AMCIS 2008 Proceedings

Americas Conference on Information Systems (AMCIS)

2008

# Towards a Semantic Data Quality Management - Using Ontologies to Assess Master Data Quality in Retailing

Jörg Becker
*European Research Center for Information*, becker@ercis.de

Martin Matzner
*University of Muenster*, martin.matzner@ercis.uni-muenster.de

Oliver Mueller
*University of Muenster*, oliver.mueller@ercis.de

Axel Winkelmann
*University of Muenster*, axel.winkelmann@ercis.de

# Towards a Semantic Data Quality Management – Using Ontologies to Assess Master Data Quality in Retailing

**Jörg Becker**
University of Muenster, Germany
European Research Center for Information Systems
becker@ercis.de

**Martin Matzner**
University of Muenster, Germany
European Research Center for Information Systems
martin.matzner@ercis.de

**Oliver Müller**
University of Muenster, Germany
European Research Center for Information Systems
oliver.mueller@ercis.de

**Axel Winkelmann**
University of Muenster, Germany
European Research Center for Information Systems
axel.winkelmann@ercis.de

## ABSTRACT

A high quality of master data is vital for process automation and IS support in enterprises. Retailers and manufacturers have started improving their master data quality on a syntactical level. Many researchers address problems of syntactical master data quality (e.g. missing values, typo errors), but only few approaches target semantic issues of master data quality management. Poor semantic data quality can lead to a misalignment between real-world phenomena and data stored in the databases. Furthermore, semantic data quality relies on the current use of data, but is not limited to the present use of the data and its impact on existing business processes. In fact it includes the likely future uses of the data as well. In this paper, we discuss the contribution of ontologies for identifying and assessing semantic master data quality problems. We develop a conceptual approach and procedure model for addressing the semantic master data quality problem. The method is applied within a scenario for automated coupon clearing in the retail industry.

## KEYWORDS

Semantic, Ontology, Data Quality, Master Data, Retail

## MOTIVATION

Today, meaningful master data is crucial for business processes in retail and industry. Retailers rely on product master data stemming from manufacturers and manufacturing companies heavily rely on data provided by retailers, e.g. to optimize the flow of goods in their own supply chain with respect to actual customer demands. Manufacturers and retailers have started standardizing and centralizing their master data in order to improve data quality (Loshin, 2005). Moreover, electronic marketplaces, collaboration in terms of ECR (Efficient Consumer Response) and CPFR (Collaborative Planning, Forecasting and Replenishment), and the establishment of centralized product master data pools have forced companies to ensure the quality of their data. Global data synchronisation networks (GDSN) and quality initiatives conducted by the consumer packaged industry and institutions like GS1 further emphasize the importance of data quality of master data (Nakatani et al., 2006). Even so, a lack of structure in the data and a highly variable (and thus often poor) data quality are major problems for master data management. Quality of data is not an "isolated function" within an information system, but instead an inherent and integral part of business management (English, 1999).

The problem of poor data quality especially affects product master data provided by manufacturing companies, as standardized guidelines for structure and content are missing (Berson and Dubov, 2007). As such, identical products are often handled inconsistently within databases and different identifiers and multiple abbreviations are used. Insufficient data quality causes a multitude of business process failures and squanders resources in terms of people, money, materials, and facilities resources (English, 1999). As a consequence, the introduction of flexible data structures beyond generic classifications of goods is not possible yet. Due to this lack of structure and quality, product master data can seldom be used in highly automated business processes, but require the involvement of humans to interpret, update, and restructure data manually (Capgemini, 2004).

Based on the semiotic levels of data, Price and Shanks (2005) propose syntactic, semantic, and pragmatic criteria for the evaluation of data quality. Furthermore, some researchers address ways of overcoming syntactical and pragmatic quality deficiencies (Redman, 1996). Only few researchers discuss concrete approaches or practical guidelines for the semantic aspects of master data quality (Cappiello et al., 2003). Furthermore, data quality relies on the current use of data, but is not limited to the present use of the data and its impact on existing business processes. Instead, the concept includes the likely future uses of the data as well (English, 1999). Enterprises' master data serve as an enabler for business ideas and services, which had not been regarded at the particular concept phase of the information system. While syntactic data quality management is well understood in theory, there are hardly any practical approaches to semantic master data quality issues.

In this paper, we develop an approach for a better semantic master data quality by discussing the contribution of ontologies for identifying and assessing semantic master data quality problems. Modeling real-world contexts in an ontology and mapping the ontology to the physical model of the database ensures a better alignment between real-word entities and database structures. Hence, real-world concepts in ontologies can be mapped to several underlying databases (with heterogeneous database schema definitions) at the same time. Furthermore, it allows the analyzing of data quality on a non-technical (semantic) level. We illustrate our approach, introducing a scenario based on coupon promotions in retail. Coupons are vouchers that entitle buyers to buy at a discount. Customers hand it over to the retailers in order to receive a discount. This is called coupon clearing. With manual, hardly automated coupon clearing processes at the point of sale retailers did not have to rely on their master data. Cash register employees were able to verify a coupon and match it to the purchase conditions that were printed on the coupon. Nowadays, self-checkout and coupon fraud force retailers to automate their coupon clearing processes which causes new demands on the semantic quality of master data (Winkelmann, 2006).

The remainder of this paper is structured as follows: In section 2, we discuss related work and explain our research method. In section 3, we describe our approach for addressing the master data quality problem with the help of ontologies. In section 4, we discuss the contribution of our work to theory and practice and give an outlook on limitations and further research in this area.

## RELATED WORK ON DATA AND INFORMATION QUALITY

### Master Data

Master data can be defined as all information that is needed for an enterprise wide "system of records" for core businesses in order to capture business transactions and measure results of these entities (Berson and Dubov, 2007; Griffin, 2005). There are various types of master data available and needed within retail companies. Retailers need master data about customers and suppliers such as name, address, banking details, etc. as well as detailed information about their products. Product master data comprises basic data (e.g. product number, UPC, EAN), listing data (listing period, assignment to assortments), purchasing data (e.g. vendor-dependent details for the deliverability), logistic, sales and POS data (Becker et al., 2001).

