INTRODUCING DATA QUALITY IN A COOPERATIVE CONTEXT

Paola Bertolazzi
Istituto di Analisi dei Sistemi ed Informatica
Consiglio Nazionale delle Ricerche (IASI-CNR)
Viale Manzoni 30, 00185 Roma, Italy
bertola@iasi.rm.cnr.it
Monica Scannapieco
Dipartimento di Informatica e Sistemistica
Università di Roma "La Sapienza"
Via Salaria 113, 00198 Roma, Italy
monscan@dis.uniroma1.it

Abstract: Cooperative Information Systems are defined as numerous, diverse systems, distributed over large, complex computer and communication networks which work together to request and share information, constraints, and goals. The problem of data quality becomes crucial when huge amounts of data are exchanged and distributed in such an intensive way as in these contexts. In this paper we make some proposals on how to introduce data quality into a cooperative environment. We define some specific quality dimensions, we describe a conceptual data quality model to be used by each cooperative organization when exporting its own data, and we suggest some methodologies for the global management of data quality.

1 Introduction

With the explosion of e-Business, Cooperative Information Systems (CIS's) ([18], [3]) are becoming more and more important in all the various relationships among businesses, governments, consumers and citizens (B2B, B2C, C2C, B2G, G2B, etc.). By use of a CIS, autonomous organizations, sharing common objectives, can join forces to overcome the technological and organizational barriers deriving from their different and independent evolutions.

In order to make cooperation possible, each organization has to make available its own data to all other potential collaborators. One possible way is that organizations agree on a common set of data they wish to exchange and make them available as *conceptual schemas* that are understood and can be queried by all cooperating organizations ([14], [12]). Technological problems deriving from the heterogeneity of the underlying systems can be overcome by using component-based technologies (such as OMG Common Object Request Broker Architecture [20], SUN Enterprise JavaBeans Architecture [16] and Microsoft Enterprise .NET Architecture [23]) to realize access to data exported by these schemas.

In a CIS it is imperative to deal with the issue of data quality, both to control the negative consequences of poor cooperative data quality, and to take advantage of cooperating characteristics to improve data quality. In fact, exchanges of poor quality data can cause a huge spread of data deficiencies among all the cooperating organizations. However, CIS's are characterized by replication of their data across different organizations. This replication can be used as an opportunity for improving data quality, by comparison of the same data at each organization.

The aim of this paper is to give some methodological suggestions to introduce data quality in a cooperative context. In our vision, organizations export not only conceptual models of their data, but also conceptual models of the *quality* of such data, therefore giving rise to many opportunities.

This paper is organized as follows. Section 2, after a brief review of the state of the art concerning data quality, introduces and defines the data quality dimensions we consider most relevant in a cooperative environment. Section 3 proposes a conceptual data quality model that can be exported by cooperative organizations. Section 4 considers a possible tailoring of the TDQM cycle [29] to a cooperative context and in Section 5 we illustrate an application scenario, the e-Government Italian initiative, which provides motivations for our work and the test bed in which we will test our approach. Section 6 concludes the paper with possible future work areas.

2 Data Quality

2.1 Related Work

The notion of *data quality* has been widely investigated in literature; among the many definitions we cite those of data quality as “fitness for use” [28] and as “the distance between the data views presented by an information system and the same data in the real world” ([21], [26]). The former definition emphasizes the subjective nature of data quality, whereas the latter is an “operational” definition, although defining data quality on the basis of comparisons with the real world is a very difficult task.

We here consider the concept of data quality as defined by a set of *dimensions*, usually considered in data quality literature as quality properties or characteristics of data (e.g. accuracy, completeness, consistency).

Many definitions of data quality dimensions have been proposed, including the identification of four categories (regarding intrinsic, contextual, representation and accessibility data aspects) for data quality dimensions [28], and the taxonomy proposed in [22], in which more than twenty data quality dimensions are classified into three categories, namely conceptual view, values and format. A survey of data quality dimensions is given in [27].

We will inherit some dimensions already proposed in literature, and we will introduce some new quality dimensions, specifically relevant in cooperative environments.

