The MUSING Approach for Combining XBRL and Semantic Web Data

~ Position Paper ~

Christian F. Leibold[1], Dumitru Roman[1], Marcus Spies[2]

[1] STI Innsbruck, Technikerstr. 21a,
6020 Innsbruck, Austria
{Christian.Leibold, Dumitru.Roman}@sti2.at

[2] LMU München, Leopoldstr. 13,
80802 München, Germany
marcus.spies@lrz.uni-muenchen.de

Introduction

Communication between the Enterprise and its customers, investors and regulatory authority as well as analysts or rating agencies is a vital factor in competition for customer and market shares. Efforts to harmonize financial markets and the financial crises in the US and Europe have led to the modernization of budget publication directives by governing bodies and the invention of international standards. Therefore, changing from a national to international reporting principles is not only of relevance for stock market traded companies. Many enterprises revisit their organizations' structures for controlling and reporting, because conventional structures have proven to be to complex and inflexible in facing the emerging requirements. In future all divisions of an enterprise need to collaborate and contribute information to the internal decision making process and external communications. The challenge at this point lies in a meaningful provision of enterprise data, tailored for the particular case and in an internationally comparable manner. This task provokes the call for an approach adopting standardized State-of-the Art knowledge from several fields like ICT and the financial industry.

In the EU funded project MUSING (Multi-Industry semantic Business Intelligence; IST Project Number FP6 – 027097) we deliver solutions and services to enable perceptive Business Intelligence, mainly show cased in three discourse domains covering (IT-) operational, financial and internationalization risks. Novel applications combine knowledge management, advanced predictive analytics and intelligent access to third party data through the integration of semantic technologies. The foundational backbone for the integration of this knowledge and project results is formalized in ontologies which themselves serve as schema for the MUSING knowledge base repository. This includes the possibility to map from/to XBRL (eXtensible Business Reporting Language, [1]) compliant data.

XBRL is an emerging XML-based taxonomy designed to allow budgeting, management, financial and regulatory reporting, controlling and auditing, and credit worthiness evaluation to share a common language. XBRL is a tagged language: by translating specific elements of the financial statements into a format that is understandable by a computer, XBRL uses a series of common definitions or ‘tags. As such, XBRL is one of the basic technology enabling the application of Basel II requirements in terms of public disclosure and reporting, as well as for on-line and real-time auditing purposes. At the time of writing, only a few European countries have established XBRL jurisdictions, and the road to capillary deployment of XBRL in the European financial system is still long. Notwithstanding that, it is widely acknowledged that XBRL is the way to follow for the next 5-year IT scenario in the field.

In this paper we report activities that are related to the instrumentalization of XBRL for semantic BI. Starting with the general picture of translating Balance sheet information into XBRL, we present a Meta model that allow the semantic integration and use of XBRL qualified data in (OWL/RDF) ontologies. The paper ends with an assessment of identified challenges and an outlook to further work.

Enabling XBRL in OWL

The Extensible Business Reporting Language (XBRL) is a building block for the MUSING developments. On our route to the semantic exploitation of XBRL for MUSING application we evaluated and combined several approaches. In the following we present only core results.

A simple starting point for the use of XBRL in a company is the reuse of information reported in balance sheets. In general, a balance sheet (BS) contains a report to the financial governing bodies (or investors). Sometimes balance sheets are only available in PDF, but XBRL compliant as the reported items are represented in the standard. In Figure 1 we show our approach to semantically exploit the information provided. We start to classify information from BS into XBRL taxonomy using natural language processing. With this we translate previously unprofitable data into information that is machine read and process able that can be reused. On the other hand we apply a model based transformation approach to deduct an OWL (and with this RDF) representation of XBRL. This is used in two ways, to a) condition the XBRL specifications for the use and access in ontologized relations and to b) make an XBRL instance available as instance of ontology concepts.

Figure 1: The MUSING approach to enable balance sheet information in BI – a transformation based approach to reuse information in ontologies (RDF and OWL)

Representation of XBRL balance sheet structures

The XBRL specification organizes balance sheet (BS) in a conceptual hierarchy that reflects reporting areas like assets, liabilities etc. in different levels of detail. This hierarchy, however, is defined only in the presentation view, it is not part of the XML BS data model. The BS items appear in the data model as juxtaposed elements without grouping or forming complex elements. It is only by reference to the presentation hierarchy that the conceptual structure is visible to the data model. This construction is very reasonable from a data warehousing (DW) perspective – in fact, BS item instances should correspond to facts in a DW fact table, while the conceptual hierarchy offers dimensions along which to perform roll-up or drill down analyses for data from multiple companies and / or reporting periods (these dimensions are captured in context data associated to the BS items). The XBRL BS monetary items can be aggregated with a suitable aggregation operator along the roll-up levels of the different reporting dimensions (often, the operator used is simply summation). The dimensions correspond roughly to the balance sheet sections.

