BUSINESS PROCESS MODELING TOWARDS DATA QUALITY ASSURANCE

An Organizational Engineering Approach

Hugo Bringel[†], Artur Caetano[*†], José Tribolet[*†]

†*CEO - Organizational Engineering Center, INOV, INESC Inovação.*

**Department of Information Systems and Computer Science, Instituto Superior Técnico, Technical University of Lisbon.*

Surface Mail Address: INESC, Rua Alves Redol 9, 1000-029 Lisboa, Portugal.

Keywords: Data Quality, Business Process Modeling, Object-Oriented Modeling, UML

Abstract: Data is handled everyday by information systems and its inherent quality is a fundamental aspect to operational and support business activities. However, inappropriate data quality may lead to economic and social problems within the organizational context. This paper addresses how to syntactically and semantically ensure data quality at information entity level. To do so, we define a business process-modeling pattern for describing the features required to ensure and validate business object data using a conceptual data quality attribute model. This pattern makes use of object-oriented concepts such as inheritance and traceability and is described as an extension to the Unified Modeling Language. A case study is presented to exemplify the proposed concepts.

1 INTRODUCTION

Assuring data quality is a complex process, in which the tradeoff between cost and quality depends on the application context and on the organization requirements.

Incorrect, incomplete or non-timely data may cause economic and social problems in organizations, which very often, only react to its consequences, rather then having a proactive attitude.

Another common issue is that data quality is not understood in a process-centric, cross-departmental perspective, but in a functional view, as a duty or competence of the information systems department.

Problems with data quality occur widely on functional organizations, where specific databases are created, forming information islands that constitute one of the mainstream causes of lack of consistency and of coherence of corporate data.

There are several approaches, but they are focused at low-level data analysis, based on computing algorithms treating implicitly data quality at the DBMS level, or they are mainly focused on quality management systems, based on ISO standards. These approaches are however not sufficient from the data consumer's point of view, (Laudon, 1986).

Regarding this, we propose a proactive approach towards guarantying data quality, considering in our proposal two levels of granularity. First, at high level when using a data quality business processes pattern, focused on activities execution upon information entities, consumed or produced by business processes, as discussed in Section 2.1. Second, at low level when assuming a data quality attribute model, as discussed in Section 2.2.

The remaining of this paper is structured as follows: In Section 2, we define the problem and propose our approach to its resolution. In Section 3, we illustrate an example of business process and data quality modeling and finally in Section 4 we present some conclusions.

2 PROPOSAL

The problem we intend to solve consists in how to guarantee that attribute values of a business entity are correct and correspond to the defined semantics. To define data quality requirements, we need multiple dimensions, depending on specific needs from different organizational levels. For instance,

Bringel H., Caetano A. and Tribolet J. (2004).
BUSINESS PROCESS MODELING TOWARDS DATA QUALITY ASSURANCE - An Organizational Engineering Approach.
In *Proceedings of the Sixth International Conference on Enterprise Information Systems*, pages 565-568
DOI: 10.5220/0002649305650568

the sales division may need inventory data to be accurate and complete, while management may need information that gathers other data quality dimensions, like reputation or timeliness of data for the decision making process.

2.1 Data Quality Modeling

In the past, data quality was often defined as non-conformance to requirements, (Crosby, 1984). However, just as there are several dimensions of product quality, such as conformance, durability or performance, in any industrial domain, data quality embraces specific characteristics, denominated quality dimensions.

In addition, data from several sources may have common dimensions in which quality may be measured. Even if data quality modeling does not depend on business process modeling, the data quality attributes should be tagged to business objects during entity/resource identification and definition.

Concluding, data quality is a multi-dimensional and hierarchical concept, (Wang, 1995).

In our perspective, it is not possible to manage different data quality perspectives of each data user using an entity-relationship model. To accomplish it, we need to model the several dimensions of data quality requirements, proposing instead, the use of role modeling, within the object-oriented paradigm.

We shall use, in data quality modeling, two concepts, namely (1) *quality parameters* and (2) quality *indicators,* forming *the quality attribute,* (Wang, 1995). Quality *parameters* relate with quality dimensions (qualitative aspects) while quality *indicators* relate with measurable attributes (quantitative aspects, normally from physical goods). For example, as quality parameters, we need to attain *confidence* and *timeliness*, and as quality indicators, we can decompose on *data source* and *creation date.* The quality parameters and indicators form the data quality *attribute.* This aggregation result in a hierarchical attribute tree that helps defining the data quality requirements. This quality attributes model provide a comprehension base to understand the characteristics that defines data quality.