Especially product master data and promotional master data are very important for automatic processes at the point of sales. Product master data include information about product name, size, price, etc. Furthermore, a unique identifier called UPC (Unique Product Code (USA)), EAN (European / International Article Number) or JAN (Japanese Article Number) is required in order to identify master data of a product. Promotional master data contains all relevant information about a promotion such as value of the promotion, validity conditions, etc. and a reference to the promoted products. For example, in order to automate a BOGOF-promotion (buy one get one free) retailers will need promotional master data about the value of the promotion ("one product for free"), the time of validity ("until July, 31st") and the product, respectively the product code of the product (e.g. "Coca-Cola 0.5 litres", EAN: 4023786889760).

### Master Data Quality

Many researchers analyzed the properties of data and information quality (in accordance to (Huang et al., 1999; Kahn et al., 2004) we use both, information and data quality, synonymously). During the last years, many authors commented on poor master data quality in industry and retail. For example, according to (Kuipers, 2004), Nestlé assumed that more than 50% of items had incorrect or redundant master data entries. For more general trends in data quality see (Agosta, 2005), who claimed that according to a survey 30 % of data warehousing maintainers suffered from poor data quality in the USA in 2005.

A common quintessence of the past two decades' research is that "data quality is a multi-dimensional concept" (cp. Pipino et al., 2002, and cited lit.), which can be distinguished into (a) the subjective perceptions of the individuals involved with the data and (b) an objective measurement based on the data set in question. Other authors use the expressions *pragmatic* (a) and *inherent* (b) quality (English, 1999). Quality as regards inherent quality is limited to the correctness of facts and a clear definition or meaning of the data. The corresponding subcategories from semiotic theory are *syntax* and *semantic* (Price and

Shanks, 2005). Quality on a pragmatic level can be defined as the "correctness of the right facts represented correctly" (English, 1999) and therefore focuses on delivery and actual usage of the data entities. All three components (definition, correctness of facts, presentation) can be separately and distinctly targeted for ensuring data quality (English, 1999), though not separately measured (Lee et al., 2002). Assuming that this argument holds, the following sections address the semantic aspects referring to the data definition component.

### Semantic Master Data Quality

The *semantic quality category* is an indicator for the alignment of the data stored in the database to the set of external phenomena they are intended to represent and which are relevant to the purpose the data is stored (i.e. the intended use of the data) (Price and Shanks, 2005). Semantic data quality can be assessed through random sampling of this representation. Semantic quality in this sense is an expression for the level of the epistemological view on data entities corresponding to an ontological view describing the external phenomena (Hirschheim et al., 1995).

In contrast to this umbrella understanding, many authors limit semantic quality aspects to the avoidance of inconsistencies arising from redundant information items as illustrated by (Codd, 1970) in his seminal paper but neglect (a) the correct mapping to real-word entities regarding the entire life cycle of both real-world entities and data and (b) the usage of data in new contexts.

(a) *Correct mapping:* This corresponds with the idea of "Information as a Product" (de Corbière, 2007; Price and Shanks, 2004; Wang, 1998; Wang et al., 1998). Information quality can be obtained by applying methods ensuring data's conformance to initial requirements specifications and specified integrity rules or its correspondence to external phenomena, which can be represented by ontologies (Price and Shanks, 2005). Like physical products, data follows a lifecycle. The development of the data is a function of the underlying information systems (which actually produce and change the enterprise's data) development cycle (Huang et al., 1999). Data entities are created, modified and deleted. In the couponing context this problem occurs for example if one product (Coca-Cola, 0.5 litres) is referred to by a follow-up identifier. Then the coupon-product-relations have to be updated, allowing the customer for getting products refered to by both identifiers.

(b) *New contexts:* Not only the master data, but also the represented entities, the business rules linking them to each other and their surrounding business context are dynamic. Consequently, actual operational usage of data may differ "substantially from that considered during system development as a result of omitted, unanticipated, or changing business requirements" (Price and Shanks, 2005). We refer to those external business requirements as *contextual data requirements*. Examples in the retail industry are extended reporting obligations for indication of ingredients (e.g. genetically modified food), hazardous materials (e.g. REACH, ROHS or WEEE) or product specific information for financial reporting requirements. Here, enterprises are faced with several new legislations and regulations (e.g. Sarbanes-Oxley, Basel II Capital Accord) which demand them to "provide, use, and report accurate, verifiable, and relevant data about their financial performance and significant material events" (Berson and Dubov, 2007).

### Meta Data for Semantic Master Data Management

For an understanding of the contextual aspects of master data, we will need to address the meta data of data entities. That meta data contains definitions (and documentation) "relating to either the business application domain or to the underlying data model that together form the IS design" (Price and Shanks, 2005) and thereby bride the identified gap.

## RESEARCH METHOD

Our research is based on the artifact of an "ontology-based semantic master data management". Hence, design science research as a qualitative research method was chosen to guide the research process, incorporating our design oriented and artifact centered research process. As Hevner et al. state: "Design science […] creates and evaluates IT artifacts intended to solve identified organizational problems" (Hevner et al., 2004). Our research outcome is based on a critical thinking, multi-method approach. There is much discussion in the literature about what is meant by critical thinking research. It is an evolving phenomenon with different understandings (Warren and Adman, 1999). Some argue that it operates through critical systems heuristics and a mode of questioning intended to reveal the normative aspects of design processes (Ulrich, 1983). Others demand a pluralistic combination of existing methodologies such as literature study, case study research or expert interviews (Jackson, 1991; Midgley, 1997; Schecter, 1991). Hence, we derived our findings from six case studies (cp. Table 1) in the retail sector, various expert interviews and our experience in master data management (Becker et al., 2007a; Becker et al., 2007b; c; Winkelmann, 2006). Case studies and expert interviews helped gaining an inside into the domains of coupon clearing and master data management.