A suitable ontology representation for this kind of DW dimensions is a part-of relationship, not a simple subclass-class hierarchy (as it is proposed in the ABRA taxonomy viewer and related tools). The parts to be related here, however, are subtables of the balance sheet facts. E.g. all asset items are aggregated in several steps along the assets dimension etc. The lowest aggregation levels for a given dimension are represented by tables containing all relevant atomic balance sheet items from the reporting area being considered. For instance, in the section assets, for the lowest level subsection unpaidCap, we have a table with the monetary items for called, uncalled and dueCapofCoop (see Figure 2).

Figure 2: Elements of the XBRL balance sheet XML structure. The element taxonomy is represented by the element naming using dotted notation. The number of dots in an element name represents its hierarchy level from top to bottom.

A similar approach to XBRL ontologies has been developed by a MUSING partner, German Research Centre for Artificial Intelligence, and is used as backbone schema for the translation of PDF2XBRL.

Integration in MUSING ontology family

As mentioned before all knowledge in MUSING is conceptualized in ontologies. Generally we apply a layered architecture that separates concerns for the specific groups of ontologies. Starting with a layer of generic and axiomatic ontologies that groups upper ontologies for world knowledge (we use Proton) and the representation of temporal axioms we integrate them by single point of access to the second layer of domain independent ontologies. Here an elaborate enterprise model is ontologized and general Business Intelligence concepts are defined. At this level as well integrates the group of standard reference ontologies that enables MUSING applications to be compliant to regulations and agreed standards (XBRL is in this group, too). Finally the lowest level contains ontologies that represent domain and application specific knowledge in specialized concepts or contextualized parameters for highly specific calculations (e.g. for risk ratings).

To generally embed XBRL in the MUSING ontology family some structural knowledge about XBRL needs to be conceptualized, which is depicted in Figure 3.

We use a model based transformation approach to translate from schema to ontology format that allows the integration and access from other concepts of the MUSING ontological family. Details and scientific elaboration on the background can be found in [3].

The namespaces XBRL "Instance", XBRL "Linkbase" are designed to **represent the structure of the XBRL schema files (instance and linkbase)** suitably for the MUSING ontologies. In the current version we have applied the translation to the HGB XBRL taxonomy, which comprises German trade legislation and Generally Agreed Accounting Principles (GAAP; [5]) and takes best practices, tax laws and doctrines into account.

Further we have transformed the structure of instance documents of XBRL, that is defined in the schema document for the namespace http://www.xbrl.org/2003/instance and is available online at [4].

The resulting XBRL ontology contains all relevant knowledge to access XBRL structural information from the linkbase and specific components reported by a company in an instance document. It explains the relation between tuples and their components. This represents the major impact of the current XBRL ontology implementation.

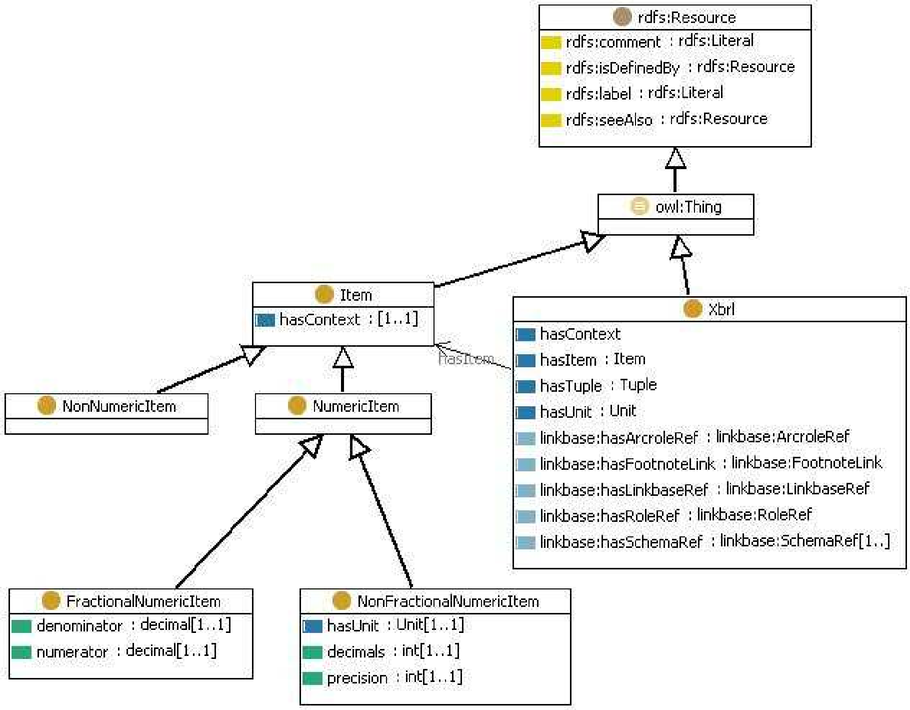

Figure 3: Visualization of the ODM compliant ontology construction principles for XBRL balance sheet structures. The levels of the XBRL taxonomy are represented from top to bottom by functional sub-properties of a has-part property.

