From Data Source Quality to Information Quality: the Relative Dimension

Laure Berti-Equille

From Data Source Quality To Information Quality : The Relative Dimension

Laure Berti
GECT, Equipe Systèmes d'Information Multi-Média
Université de Toulon et du Var
B.P. 132, F-83957 La Garde Cedex, FRANCE
berti@univ-tln.fr

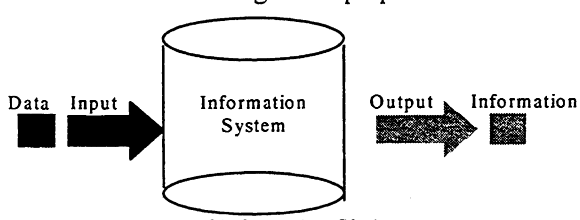
Abstract

This paper presents our current work concerning the characterization of the Relative Information Quality concept. In the context of over- and disinformation stressed by Push and Pull Technologies improvements, Information Quality judgment is very relative to Data Supervisors and to Information Customers. We present the methods we are currently being implementing and testing for user profile acquisition, flexible and personal Information Quality configuration and Computer-aided Relative Information Quality Estimation. These functionalities will extend the VIGIWARE system we previously presented during the 1997 Conference on Information Quality as an Information System specifically dedicated to Technological Watch enabling the management and recommendation of contradictory multi-source information and their quality.

1. Introduction

Effective communication of usable information to the decision maker is crucial to the competitive advantage of the enterprise and central to its decision support systems. Recent studies have emphasized the need for data quality and the need for a richer information context*. Beyond the borders of the enterprise, Multimedia Information Technology is now at the stage where it may be able to create a new infrastructure to facilitate the

* In this paper, we indiscriminately use the terms "Data" and "Information" in spite of a shade of meaning due to their respective position in the Information Chain figure we propose :



Extract of Information Chain

deployment of knowledge in service to Society. Such an infrastructure must be based on new research on methods of information representation, involving ontologies, frames, semantic networks, annotation techniques, combinations of these, and other newer approaches for representing, storing and transmitting high-quality information (Strong, 1998). New theories of information measurement, summarization, expression and co-expression must be developed together with technologies that enable modality and media independence. For the largest audience, the trend is now emerging with new universal access initiatives and human-centered systems research and development to increase accessibility and usability of computing systems and communications networks. In the context of information mass-production, it is becoming clearer and clearer that Information Quality will become one of the keystones of the "Information Consumer Society" and its impacts will be all the more tremendous that it will be measured by the degree to which all citizens may both provide and access high-quality information, and thus, share regardless of ability or technology, and independent of place or time. Nevertheless, a consistent gap still remains between users expectations regarding Information Quality and the perceived quality of the information they are using. The modern quality paradigms emphasize the importance of Customer Satisfaction as a driver to the Information Quality improvement process. Among the different management approaches aimed at satisfying all (internal / external, contradictory, changeable) customer requirements, needs and expectations, the practical and common sense TQM concept of Customer Focus remains an implicit problem. As a matter of fact, « poor quality data appears to be the norm rather than the exception » (Redman, 1996). The relative lack of consensual definitions and commonly accepted standards for Information Quality dimensions may be a consequence of the compartmentalization between the different application domains where Information Systems are involved and where Information Quality is viewed from a very particular angle. Beyond error handling, another difficulty comes from the complexity of surrounding the subjective dimension, which conditions for each user his perception of the quality of an information item and on which is based his satisfaction.

The purpose of this paper is to stress three aspects of the Relative Information Quality concept we define and whose complexity essentially depends on both contradictions between multi-user requirements and contradictions between multi-source information. In the continuation of the VIGIWARE project we initialize three years ago (Berti, 1997), we present one of the current orientations of our work for what concerns the development of user-centered tools for Information Quality improvement in a critical area of Information

Systems dedicated to Push Technologies for Technological Watch. These tools are involving techniques for information fusion, integration of knowledge-based methods, and personal recommendation of value-added information. Section 2 presents an overview of methodological and conceptual approaches concerning Information Quality across different application domains of Information Systems. Section 3 describes the Relative Information Quality as a new category grouping subjective and relative Information Quality parameters and presents the methods enabling the flexible characterization and evaluation of Relative Information Quality by users. Section 4 shows the importance of Relative Information Quality in the context of Technological Watch. Conclusions are given in Section 5.

2. Trans-disciplinary Overview of Information Quality

2.1. Methodological Approaches

Related Work

Quality Engineering methods, such as TQM (Total Quality Management), SQC (Statistical Quality Control), and QFD (Quality Function Deployment) are commonly used by many product design and manufacturing disciplines. They are entering in the field of Information Quality with different derived approaches and methodologies. Current approaches and prospective Data Quality improvement methods recommended for Software Engineering data collection and analysis include :

- Engineering (or reengineering) the data collection process and entry processes for Data Quality,
- Human factors in data collection and manipulation for Data Quality,
- Joint Information System and Process Design for Data Quality,
- Data editing, error localization and imputation techniques,
- Sampling and inspection methods,
- Data tracking : follow a random sample of records through the process to trace the root sources of error in the data collection and reporting process.