| Case Study | Company information | Project conducted |
|---|---|---|
| Retailer | Wholesaler with 400 employees, 50 wholesale markets and approximately 30,000 products. | Process analysis, master data reorganisation and new ERP selection and analysis. |
| Retailer | Major retailer in the jewelry sector with 200 stores and 2,000 employees all over Germany. | Reorganization of controlling system. Implementation of a data warehouse. |
| Facility management company | Full service provider for all aspects of real estate. Management of more than 35,000 properties; 6.800 employees. | Implementation of a new master data concept. |
| Manufacturer | Small-sized, one-site manufacturing company with 125 employees, development of metal products with a high degree of manual processes. | Process automating and master data reorganization, implementation of a production planning system. |
| German coupon clearing house | 10 employees, covers approximately half of the German and some of the European retail market with its automatic coupon processing solutions | Development of in-store coupon processing solutions. Development of a centralized master data pool for promotional master data. |
| Manufacturer of European product master data pool | Manufacturer of solutions for managing structured and unstructured data, multimedia content as well as transaction data from a large variety of sources. | Thorough analysis of existing master data pools and its syntactical quality measures. |

**Table 1: Case studies resp. projects in the context of master data management**

## ONTOLOGY BASED SEMANTIC MASTER DATA MANAGEMENT

In this section we describe our approach to master data quality. According to Price and Shanks (2005), as illustrated in before, syntactic data quality can be assessed through integrity checking. Semantic data quality can be assessed through random sampling. Finally, pragmatic data quality – as it depends on user perceptions – can only be assessed through empirical methods such as surveys or interviews. In the following we present a systematic approach for measuring semantic data quality, which follows a more systematic approach than 'random sampling'. The basic idea is to understand an information system as a representation of a real-world system (Weber, 1997). A state of the real-world system at a certain time is represented by the data stored in the information system (Wand and Wang, 1996). As information systems only represent a partial, simplified view of the real world, two questions arise when trying to assess semantic data quality:

- Question 1: Which real-world phenomena should be represented in the information system (i.e. which entities are relevant)?
- Question 2: How should relevant phenomena be represented in the information system (i.e. which attributes and relationships are relevant)?

Answering these questions is a prerequisite to assessing the data quality of a given information system on a semantic level. For doing so, we propose to build a conceptual model to capture the semantics of that part of the real-world we want to represent and to compare this conceptual model to the physical data model of the information system (cp. Figure 1).

One approach to capture the semantics of a domain is to use ontologies. The term ontology originates from philosophy and refers to the study of being or existence, i.e. it is concerned with the question: 'What exists?' Or to be more precisely (Sowa, 2000): 'What categories of things exist?' Hence, a domain ontology explains the types of things in a domain. In the context of information systems an ontology can be defined as "an abstract simplified view of the world that we wish to represent for some purpose" (Gruber, 1993). It consists of a set of representational primitives with which to model the domain of discourse (in the following see (Gruber, 2008)). These representational primitives are typically concepts (or classes), attributes (or properties) of these concepts and relationships between concepts. The definitions of the representational primitives include information about their meaning and constraints on their logically consistent application. In the context of database systems ontologies can be regarded as a level of abstraction of data models and implementation issues.

Having constructed an ontology for a certain domain of discourse of the real world we can compare this conceptual model to the physical data model as well as to the actually stored data of the information system. Any differences might be a possible data deficiency. Figure 1 depicts our approach which can be subdivided into the four phases *analysis*, *modeling*, *mapping*, and *assessment*. Each of these phases contains activities which are connected through artefacts – either as inputs or outputs. In contrast to typical software or database engineering approaches our approach presumes a given information system and maps a conceptual model of the real-world to the elements of this system to assess its semantic quality.

**Analysis**

We start by defining the domain of discourse and its scope (i.e. boundaries). Exemplary data quality domains in the area of retailing are IT-enabled business strategies (e.g. e-commerce), business processes or practices which rely on quality data (e.g. CPFR, ECR, RFID, couponing promotions) or legal regulations which state requirements on data management (e.g. SOX, Basel II).

In the following we will use coupon promotions as a running example. Coupons are a marketing instrument in form of a printed or electronic voucher, which allows either a direct or indirect discount to be obtained if the redemption conditions of the voucher are met (Winkelmann, 2006). In the past, employees had to redeem and clear coupons at the point of sale (POS) manually. Today, automatic coupon clearing becomes ever more important since it avoids the disadvantages (time consumption, risk of failure and fraud) of manual and semi-automatic clearing and enables an automated processing at the POS without any need to intervene (e.g. at self-checkout counters). The same holds for issuing coupons through the POS system. It is possible to print a coupon in combination with the sales slip. For example, it is possible that a customer buys a product of a certain manufacturer and receives a coupon that allows for a rebate with his next purchase of a different product of the same manufacturer. Automated couponing heavily relies on effective and efficient IT support. Hence, the quality of master data is a necessary prerequisite for automated processing of couponing promotions.