Data quality issues have been addressed in several research areas, e.g. quality management in information systems, data cleaning, data warehousing, integration of heterogeneous databases and web information sources. Based on the analogy between data and manufacturing products an extension of Total Quality Management (TQM) to data is proposed in [29]: Total Data Quality Management (TDQM). Four phases are recognized as necessary for the managing of the Information Product (IP): definition, measurement, analysis and improvement. In this last the Information Manufacturing Analysis Matrix [1] can be used. We here consider the TDQM approach and its extension to CIS.

To our knowledge, many aspects concerning data quality in CIS have not yet been addressed; however, when dealing with data quality issues in cooperative environments, some results already achieved for traditional and web information systems can be "borrowed". In CIS’s, the main data quality problems are:

- Assessment of the quality of data exported by each organization.
- Methods and techniques for exchanging quality information.
- Improvement of quality.
- Heterogeneity, due to the presence of different organizations, in general with different data semantics.

Results achieved in the data cleaning area ([6], [9], [7]), and the data warehouse area ([25], [9]) can be adopted for the Assessment phase. Heterogeneity has been widely addressed in literature, especially with respect to schema integration issues ([2], [8], [24], [11], [4]).

Quality improvement and methods and techniques for exchanging quality information have been only partially addressed in literature (e.g., [15]). This paper particularly addresses the exchange of quality information by proposing a conceptual model for such exchanges, and makes some suggestions on quality improvement based on the availability of quality information.

2.2 Data Quality Dimensions

Two categories for data quality dimensions can be distinguished. *Intrinsic data quality dimensions* characterize properties inherent to data, i.e., which depend on the very nature of data. *Process specific data quality dimensions* describe properties that depend on the cooperative process in which data are exchanged.

2.2.1 Data Intrinsic Dimensions

Only the most important dimensions [26] and those we consider most relevant in a cooperative environment are discussed here. These are:

- accuracy,
- completeness,
- currency,
- internal consistency.

Standard literature definitions for these are assumed (e.g. [22]).

2.2.2 Process Specific Dimensions

The need for context-dependent data quality dimensions has been recognized [28]. In CIS, the cooperative process provides the context and data quality dimensions are related to data evolution in time and within the process. We have therefore chosen and adapted some of the dimensions proposed in [28] (timeliness and source reliability), and in addition propose a new dimension dependent on the specificity of our context (importance).

Process specific dimensions are tied to specific data exchanges within the process, rather than to the whole process. Hence, in the following definitions, we consider a *data exchange* as a triple `<source organization i, destination organization j, exchange id>`, representing the cooperating organizations involved in the data exchange (i.e. source and destination organizations) and the specific exchange[1]. Moreover, in the following, we will refer to *schema element* meaning, for instance, an entity in a Entity-Relationship schema or a class in an object oriented schema expressed in Unified Modeling Language (UML) [19].

[1] The `exchange id` has the role of identifying a specific data exchange between two organizations, as they may be involved in more than one exchange of the same data within the same cooperative process.

Timeliness

- ⇒ *The availability of data on time, that is within the time constraints specified by the destination organization.*

For instance, we can associate a low timeliness value for the schedule of the lessons in a University if such a schedule becomes available on line after the lessons have already started. To compute this dimension, each organization has to indicate the *due time*, i.e., the latest time before which data have to be received. According to our definition, the timeliness of a value cannot be determined until it is received by the destination organization.

Importance

- ⇒ *The significance of data for the destination organization.*

Consider organization B, which cannot start an internal process until organization A transfers values of the schema element X; in this case, the importance of X for B is high.

Importance is a complex dimension whose definition can be based on specific indicators measuring:

- the number of instances of a schema element managed by the destination organization with respect to a temporal unit;
- the number of processes internal to the destination organization in which the data are used;
- the ratio between the number of core business processes using the data and the overall number of internal processes using the data.

Source Reliability

- ⇒ *The credibility of a source organization with respect to provided data. It refers to the pair* `<source, data>`.

The dependence on `<source, data>` can be clarified through an example: the source reliability of the Italian Department of Finance concerning `Address` of citizens is lower than that of the City Councils; whereas for `SocialSecurityNumber` its source reliability is the highest of all the Italian administrations.

3 Data and Data Quality Models

3.1 Data Model

All organizations involved in a CIS need to export their data according to some specific schemas; these are referred to as *cooperative data schemas*.

These are class schemas defined in accordance with the ODMG Object Model [5]. Specifically they describe types of exchanged data items, wherein types can be:

- classes, whose instances have their own identities;
- literals, when instances have no identities, and are identified by values.