Many methodologies have been proposed to facilitate the implementation of an organization's overall Data Quality Policy or to improve specific (sub)processes in Information Manufacturing and Mining. Some of them are shortly described in Table 1.

Authors & References	*Methodologies*	*Objectives*
(Wang, 1998) (Wang, Kon, Madnick, 1993)	TDQM Total Data Quality Management	1) Defining Information Quality : Information Product Quality specifications are an integral component of Information System Design process, 2) Measuring Information Quality : IQ Metrics are designed to track percentages of errors, timeliness, completeness, consistency..., 3) Analyzing : Methods (SPC, pattern recognition, Pareto Chart analysis) investigate the cause for IQ problems, 4) Improving Information Quality
(Redman, 1996)	- Data Quality Requirements - Managing Information Chains - Strategies for improving Data Accuracy	- Translating the voice of the customer into the language of the process involving QFD, Impact, Translation and Process Performance or Systems/Process Matrices at the Design Stage - Improving the processes in Information Chains - Background for developing an operational plan for data quality improvement
(Dvir, Evans, 1996)	InfoQual	Facilitating the completion of customers Information Quality needs and expectations in an Information Quality project
(Meyen, Willshire, 1997)	DQEF Data Quality Engineering Framework	Addressing the temporal aspect of Data Quality environments : 1) Defining a Data Quality Engineering Model (DQEM) 2) Defining measures of DQEF success 3) Verifying DQEF on selected portions of the existing database 3) Validating the framework
(Ballou, Tayi, 1989)	Methodology for Allocating Resources for Data Quality Enhancement	Using effectively resources committed to data integrity maintenance via heuristic for balancing the costs of parameter estimation and data quality maintenance.
(Gardyn, 1997)	Quality Model Framework for a Data Warehouse	Linking the observable Data Warehouse properties and Data Quality Dimensions
(Paradice, Fuerst, 1991)	Management Information System (MIS) Data Quality Methodology	Formulating the error rate of stored Management Information Systems records

Table 1. Some Methodologies and Frameworks for Information and Data Quality

Adaptation of STEP methodology

Although it seems obvious that Information Quality depends on the good (or mis-) understanding that the Information Producer may have about the data specifications, it is necessary to choose the methodology which may facilitate this understanding for the complete definition of potential users' Information Quality expectations. The risks of omission, redundancy and divergence of interpretations are omnipresent during the step of collecting user's requirements (interviews, surveys...). The analysis and synthesis of the resulting requirements inventory are necessary to establish the order of priorities and data modeling. But, the defective analysis may generate the choice of a data structuration that may be not adapted for the expected usage or may be more expensive than another solution. Finally, it's essential to specify information about Information Production Process and about what is related to the Information Service (documentation).

Because Technological Watch focuses on advanced and competitive Industrial products, we decided to adapt and use the methodology of STEP - Standard for The Exchange of Product - data model (ISO 10303), for what concerns the development of Data and Data Quality Specifications. The main reasons of this choice are that STEP has been developed to meet industry requirements for standard data specifications that support :

- long term storage and retention of information about the Information Products (meta-information and documentation),
- independence of data and meta-data from the software tools which create or consume information and meta-information,
- communication of (meta-)information between departments, disciplines and enterprises.

A detailed description of the methodology can be found in (Fowler, 1995).

Within STEP, elements of Data Specifications are taken to be the representation of facts about objects in the real world (*Industrial Products*). The basis of STEP Data Specifications lies in a framework for Industrial Product data modeling that is based on a classification of the types of data that describes Industrial Products. This classification identifies five major types of data which also constitute the different steps for specifications (see Table 2 and 3). Note the distinction between *Industrial Product* and *Information Product* : actually, in the context of Technological Watch, *Information Products* describe an *Industrial Product.* Table 2 presents the Data Specifications for an Industrial Product.

Type of Data	*Description*
Application context	Data that defines the types of Industrial Product, disciplines, and life-cycle stages. The use of an application context allows data that represents an "as designed" Industrial Product to be distinguished from an "as built" Industrial Product.
Industrial Product Definition	Data that identifies the Industrial Product, including variants and categories, and the defined life-cycle views of the Industrial Product. Industrial Product definition data also include what is related to the structure of the Industrial Product.
Industrial Product Property Definition	Data that characterizes the Industrial Product by its properties, independent of the representation of properties. For example, it is possible to identify the shape of an object, or aspects of the shape as a property of the real object, without providing a detailed description of shape using a CAD model.
Industrial Product Property Representation	Data that represents the properties of the Industrial Product, including multiple representations of the same property : a 3D CAD model, a technical illustration or an engineering drawing are different representations of the same shape.
Industrial Product Property Presentation	Data that defines the presentation of product information to support human communication. The shape of an object (the property) is represented by coordinate values, curves, surfaces. This representation is presented by assigning colors, line fonts and displaying the resulting picture on the workstation.