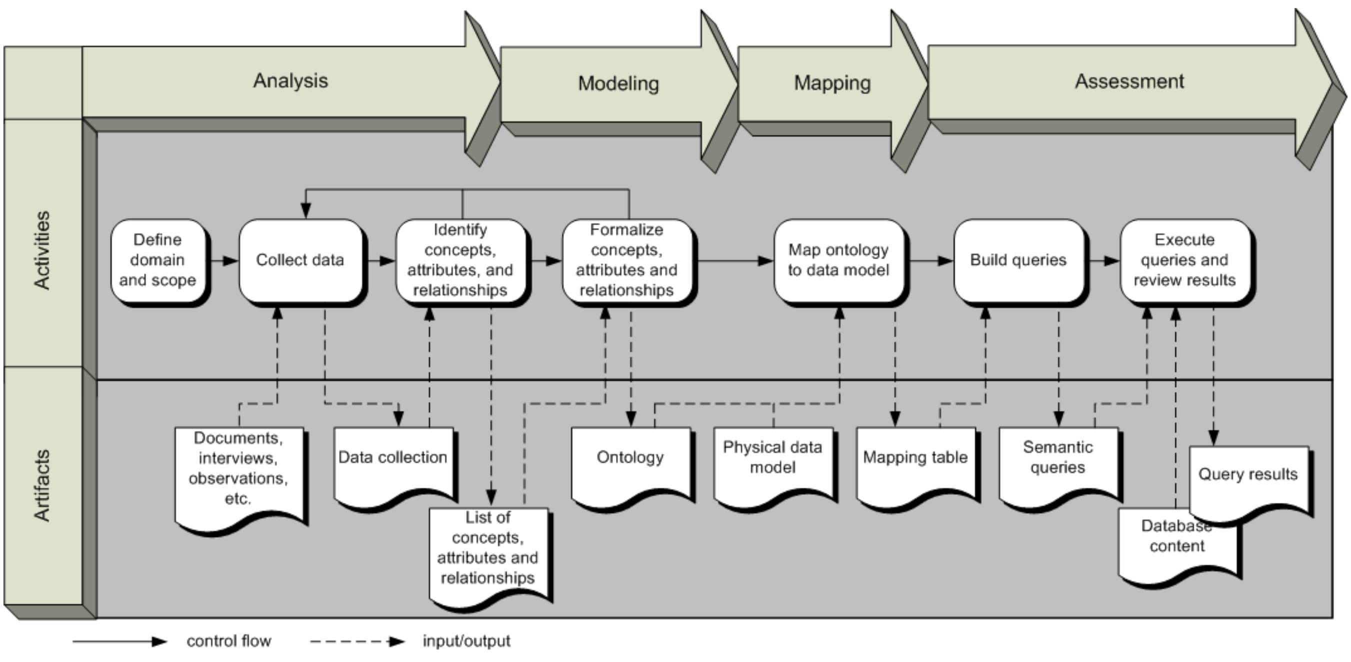


**Figure 1: Procedure model of our approach to semantic data quality management**

It is advisable not only to identify the domain but also to limit the scope carefully. A common technique is to use so-called motivating scenarios and competency questions (Gruninger and Fox, 1995; Ushold, 1996; Ushold and Gruninger, 1996). The motivating scenarios have the form of stories, examples or problem statements, but also include some intuitively relevant solutions. These solutions often provide a first idea of the intended semantics of the ontology. From the motivating scenarios a set of questions arises, so-called competency questions. These questions place demands on the underlying ontology. The primary purpose of competency questions is not to generate ontological commitments; they are rather intended to act as a litmus test for an iterative modeling and evaluation of the ontology (Noy and McGuinness, 2008), e.g.: Does the ontology in its current version contain enough information to answer all competency questions? Is the level of detail appropriate to effectively and efficiently answer the competency questions?

Here is an exemplary motivating scenario for our running example: A retailer wants to promote a new stain remover. He aims at targeting customers who buy products with specific ingredients such as tomatoes, because spots of tomatoes are hard to clean. A coupon for a free package of stain remover shall be printed at the POS, every time a customer buys a product which contains tomatoes, e.g. ketchup, pasta sauce, tomato soup. The promotion is limited to two weeks.

Exemplary competency questions which can be used to decide about the expressiveness of the to-be-constructed ontology:

- What products are made of tomatoes?
- What EANs (European Article Number) belong to products with tomatoes?
- What EANs belong to the stain remover (Frequently, a product has more than one EAN – depending on the point of production, etc.)?
- Is the promotion still running?

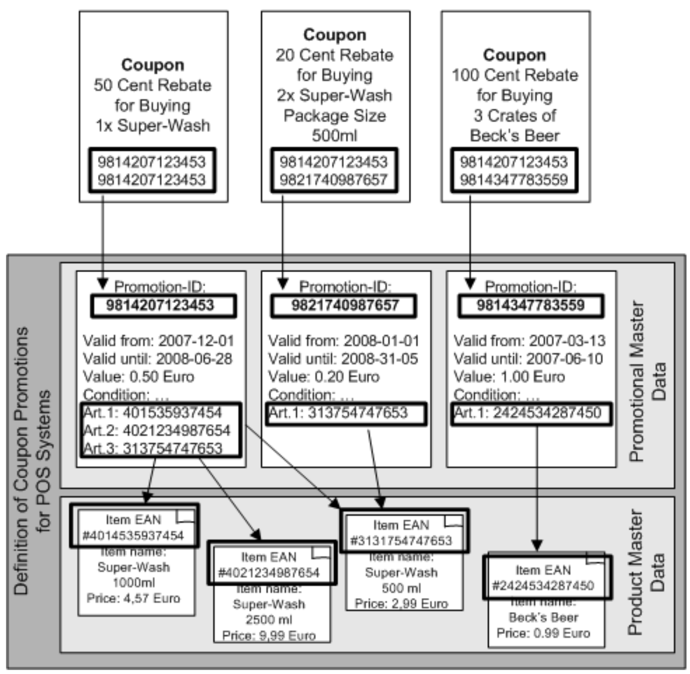


**Figure 2: The roles of coupon, barcode, and master data in automatic coupon clearing**

Once the domain (including its scope) has been defined, data collection follows. Numerous qualitative empirical methods can be used for data collection. Important data sources which may act as input for this task are document analysis (e.g. of laws, regulations, business documents, archival records, reports, meeting minutes), semi-structured, open-ended or focused interviews (e.g. with domain experts, employees) and direct or participant observations (e.g. of employees). The output is a consolidated data collection, e.g. in form of a report, which serves as the input for all further analysis. On the basis of this data collection, central concepts, attributes and relationships inherent in the analyzed domain can be identified. To do this, it might be helpful to first brainstorm important terms of the domain and not to worry too much about classification and overlap (Ushold, 1996). In subsequent steps the collected terms are classified as concepts, attributes or relationships and refined to reduce overlaps. During this task it might be necessary to go back to the previous step of data collection if you are lacking information about certain aspects or if you want quantitatively evaluate identified concepts, attributes and relationships, e.g. by undertaking surveys. The final output of the analysis phase is a semi-structured list of important concepts, attributes and relationships of the analyzed domain.