2.2 Business Process Modeling Approach towards Data Quality

In this paper we use an object-oriented business process modeling framework which is used to model the interaction between process activities, business goals, resources and information systems (Vasconcelos, 2001).

To solve the problem stated above, we propose the use of a business process pattern to ensure data quality in an organization, making use of defining data quality attributes upon information entities considering different meanings on each business perspective.

In this paper, we introduce the notion of *Tertiary Organizational Processes*. Tertiary processes represent activities that cross over operational and support processes planes, interacting within active business entities, with the purpose of achieving some special purpose objectives. The modeling of tertiary processes is facilitated by the introduction of the concept of an "orthogonal" plane, relative to the operational and support business process plans.

We therefore consider three layers in which lies the organization modeling of any business process.

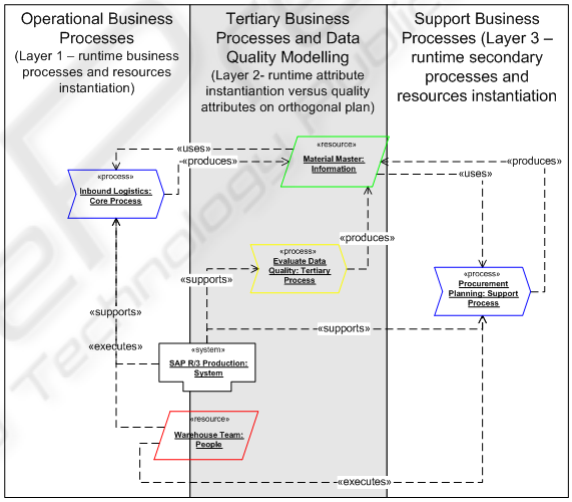

Figure 1 – Process diagram on business (core and support) processes and data quality processes interaction

Figure 1, depicts the main conceptual structure and layer separation we use on our approach.

The first layer models the operational business process activities, whereas the third layer models the support process activities. The middle layer deals with data modeling and data quality evaluation (and assurance) activities, where designed classes of objects "resources" are instantiated from different perspectives, the quality (tertiary) processes *versus* operational or support processes.

2.2.1 The Resource Stereotype

In business process modeling, resources are objects within the business that processes manipulate. The resource types are represented as classes while resource instances are represented as objects. The business object in discussion is the resource stereo-

type, which we focus in the *informational entity* as specialization of a resource class. The informational entity is modeled at two different levels: business process and class level.

The business process level is where the information entity is modeled as a resource stereotype and where it interacts with a process. The class level is where data is modeled in object-oriented classes, and where its attributes are specified. There is an inherent complexity at this level, because, for the same data, we have overlapping and crosscutting requirements (expressed by attributes and methods) between core business processes, support business processes and tertiary processes – such as the data quality processes.

We shall also consider predefined quality classes, with the data quality attributes of conceived business objects. These classes compose the quality attributes into the information entity.

We are thus extending the concept of information entity by using new attributes devoted to quality, beyond the usual and basic pre-defined data attributes.

2.2.2 Data Quality Process Pattern

We propose the use of a data quality pattern at business process level, based on (English 1999), which integrates best practices from the quality management universe. This pattern consists on a business process model that can be reused through adaptation in specific organizational scenarios. Figure 2 depicts the flow between pattern's top activities. It is important to note that this pattern is composed of tertiary process and applies to an orthogonal organization plan, as previously discussed on Section 2.1.4

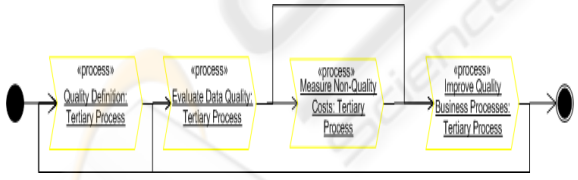

Figure 2 – A data quality business processes pattern

The "Quality and Architecture Definition" and "Evaluate Data Quality" activities focus on the evaluation of data quality, describing how the organization deals with data quality and what processes and resources are involved. The interaction between entities or resources, either consumed or produced by these activities, is represented at lower granularity level to make possible the activities autonomy and promote this pattern reuse.

Two processes may interact with shared information entities; however, to capture the information entity interaction in different contexts, we propose using role-modeling concepts (v. Section 2.3).

2.3 Role-based Modeling

A role represents some unit of responsibility or behavior. Actors play different roles while performing business process activities. Roles can be considered types in the sense they describe the behavior that is carried out by an instance of that role by a specific actor. Therefore, there may be multiple instances of the same role when a process is enacted. A single actor may also play multiple roles. Role models can be instantiated, aggregated and generalized.

