Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2009 Proceedings

Americas Conference on Information Systems (AMCIS)

2009

An Ontology Based Approach to Data Quality Initiatives Cost-Benefit Evaluation

Gianluigi Viscusi
Universita di Milano Bicocca, gianluigi.viscusi@epfl.ch

Andrea Maurino
Universita di Milano Bicocca, maurino@disco.unimi.it

Simone Grega
Universita di Milano Bicocca, grega@disco.unimib.it

An Ontology Based Approach to Data Quality Initiatives Cost-Benefit Evaluation

Gianluigi Viscusi
Dipartimento di Informatica Sistemistica e Comunicazione (DISCo)
Università di Milano Bicocca
viscusi@disco.unimi.it

Andrea Maurino
Dipartimento di Informatica Sistemistica e Comunicazione (DISCo)
Università di Milano Bicocca
maurino@disco.unimi.it

Simone Grega
Dipartimento di Informatica Sistemistica e Comunicazione (DISCo)
Università di Milano Bicocca
grega@disco.unimib.it

ABSTRACT

In order to achieve higher data quality targets, organizations need to identify the data quality dimensions that are affected by poor quality, assess them, and evaluate which improvement techniques are suitable to apply. Data quality literature provides methodologies that support complete data quality management by providing guidelines that organizations should contextualize and apply to their scenario. Only a few methodologies use the cost-benefit analysis as a tool to evaluate the feasibility of a data quality improvement project. In this paper, we present an ontological description of the cost-benefit analysis including the most important contributes already proposed in literature. The use of ontologies allows the knowledge improvement by means of the identification of the interdependencies between costs and benefits and enables different complex evaluations. The feasibility and usefulness of the proposed ontology-based tool has been tested by means of a real case study.

Keywords

Data Quality, Cost-benefit analysis, Metric, Data Quality Assessment, Ontology

INTRODUCTION

In this paper we discuss an ontology based approach to cost-benefit analysis aiming to provide a support for the data quality (DQ) experts in the assessment and improvement phases of DQ programs by means of an ontology-based tool. Here we adopt a light-weight perspective on ontology (Guarino, 1998); first we consider the existing classifications in order to define an ontology that classifies costs and benefits by identifying significant relationships among them. The utilization of such an ontology aims to support the discovery of significant knowledge in order to enable DQ experts to identify and measure costs and benefits. The tool allows users to identify costs by automatically proposing suitable metrics applicable in the assessment phase. The DQ experts can improve the ontology by adding new cost or benefit instances and associated metrics. The ontology and the tool proposed have been conceived to support methodology for the planning of eGovernment initiatives when dealing with DQ issues. In detail, the contribution of this paper supports the GovQual methodology (Batini, Viscusi et al., 2009) in the identification of costs associated with poor DQ and in the evaluation of the costs and benefits associated with the improvement processes.

The paper is organized as follows. The background and motivations Section reviews the main contributions in the DQ literature. In the GovQual methodology Section we explain in detail the steps of the methodology for the planning of eGovernment initiatives that we aim to support in cost-benefit analyses. The Cost-Benefit Ontology Section describes the ontology we created; and in the Metrics for Cost and Benefit Evaluation Section we describe the metrics we proposed for the evaluation of costs and benefits. Finally, we present the use of the tool in a real case study. Conclusions and future work conclude the paper.

BACKGROUND AND MOTIVATIONS

Cost-benefit analysis is an arduous task in many domains (Mishan and Quah, 2007). In particular, there is widespread awareness that the costs of DQ are relevant for companies at the strategic level (Ge and Helfert, 2007; Huang, Y. et al., 1999; Redman, 1996). Nevertheless, the economic benefit of DQ has been rarely investigated (Neely, 2005), and only recently in the literature have contributions provided tools and guidelines to support the choice of a DQ improvement process by conducting a cost-benefit analysis (Remenyi, Money et al., 2000; Yang, Pipino et al., 2006) and framework, adopting an economic perspective for the DQ assessment within a specific context (Even and Shankaranarayanan, 2007; Even, Shankaranarayanan et al., 2007). Furthermore, DQ evaluation is relevant if considered as part of a wider and complex context such as the planning of eGovernment initiatives, where the lack of tools and guidelines is critical for the DQ impacts on the effectiveness of the initiatives (Heeks, 2005). How to identify, categorize and measure DQ costs is still treated by few studies. The existing proposals range from classifications provided for costs and benefits to methodologies for performing the cost benefit analysis process (Eppler and Helfert, 2004). In the following, we focus on classifications and in particular we consider three detailed classifications for costs that appear in English (1999), Loshin (2004), and Eppler-Helfert (2004).

In the English classification (English, 1999) *costs caused by low DQ* are analyzed and divided into three categories:

- *Process failure costs*, when poor quality information causes a process not to perform properly.
- *Information scrap and rework*, when information is of poor quality and requires several types of defect management activities.
- *Lost and missed opportunity costs* that correspond to the revenues and profits not realized because of poor information quality.