### Modeling

The purpose of this phase is to formalize the findings of the foregoing qualitative analysis by actually building the domain ontology itself. This typically involves encoding the identified concepts, attributes and relationships using a formal language, e.g. OWL (Web Ontology Language), RDF (Resource Description Framework) or UML (Unified Modeling Language), and an ontology modeling tool (e.g. Protégé).

Using the list of identified ontology constructs as a starting point for formalization, a reasonable approach is to first model the concepts of a domain and to organize them into a hierarchical taxonomy. Afterwards the internal structure of the concepts is specified by adding attributes (and their value types). The next step is defining relationships (including cardinalities) among concepts. The last (optional) step includes creating individual instances of concepts and filling in attribute values as well as relationships to other instances.

Figure 3 shows the ontology of our running example in OWL. Its central concepts are coupons, promotions, conditions, rebates (discounts and rebates in kind), products, and ingredients. Additionally, relationships between concepts do exist, e.g. coupons belong to a promotion, products contain ingredients and products are assigned to EANs. The ontology also contains instances of concepts. For example, the product 'Mama Mia Pasta Sauce' contains the ingredient 'tomato' and has the EAN '300819814711'. Likewise, relationships are defined on instance level. The coupon 'Coupon_4711' has the condition 'Tomato → Stain_Remover' which has the premise ingredient 'tomato' and the conclusion 'get_free_article' of product 'Super_Clean_Stain_Remover'.

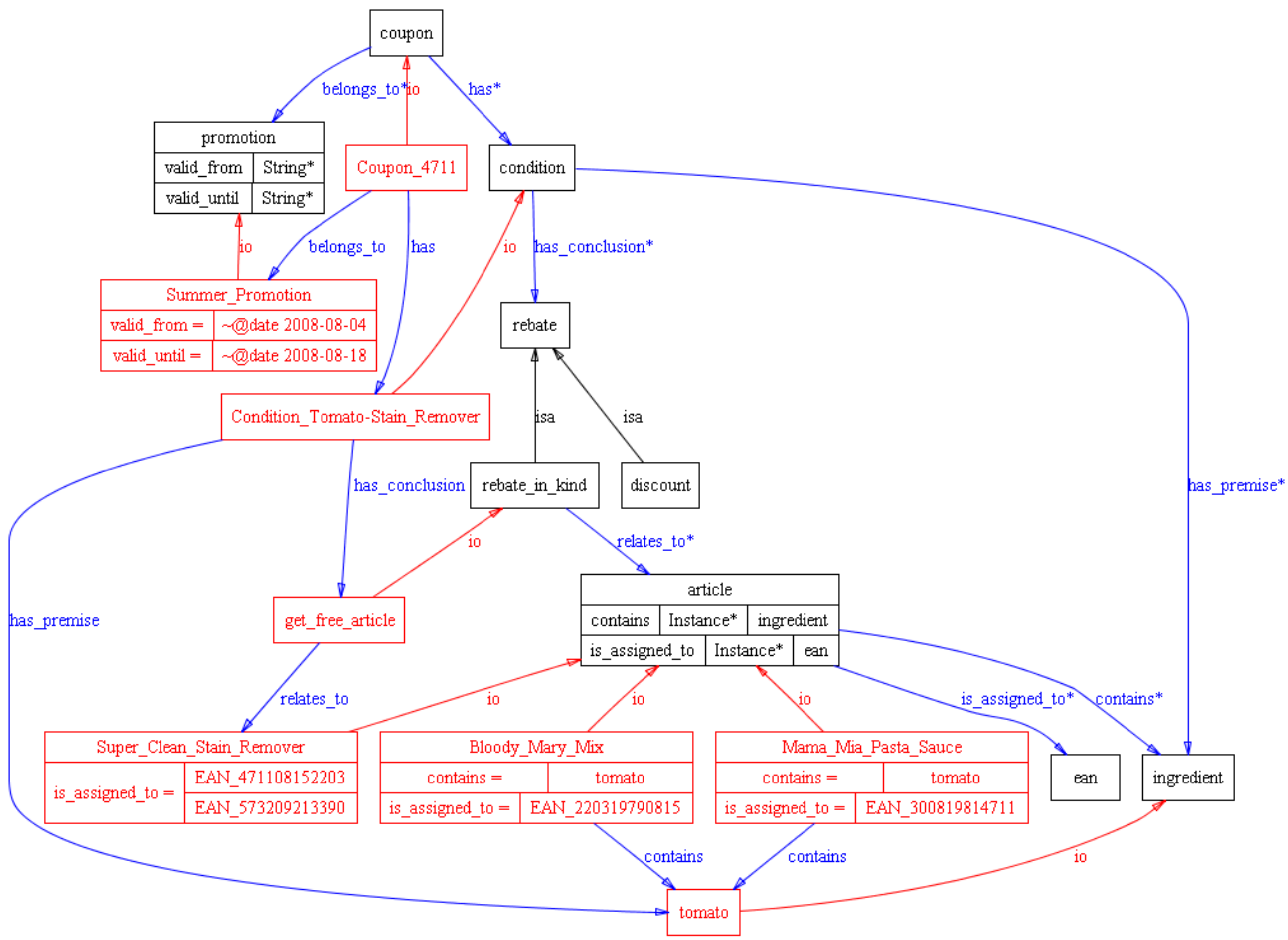


**Figure 3: Exemplary ontology including instances for couponing promotions in retail**

The output of the modeling phase is a semi-formal representation of the business needs regarding a specific context (e.g. couponing promotions). It abstracts from technical issues which are addressed by logical or physical data models.

