

Economic and Financial Information Management and the Semantic Web

Rubén Lara and Borja Foncillas

Tecnología, Información y Finanzas (TIF), Madrid, Spain.
<http://www.grupoanalistas.com/>
{rlara, bfoncillas}@afi.es

Abstract. Economic and financial information has critical value for all kinds of institutions. An effective generation, gathering, integration and analysis of such information is a key success factor for any business and, therefore, requires techniques that reduce the current burden of dealing with big volumes of rich information. While the information integration problem is common to all business areas, it becomes apparent in the financial area, where information coming from multiple sources has to be homogenized in order to facilitate analysis and to provide a uniform view of the business e.g. via a balance scorecard. In addition, financial information is subject to official supervision e.g. by central banks. As a consequence, the required information has to be obtained, homogenized following the supervisor rules, and validated, guaranteeing its quality. In this paper, we present part of our experience in the use of semantic Web technologies for the management of economic and financial information and the problems encountered. In addition, we briefly introduce XBRL, a language that has gained a big momentum for business reporting, and compare the paths followed by the XBRL community and the semantic Web community.

1 Introduction

The field of economy and finance is a conceptually rich domain where information is complex, huge in volume and a highly valuable business product by itself [2]. While some business sectors make an extensive use of this kind of information e.g. credit institutions, insurance companies or investment firms, all other sectors also need to manage and analyze their own financial information and, furthermore, the quality of such management and analysis becomes a key business success factor. The volumes, complexity and value of economic and financial information makes it a particularly interesting target for the application of novel information integration techniques. For this reason, an evaluation of what the semantic Web has to offer in this context can serve as a relevant feedback to the semantic Web community, and as a valuable information for businesses who might decide to promote or adopt semantic Web technologies.

Analyzing economic and financial information, providing value-added analysis of such information, and helping our customers to better manage and exploit their information assets are precisely our core activities. Tecnología, Información y Finanzas (TIF) is part of Grupo Analistas, a corporation that generates high-quality economic and financial information (equity research notes, newsletters, analysis, sector reports, recommendations), and provides technology solutions for information consumers to access,

manage, integrate, exploit and publish this information. The consumer profile of this information is diverse, including financial institutions, banks, SMEs that use the information in decision making and foreign trade activity, and distributors who publish the information in first-rank printed and digital media. An important group of professionals and domain experts in the company is in charge of daily generating a wide range of economic and financial information, including economic, market, bank, and financial analysis, commercial fair reports, import/ export offers and news and manuals, among others. Banking consultancy is also part of our services, which includes the integration and analysis of heterogeneous information, and the provision of management information services that support decision-making by offering a coherent and unified view of key financial data.

Due to the nature of our services, we have investigated semantic Web technologies as a potential means to improve the solutions we offer to our customers and to improve our internal information management and consumption processes.

In the following, we will summarize the results of our experience (Section 2) and conduct a discussion around the eXtensible Business Reporting Language (XBRL)[1] (Section 3). The aim of the discussion is to compare the paths followed by the XBRL community and by the semantic Web community, trying to extract conclusions on what industry is missing from the semantic Web approach.

2 Experience on Applying Ontologies to Economic and Financial Information Management

In [2] we presented a research project done in cooperation with Universidad Autónoma de Madrid (UAM)¹ to develop an ontology-based platform which adds explicit semantics to part of our information, and uses such semantics for visualization and search purposes. We developed, as the basis for this work, an ontology to explicitly describe the chosen part of our business domain.

As a result, we experienced with the development of ontologies, the annotation of existing content and the exploitation of the added annotations; details can be found in [2]. During the project we found that some tools were not ready for industrial development, and efficiency was still an open problem.

One of the most important steps we accomplished was the creation of a domain ontology from scratch, as no economic or financial ontology was available by that time. However, things have not changed much since then: ontologies that reflect the real consensus of entities producing and consuming information in the same domain are hardly available, which eliminates the possibility of expressing information in some agreed terms and, therefore, seriously reduces the benefits of using ontologies and semantic Web technologies.

The major conclusion of the project has been two-fold:

- Making a commonly used conceptualization of a domain explicit has some benefits, such as providing a reference for communication (both among persons and computers) and helping to improve data quality based on this conceptualization.

¹ <http://www.uam.es/>

- The major promise of ontologies and the semantic Web has been making (semi)automatic interoperation possible. However, the biggest challenge has not been solved: facilitating the definition of **real** ontologies i.e. ontologies that really reflect the consensus of a critical mass of people so that they can be actually shared and reused.