New classes can be defined as collections of objects (as instances are objects) or as structured literals, as a record of literals.

3.2 Data Quality Model

This describes the conceptual data quality model that each cooperating organization must define in order to export the quality of its own data.

A *cooperative data quality schema* is a UML class diagram associated to a cooperative data schema, describing the data quality of each element of the data schema. It can be divided into two types, intrinsic and process specific, described in the following sections.

3.2.1 Intrinsic Data Quality Schemas

Intrinsic data quality dimensions can be modeled by considering specific classes and structured literals called here *dimension classes* and *dimension structured literals.*

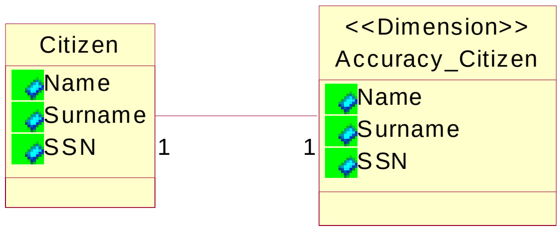

Figure 1. Example of an intrinsic data quality schema.

Each data quality dimension (e.g., completeness or currency) is modeled by a specific class or structured literal. These represent the abstraction of the values of a specific data quality dimension for each of the attributes of the data class or of the data structured literal to which they refer, and to which they have a one-to-one association.

A dimension class (or a dimension structured literal) is represented by a UML class labeled with the stereotype `<<Dimension>>` (`<<Dimension_SL>>`), and the name of the class should be `<DimensionName_ClassName>` (`<DimensionName_SLName>`).

An *intrinsic data quality schema* is a UML class diagram, the elements of which are: dimension classes, dimension structured literals, the data classes and data structured literals to which they are associated and the one-to-one associations among them.

Consider the class `Citizen`. This may be associated to a dimension class, labeled with the stereotype `<<Dimension>>`, the name of which is `Accuracy_Citizen`; its attributes correspond to the accuracy of the attributes `Name`, `Surname`, `SSN`, etc. (see Figure 1).

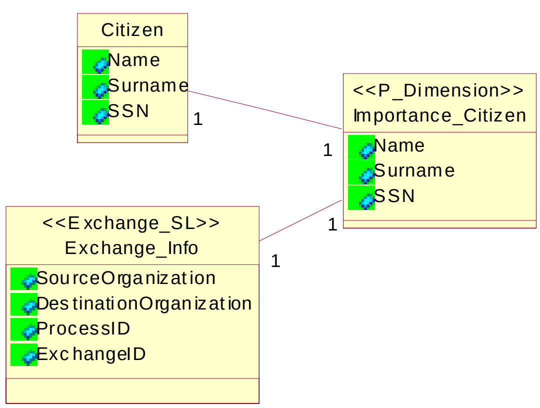

Figure 2. Example of a process specific data quality schema.

3.2.2 Process Specific Data Quality Schemas

Tailoring UML in a way similar to that adopted for intrinsic data quality dimension, we introduce *process dimension classes* and *process dimension structured literals*, which represent process specific data quality dimensions, just as dimension classes and dimension structured literals represent intrinsic data quality dimensions.