### Mapping

In order to analyze the semantic data quality of a given information system, a mapping between the domain ontology and the underlying database must be conducted (cp. Figure 4). This mapping links the semantics of the ontology to the physical data model and actual tuples stored in the database. For each concept (incl. attributes) and relationship of the ontology a counterpart in the underlying physical data model has to be identified. Necib and Freytag (2005) identify three types of such mappings: First, a mapping between ontology concepts (incl. attributes) and database relations (i.e. tables or views). For example, the ontology concept ‘article’ could be mapped to a database relation ‘items’. Second, mappings between ontology relationships and database relations. For example, the ontology relationship ‘contains’ could be mapped to a database relation ‘recipe’ containing ingredients of products. This link might be a relation (in case of a many-to-many cardinality) or a simple foreign key constraint (in case of a one-to-many cardinality). Third, a mapping between ontology concepts (incl. attributes) or instances of concepts and database attribute values. For example, the ontology instance ‘tomato’ could be mapped to the value ‘true’ of an attribute ‘contains_tomato?’ in the database relation ‘items’.

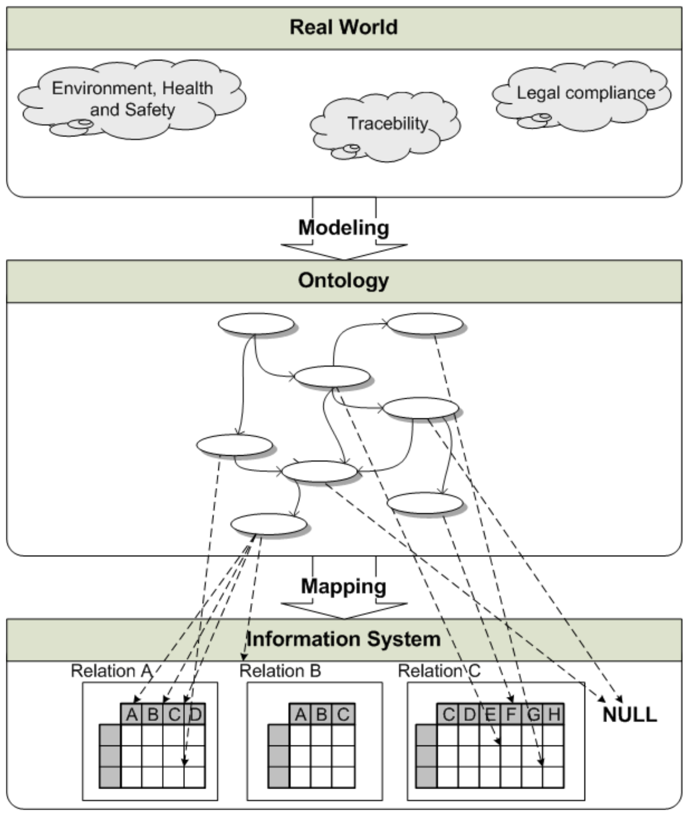


**Figure 4: Presented approach to semantic data quality management**

In some cases it will not be possible to identify a counterpart for an ontology primitive straight away. This might be the case, because a type of phenomenon of the real world is not represented in the information system. For example, a standard ERP system might not include all relations or attributes required for retail (e.g. promotions or coupons). This is called a data model deficiency (as opposed to a data deficiency) which can only be resolved by modifying or enhancing the data model (Shanks, 1999).

The output of the mapping phase is a mapping table containing ontology primitives on the one side, and relations or attribute values of the database schema on the other side.

### Assessment

The last phase comprises the actual quality assessment which involves building semantic queries, executing these queries and analyzing the query results. Due to the previous construction of a domain ontology and the mapping of ontology primitives to the physical data model, analysts can build queries on a semantic level without bothering about implementation details of the underlying database. These semantic queries will then be translated into SQL queries based on the mapping information provided. Executing these queries can answer the question whether all attributes and relationships that are required in a certain context, such as couponing, are filled. In contrast to answering such questions by using SQL queries, our approach overcomes typical interoperability problems, e.g. inconsistent naming of relations, attributes, and attribute values by moving away from technical issues. Figure 5 shows a screenshot from a first prototype incorporating the presented approach to semantic master data quality assessment. The ontological underlying is part of a larger promotion platform prototype (Winkelmann et al., 2008). The prototype supports the cooperation between industry and retail in the field of promotional master data. The green bars at the right hand side of the screenshot represent the degree of coverage between the ontology and the physical database model and content. The quality measure in this first prototype is very simple and is revealed by a linear model that includes the counting of inconsistencies and wrong namings.