Considering the Loshin classification (Loshin, 2004) here the costs of low DQ are classified by their different domain impacts, respectively:

- the *operational domain*, which covers the aspects of the system for processing information and the costs of maintaining the operation of the system;
- the *tactical domain*, which attempts to address and solve problems before they arise;
- the *strategic domain*, which stresses the decisions affecting the longer term.

Finally, Eppler-Helfert (2004) classified specific costs mentioned in the literature, such as higher maintenance costs and data re-input costs, into two major classes of costs, namely costs due to poor DQ and improvement costs. Costs due to poor DQ are categorized in terms of their measurability or impact, resulting in direct vs. indirect cost classes. Direct costs are those monetary effects that immediately arise from low DQ, while indirect costs arise from the intermediate effects. Improvement costs are categorized with the information quality process. Benefits are classified into three categories:

- *Monetizable*, when they correspond to values that can be directly expressed in terms of money.
- *Quantifiable*, when they cannot be expressed in terms of money, but one or more numeric indicators exist that measure the benefits.
- *Intangible*, when a type of benefit cannot be expressed by a numeric indicator.

The classifications discussed introduce different types of costs taxonomies. Nevertheless, previous works do not exploit possible interdependences between cost factors, which can be useful in supporting the DQ experts in the decision process. To this end, a formal ontology can be developed from the discussed classifications. In this paper, we aim for support cost-benefit analysis in the DQ field by defining an ontology that helps in the evaluation of the costs and benefits elements and in the discovery of significant relationships among costs. The proposed ontology has been conceived to support a DQ management program in the planning phases of GovQual (Batini et al., 2009), a multidisciplinary methodology conceived in

order to support the planning of eGovernment initiatives, where DQ has a central role. In the following we provide a general discussion of the methodology.

GOVQUAL AT A GLANCE

The general idea of GovQual is that the planning process should be driven by social, economic, juridical and technological issues considered in their strict relationship. According to this vision, the planning activity results in the choice of projects that (starting from the present qualities of the system) better fit the achievement of new target qualities. Figure 1 shows the high-level representation of the methodology with the inputs, outputs, and phases. Main inputs are the *social context* where the planning activity takes place, the *legal framework* of the considered context (country, region, etc), the *political vision* and objectives enacted by the central government and local authorities and the *existing technological solutions*. Besides these issues, a set of general *socio-economic/legal/technological indicators* facilitates the measurement of the quality level of the overall system and establishes new quality targets. These indicators encompass DQ dimensions.