Table 2. Data Specifications for an Industrial Product (Step 1)

Analogically, Data Quality Specifications are taken to be the representation of judgments about the *Information Product* describing the *Industrial Product.* These meta-data are specified with the same schema by the Data Supervisor, expert in the Technological Application Domain. Table 3 presents Meta-Data Specifications concerning the quality of the Information Product which describes the Industrial Product.

Type of Meta-Data	*Description*
Application Context of Quality Evaluation	Meta-data that defines the purpose for which the Information Product quality is evaluated : who's the Data Supervisor ? ...
Information Product Definition	Data that identifies an instance of the Data Model describing the Industrial Product (including Step 1, Table 2)
Information Product Quality Criteria Definition	Meta-data that characterizes the quality of the information product with different quality criteria.
Information Product Quality Criteria Representation	Meta-data that represents the criteria of the Information Product Quality, including multiple representations of the same criteria measurement.
Information Product Quality Criteria Presentation	Meta-data that defines the presentation of the Information Product Quality to support human communication.

Table 3. Meta-Data Specifications for Information Product quality (Step 2)

The Figure 1 schematically shows our methodological approach.

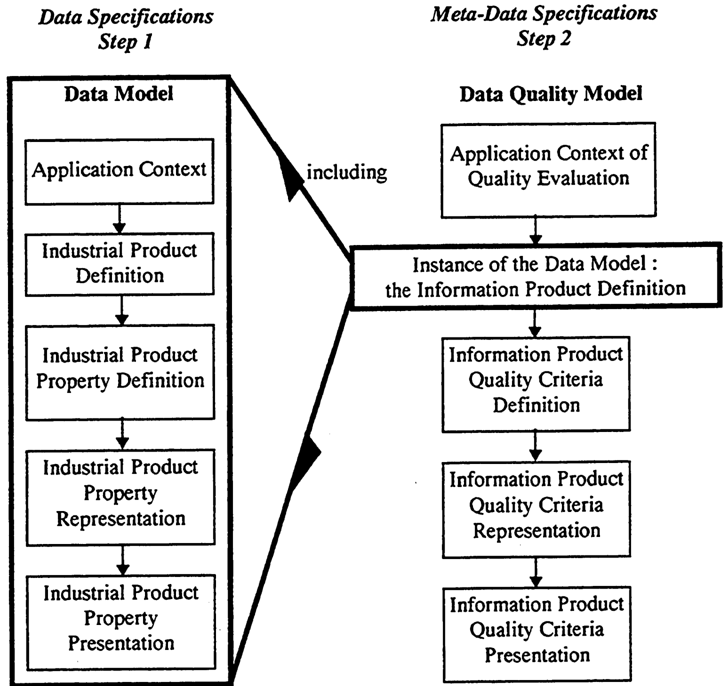


Figure 1. Data Model and Data Quality Model with adapting the STEP Methodology

2.2. Multidimensional Characterization of Information Quality

In the last couple of years, an increasing body of research literature focused on the Information Quality characterization and measurement (Wang, Storey, Firth, 1995). Numerous multidisciplinary researches are addressing the problem of Information Quality Dimensions and it is widely agreed that the complete Information Quality notion must be decomposed into several dimensions and grouped into different categories. It is assumed that these dimensions should be used in the process of designing, testing and selecting or evaluating Information Quality and systems as it is suggested in many methodologies.

Managerial and Organizational Information Quality

The first step was the dichotomy of the Information Quality concept into several categories, dimensions, aspects and attributes. Although this step was commonly accepted, its results showed that a consensus for Information Quality characterization and terminology is hard to find, especially with the differences of culture, language, education, interpretation and application domains...

Authors & References	*IQ Dichotomy*
(Zmud, 1978)	4 Quality Dimensions, 25 Factors
(Delen, Rijsenbrij, 1992)	4 Quality Dimensions, 21 Aspects, 40 Attributes
(Wang, 1998) (Strong, Wang, 1994)	4 Categories, 179 Attributes
(Redman, 1996, chap.13)	15 Characteristics of Quality Dimensions for a Conceptual View, 4 Quality Dimensions of Data Values 8 Quality Dimensions of Data Representation
(Gardyn, 1997, p. 278)	5 Quality Dimensions

Table 4. Some Propositions of IQ Characterization

At the trans-disciplinary level, the consensus for the definitions of the notion and dimensions of Information Quality is all the more difficult to find between the different application domains as nature and formats of information are very specific. In this connection, the next section briefly presents the specific approach for Geographic Information Quality.

Geographical Information Quality

A lot of research and standardization programs (Günther, Voisard, 1997) are currently being carried out in the context of geographic and environmental information in order to promote the exchanges of data, to use meta-data for facilitating data transfers, and to improve geographic data integration and update. Seven main meta-data about Geographic Information Quality have been defined in different standards (see Table 5) as follows:

Data Lineage includes « a description of the source material from which the data where derived, and the methods of derivation, including all transformations involved in producing the final digital files » (NIST , 1994, p.21).