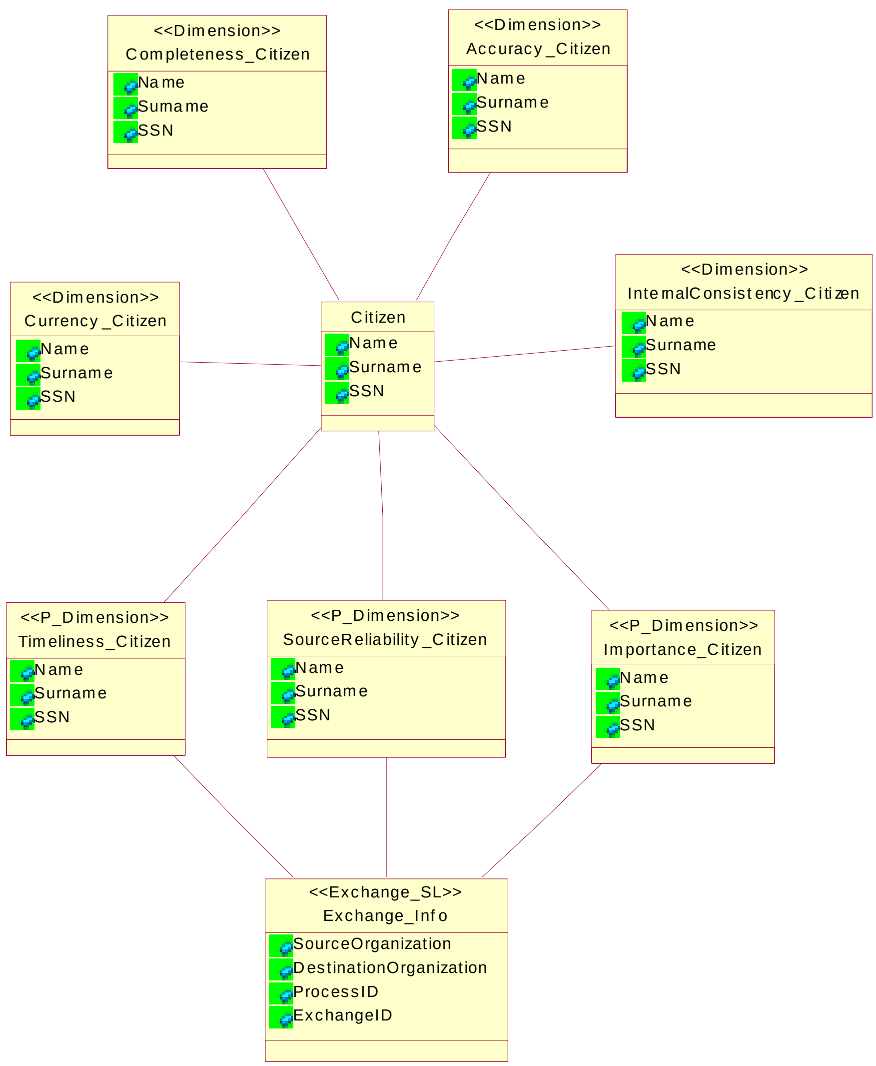

Figure 3. A cooperative data quality schema referring to the `Citizen` class. All the associations are 1-ary.

Process dimension classes and literals are represented by the UML stereotypes `<<P_Dimension>>` and `<<P_Dimension_SL>>`. The name of the class should be `<P_DimensionName_ClassName>` (`<P_DimensionName_SLName>`).

Also necessary is an *exchange structured literal* to characterize process dimension classes (and structured literals). As described in Section 2.2, process specific data quality dimensions are tied to a specific exchange within a cooperative process. This kind of dependence is represented by exchange structured literals. They include the following mandatory attributes:

- source organization,
- destination organization,
- process identifier,
- exchange identifier.

Exchange structured literals are modeled as UML classes stereotyped by `<<Exchange_SL>>`.

A *process specific data quality schema* is a UML class diagram, the elements of which are: process dimension classes and structured literals, the classes and structured literals to which they are associated, exchange structured literals and the associations among them. Figure 2 gives an example.

The considerations discussed in this section are summarized in Figure 3, in which a cooperative data quality schema describes the quality of both intrinsic and process specific dimensions for the `Citizen` class: the intrinsic data quality dimensions (accuracy, completeness, currency, internal consistency) are labeled with the stereotype `<<Dimension>>`, whereas the process specific data quality dimensions (timeliness, importance, source reliability) are labeled `<<P_Dimension>>`, and are associated to the structured literal `Exchange_Info`, labeled `<<Exchange_SL>>`.

4 TDQM$_{CIS}$: a Cycle for Quality Treatment in Cooperative Environments

The Total Data Quality Management (TDQM) cycle has been proposed with the aim of providing users with high data quality by considering data as a manufactured product [29]. In this section we show the first steps towards a tailoring of the TDQM cycle to cooperative environments. The TDQM cycle consists of the following phases:

- definition - the identification of data quality dimensions and of the related requirements;
- measurement - producing quality metrics. These provide feedback to data quality management and allow the comparison of the effective quality with pre-defined quality requirements;
- analysis - identifying the roots of quality problems and then studying their relationships;
- improvement - information quality improvement techniques.

These four phases have been redesigned in the context of CIS's, giving rise to the *cooperative TDQM cycle (TDQM$_{CIS}$)*, applicable in the practical cases deriving from the Italian e-Government initiative described in Section 5.