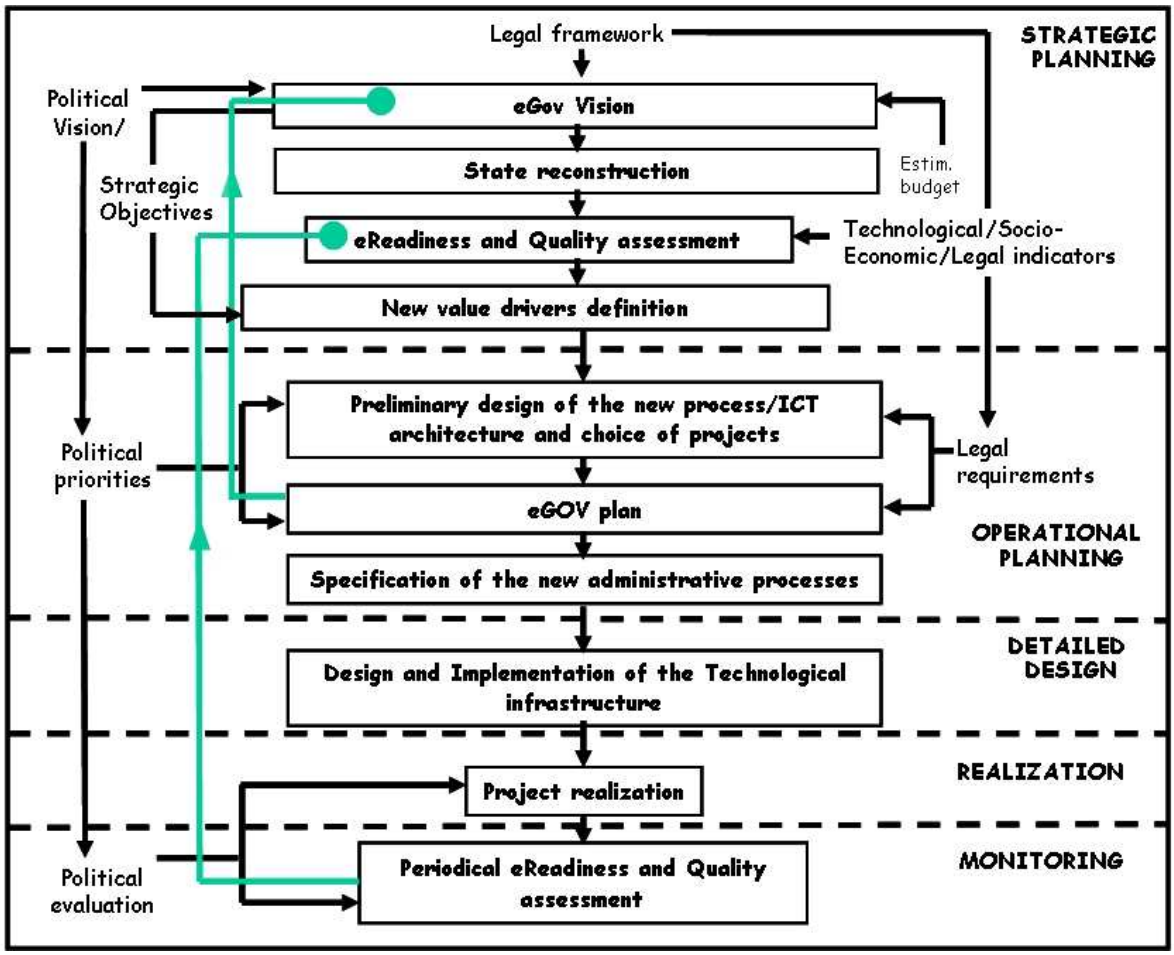

Figure 1. The phases of the GovQual Methodology

There are five main macro-phases of the methodology : (1) *strategic planning*, (2) *operational planning*, (3) *detailed design*, (4*) realization*, (5) *monitoring*. *Strategic planning* is the core macro-phase of the GovQual methodology and is composed by four phases: (1) *state reconstruction*, (2) *e-Readiness and quality assessment*, (3) *new value drivers definition*, and (4) *preliminary design of the new process/ICT architecture and choice of projects* (Batini et al., 2009). Here we focus on the critical issue emerging from the experiences of the application of the methodology, that is to support the *quality assessment*

phase providing tools for DQ analysis. In this phase a cost-benefit analysis is needed to fix DQ targets and to select the most suitable improvement process, further refined in the detailed design phase. The costs related to DQ management may be classified as *non quality costs* and *quality costs*. The former are all the costs derived by low quality levels and are mainly related to the re-execution of the failed processes and to the processes correction. The latter are all the costs that an organization invests in improvement activities (e.g., the licence costs of a data cleaning tool).

The first role of the proposed ontology is to support the evaluation of *non quality costs*, providing directions for the planning of improvement initiatives in the operational planning macro-phase (see Figure 1). The results of the evaluation of costs of non quality data contribute in the definition of the quality target and of the economic threshold for the improvement plan by relating DQ assessment to the different layers considered in the GovQual methodology (that is, organizational, juridical, economic, and sociological layers). The aims of the quality assessment and of the new value drivers definition phases is to identify the most suitable improvement initiatives as part of the overall eGovernment initiatives resulting from the strategic planning macro-phase. Thus, DQ experts evaluate all the possible benefits according with DQ improvement benefits ontology, and they quantify (and, if possible, monetize) all detected costs.

THE COST-BENEFITS ONTOLOGY

Figure 2 shows the top level ontology describing costs and benefits related to DQ (in the following, *OntoCB*). We represent the available classifications through the concept *cost/benefit classification*. This concept allows DQ expert to adopt an already existing classification (e.g. Loshin classification) or to create a new one. A *cost classification* is related to *cost items*. For each existing classification, we add a specific concept (e.g. the English Costs perspective) as child of a cost item; then, we associate the instance of cost classification to this specific cost item by means of an axiom, formally:

$$CostX \sqsubseteq \exists\, hasClassification \ni X$$

Where X is the name of the cost classification (e.g. English). Cost items can be *domain independent* or *domain dependent*. Each cost item has a textual description and is related to a *source* storing its bibliographic information. As described in the previous Section, cost items can also be classified as *non quality costs* and *quality costs*. Costs already described in literature are represented as a hierarchy of concepts, which are children of the cost item concept. Cost items are characterized by three important properties (i) *equivalence*, (ii) *narrow* and (iii) *broader*. The first property describes that an instance of the cost item is equivalent to another one. This property is applied when the same costs are proposed in different cost classifications. For example, *verification costs* are considered in both English and Eppler-Helfert classifications. Narrow and broader terms are two important properties representing respectively a more specific cost item, and a more general cost item. For example, the *business review work cost* proposed in the English classification is a broader term of the *rework cost* proposed by Loshin. The use of these properties is important in order to help the DQ expert to navigate OntoCB in a more useful way. The DQ expert can also add a new cost item concept as a child of the top concept cost item.