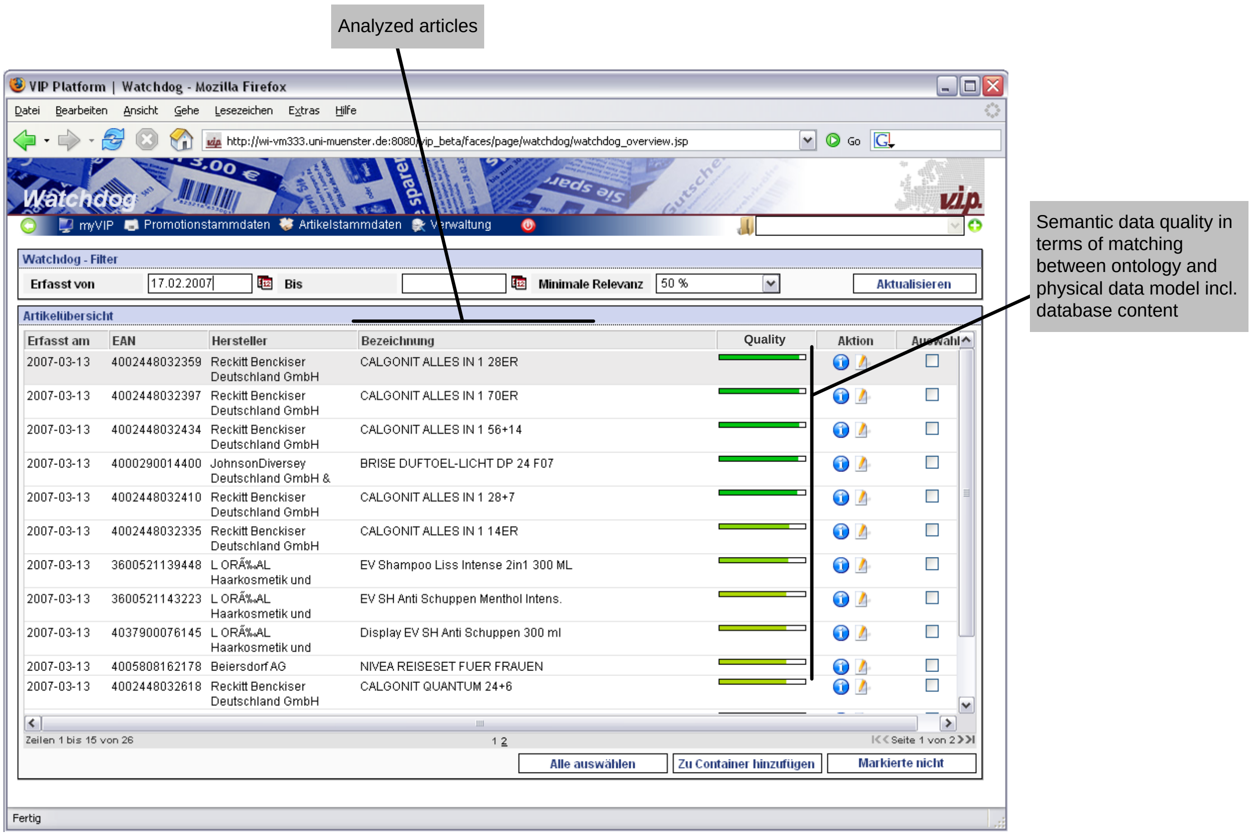


**Figure 5: First prototype incorporating the presented approach**

## CONCLUSION

### Contribution

While syntactic data quality management is well understood (though not satisfactorily solved in practice yet), little practical approaches to semantic quality issues have been proposed so far. Hence, our approach contributes to the discussion by applying ontologies for semantic master data management. The approach can help to solve a variety of issues in the area of semantic data quality. Our approach delivers contributions at two different levels:

(A) data model level

- Modeling real-world contexts in an ontology and mapping the ontology to the physical model of the database ensures a better alignment between real-word entities and database structures.
- Real-world concepts in ontologies can be mapped to several underlying databases (with heterogeneous database schema definitions) at the same time. Therefore, the conceptual approach can for example be applied for database structures of dislocated ERP system implementation scenarios.

(B) data content

- Business users are able to assess the quality of the actual data base *content* by building semantic queries, which abstract from actual implementation details (e.g. are all ingredients maintained?).
- The use of ontologies to capture the semantics of the underlying databases allows for the analysis of data quality without bothering about technical issues.

One actual application of our contributions can be the assessment of new business ideas or the preparation of upcoming business innovations or legal requirements by identifying data needs with the help of a reference ontology. By conceptualizing real-world entities that are not used in today's business but might be of interest tomorrow (e.g. automated promotions on ingredient basis) potential future data deficiencies can be identified. Furthermore, data quality dimensions defined within the ontology can be used to develop quality audit guidelines and procedures.

### Limitations

The results within this paper are based on a limited number of case studies and expert interviews. Due to the interpretive nature of this research we cannot claim that semantic data management has been fully solved. We understand our approach as a starting point towards a semantic data quality management.

### Future Research

We propose to use the procedure model we have introduced in this paper as a sensitizing device to conduct further research in this area of data quality management. Furthermore, our concept has to be applied to various domains in order to evaluate and further develop our approach. We believe that semantic master data quality management will play an important role within business process and IS management in the future in order to further automate business processes and to better align business requirements and data.

## REFERENCES

1. Agosta, L. (2005) Trends in Data Quality, *DM Review*, 15, 2, 34-35.
2. Becker, J., Janiesch, C., and Pfeiffer, D. (2007a) Context-Based Modeling: Conceptualization of a Novel Modeling Approach and Application for the Design of Business Documents, *Proceedings of the 11th Pacific Asia Conference on Information Systems (PACIS 2007)*, Auckland, New Zealand, 143.
3. Becker, J., Janiesch, C., and Pfeiffer, D. (2007b) Reuse Mechanisms in Situational Method Engineering, *Proceedings of the IFIP WG 8.1 Working Conference on Situational Method Engineering: Fundamentals and Experiences*, Geneva, Switzerland, 79-93.
4. Becker, J., Janiesch, C., and Pfeiffer, D. (2007c) Towards More Reuse in Conceptual Modeling: A Combined Approach Using Contexts, *Proceedings of the 19th International Conference on Advanced Information Systems Engineering (CAiSE 2007)*, Trondheim, Norway, 81-84.
5. Becker, J., Uhr, W., and Vering, O. (2001) Retail Information Systems Based on Sap Products, Springer, Berlin.
6. Berson, A., and Dubov, L. (2007) Master Data Management and Customer Data Integration for a Global Enterprise, McGraw-Hill, New York.
7. Capgemini (2004) Internal Data Alignment: Learning from Best Practices. Results of an Internal Data Alignment Survey,
8. Cappiello, C., Francalanci, C., and Pernici, B. (2003) Time-Related Factors of Data Quality in Multichannel Information Systems, *Journal of Management Information Systems*, 20, 3, 71-91.
9. Codd, E.F. (1970) A Relational Model of Data for Large Shared Data Banks, *Communications of the ACM*, 13, 6, 377-387.
10. de Corbière, F. (2007) Interorganizational Information Systems and Data Quality Improvement: The Case of Product Information in the French Large Retail Industry, *Proceedings of the 12th International Conference on Information Quality (ICIQ)*, Cambridge, MA.
11. English, L.P. (1999) Improving Data Warehouse and Business Information Quality, Wiley, New York
12. Griffin, J. (2005) The Master Data Challenge, *DM Review*, 15, 5, 85.
13. Gruber, T. (1993) A Translation Approach to Portable Ontology Specifications, *Knowledge Management Acquisition*, 5, 2, 199-220.
14. Gruber, T. (2008) Ontology, in Liu, L. and M. Tamer Öszu (Eds.) *Encyclopedia of Database Systems*, Springer.
15. Gruninger, M., and Fox, M.S. (1995) Methodology for the Design and Evaluation of Ontologies, *Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI-95*, Montreal.