There are five phases to the *(TDQM$_{CIS}$)*: Definition, Measurement, Exchange, Analysis and Improvement. They are illustrated in Figure 4. Like the TDQM cycle, TDQM$_{CIS}$ is a continuous cycle, in the sense that it must be applied in an iterative way.

4.1 TDQM$_{CIS}$ Definition

In the TDQM cycle the Information Product (IP) is defined at two levels: its functionalities for information consumers and its basic components, represented by the Entity-Relationship schema.

In the TDQM$_{CIS}$ cycle quality data is associated to an IP and specified in terms of intrinsic and process specific dimensions. Both IP's and their associated quality data need to be exported by each cooperative organization through cooperative data and quality schemas, as described in Section 3.

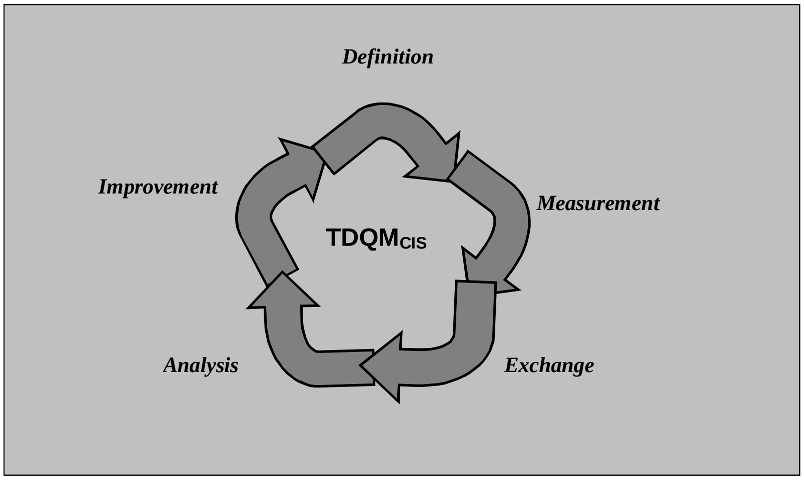

Figure 4.The phases of the TDQM$_{CIS}$ cycle.

What organizations have to export is driven by the cooperative requirements of the processes they are involved in. The definition phase therefore also needs to:

- model cooperative processes;
- specify cooperative requirements in terms of what data must be exported and what quality information is needed in each cooperative process.

Note that our focus is a business-to-business context, in which the consumers of exported data are members of the same CIS as the exporter.

4.2 TDQM$_{CIS}$ Measurement

Two measurement types are made:

- Static: source reliability and all intrinsic data quality dimensions are measured statically, i.e. each cooperating organization assesses the quality of its data once using traditional methods (for example the statistical methods proposed in [17]. Data quality values must be computed with respect to the conceptual specification of the defined cooperative data quality schemas. There should also be a general agreement on the metric scales used for data quality dimension measurements.
- Dynamic: only timeliness is measured dynamically, i.e. during execution of the cooperative process. To calculate timeliness each organization must indicate the due time, as described in Section 2.2.2.

The importance dimension is not measured at all: values must be specified by each organization, on the basis of how important exchanged data are for the cooperative process. Moreover, importance is used to evaluate data quality measurements of the other dimensions, as it will be clarified in the description of the analysis phase.

4.3 TDQM$_{CIS}$ Exchange

This phase is additional to the standard TDQM cycle. It is related to the quality of data exchanged among cooperating organizations and includes the exact definition of a transferred unit (TU). Quality data to be transferred include intrinsic dimension values and source reliability values. Importance and timeliness are calculated by the destination organization. With respect to the specified data and quality conceptual models, we distinguish the following types of TU:

- **Type a**: a single attribute value X with its associated quality data, consisting of the values of all the data quality dimensions calculated in the static measurement phase (see Figure 5, which must be completed with the values of the dimensions for X).

Attribute Value	Accuracy Value	Completeness Value	Currency Value	Internal consistency Value	Source reliability Value
X	..	..	..	..	..

Figure 5.Transferred Unit of type a.

- **Type b:** a composite (i.e. multi-attribute) unit with its associated quality. Note that our conceptual model makes a distinction between classes and literals, but the composite unit effectively transferred includes a class instance with all the associated literal instances. Quality data include the values of all the transferred data quality dimensions for each of the attribute values of the composite unit. In Figure 6, the type b TU related to a composite unit including three attribute values (X,Y,Z) is shown.