Non quality cost items are related to an *error* which are domain dependent (e.g. value absent in a given tuple). Errors can be associated to one ore more *DQ dimensions* (e.g. completeness, consistency, etc.) and to one or more *information chunk*s (e.g. department table). Cost items are evaluated on the basis of *metrics*. Metric formulas are calculated by means of the *resources* they need. Resources can be *data* (e.g. the number of tuples considered), *consultant* (e.g. the cost of external specialists), *employee* (e.g. hourly cost of internal employee involved), *media* (e.g. the number of communication media adopted, such as e-mail or phone), *software* (e.g. cost deriving from adoption of specific DQ software) and *time*, that is the number of hours spent in performing an activity. The DQ expert has to derive a domain dependent metric from the available domain independent ones. Focusing on the benefit ontology, *benefits* are organized in *benefit items* that can be related to DQ costs. The use of OntoCB as the supporting tool is justified due to the possibility of inferring new knowledge starting from the explicitly declared one. OntoCB exploits standard reasoning tools (Denny, 2004) for detecting:

- New equivalence properties among cost items,
- New narrow properties among cost items,
- New broader properties among cost items.

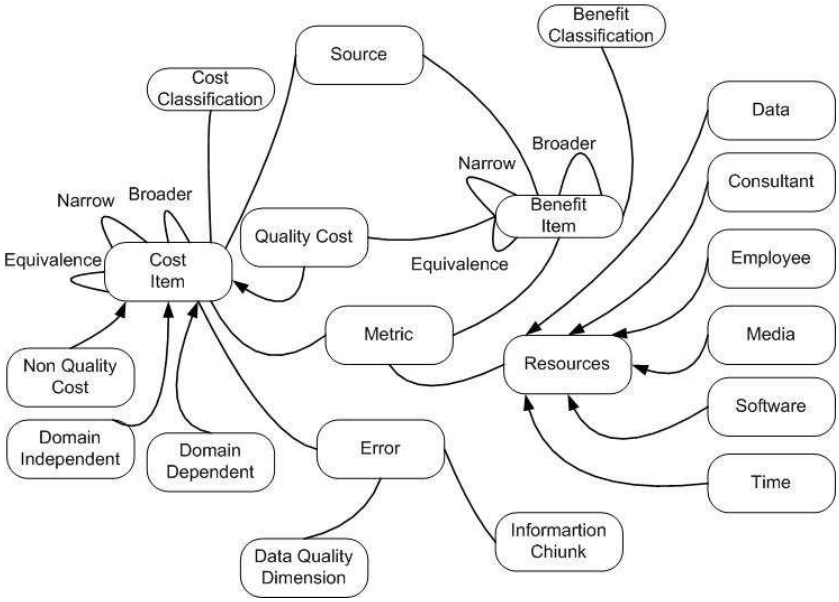

Figure 2 – OntoCB schema

A relevant issue in the adoption of ontologies is the visualization of the results. In this paper we adopt ontosphere3D (Bosca and Bonino, 2004), a Protégé plug-in for ontology navigation where information is presented on a 3-dimensional view-port enriched with several visual cues (such as the colour or the size of visualized entities). User interface features direct manipulation operations such as zooming, rotating and translating objects. The ontology based approach exploits the metrics discussed in the following Section.

METRICS FOR COST-BENEFIT EVALUATION

In this paper, we propose domain independent metrics for the cost item described in the previous sections, covering about 65% of the cost items proposed in the literature. These costs can be evaluated with the following general formula:

$$\sum i(\text{activityTime}*\text{HourlyCost}*\text{NumberEmployeeInvolved})+\text{supportCost} \quad (2)$$

i= wrong tuples

According to (2), the above mentioned costs can be evaluated by the sum, for each wrong tuple, of the product among the time spent for performing the activity (e.g. detection, correction, prevention), the hourly cost of employees and the number of employees involved in the activity. We also consider other costs, called *supportCost*, which can be caused by the performed activity (e.g., the cost of acquisition of new data used for correcting the wrong tuple or the cost of a software tool for data cleaning). Many metrics, such as the one related to the *software rewrite* cost proposed by English, can be found in the literature where several approaches for software development have been defined, such as the function point ones (Conte, Dunsmore et al., 1986).

$$\sum i(\text{deploymentTime}*\text{HourlyCost}) \quad (3)$$

i=new or modified function points

Formula (3) defines the cost of software rewrite as the sum of the cost to rewrite a single function point that is calculated as the product of the time for modifying one function point for the hourly cost of each software programmer. Other metrics are straightforward. For example, *data cleansing software* cost proposed by English can be measured by considering the cost of the license for buying such a specific tool. Another example is the *training cost* proposed by Eppler-Helfert. For about 35% of the cost items proposed in literature it is difficult to propose a meaningful metric such as the *increased difficulty* cost of the Loshin classification or the *recovery costs of unhappy customer* of the English classification. In this case it is possible to propose a proxy metric that measures a cost item which is very close to the original. The proxy metrics evaluate part of the

whole cost, thus they provide a fraction of the cost item. In this case it is possible to estimate the percentage of the cost item that the proxy costs represents. For example, while it is quite difficult to evaluate the *cost of wrong decision* (proposed by Eppler-Helfert), it is possible to measure resources (including time, people) spent for producing the wrong decision. The DQ expert can use this general metric for defining a specific new metric. By using an ontological approach the DQ expert can use this metric not only for the costs proposed by Loshin, but also for all cost items for which the equivalent or narrow properties is verified. For example, formula (2) can also be used for all direct costs proposed in the Eppler-Helfert classification.