Positional Accuracy is « the degree of compliance to the spatial registration standard » (Ibid. p.22).

Completeness includes « information about selection criteria, definitions used and other relevant mapping rules, the relationship between the objects represented and the abstract universe of all such objects. It describes the exhaustiveness of a set of features » (Ibid. p.24).

Logical Consistency describes the number of features, relationships, or attributes that have been correctly encoded in accordance with the integrity constraints of the feature data specifications.

Semantic Accuracy refers to the pertinence of meaning of the geographic object rather than to the geometrical representation and is an element of the evaluation of "fitness for use" that users have to perform before using spatial data.

Temporal Information describes the date of observation, type of update (creation, modification, deletion, unchanged), and validity periods for spatial data records.

A quick survey on the usual understanding of geographical quality meta-data in the different existing standards shows us the extend of undertaken works concerning Geographic Data Quality (Morrison, Guptill , 1995) (Fisher and al., 1995).

Standards	*References for Data Quality*	*Standard Description*
SDTS	(NIST, 1994) Section 1	Spatial Data Transfer Standard
FGDC	(FGDC, 1997) Group 2	FGDC Content Standard for Digital Geospatial Metadata defines meta-data to provide a «common set of terminology and definitions for documentation related to these meta-data».
CEN/TC 287	(CEN, 1996)	The European Technical Committee 287 for Normalization defines meta-data as « data about datasets »
ISO/TC 211	(ISO, 1997) Section 2	Resulting from ISO 1546 Standard, ISO/TC/211 specifies meta-data as « data about the content quality, condition, and other characteristics of data » to describe a complete dataset.
OpenGIS	(OpenGIS, 1997)	OpenGIS Meta-data specifications rely on other existing standards especially ISO and FGDC. But the difference is in the expected usage of metadata.

Table 5. Standards including Metadata about Geographic Data Quality

3. A New Category : the Relative Data Quality

3.1. Description

We previously presented different approaches for characterizing Information Quality which either use scientifically grounded methodologies in order to rigorously define Information Quality Dimensions, or create a universal standard set of operational Quality Dimensions. Our practical approach follows a third research avenue suggested in (Wang, Storey, Firth, 1995) and proposes to assist Information Product Consumers (Data Supervisors or Information Customers) for defining, specifying or evaluating Information Quality Dimensions. It mainly focuses on the implementation issues of Relative Information Quality Dimensions.

We consider that the concept of Information Quality is quadricephalous and we propose a decomposition of Information Quality into the four components presented in Figure 2 : 1) the component concerning the Quality of Data Management by the Information System (in terms of internal data storing, processing...), 2) a component concerning the Quality of Data Representation (modeling, and external views), 3) an intrinsic component : the set of "absolute" characteristics of Data Quality (accuracy, believability...), and 4) a threefold Relative Data Quality component, including a first subcomponent relative to users, a second contextual subcomponent : the set of characteristics depending on a frame of reference such as time (variability, timeliness...) or knowledge state (completeness...), or depending on the application context, and the third Relative Data Quality subcomponent is relative to the quality of other homologous data extracted from different sources. This last subcomponent is evaluated by comparing and cross-checking data which describe the same reality.