Attribute Value	Accuracy Value	Completeness Value	Currency Value	Internal consistency Value	Source reliability Value
X	..	..	..	..	..
Y	..	..	..	..	..
Z	..	..	..	..	..

Figure 6.Transferred Unit of type b.

4.4 TDQM$_{CIS}$ Analysis

This phase differs from its correspondent in the TDQM cycle, as an analysis step is introduced during the execution of the cooperative process. In particular we distinguish:

- the analysis phase of organization A which sends the TU and
- the analysis phase of organization B, which receives the TU.

A's analysis is similar to the classical analysis phase in the TDQM cycle: the internal processes are analyzed and the causes of poor quality are determined. B's analysis phase is discussed in detail below.

4.4.1 Destination Organization Analysis

This is performed during the execution of a cooperative process. Organization B receives from A a TU including the values of the intrinsic data quality dimensions and the source reliability.

B has three tasks:

- Calculate timeliness as the difference between the due time and the arrival time.
- Interpret the TU's intrinsic quality values. We evaluate dimension values, such as accuracy or completeness, on the basis of importance and source reliability. All intrinsic

data quality values can be weighted with their associated importance and source reliability values, using a weighting function chosen by organization B. For instance, a "low" source reliability for an attribute X of the TU should be weighted with the result of a "high accuracy" for X's value. The evaluation of timeliness is affected only by importance - source reliability is not relevant. If importance is "high" but the data are not delivered in time, they will be probably discarded by the receiving organization. The interpretation and evaluation phases of TU data quality may also include the calculation of data quality values for the entire TU, starting from the values of dimensions related to each of attributes included in the TU. Though this problem is not in the scope of this paper, we can say that with DIM being a specific dimension, TU a transferred unit, and xi an attribute of TU, the quality value of the dimension DIM for TU is a function F of the value of DIM for each attribute xi of TU, that is:

$$Q_{TU}\ (DIM) = F_{xi \in TU}\ (DIM)$$

For each dimension a particular function F can be chosen. We also observe that on the basis of this analysis B can choose to accept or reject the TU.

As an example let us consider the `Citizen` class with the attribute `Name`, `Surname`, `SSN`, shown in Figure 1. If we consider an "average" function for accuracy and a "boolean-and" function for completeness, we have:

$Q_{Citizen}(Accuracy) = AVERAGE_{\ Name,Surname,SSN}\ (Accuracy)$

$Q_{Citizen}(Completeness) = BOOLEANAND_{\ Name,Surname,SSN}\ (Completeness)$.

- Send a TU to another organization. B's analysis phase introduces an activity typical of cooperating systems where an organization is at the same time both a consumer and a producer of an IP. In this case B receives a TU X from A, performs a task based on X and then sends a new TU Y to C (see Figure 7).

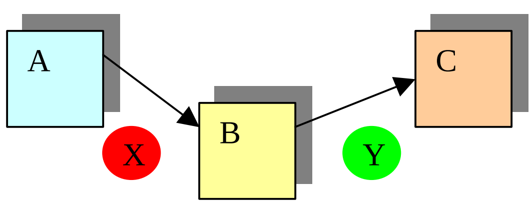

Figure 7.Cooperative exchanges among three different organizations.

Y can be seen as "derived" from X in some way. If the more general case in which both X and Y are type b TU's is considered, the following cases can be distinguished:

1. *B sends X to C without modifying it (*Y *=X).* If B had specified a due time for B then a value of the timeliness is calculated. All other quality dimension values remain unchanged and are sent to C.
2. *B changes some attribute values and sends* Y *to C.* Here we consider only one attribute value change; other cases can be easily reduced to this. In this case, let `X.ai` be the changed attribute. For each intrinsic quality dimension, except consistency, the values previously calculated by B in the measurement phase are replaced.

Consistency must be recalculated as we consider an internal type of consistency[2]. The value for source reliability must be changed from that of A to that of B.

3. *B uses X to produce* Y *and sends* Y *to C.* Y is a different TU, so B must calculate the values of all the transferred data quality dimensions. In relation to the possible ways of calculating these attributes, we can distinguish the following cases:
   - If an attribute of Y is obtained by arithmetic operations starting from attributes of X, possible ways of combining the values of the quality for the different dimensions are proposed in [1].
   - If the value of an attribute Y.ai is extracted from a database of B on the basis of the attribute value X.ai then:
     - The accuracy of Y.ai depends on the accuracy of X.ai, with respect to semantic aspects [3].
     - All other data quality dimension values are known from the measurement phase.