CASE STUDY

In the following, we present a case study in which the GovQual methodology and OntoCB have been applied. The case study concerns the analysis of costs and benefits in a set of Italian local public administrations and we address DQ issues related to *the fulfilment of the formative obligation in the Piedmont Region*. The Italian legislation (law n.144 of May 1999) states that young people have to be inserted into a formative path until they are 18 years old; local administrations are responsible for keeping this formative obligation. The analysis we conducted allowed the execution of the assessment and DQ improvement process of the *Formative Obligation Register (FOR)*. FOR contains data about the students and provides a region-level centralized register for monitoring them. Moreover, it controls the *scholar dispersion* phenomenon, helping and taking up the insolvent subjects to a formative choice, thus conforming to the legal obligation. According to law, there are three possible formative paths that young people between fifteen and sixteen years old can follow:

- to continue studying;
- to attend lectures of *Professional Education* provided by the Region in order to obtain professional qualification;
- to apply for an *Apprenticeship* in order to obtain a certificate that attests the acquired professional competence.

According to this legal framework each Italian Region curbs the *scholar dispersion* by a periodical extraction of the data related to the students who are under the age of the formative obligation, but does not show as attending any professional formative or apprenticeship courses. This operation can be performed by verifying if the student appears in any available informative source or not possible. Then, the *Employment Market Places* will reach and contact the critical subjects to apply every orientation action provided by the law. The main problem is related to *false missing* subjects, i.e. a student does not attend any formative channel, but it is erroneously verified as a regular participant. The causes for false missing subjects can be: i) registration errors (e.g. the student is not registered in any database or his personal data are duplicated) and ii) omitted registration (e.g. data about "death" or "transfer out of the region" is not registered). The goal of Piedmont Region is to verify the *true missing* subjects and to contact the students in order to resolve their reintroduction in a formative way.

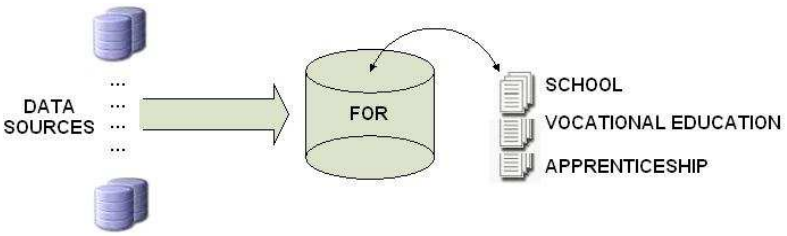

Figure 3 – Reconstruction of the informative flows

In the *state reconstruction* of the GovQual methodology, we analyze and reconstruct the informative flows (see Figure 3) that feed the FOR. In particular, FOR centralizes the information related to students that is provided by different sources (formative channels and local registers). Significant characteristics of the FOR are:

- daily feeding;
- each source has its own register of students with related personal data;
- each student should appear only once and he/she is identified by the taxpayer's code number.

In the *quality assessment* phase, we assess the FOR and evaluate the non quality costs. The obtained list of costs is identified and classified by using OntoCB. In the following, we propose a significant selection of the obtained results:

- *Costs of the unnecessary performed researches:* are sustained whenever an individual data is incorrect and additional inquires about it are necessary. These costs are due to a working process failure and they are irrecoverable. For these reasons, browsing the *Cost items* through the English perspective, we classify these costs as *Irrecoverable costs* item of the *Process failure costs* hierarchy. Figure 4A shows an example of browsing with Ontosphere3D carried out to define the following costs:
- *Costs of data re-entry performed by the Employment Market Places*: are verified when an erroneous data contained in the FOR must be corrected; in this case, the employees proceed to data re-entry after their examination. In this case, we browse OntoCB through the Eppler-Helfert perspective and classify these costs as a *Re-entry costs* item of the *Direct costs* hierarchy.
- *Costs of erroneous phone announcement:* are sustained whenever an employee of the Employment Market Places phones to *false missing* subjects and their data must be rejected. Browsing OntoCB, we do not discover this type of cost in the proposed classifications. However, these costs are mainly due to the phone calls performed by the employees and the data scrapped. For these reasons, we insert a new cost item in the *Domain dependent* hierarchy and join it with the *Information scrap and rework cost* hierarchy by the *Narrow property*.
- *Costs of data verification performed by the Employment Market Places:* are verified when data is inconsistent (e.g. a male student has associated a female name). In this case, it is necessary to verify these data with additional researches. Browsing OntoCB through the Eppler-Helfert perspective, we discover the *Verification costs* item in the *Direct cost* hierarchy. Observing the relationships presented in Figure 4B, it is worth noting that the *Verification cost* item is also used in the English classification. In particular, it is joined with the *Information scrap and rework cost* hierarchy.