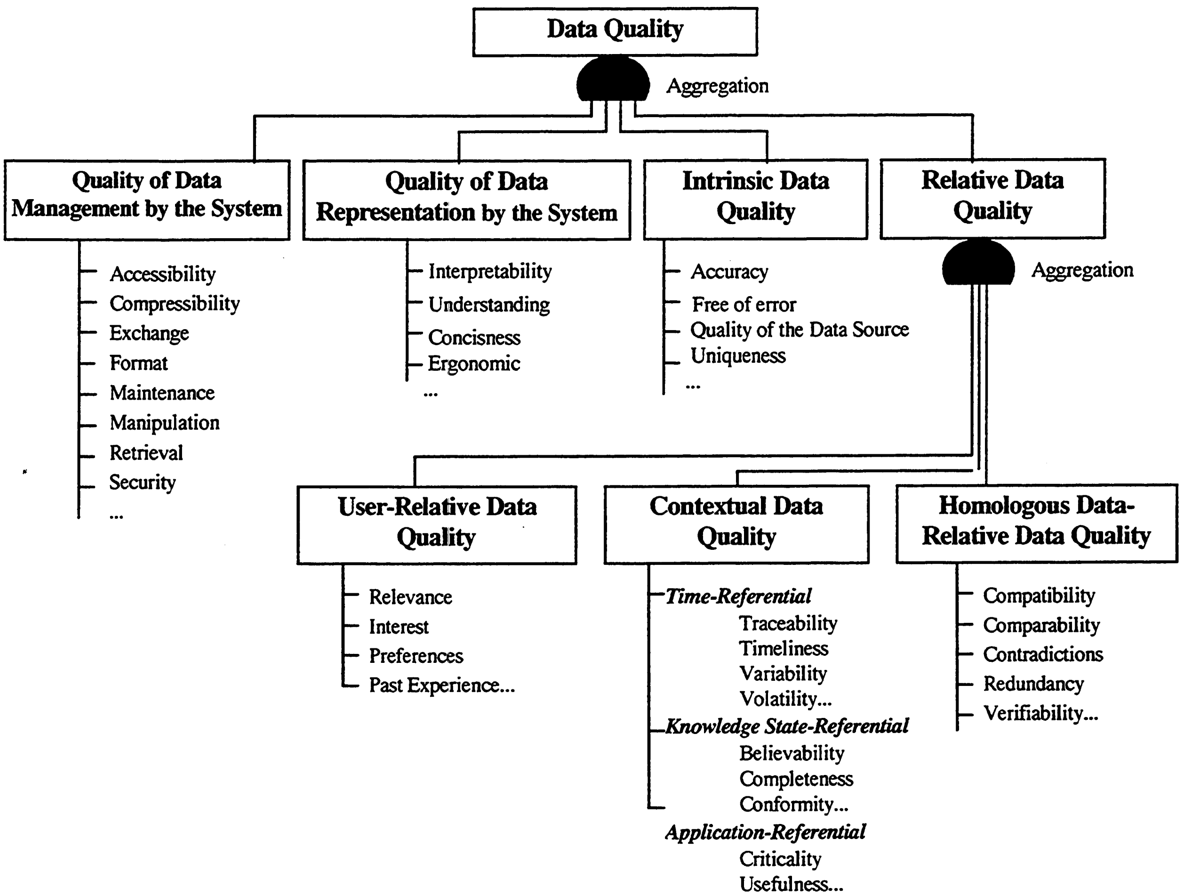


Figure 2. Aggregation of Data Quality Categories

Because we consider that Relative Information Quality is crucial for multi-source data and multi-user requirements, we focused on characterizing this vague and subjective notion as an important category of Information Quality. Some of the Information Quality Dimensions cannot be precisely and rigorously define and agreed universally. Although proposed definitions are relevant and operational in many cases, they may be understood and interpreted differently in accordance with the current practice. Actually, Data Quality criteria definitions are unequivocal : their meaning depends on the context or the task at hand. We believe that each shade of meaning could eventually be better specified by users with tools which should facilitate the non-intrusive acquisition of user profile concerning the Relative Information Quality. Consequently, our objective was to develop user-centered tools and to adapt as far as possible specific mechanisms and interfaces for high-quality information recommendation in the context of multi-source and contradictory data.

3.2. Techniques for Characterizing User-Relative Information Quality

One of the (much debated) orientation of our project concerns our work with specialists in Ergonomical Psychology. First, our approach was to adapt, to implement different methods used in Social Sciences and Micro-Psychology and to apply them for defining on-line individual views of Relative Information Quality expected by the user for his punctual (or frequent) tasks. The objective is to build tools for extracting the implicit Information Quality requirements conditioning user satisfaction, via new interfaces for delimiting, characterizing, ranking and evaluating the subjective criteria of Relative Information Quality. The leitmotiv was « To allow the user to "configure" his software environment for what concerns the Quality of Information he wants to receive from Push Information Service Providers ». This approach could be associated to advanced functionalities of some rating and labeling services on the Internet, which propose systems for "flexible" blocking information reception based on labels embedded in the broadcast stream (PICS, Platform for Internet Content Selection (Resnick, Miller, 1996) (Resnick, 1998)). The next subsections will briefly present some of the techniques we are implementing for enabling users to define, delimit, organize subjective IQ Dimensions into a personal hierarchy and evaluate Relative Information Quality with advanced interfaces currently under development with Tcl/Tk.

Attributes Constellation

This technique enables the user to graphically represent spontaneous (or forced) mental associations concerning the Relative Information Quality concept. Each user defines (or chooses within a list) the most important Information Quality criteria. The Attributes Constellation is a diagram built for a targeted audience. About 40 persons are interviewed and they associate different concepts with the central concept. In the Figure 3, the associated concepts (attributes) are located at the distance $d = 1/(1+\log(P_{ij}))$ from the central concept of Relative Information Quality. Pij is the probability of occurring concept associations. The size of the attribute circles may be chosen in taking into account the psychological importance of each attribute.