4.5 TDQM$_{CIS}$ Improvement

A cooperative environment offers many opportunities for actions that can improve the quality of data shared by cooperating organizations and exchanged in cooperative processes.

The data quality measurement phase enables organizations to understand and address the quality weaknesses of their data on the basis of a comparison with the quality of the same data owned by other organizations. As already observed, the quality of the data held by an organization must be "filtered" according to the source reliability dimension. Source reliability values of cooperating organizations may be set by a *source reliability manager*, which might be one of the CIS members or an external organization. Its main role should be to certify the source reliability of each data exchange within a cooperative process and supply such information to the destination organization on request.

Some other improvements can be made on the basis of the analysis phase performed by receiving organizations. Evaluating and interpreting the quality of delivered TU's gives the opportunity of sending accurate feedback to the source organizations, which can then implement corrective actions to improve their quality.

Another important opportunity for improvement derives from the dynamic evaluation of timeliness during cooperative process executions. It may be possible to trace the timeliness values for each of the organizations involved in a specific process execution, thus identifying the most critical exchanges with respect to the timeliness of the whole process.

5 An e-Government Application Scenario

The approach presented in this paper will be validated in the Italian e-Government initiative [13]. In Italy, in 1993, the Italian Parliament created the Authority for IT in Public Administration (Autorità per l'Informatica nella Pubblica Amministrazione, AIPA) with the aim of promoting

[2] Consistency implies that two or more values do not conflict with one other. By referring to internal consistency we mean that all values compared to evaluate consistency are internal to a specified schema element.

[3] Semantic accuracy can be seen as the proximity of a value `v` to a value `v'` with respect to the semantic domain of `v'`; we can consider the real world as an example of an semantic domain. For example if `X.ai` is the key to access to `Y.ai`, it may cause access to an instance different from the semantically correct instance: if `X.ai=Josh` rather than correctly `X.ai=John`, `Josh` can be a valid key in the database of B so compromising the semantic accuracy of `Y.ai`.

technological progress, by defining criteria for planning, implementation, management and maintenance of information systems of the Italian Public Administration[4]. Among the various initiatives undertaken by AIPA since its constitution, the Unitary Network project is the most important and challenging.

This project has the goal of implementing a “secure Intranet” capable of interconnecting public administrations. One of the more ambitious objectives of the Unitary Network will be obtained by promoting cooperation at the application level. By defining a common application architecture, the Cooperative Architecture, it will be possible to consider the set of widespread, independent public administration systems as a Unitary Information System of Italian Public Administration (as a whole) in which each member can participate by providing services *(*e-Services) to other members ([14], [13]). The Unitary Network and the related Cooperative Architecture are an example of CIS. Similar initiatives are currently undertaken in the United Kingdom, where the e-Government Interoperability Framework *(*e-GIF) sets out the government’s technical policies and standards for achieving interoperability and information systems coherence across the UK public sector. The emphasis of these approaches is on data exchanges, and is therefore focused on document formats (as structural class definitions). The approach presented here aims at introducing a methodology so that organizations can also exchange the quality of their data, and obtain feedback on how data quality can be improved.

6 Concluding Remarks and Future Work

This paper has proposed a possible way to deal with the issue of data quality in a cooperative environment. The importance of introducing specific data quality dimensions was dealt with first. A conceptual modeling language, to represent the quality of the data exported by cooperating organizations was then obtained by the tailoring of the Unified Modeling Language. Finally we discussed how TDQM cycle might be adapted for a cooperative context.

The future directions of our work will principally address a more specific definition of the tailoring of the TDQM cycle, and a validation of our ideas in the context of the Italian e-Government initiatives.

7 Acknowledgements

The authors would like to thank Carlo Batini (AIPA), Barbara Pernici (Politecnico di Milano) and Massimo Mecella (Università di Roma “La Sapienza”) for important discussions about the work presented in this paper.

8 References

[1] Ballou D. P., Wang R. Y., Pazer H., Tayi G. K., Modeling Information Manufacturing Systems to Determine Information Product Quality, Management Science, vol. 44, no. 4, 1998.

