Christophe Debruyne, Pieter De Leenheer, and Robert Meersman. Empowering enterprise data governance with BSG. In M. A. Musen and Ó. Corcho, editors, K-CAP, pages 197-198. ACM, 2011

Empowering Enterprise Data Governance with BSG

Christophe Debruyne
Semantic Technology and Applications Research Lab, Vrije Universiteit Brussel, Brussels, Belgium
chrdebru@vub.ac.be

Pieter De Leenheer
Business Web & Media, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
Collibra nv/sa, Brussels, Belgium
p.g.m.de.leenheer@vu.nl

Robert Meersman
Semantic Technology and Applications Research Lab, Vrije Universiteit Brussel, Brussels, Belgium
meersman@vub.ac.be

ABSTRACT

Domain rules are important for businesses to obtain good data governance. Although efficient for storing and processing data, the use of popular semantic technologies alone does not suffice. As the Web is gaining a prominent role for enterprises (and communities in general), appropriate methods and tools are required for data governance, with a proper emphasis on facts in natural language. This paper presents Business Semantics Glossary that supports the a method called Business Semantics Management.

Categories and Subject Descriptors

I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods

General Terms

Management, Standardization

1. INTRODUCTION

The Linked Data (LD) initiative is an important first step to unlock hidden data, and make Web-based access to it scalable. Yet in order for a LD service market to flourish, one has to consider the governance aspects. Especially in a business context, domain rules are often required [6]. To illustrate the point consider the problem of providing (linguistically, lexically) *identifying references* for concepts/entities. In LD, the primary reference structure for concepts and their instances is a URI rather than a perhaps more "communication-oriented" reference scheme based on (the agreement on) the combination of *identifying attributes*.

Agreements are made by communities, stakeholders with a common goal representing *autonomously developed* information systems. Community involvement thus is essential for system and enterprise interoperability. Reaching a common agreement between many stakeholders proves to be difficult, and thus a methodology for communities to develop and maintain a representation of their world is needed [2]. Such methods can learn from following DB modeling principles.

Firstly, the non-involvement of non-tech savvy domain experts is not longer an excuse. Wiki technology has been put forward as a mean to reach agreement and share knowledge about different subjects. Secondly, Analyzing natural language discourse. Fact-oriented (database) design methods such as NIAM [7] and ORM [3] already showed that the closer the link between human natural language communication and the system and/or business communication that results from it, the more likely such systems will work as intended by their various stakeholders. This is particularly important for interfaces where humans, systems and businesses interact, as the human discourse needs to be mapped meaningfully onto application symbols. These techniques furthermore allow scalable solutions to ontology engineering through a *separation of concerns* - as done in databases - by separating the schema level from the instance level. As a consequence, applications become minimally sensitive to changes in data representation. Thirdly, employing legacy data, output reports, and interviews with domain experts as fulcrum for leveraging validation. In the case of ontology engineering: lift data models into ontologies by removing application specific context (e.g. non-conceptual identifiers such as an automatically incrementing key).

Copyright is held by the author/owner(s).
K-CAP'11, June 26–29, 2011, Banff, Alberta, Canada.
ACM 978-1-4503-0396-5/11/06.

2. BUSINESS SEMANTICS MANAGEMENT

Business Semantics Management (BSM) [1] draws from best practices in ontology management and ontology evolution. The representation of business semantics was originally based on the DOGMA [4] approach, which allowed the application world to be associated with a lexical world relying on the fact that the knowledge building blocks expressed in natural language are easily obtained and agreed upon. Recently, BSM adopted Semantics of Business Vocabulary and Business Rules (SBVR) [5], an OMG standard pushed by the fact-oriented modeling community and fully compatible with DOGMA. BSM consists of two complementary cycles: *semantic reconciliation & application.*

In semantic reconciliation, business semantics are modeled by extracting, refining, articulating (e.g. providing definitions) and consolidating fact-types from existing sources. Ultimately, this results in a number of consolidated language-neutral semantic patterns that are articulated with informal meaning descriptions (e.g. WordNet senses) and that are reusable for constructing various semantic applications.

In semantic application, existing information sources and services are committed to a selection of semantic patterns. This is done by selecting the relevant patterns, constraining their interpretation and finally mapping (or committing) the selection on the existing data sources. In other words, a commitment creates a bidirectional link between the existing data sources and services and the business semantics that describe the information assets of an organization. The existing data itself is not moved nor touched.