By understanding the behavior of processes, we are providing the means to reuse it and adapt its organizational concepts.

2.3.1 Role modeling on data quality

Since role modeling allows the behavior of resources to be clearly separated and identified, we can have different contexts of data quality as attributes of class that represents a business object. Resources are specialized so that its attributes and methods allow handling its quality features. Processes concerning quality attributes instantiate predefined classes and set values to its quality attributes. The quality attributes, using predefined combinations of quality parameters and indicators, allow judging the data quality.

This approach leads to a better understanding and confidence in data since quality information is kept within a resource (informational entity type), which facilitates the data quality evaluation process.

3 CASE STUDY

This case study results from a research project on a real organization. This scenario illustrates the business processes of inbound logistics in a large warehouse. The targeted company handles an average of 22.000 products and performs a few dozens of daily inbound transactions.

The process starts when the materials arrive at the warehouse. The process activities start with the "Materials Checking", "Materials Unloading", "Data Input" and ends with the "Material Storing" in the warehouse facility.

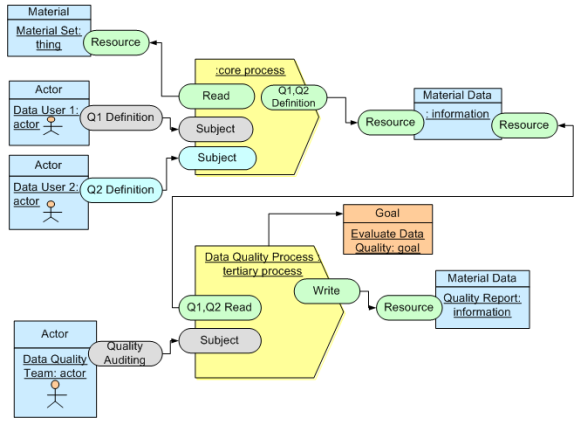

Figure 3 – Data quality evaluation using role modeling

Figure 3 depicts how to ensure data quality at business process level making use of role modeling.

Regarding the "Data Quality Evaluation" (Figure 2) sub-process, it is here (Figure 3) represented as "Data Quality Process", while the "Core Process" represents a logistics operational business process.

An oval represents the role associated to a class. It is a shorthand modeling for aggregating the role class with the base class.

In this example, "Material" corresponds to the "Material Set" used in the "Core Process". "Data Users" are the actors, which specify data quality requirements as part of a "Data Quality Definition" process. In this process, they define their respective data quality requirements, for instance: as Q1= {Timeliness, Completeness}; Q2= {Accuracy, Currency} for the "Material Set". The "Core Process" produces updated "Material Data" informational entity, later audited using of a sampling set of materials. A "Data Quality Team" acts in the data quality requirements auditing. The "Data Quality Process" result is a "Quality Report" for later analysis. The overall process goal is to "Evaluate Data Quality" generated by "Core Process".

In this way, we have ensured data quality, using role modeling to manage different context and semantics of data quality, combined with the core or support business processes, which have data quality support in their informational entities.

4 CONCLUSIONS

This paper proposes a data quality pattern to model the data quality intrinsic to business processes. This pattern can be used for data improvement on any organization, and makes use of a set of business processes to syntactically validate the data according to a model, which depicts qualitative and quantitative data quality attributes. Role modeling is used to manage different quality contexts quality from different data users, assuring data quality at semantic level.

This contribution addresses data quality from a organizational engineering perspective using business process modeling, leveraging data confidence and promoting continuous data quality improvement.

REFERENCES

Crosby, P.B., 1984. *Data Quality Without Tears*. McGraw-Hill.

English, L.P., 1999. *Improving Data Warehouse and Business Information Quality, methods for reducing costs and increasing profits*. Wiley.

Eriksson, H.E., Penker, M., 2000. *Business Modeling with UML: Business Patterns at Work*. OMG Press.

Laudon, K.C., 1986. *Data Quality and Due Process in Large Interorganizational Record Systems*. Communications of the ACM, 4-11.

Vasconcelos,A, Caetano,A., Neves,J., Sinogas, P., Mendes R., Tribolet, J., 2001. A Framework for Modeling Strategy, Business Processes and Information Systems. *5th IEEE International Conference on Enterprise Distributed Object Computing*. IEEE Press. Seattle, USA.

Wang, R.Y., Reddy, M.P., Kon, H.B., 1995. *Toward quality data: An attribute-based approach*. Decision Support Systems 13, pp.349-372.