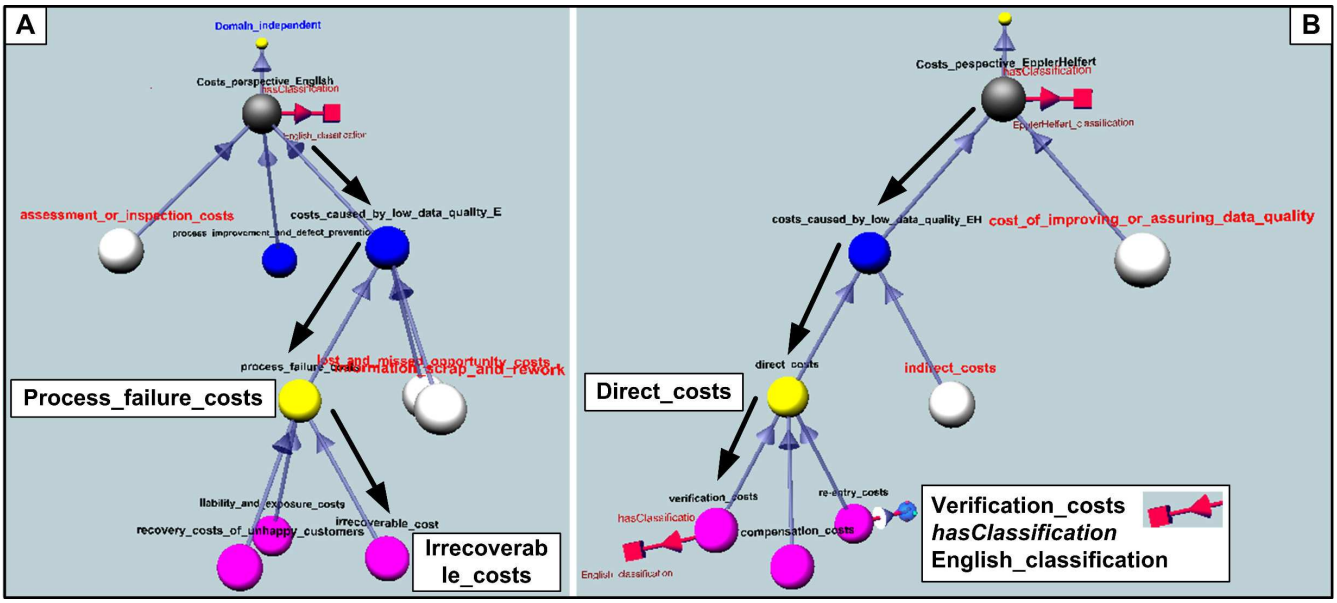

Figure 4 – Browsing OntoCB to classify the *Irrecoverable costs* and *Verification costs*

In order to identify the benefits obtained through the application of the DQ improvement process, we perform the same activities described above for the quality costs identification. In the following, we propose a selection of the obtained results considering only the *Monetizable* benefits:

- *Revenue increase by school fees:* represents the revenues obtained by school fees thanks to the growth of the number of students who attend the public school. Browsing the *Benefit items,* we classify these benefits as the *Revenue increase* item of the *Monetizable* hierarchy.
- *Cost decrease:* represents every cost decrease obtained by the DQ improvement process on the FOR. Also these benefits are included in the *Monetizable* hierarchy as *Cost decrease* item.

At this stage, we evaluate costs and benefits listed above with the metrics defined in the OntoCB. In order to apply these metrics, we insert the useful domain information. Browsing the OntoCB, it is possible to define particular instances of the *Resource* categorizations; such as data, consultant, employee, media and software time. In the following, we report the activities performed to select the suitable metrics and to apply them to the presented costs and benefits, that is:

- *Re-entry costs:* starting from the *Re-entry costs* item, we identify the associated metric through the *HasMetric* property. As shown in Figure 5, we are able to identify the elements involved in the metric formula. In particular, the Re-entry costs formula is composed of the following data:
- *HourlyCost*: cost of the employee (an instance of the *Resource– Hourly cost* item);
- *ReEntryTime*: needed time to re-insert and update a record (an instance of the *Resource – Time* item);
- *EmployeeNumber*: number of the employees involved in the re-entry operations (an instance of the *Resource – employee*).
- *Wrong data*: number of wrong tuples in the database (an instance of the *Resource – Data* item);