Figure 1: Screenshot of Home Address in the Location vocabulary. Even though the concept and its relations look like natural language, one can automatically generate a formal specification (e.g. in RDFS). This screenshot shows the fact types (or lexons), descriptions in natural language, the concept's place in a taxonomy and the steward of that term. The Location vocabulary is part of the School Indications speech community, in turn part of the Flemish Public Administration community.

3. BUSINESS SEMANTICS GLOSSARY

Business Semantics Glossary (BSG), see Figure 1, is a web-application aimed at both business as well as technical users. It lets people collaboratively manage their business semantics according to the BSM methodology. BSG is based on the Wiki paradigm that is a proven technique for stakeholder collaboration and is essential for evolving business semantics. Governance models are built-in and user roles (e.g. steward, stakeholder) can be applied to distribute responsibilities and increase participation. The software takes care of the audit trails who changed what, when and why.

Once semantic applications are running, it must be possible to monitor and feed unexpected side effects or failures back, calling for a new iteration of BSM. The BSG is the vehicle that serves the reconciliation of the newly scoped concepts. The BSM cycle is repeated until an acceptable balance of differences and agreements is reached between the stakeholders that meets the requirements of the semantic community. Gradually, closed divergent metadata sources are replaced with metadata sources that follow an open standard, and are kept coherent via BSG.

After a consensus has been obtained using BSM with the glossary, the terms and relations in the ontology can be *implemented* in other formalisms such as OWL and RDFS. BSM is thus *not* an alternative representation for ontologies on the Semantic Web, but a method, tool and representation *for users* to reach an agreement on their world that *preceeds* the implementation in Semantic Web languages.

SBVR's structure allows implementing a business semantics system that takes into account the existence of multiple perspectives on how to represent concepts by means of vocabularies (a set of terms and fact types drawn from a single language to express concepts), and includes the modeling of a governance model to reconcile these perspectives pragmatically in order to come to an ontology that is agreed and shared (by means of communities and speech communities) [1]. A *semantic community* is a group of stakeholders having a body of shared meanings. Stakeholders represent an organization or a business unit (and their autonomously developed information system). A *speech community* is a sub-community having a shared set of vocabularies to refer to the body of shared meanings; it groups stakeholders and vocabularies from a particular natural language, e.g. jargon.

4. CONCLUSIONS

RDF and LD brought us one step closer to a Semantic Web. For businesses and organizations (and their communities) to flourish on this new service market, agreements on the data and their domain rules need to be obtained by the community before the ontologies are implemented in RFDS and OWL. For this an appropriate method and tool are needed such as the Business Semantics Method and Glossary presented in this paper. The method and tool adhere to the three principles presented in the first Section: involvement of non-tech savvy experts, natural language discourse with facts and the use of external sources as references. When the community reaches an agreement after each iteration, the model created by the community with facts in natural languages can then be implemented in other formalisms ideal for machine processing.

5. REFERENCES

[1] De Leenheer, P., Christiaens, S. & Meersman, R. Business semantics management: A case study for competency-centric HRM. *Computers in Industry*, 61(8):760–775, 2010.

[2] de Moor, A., De Leenheer, P. & Meersman, R. DOGMA-MESS: A meaning evolution support system for inter-organizational ontology engineering. In *Proc. of the 14th Int. Conf. on Conceptual Structures (ICCS 2006)*, vol. 4068 of *LNCS*, pp 189–203. Springer, 2006.

[3] Halpin, T. *Information Modeling and Relational Databases*. Morgan Kaufmann, San Francisco, CA, USA, 2008.

[4] Jarrar, M. & Meersman, R. Ontology engineering - the DOGMA approach. In Dillon, Chang, Meersman & Sycara, eds, *Advances in Web Semantics I*, vol. 4891 of *LNCS*, pp 7–34. Springer, 2009.

[5] OMG. Semantics of business vocabulary & business rules, v1.0. http://omg.org/spec/SBVR/1.0/, 2009.

[6] Spyns, P., Meersman, R. & Jarrar, M. Data modelling versus ontology engineering. *SIGMOD Record Special Issue*, 31 (4):12–17, 2002.

[7] Wintraecken, J. *The NIAM Information Analysis Method, Theory and Practice*. Kluwer Academic Publishers, 1990.