16. Hevner, A.R., March, S.T., Park, J., and Ram, S. (2004) Design Science in Information Systems Research, *MIS Quarterly*, 28, 1, 75-105.
17. Hirschheim, R., Klein, H.K., and Lyytinen, K. (1995) Information Systems Development and Data Modeling: Conceptual and Philosophical Foundations, Cambridge University Press, Cambridge, MA.
18. Huang, K.-T., Lee, Y.W., and Wang, R.Y. (1999) Quality Information and Knowledge, Prentice Hall PTR, Upper Saddle River, N.J.
19. Jackson, M.C. (1991) Systems Methodology for the Management Sciences, Plenum, New York.
20. Kahn, B., Pierce, E., and Melkas, H. (2004) Iq Research Directions, in Chengalur-Smith, I., L. Raschid, J. Long and C. Seko (Eds.) *Proceedings of the 9th International Conference on Information Quality*, Cambridge, MA, 326-332.
21. Kuipers, P. (2004) Data in Dire Need of a Spring Clear, *Elsevier Food International*, 3, 74-79.
22. Lee, Y.W., Strong, D.M., Kahn, B.K., and Wang, R.Y. (2002) Aimq: A Methodology for Information Quality Assessment, *Information & Management*, 40, 2, 133-146.
23. Loshin, D. (2005) Master Data Standards and Data Exchange, *DM Review*, 15, 8, 72-77.
24. Midgley, G. (1997) Mixing Methods: Developing Systemic Intervention, in Mingers, J. and A. Gill (Eds.) *Multimethodology: The Theory and Practice of Combining Management Science Methodologies*, Wiley, Chichester.
25. Nakatani, K., Chuang, T.-T., and Zhou, D. (2006) Data Synchronization Technology: Standards, Business Values and Implications, *Communications of the AIS*, 17, 44, 962-995.
26. Necib, C.B., and Freytag, J.-C. (2005) Query Processing Using Ontologies, *Proceedings of the Proceedings of the 17th Conference on Advanced Information Systems Engineering (CAISE'05)*, Porto, Portugal.
27. Noy, N.F., and McGuinness, D.L. (2008) Ontology Development 101: A Guide to Creating Your First Ontology., Stanford University, Stanford, CA.
28. Pipino, L.L., Lee, Y.W., and Wang, R.Y. (2002) Data Quality Assessment, *Communications of the ACM*, 45, 4, 211-218.
29. Price, R., and Shanks, G. (2004) A Semiotic Information Quality Framework, *Proceedings of the Proceedings of the IFIP TC8/WG8.3 International Conference*, 658-672.
30. Price, R., and Shanks, G. (2005) A Semiotic Information Quality Framework: Development and Comparative Analysis, *Journal of Information Technology*, 20, 2, 88-102.
31. Redman, T.C. (1996) Data Quality for the Information Age, Artech House, Boston, MA.
32. Schecter, D. (1991) Critical Systems Thinking in the 1980s: A Connective Summary, in Flood, R. L. and M. C. Jackson (Eds.) *Critical Systems Thinking: Directed Readings*, Wiley, Chichester.
33. Shanks, G. (1999) Semiotic Approach to Understanding Represntation in Information Systems, *Proceedings of the Proceedings of the Information Systems Foundations Workshop - Ontology, Semiotics and Practice*.
34. Sowa, J.F. (2000) Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks Cole, Pacifc Grove, CA.
35. Ulrich, W. (1983) Critical Heuristics of Social Planning, Haupt, Bern, Switzerland.
36. Ushold, M. (1996) Building Ontologies: Towards a Unified Methodology, *Proceedings of the Proceedings of Expert Systems 1996, the 16th Annual Conference of the British Computer Society Specialist Goup on Expert Systems*, Cambridge.
37. Ushold, M., and Gruninger, M. (1996) Ontologies: Principles, Methods and Applications, *Knowledge Engineering Review*, 11, 2.
38. Wand, Y., and Wang, R.Y. (1996) Anchoring Data Quality Dimensions in Ontological Foundation, *Communications of the ACM*, 39, 11, 86-95.
39. Wang, R.Y. (1998) A Product Perspective on Total Data Quality Management, *Communications of the ACM*, 41, 2, 58-65.
40. Wang, R.Y., Lee, Y.W., Pipino, L.L., and Strong, D.M. (1998) Manage Your Information as a Product, *Sloan Management Review*, 39, 4, 95-105.
41. Warren, L., and Adman, P. (1999) The Use of Critical Systems Thinking in Designing a System for a University Systems Support Service, *Information Systems Journal*, 9, 223-242.
42. Weber, R. (1997) Ontological Foundations of Information Systems, Coopers and Lybrand, Melbourne.
43. Winkelmann, A. (2006) Integrated Couponing. A Process-Based Framework for in-Store Coupon Promotion Handling in Retail., Berlin.
44. Winkelmann, A., Beverungen, D., Janiesch, C., and Becker, J. (2008) Improving the Quality of Article Master Data Specification of an Integrated Master Data Platform for Promotions in Retail, *Proceedings of the 16th European Conference on Information Systems*, Galway, Ireland, to appear.