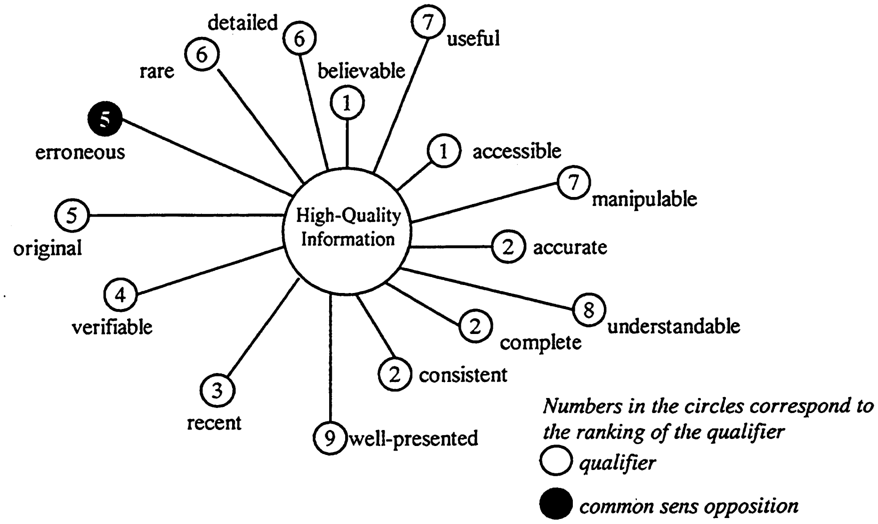


Figure 3. Attributes Constellation for High-Quality Information Concept

Analysis of Correspondences Matrix

This method consists in establishing the degree of correspondence between variables two by two : the similarity matrix is exploited by factor analysis. Generally, impact, similarity or correlation between concepts are expressed in a Correspondences Matrix in a clearer way than the precise definitions of these concepts. The principle of the graphic method uses the visual artifice in taking advantage of the imprecision and vagueness of the vision : the size of a circle will express the importance of the correspondence on a 3, 5, or 7 degrees scale. This method can also enable the recognition of latent patterns.

Source / Information	Original Nationality	Confidentiality	Specialization	Descriptive Breadth	Descriptive Depth
Accuracy	∘	○	◯	◯	◯
Believability	○	○	○	◯	○
Objectivity	○	○	∘	○	∘
Traceability	○	∘	○	∘	∘
Completeness	∘	○	◯	◯	◯
Validity	○	◯	◯	∘	∘

Scale with 3 degrees: ∘ ○ ◯

Figure 4. Generic Correspondences Matrix Between 5 Characteristics of Information Sources and 6 Quality Criteria of their Information Content

Trade-off Method

The trade-off method is based on putting someone in a position to choose between several possibilities (2, 3, 4...) to which he has to allocate a global weight of resource. In position of "mental tension", he has to " sacrifice " some elements in the list ; hence, a " micro-anguish " which may generate a kind of creativity.

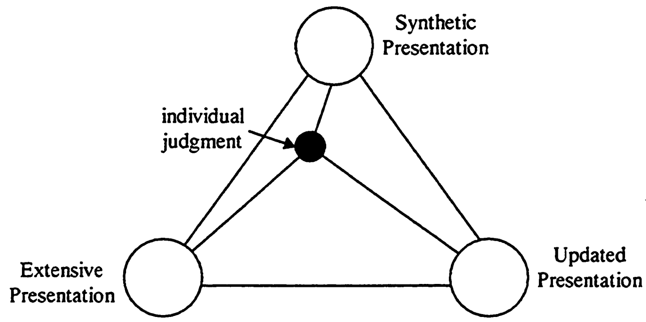


Figure 5. Trade-off for the Best Visualization of an item of Information

Method of Symbolic Equations

The method of symbolic equations is based on the notion which the observant (user) may feel about the existing relations between the subjective concept dimensions. His "feeling" may be sometimes clearer, more pregnant than his ability to define precisely each dimension. He is able to catch the relations of proportional, exponential growth... better than the dimensions that he puts in relation. This fragmentary knowledge may be made more concrete within *symbolic equations*, by allocating algebraic letters for naming each dimension and by writing their "sensed" relations. This perilous method, which is mainly based on induction and intuition, is an experimental method for searching rather than a rational method. In the context of Knowledge Engineering, we use it for extracting "implicit" rules and patterns concerning Relative Information Quality.

For example : The propensity to Communicate an information item (C) varies proportionally to the critical Importance of the information (I) and to its "Privatization" (P) and inversely proportionally to the sum of the Access cost (A) and the Broadcasting cost (B). This subjective rule can be presented by the following symbolic equation :

$$C = \frac{I \times P}{A + B}$$

Semantograms

Semantograms are the graphic presentations of concepts on a paper sheet or on the screen by mean of circles with the name of each concept. The user places them by according to his own "mental configuration" which "copies" a certain image of the connections that these concepts may have together, with a certain kind of neighborhood or relative distances. This configuration represents a "mental world" whose units of measure are not precisely defined, but which really has topological relations : neighborhood, distance, overlapping, number of necessary steps to go from the concept A to the concept Z through other concepts, notion of detours, size of concepts (very or not important concepts). The way that the user has linked concepts together may present them in a didactic chain or explanation chain. The respective coloration of concepts may involve a symbolism which is not necessarily arbitrary or aesthetic, but which may suggest latent factors of expressiveness. A semantogram can be used for defining or completing Meta-Data Specifications because it can show and express the particular data quality perspective of a Data Supervisor. It can make understand his data quality vision and evaluation.

