from quantitative point of view. Adequacy includes the capability of data to represent the context observed by users.

Example: data are adequate if there are all attributes that user needs and if existing attributes have an assigned value where user is expecting one

5.3.2.4 Effectiveness compliance

The capability of data to adhere to standards, conventions or regulations in force and similar rules relating to effectiveness.

5.3.3 Productivity

5.3.3.1 Definition of productivity

The capability of data to enable users to carry out his task in an efficient way by using an appropriate amount of resources.

5.3.3.2 Relevance

The extent to which data is applicable and helpful for the task at hand [5], [11].

Data are relevant for users when they satisfy the scope and the goal of his information needs.

5.3.3.3 Interpretability

The extent to which the real meaning of data is easy for users to comprehend.

Note: Metadata can help the comprehension of data [9]

5.3.3.4 Productivity compliance

The capability of data to adhere to standards, conventions or regulations in force and similar rules relating to productivity.

5.3.4 Safety

5.3.4.1 Definition of safety

The capability of data to achieve an acceptable level of risk of harm to people, business, property or the environment in a specified context of use.

5.3.4.2 Safety relevance

The extent to which data is related to safety.

Example : data concerning blood group are relevant to users of health care computer systems.

5.3.4.3 Safety compliance

The capability of data to adhere to standards, conventions or regulations in force and similar rules relating to safety.

5.3.5 Satisfaction

5.3.5.1 Definition of satisfaction

The capability of data to satisfy users in a specified context of use.

5.3.5.2 Credibility

The extent to which data satisfy user's needs and it is regarded as true and credible by them [5].

5.3.5.3 Accessibility

The capability of data to be accessed, particularly by people who need supporting technology or special configuration because of some disability.

5.3.5.4 Satisfaction compliance

The capability of data to adhere to standards, conventions or regulations in force and similar rules relating to satisfaction.

6 Approach to Data Quality

6.1 Data Quality Approach Diagram

The following diagram shows an example of a data life cycle beginning with the Data Quality model customization.

Four stakeholders or contributors to data quality are represented : a data quality administrator, a data producer, a data acquirer, and a data user.

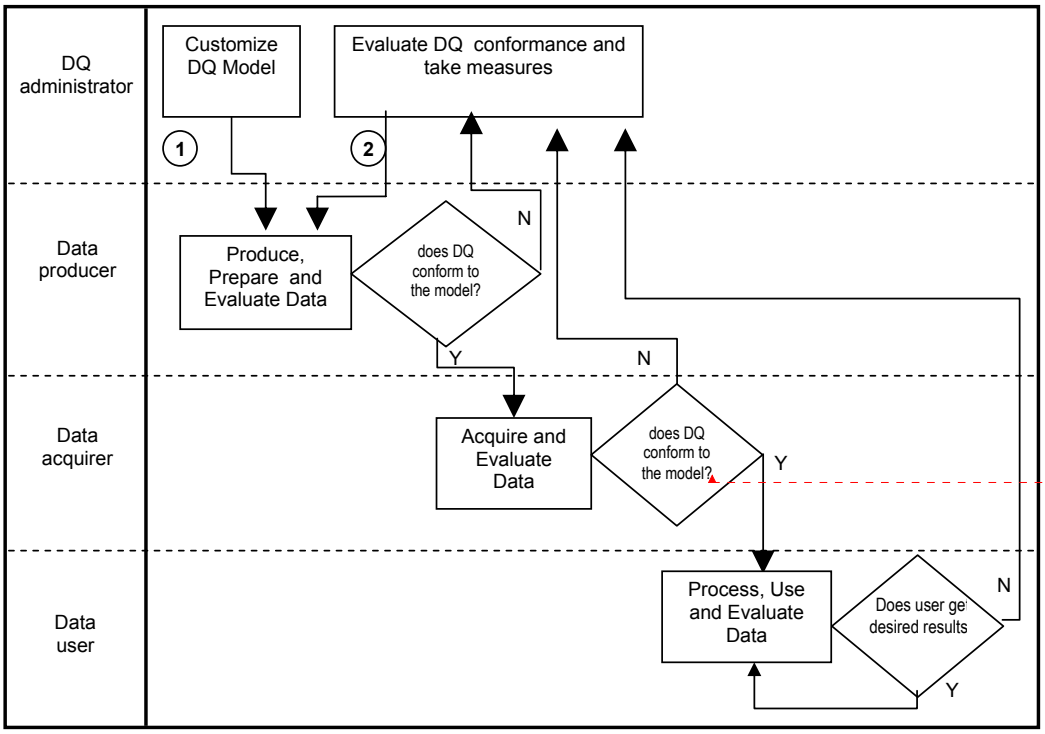

Figure 1 Example of a Data Quality Approach

6.2 Data Quality Administrator

The data quality administrator can be an individual or an organization who has the responsibility of managing the quality of an organization's data. The data quality administrator shall:

- ✓ Adapt the standard Data Quality model to the organization's needs (1)
- ✓ Evaluate all the non conformity warnings received from other stakeholders. These warnings can generate three kinds of measures :

  a) Data correction: in this case the data quality administrator asks the data producer to correct the data (2)

  b) Technology adjustment: this consists of the improvement of hardware or software capabilities;

  c) Organizational measures, i.e.: user's training to avoid errors in the data entry process.

Formatted: Font: Arial Narrow

Formatted: English (U.K.)

Formatted: Font: 8 pt

Formatted: Font: 8 pt

6.3 Data Producer

The data producer can be an individual, an organization, an agency, an institution, which has responsibility for generating data. The data producer shall:

- produce data compliant with the quality characteristics defined in Table 1 and Table 2;
- transmit data in accordance with procedures defined;
- take actions according to data quality administrator requests in case of incorrect, incomplete or missing data.

6.4 Data Acquirer

The data acquirer may be an individual, an organization, agency, institution, which has the responsibility of acquiring data into the Computer System. The data acquirer has also the responsibility of evaluating the quality of the data received from the data producer according to the organization's Data Quality model characteristics that can be evaluated in the processes in which he operates. Data Quality evaluation includes risk analysis to determine the adequacy of data quality characteristics including correctness, completeness, understandability, accuracy precision and relevance for the intended use.

It may be unwise and costly to request that data more precise or accurate than is necessary.

Example 1: A banking transaction to transfer funds from one account to another may be adequate if defined to the nearest two digits of precision, e.g., $1000.00, whereas a currency conversion may require precision of more than 2 digits, e.g. 1 Euro = x.xxxx USD.

Example 2: A location requiring only approximate description could be described as:

North Latitude : 47 degrees, 14 minutes
West Longitude: 24 degrees, 23 minutes

whereas a request for a more precise definition for the same location would be:

North Latitude : 47 degrees, 14 minutes, 15 seconds
West Longitude; 24 degrees, 23 minutes, 46 seconds

6.5 Data User

The data user is the stakeholder which is affected by the quality of the data.

The data user may be the processor of the data or the recipient of the processed data. The data user has also the responsibility of evaluating the quality of the data he manages according to the organization's Data Quality model characteristics that can be evaluated in the processes in which he operates. Data Quality evaluation includes risk analysis to determine the adequacy of data quality characteristics.

7 Bibliography

[1] ISO/IEC 9126-1: 2001 Software engineering — Product quality — Part 1: Quality model

[2] L. P. English, Improving Data Warehousing and Business Information Quality, Wiley, 1999

[3] D. Natale, M. Scannapieco, P. Angeletti, T. Catarci, G. Raiss, "Qualità dei dati e standard ISO/IEC 9126: Analisi critica ed esperienze nella Pubblica Amministrazione Italiana", Workshop AIPA "Sistemi in rete nella Pubblica Amministrazione" with the participation of Sogei-Società Generale d'Informatica S.p.A. and the University of Rome "La Sapienza", (in conjunction with The VLDB Very Large Data Base Conference 2001) Rome September 2001.Esempi di misurazioni

[4] Redman T.C.: Data Quality for the Information Age. Artech House, 1996

[5] Wang R.Y., Strong D.M.: Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems, vol. 12, no. 4, 1996.

[6] ISO/IEC 19113 "Geographic Information – Quality principles" prepared by ISO/TC 211

[7] Sarbanes-Oxley Act; Data Quality Act of 2001, Public Law 106-554

[9] ISO/IEC 11179 "Information Technology – Metadata registries (MDR)"

[10] Huang, K.T., Lee, Y., Wang, R. Quality Information and Knowledge. Prentice Hall, Upper Saddle River, 1999

[11] Pipino, L., Lee, Y., Wang, R. "Data Quality Assessment" Communications of the ACM April 2002/Vol. 45, No. 4

[12] Olson, J. E. Data Quality: the accuracy dimension. Ed. Morgan Kaufmann Publishers. 2003