While the XBRL taxonomy expresses the aggregation levels only implicitly, an ontology representation compliant with the ontology definition meta model (ODM, [2]) makes the dimensions for a data warehouse application obvious. The proposed representation, therefore, provides added value compared to the XML schema definitions.

Rationale

The position of MUSING in XBRL related activities can be summarized as follows:

1. XBRL taxonomies can be represented in ontologies rather than being confined to the presentation layer as today. The benefit is that taxonomies become more manageable, it becomes easier to integrate new concepts and relationships from different taxonomies, and it is easier to represent and use them in applications. The integration can also be extended to cover taxonomies from other standardisation bodies that are not part of the XBRL process, like the NACE taxonomy for company activities or NUTS for regional information. More generally, this representation allows to use and foster information representation standards for the integration of data and knowledge representation as proposed by the OMG ODM (ontology definition metamodel).

2. XBRL aggregation, reclassification and computation rules can be explicitly represented in ontologies using a standard rule representation language. The rules could be attached to classes representing regulatory systems (like national legislations). This is a quite challenging task compared to an ad-hoc

solution, but, if eventually achieved, it makes adaptation of XBRL to local or regional standards much easier. In addition, explicit representation of rules allows for better change management.

3. XBRL references to legal requirements and rules can be moved from the layer of physical information links to a logical representation layer that allows for comparisons between different requirements and a logical structure to the representation of the requirements (e.g., in terms of relevant business practices, or auditing rules, etc.). This will relate our activities to the OMG semantics of business vocabularies and business rules (SBVR) standard.

4. An important development in XBRL is the inclusion of taxonomies and other reporting structures for intangible assets (e.g. company knowledge, customer capital). The usage of ontologies can support this development, since it allows for a common conceptual integration of intangible and conventional "tangible" assets etc. reporting. A major step in this direction has been done in the last version of the MUSING ontologies by integrating a conceptual enterprise model derived from several quasi-standard initiatives in enterprise modelling.

5. An important development in IT operational risks (IT Op-Risk) is that Risk mapping can be represented formally, and in an explicit way. This allows the comparison of risk estimation in different companies. This in turn allows the construction of shared pooled risk databases which can be used as sources of “external data” rather useful to assess and predict risks.

6. One of the most important activities in Business Continuity Management is Risk identification and estimation. Therefore, by providing standards for IT Op-Risk, we also provide standards for the “quantitative” part of Business Continuity management

7. Intangible capital and non financial indicators. The challenge for many executives is to see the link between intangibles and business performance — providing an addendum to the annual report does not build this link. This is where Business Intelligence applications might play a key role by enabling organizations to build planning models that link financial and nonfinancial business drivers, thus giving executives a way to understand the potential impact of intangibles on financial performance.

Outlook & Challenges

Some technical details still are unresolved (e.g. datatype properties in the ontology do not meet all necessary restrictions given by the XBRL definition). Nevertheless, work-arounds in the specific contexts have been developed.

Internationalization: Can we apply our approach for an integrated use of information reported in several “GAAP” and IAS/IFRS taxonomy dialects? How can we translate between them? Is that a mentionable requirement at all?

Purpose identification: Does it make sense to build up a knowledge base that reasons over billions of information items? (this would be only a couple of thousand balance sheet instances). Meaning, do we need to have everything from a reported XBRL instance available for reasoning?

Problem of focus: of the reuse of the information from all analyzed balance sheets or - preferably – extract information from the balance sheets and populate associated ontology concepts.

References

[1] http://xbrl.org/frontend.aspx?clk=LK&val=20

[2] Object Management Group. (2006). Ontology Definition Metamodel Specification: Object Management Group.

[3] Spies, M. (2009). An ontology modelling perspective on business reporting languages. Information Systems, Elsevier, (in press, DOI information: 10.1016/j.is.2008.12.003).

[4] http://www.xbrl.org/2003/xbrl-instance-2003-12-31.xsd

[5] http://www.abra-search.com/ABRASearch.html?locale=de&taxonomy=de-gaap-ci-2007-12-01