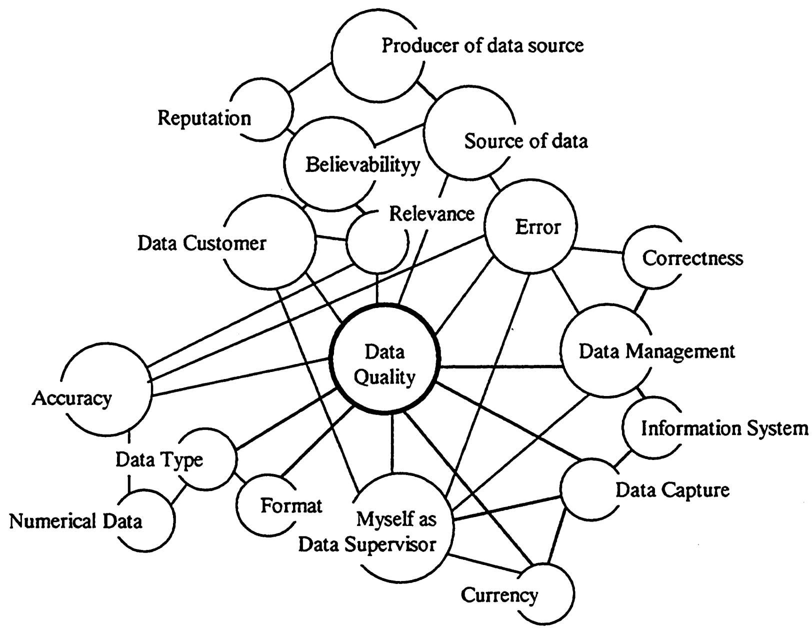


Figure 6. A Data Quality-Centered Semantogram made by the Data Supervisor

We admit that such schemas have a considerable part of subjectivity, and they are easy to refute, but they also are very rich in inspiration for those who are using them and those who are looking at them.

4. Importance of Information Quality in Technological Watch

4.1. Application Domain Description

In the context of Technological Watch, our approach is in the continuation of the VIGIWARE project we described in (Berti, 1997), and it is mainly related to user profile acquisition and personal information filtering and recommending. The Information Chain hinges on four preponderant steps carried out by experts with the role of Data Supervisors :

- selecting Information Sources such as technical reports, proceedings, commercial brochures... diffused by competent Information Producers. The selection of producers is based on very subjective high-quality criteria : credibility, good reputation to keep, interest to disclose information, domain competence, substantial financial supports, dependencies between Information Sources...

- collecting dynamically relevant information,
- cross-checking, valuing information and Information Quality assessing and tagging,
- recommending certified and value-added information to Information Customers (decision makers).

In the final stage of Information Chain, the Technological Watch Group, stamps the information with a certain evaluation degree for each subjective Information Quality criterion. Doing so, the stakes of Information Quality tagging are important for Information Producers and Consumers, and the deontology of Information Service Providers is all the more necessary.

Beyond the development of an Information System specifically dedicated to Technological Information Push, the first underlying objectives for providing Technological Watch assistance are :

- defining a conceptual framework by identifying the primary users' requirements of data handling for Technological Watch,
- identifying the sources, nature and content of *information* necessary to develop information inputs to and outputs from,
- identifying the sources nature and content of *knowledge* necessary to develop "intelligent" functionalities for Information System and assistant tools.

Contextual Information Quality with the Application Referential

The nature of Technological Information is complex : textual, informal, vague, incomplete, rare or overabundant, contradictory, non authenticated. Consequently, the value of Technological Information has an important part of subjectivity due to evaluation made by experts as Data Quality Supervisors. The main constraint is their necessary omnipresence during the data collection, capture and information recommending. Thus, the Information System and assistant tools must support an intensive and efficient experts/system interaction.

Multi-source Information Quality

Because the VIGIWARE system is intended to provide personal adaptive information analysis and recommendation through dissemination of news and relevant Information Sources, we necessarily consider multiple Information Sources and we need to cross-check, compare, aggregate homologous and often contradictory items of information. More information leads to better inferences and decisions. Actually, the truth doesn't exist in itself but it is built. Evaluating Relative Information Quality by an iterative comparison process is an interesting way to do so.

4.2. Development and Extension of the VIGIWARE System

Characterizing the Relative Information Quality and assisting Information Quality evaluation was our first motivation for the development of interfaces implementing user-oriented methods. Satisfying Information Customers was the second one. Interface and tools we are currently developing involve some of the previously presented methods and techniques and need to be validated with the demonstration of their viability. That constitutes the convergence point of our combined efforts for extending the functionalities of the VIGIWARE system.

5. Conclusions

The main idea of this paper was to stress that the relative aspect of Information Quality and to propose methods and techniques to characterize it. This idea is different from most of the views expressed in the Data and Information Quality literature and it may be much debated, and, thus, very risky. In the context of Technological Information Push Service, driven by the Customer Voice, we need to explore new research orientations of Information and Knowledge Engineering specifically applied to the relative aspect of Information Quality. Needless to say, that the only ambition is here to propose Information Quality assistance tools for users : Data Supervisors and Information Customers.