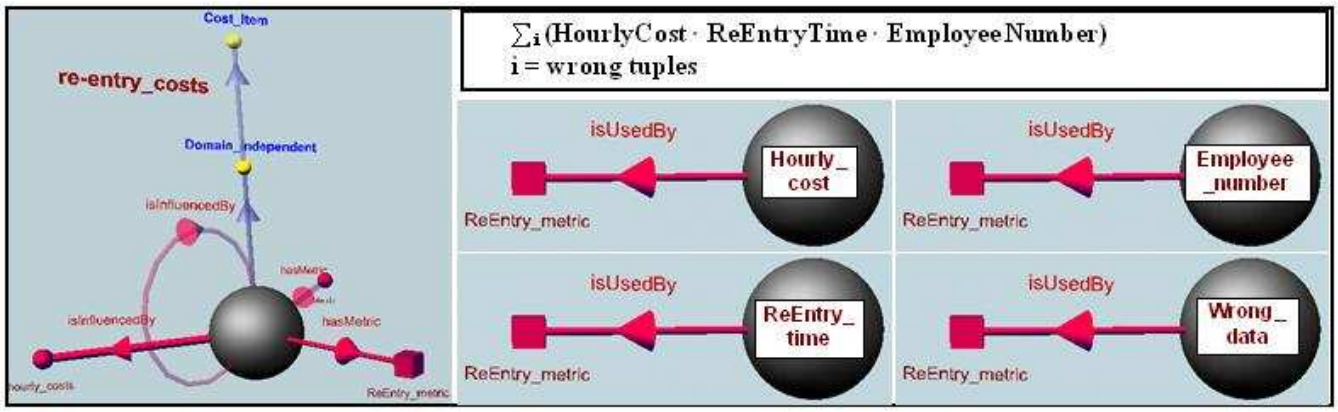

Figure 5 – ReEntry metric and involved information

Furthermore, we instantiate (see Table 1) each information composing the metric through the *IsUsedBy* property. This activity is repeated for the other costs obtaining the results shown in Table 1.

COST	CONCEPT	INSTANCE
ReEntry $\sum_i$ (HourlyCost · ReEntryTime · EmployeeNumber) i = wrong tuples	Hourly_cost	12 €/h
	ReEntry_time	4 minutes a record
	Employee_number	1
	Wrong_data	14.000
	ReEnty_costs	11.200 €
Verification $\sum_i$ (HourlyCost · VerificationTime · EmployeeNumber) i = wrong tuples	Hourly_cost	12 €/h
	Verification_time	1 h a record
	Employee_number	1
	Wrong_data	14.000
	Verification_costs	168.000 €
Irrecoverable	Hourly_cost	12 €/h
	Investigation_time	1 h

$\sum_i$ (HourlyCost · InvestigationTime · EmployeeNumber) + irrCost · α i = wrong tuples α=0,1	Employee_number	1
	Wrong_data	14.000
	Activity_number	1
	Irrecoverable_costs	168.000 €
Phonecall $\sum_i$ (HourlyCost · PhoneTime · EmployeeNumber) i = wrong tuples	Hourly_cost	12 €/h
	Phone_time	10 minutes
	Employee_number	1
	Wrong_data	14.000
	Phonecall_costs	28.000 €

Table 1 – Costs quantification

The same activities performed to quantify the costs are applied to the benefits related to the improvement process that we propose. In particular, observing the properties referred to the *Revenue increase* item, we verify that a particular metric to calculate this benefit is not defined. In fact, the metric for the *Revenue increase* benefit is strongly dependent on the domain where it must be applied. Therefore, we need to define a new metric that depends on the school fees paid by the students. We create an instance of the generic concept *Metric* and associate it with the *Revenue increase* item by the *HasMetric* property. Then, we create a new cost item in the *Domain dependent* category, named *IndividualFees_cost*. In order to calculate the *Revenue Increase* benefit, we report the necessary assumptions made in the case study: (i) the augment of the collateral expenses is not considered; (ii) the families with no sufficient economic resources are about 3%. Therefore, these families take advantage of the school fees exemption (i.e. exemption threshold); (iii) we assume that our improvement tecnique is able to correct 50% of wrong data (i.e. 7.000 records), where 70% refers to retrieved students which do not attend any formative obligation (i.e. 4.900 retrieved students). Table 2 shows the metrics and data used to calculate the *Revenue increase* benefit.

BENEFIT	CONCEPT	INSTANCE
Revenue Increase (RetrievedStudents_number - RetrievedStudents_number · Exemption_threshold) · IndividualFees_cost	IndividualFees_cost	94 €
	RetrievedStudents_number	4.900
	Exemption_threshold	3%
	Revenue_Increase	446.782 €

Table 2 – Revenue Increase benefit quantification

As shown in Table 2, in addition to the *IndividualFees_cost* item, we insert the information about the estimated number of retrieved students obtained applying a DQ improvement process and the value of the exemption threshold as instances of the *Resource* categorizations. Besides, the metric to calculate the *Cost decrease* benefit is independent from the domain, and it is easily applied. Starting from the *Cost_decrease* item, we verify if the associated metric is calculated as the difference between the total of costs measured before the application of a DQ improvement process and the costs measured after it. Table 3 shows the obtained results.