[13] Wang, R., Storey V.C., Firth, C. F. "A framework for analysis of Data Quality Research". IEEE Transactions on Knowledge and Data Engineering, Vol 7(4). (1995) Pp 623-640

[14] Azuma M.: Square, the Next Generation of the ISO/IEC 9126 and 14598 International Standard Series on Software Product Quality. In Proceedings of the 12th European Software Control and Metrics Conference, London, 2001

[15] Ballou, D. P. and Pazer H. L., Modeling Data and Process Quality in Multi-input, Multi-output Information Systems. Management Science, vol. 31,no 2, 1985

[16] Batini C., Lenzerini M., Navathe S.B. (1984): A comparative analysis of methodologies for database schema integration. ACM Computing Survey, vol. 15, no. 4, 1984

[17] Calvanese D., De Giacomo G., Lenzerini M., Nardi D., Rosati R.: Information Integration: conceptual Modeling and Reasoning Support, In Proceedings of the 6th International Conference on Cooperative Information Systems (CoopIS'98), New York City, NY, USA, 1998

[18] Crosby P.B.: Quality without Tears. McGraw-Hill, New York, 1984

[19] Gertz M.: Managing Data Quality and Integrity in Federated Databases. Second Annual IFIP TC-11 WG 11.5 Working Conference on Integrity and Internal Control in Information Systems, Airlie Center, Warrenton, Virginia, 1998

[20] ISO/IEC 2382-1 : Information technology – vocabulary – part 1 : fundamental terms

[21] ISO/IEC 9126: Information Technology- Software Product Evaluation- Quality Characteristics and Guidelines for their Use

[22] Kriebel, C. H. and O. Mikhail, Dynamic Pricing of Resources in Computer Networks. Logistics, 1980

[23] Madnick S.: Metadata Jones and the Tower of Babel: The Challenge of Large – Scale Semantic Heterogeneity. Proceeding of the 3rd IEEE Meta-Data Conference (Meta-Data '99), Bethesda, MA, USA, 1999

[24] Natale D.: Qualità e Quantità nei Sistemi Software. Franco Angeli, 1995

[25] Ulmann J.D.: Information Integration using Logical Views. Proceedings of the International Conference on Database Theory (ICDT '97), Greece,1997

[26] M. Scannapieco, T. Catarci, Data quality under the Computer Science perspective, University of Rome. 2004

[27] Bovee M., Srivastava R. P., Mak B.R.: A Conceptual Framework and Belief-Function Approach to Assessing Overall Information Quality. In Proceedings of the 6th International Conference on Information Quality, Boston, MA, 2001.

[28] Horwan Veregin – Data Quality Measurement and Assessment

[29] J. Craig Cleaveland : An introduction to data types, 1986

8) ANNEX A : Correspondence between the characteristics defined in ISO/IEC 9126-1 and the characteristics in ISO/IEC 25012

INTERNAL / EXTERNAL QUALITY			
Characteristics	**Subcharacteristics**		**Notes**
9126-1/25012	**9126-1**	**25012**	
Functionality	Suitability (1)	Consistency (1)	(1) Suitability is expanded in new subcharacteristics
		Currency (1)	
		Completeness (1)	
		Precision (1)	
	Accuracy	Accuracy	
	Interoperability	Interoperability	
	Security	Security	
	Functionality compliance	Functionality compliance	
Reliability		Availability	
	Maturity		
	Fault tolerance		
	Recoverability	Recoverability	
	Reliability compliance	Reliability compliance	
Usability	Understandability	Understandability	
	Learneability (2)		(2) Learnability is outside the scope (it is a characteristic of data definition, not of data itself)
	Operability	Manageability	
	Attractiveness	Attractiveness	
	Usability compliance	Usability compliance	
Efficiency	Time behaviour		
	Resource utilization	Resource utilization	
	Efficiency compliance	Efficiency compliance	
Maintainability	Analysability	Analysability	
	Changeability	Changeability	
	Stability		
	Testability		
	Maintainability compliance	Maintainability compliance	
Portability	Adaptability	Adaptability	
	Installability		
	Co-existence		
	Replaceability		
	Portability compliance	Portability compliance	

QUALITY IN USE		
Characteristics	**Subcharacteristics**	
9126-1/25012	**9126-1**	**25012**
Effectiveness	-	Timeliness Adequacy Effectiveness compliance
Productivity	-	Relevance Interpretability Productivity compliance
Safety	-	Safety relevance Safety compliance
Satisfaction	-	Credibility Accessibility Satisfaction compliance