This work has been supported by AERO (France) under contract PA/2594.0107.

6. References

(Berti, 1997) Berti, L., "Out of Overinformation by Information Filtering and Information Quality Weighting", Proceedings of 1997 Conference on Information Quality, (1997), pp. 187-193.

(Ballou, Tayi, 1989) Ballou, D. P., Tayi, K. G., "Methodology for Allocating Resources for Data Quality Enhancement", Communications of ACM, 32, 3, (1989), pp. 320-329.

(CEN, 1996) European Committee for Normalization, Geographic Information - Data Description - Quality, Draft, (1996)

(Delen, Rijsenbrij, 1992) Delen, G. P. A., Rijsenbrij, D. B. B., "A Specification, Engineering and Measurement of Information Systems Quality", Journal of Systems Software, 17, 3, (1992), pp. 205-217.

(Dvir, Evans, 1996) Dvir, R., Evans, S., "A TQM Approach to the Improvement of Information Quality", Proceedings of 1996 Conference on Information Quality, (1996).

(FGDC, 1997) Federal Geospatial Data Committee, "Content Standard for Digital Geospatial Metadata", (1997), http://www.fgdc.gov/Metadata/Metadata.html.

(Fisher and al., 1995) Fisher, P. F., Maguire, D., Goodchild, M. F., Rhind, D. W., "Spatial Data Sources and Data Problems", Geographical Information Systems, vol. 1, chap. 13, Longman Scientific & Technical, (1995), pp. 175-189.

(Fowler, 1995) Fowler, J., "STEP for Data Management, Exchange and Sharing", Technology Appraisals, (1995), http://www.pdt1.demon.co.uk/standard/step.

(Gardyn, 1997) Gardyn, E., "A Data Quality Handbook for a Data Warehouse", Proceedings of 1997 Conference on Information Quality, (1997), pp. 267-290.

(Günther, Voisard, 1997) Günther, O., Voisard, A., Klas, W., Sheth, A., "Metadata in Geographic and Environmental Data Management", McGraw Hill, (1997), pp. 1-31.

(ISO, 1997) International Standard Organisation, ISO Standard 15046-15, Geographic Information - Metadata, Second Draft, (1997).

(Meyen, Willshire, 1997) Meyen, D., Willshire, M. J., "A Data Quality Engineering Framework", *Proceedings of 1997 Conference on Information Quality*, (1997), pp. 95-116.

(Morrison, Guptill, 1995) Morrison, J. L., Guptill, S. C., Morrison, J. L., "Elements of Spatial Data Quality", International Cartographic Association (ICA), Commission on Spatial Data Quality, Pergamon, (1995).

(NIST, 1994) National Institute of Standards and Technology, Federal Information Processing Standard Publication 173, (1994).

(OpenGIS, 1997) Gardels, K., "OpenGIS and On-line Environmental Libraries", SIGMOD Record, ACM Special Interest Group on Management of Data, 26, 1, (1997), pp. 32-38.

(Paradice, Fuerst, 1991) Paradice, D. B., Fuerst, W. L., "An MIS Data Quality Methodology Based on Optimal Error Detection", Journal of Information Systems, (1991), pp.48-66.

(Redman, 1996) Redman, T. C., "Data Quality for the Information Age", Artech House Publishers, (1996).

(Resnick, Miller, 1996) Resnick, P., Miller, J., "PICS : Internet Access Controls Without Censorship", Communications of the ACM, 39, 10, (1996), pp. 87-93.

(Resnick, 1998) Resnick, P., "W3C Platform for Internet Content Selection", (1998), http://www.w3.org/PICS/.

(Strong, 1998) Strong, G., "Human-Centered Systems Research and Development in the United States", (1998), http://www.cise.nsf.gov/iis/hci_home.html.

(Strong, Wang, 1994) Strong, D., Wang, R. Y., "Beyond Accuracy : What Data Quality Means to Data Consumers", *Tech. Rep. TDQM-94-10*, Massachusetts Institute of Technology, (1994).

(Wang, 1998) Wang, R. Y., "A Product Perspective on Total Data Quality Management", Communications of ACM, 41, 2, (1998), pp. 58-65.

(Wang, Kon, Madnick, 1993) Wang, R. Y., Kon, H. B., Madnick, S. E., "Data Quality Requirements Analysis and Modeling", Proceedings of the Ninth International Conference on Data Engineering, (1993), pp. 670-677.

(Wang, Storey, Firth, 1995) Wang, R. Y., Storey, V. C., Firth, C. P., "A Framework for Analysis of Data Quality Research", IEEE Transactions on Knowledge and Data Engineering, 7, 4, (1995), pp. 623-640.

(Zmud, 1978) Zmud, R., "Concept, Theories and Techniques : An Empirical Investigation of the Dimensionability of the Concept of Information", Decision Design, 9, 2, (1978), pp. 187-195.