COSTS	Value obtained before the improvement	Value obtained after the improvement	Difference
ReEntry	11.200 €	5.600 €	5.600 €
Verification	168.000 €	84.000 €	84.000 €
Irrecoverable	168.000 €	84.000 €	84.000 €
Phonecall	28.000 €	14.000 €	84.000 €

TOTAL	375.200 €	187.600 €	187.600 €

Table 3 - Cost decrease benefit quantification

CONCLUSION

Cost-benefits analysis is an important task in the DQ management program. Several cost classifications have been proposed in literature, but little attention is placed in the providing of operative tools supporting the DQ expert. In this paper we propose an Ontology-based tool for supporting the DQ expert in the measurement of costs and benefits, in particular in the context of the planning of eGovernment initiatives carried out by means of the GovQual methodology. There are several future works. From a methodological view point we plan to further enhance the GovQual methodology to better cover the cost benefit evaluation. We also extend OntoCB in order to help the DQ expert to improve the automatic identification of the most critical DQ dimensions. This goal can be achieved by means of query language such as SPARQL (Prud'hommeaux and Seaborne, 2007), an RDF query language that allows extracting the DQ dimensions associated to errors causing the most relevant costs. Finally, new domain independent and also domain dependent metrics will be added to improve OntoCB.

ACKNOWLEDGMENTS

The work presented in this paper has been partially supported by the Italian FIRB projects NeP4B - Networked Peers for Business(RBNE05XYPW) and eG4M - eGovernment for Mediterranean Countries (RBNE0358YR).

REFERENCES

1. Batini, C., Viscusi, G., and Cherubini, D. (2009). GovQual: A quality driven methodology for E-Government project planning. *Government Information Quarterly, 26*, 106-117, Elsevier.
2. Bosca, A., and Bonino, D. (2004). *OntoSphere: more than a 3D ontology visualization tool.* Paper presented at the Second Italian Semantic Web Workshop.
3. Conte, S., Dunsmore, H., and Shen, V. (1986). *Software engineering metrics and models*. Redwood City, CA, USA Benjamin-Cummings Publishing Co., Inc. .
4. Denny, M. (2004). Ontology Tools Survey Retrieved February, 25, 2008, from http://www.xml.com/lpt/a/1447
5. English, L. (1999). *Improving Data Warehouse and Business Information Quality*: Wiley & Sons.
6. Eppler, M., and Helfert, M. (2004). *A classification and analysis of data quality costs.* Paper presented at the Ninth International Conference on Information Systems (ICIQ-04), Boston.
7. Even, A., and Shankaranarayanan, G. (2007). Utility-driven assessment of data quality. *SIGMIS Database, 38*(2), 75-93,
8. Even, A., Shankaranarayanan, G., and Berger, P. D. (2007). Economics-Driven Data Management: An Application to the Design of Tabular Data Sets. *Knowledge and Data Engineering, IEEE Transactions on, 19*(6), 818-831,
9. Ge, M., and Helfert, M. (2007). *Develop a Research Agenda: A Review of Information Quality Research.* Paper presented at the 12th International Conference on Information Quality (ICIQ), November 9-11, 2007, MIT USA.
10. Guarino, N. (1998). Formal ontologies and information systems. In N. Guarino (Ed.), *Proceedings of FOIS'98.* Amsterdam: IOS Press.
11. Heeks, R. (2005). *Implementing and managing egovernment : an international text* (1st ed.). London; Thousand Oaks, CA: SAGE Publications.
12. Huang, K., Y., L., and Wang, R. (1999). *Quality Information and Knowledge.* . Upper Saddle River, NJ: Prentice Hall.
13. Loshin, D. (2004). *Enterprise knowledge management - the data quality approach.* : Morgan Kaufmann
14. Mishan, E. J., and Quah, E. (2007). *Cost-Benefit Analysis*: Routledge.
15. Neely, P. M. (2005). *The Product Approach to Data Quality and Fitness for Use: A Framework for Analysis.* Paper presented at the The 2005 International Conference on Information Quality (MIT IQ Conference),, MIT, Cambridge, MA, USA.
16. Prud'hommeaux, E., and Seaborne, A. (2007). W3C Recommendation 15 January 2008 Retrieved February, 25, 2009, from http://www.w3.org/TR/rdf-sparql-query/
17. Redman, T. C. (1996). *Data quality for the information age*. Boston, MA: Artech House.
18. Remenyi, D., Money, A., Sherwood-Smith, M., and Irani, Z. (2000). *The effective measurement and management of IT costs and benefits*. Oxford: Butterworth-Heinemann.
19. Yang, W. L., Pipino, L. L., D. Funk, J., and Wang, R. Y. (2006). *Journey to Data Quality*: MIT Press.