[2] Batini C., Lenzerini M., Navathe S.B.: A Comparative Analysis of Methodologies for Database Schema Integration, ACM Computing Survey, vol. 15, no. 4, 1984.

[3] Brodie M.L.: The Cooperative Computing Initiative. A Contribution to the Middleware and Software Technologies. GTE Laboratories Technical Publication, 1998, available on-line (link checked July, 1st 2001): http://info.gte.com/pubs/PITAC3.pdf.

[4] See AIPA’s web site, http://www.aipa.it for details.

[4] 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.

[5] Cattell R.G.G., Barry D. et alii: The Object Database Standard: ODMG 2.0.Morgan Kauffmann Publishers. 1997.

[6] Elmagarmid A., Horowitz B., Karabatis G., Umar A.: Issues in Multisystem Integration for Achieving Data Reconciliation and Aspects of Solutions. Bellcore Research Technical Report, 1996.

[7] Galhardas H., Florescu D., Shasha D., Simon E.: An Extensible Framework for Data Cleaning, in Proceedings of the 16th International Conference on Data Engineering (ICDE 2000), San Diego, California, CA, 2000.

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

[9] Hernadez M.A., Stolfo S.J. : Real-world Data is Dirty: Data Cleansing and The Merge/Purge Problem. Journal of Data Mining and Knowledge Discovery, vol. 1, no. 2, 1998.

[10] Jeusfeld M.A., Quix C., Jarke M.: Design and Analysis of Quality Information for Data warehouses. In Proceedings of the 17th International Conference on Conceptual Modeling (ER'98) , Singapore, 1998.

[11] 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.

[12] Mecella M., Batini C.: Cooperation of Heterogeneous Legacy Information Systems: a Methodological Framework. Proceedings of the 4th International Enterprise Distributed Object Computing Conference (EDOC 2000), Makuhari, Japan, 2000.

[13] Mecella M., Batini C.: Enabling Italian e-Government Through a Cooperative Architecture. In Elmagarmid and McIver 2001.

[14] Mecella M., Pernici B.: Designing Wrapper Components for e-Services in Integrating Heterogeneous Systems. To appear in VLDB Journal, Special Issue on e-Services, 2001.

[15] Mihaila G., Raschid L., Vidal M.: Querying Quality of Data Metadata. In Proceedings of the 6th International Conference on Extending Database Technology (EDBT’98), Valencia, Spain, 1998.

[16] Monson-Haefel R.: Enterprise JavaBeans (2nd Edition). O'Reilly 2000.

[17] Morey R. C.: ‘Estimating and Improving the Quality of Information in the MIS, Communication of the ACM, vol. 25, no. 5, 1982.

[18] Mylopoulos J., Papazoglou M. (eds.): Cooperative Information Systems. IEEE Expert Intelligent Systems & Their Applications, vol. 12, no. 5, September/October 1997.

[19] Object Management Group (OMG): OMG Unified Modeling Language Specification. Version 1.3. Object Management Group, Document formal/2000-03-01, Framingham, MA, 2000.

[20] Object Management Group: The Common Object Request Broker Architecture and Specifications. Revision 2.3. Object Management Group, Document formal/98-12-01,Framingham, MA, 1998.

[21] Orr K.: Data Quality and Systems Theory. In Communications of the ACM, vol. 4, no. 2, 1998.

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

[23] Trepper C.: E-Commerce Strategies. Microsoft Press, 2000.

[24] Ulmann J.D.: Information Integration Using Logical Views, in Proceedings of the International Conference on Database Theory (ICDT ‘97), Greece, 1997.

[25] Vassiliadis P., Bouzeghoub M., Quix C.: Towards Quality-Oriented Data Warehouse Usage and Evolution. In Proceedings of the International Conference on Advance Information Systems Engineering (CAiSE’99), Heidelberg, Germany, 1999.

[26] Wand Y., Wang R.Y.: Anchoring Data Quality Dimensions in Ontological Foundations. Communication of the ACM, vol. 39, no. 11, 1996.

[27] Wang R.Y., Storey V.C., Firth C.P.: A Framework for Analysis of Data Quality Research. IEEE Transaction on Knowledge and Data Engineering, Vol. 7, No. 4, 1995.

[28] 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.

[29] Wang R.Y.: A Product Perspective on Total Data Quality Management. In Communication of the ACM, vol. 41, no. 2, 1998.