3 XBRL: a Lost Opportunity or a lesson to be learnt?

In parallel to the semantic Web research, the eXtensible Business Reporting Language has gained momentum as a means of exchanging business reporting information. In the following, we briefly introduce XBRL, the acceptance it is gaining in the financial domain, and we relate it to semantic Web research.

3.1 XBRL in a nutshell

XBRL is a language that builds on top of XML and XML Schema to provide users with a standard format in which information can be exchanged, enabling the automatic extraction of information by software applications [1]. For that purpose, XBRL defines **taxonomies**, which provide the elements that will be used to describe information, and **instances**, which provide the real content of the elements defined.

Taxonomies make use of five different types of XLink linkbases, namely: definition linkbases, calculation linkbases, presentation linkbases, label linkbases and reference linkbases. The first three types contain different kinds of relations between elements, whereas the last two types contain documentation of elements.

Definition links describe relations among concepts in a taxonomy, such as general-special relations, that provide information on what an element actually is e.g. the specialization of some other concept. Calculation linkbases provide information on how some elements are calculated in terms of some other elements, which can be exploited for data validation. Presentation linkbases contain relations such as parent-child that are exclusively used for presentation purposes e.g. a given element will be shown as the child of some other.

The last two types of links do not define relations among elements but document elements in a taxonomy. Label links provide labels in natural language with the purpose of facilitating the understanding of data by a human user. XBRL comes with multilinguality support and enables the user to associate labels in different languages to the same element. Reference links point to legal or other type of documentation that explains the meaning of a given taxonomy element.

Instances are then created according to the elements defined in a given set of taxonomies and linkbases, constituting what is called a Discoverable Taxonomy Set (DTS)[1], and therefore structured, documented and interpreted according to such DTS. This gives data an agreed meaning in a similar way ontologies do.

3.2 XBRL in the economic and financial domain

XBRL is intended to cover the needs of a particular domain, and it has evolved with the needs of such domain in mind. At the time of writing, XBRL is being promoted by

public institutions such as the Committee of European Banking Supervisors (CEBS)², which includes high level representatives from the banking supervisors and central banks of the European Union. CEBS has promoted the creation of working groups that have the mission of defining XBRL taxonomies to be later adapted and used for the financial reporting that banks and other institutions have to submit periodically to the banking supervisors. One of these working groups is COREP³, devoted to the creation of an initial taxonomy for common solvency ratio reporting in the context of the new Basel II capital agreement⁴.

COREP has already led to the creation of initial taxonomies that will be reviewed when the directive concerning the new capital agreement of the EU is final, and later adapted (if necessary) by the national supervisors but keeping a common ground. Grupo Analistas, as a member of the COREP group, is contributing to the development of this initiative.

XBRL is not only being promoted by CEBS, but also by other public institutions such as CNMV⁵, the supervisor of the Spanish stock market, and a number of taxonomies are already approved, such as the US GAAP taxonomies. In addition, it can be seen that important software vendors e.g. Fujitsu⁶ have already paid important attention to XBRL and some have even published their financial data in XBRL e.g. SoftwareAG⁷ or Microsoft⁸.

3.3 XBRL and the Semantic Web

While XBRL is more domain specific than semantic Web technologies, they have a fairly similar goal: making the semantics of information explicit so that information exchange can be eased. Towards this goal, XBRL seems to have focused on a language heavily based on existing standards, adding as little as necessary to provide a usable language for business reporting, while the semantic Web community has proposed and evolved a number of (sometimes complementary and sometimes competing) standards to solve the semantic problem for any kind of domain, without committing to any particular application area.

Even though it might seem that the scope of XBRL is more limited than the scope of semantic Web research, XBRL is being more effective than semantic Web technologies on having a real impact. From our point of view, the major reason for this is that XBRL has made an effort to have an application domain and to identify the needs of technology users. If a domain is to be formalized and consensus is to be reached, that can never happen without having in place the entities and people that should reach that consensus. On that regard, the semantic Web has advanced in some particular topics but has forgotten its pre-requisite: finding **shared** conceptualizations i.e. ontologies.

² <http://www.c-ebs.org/>

³ <http://www.corep.info/>

⁴ <http://europa.eu.int/rapid/pressReleasesAction.do?reference=IP/04/899&language=en&guiLanguage=en>

⁵ <http://www.cnmv.es/>

⁶ <http://xml.fujitsu.com/xbrl/>

⁷ <http://xbrl.softwareag.es/xbrlSageSite/index.html>

⁸ <http://www.microsoft.com/msft/xbrlinfo.mspix>

Currently, really agreed, ready-to-use ontologies are hardly available, while XBRL is already succeeding on defining agreed taxonomies and on bringing together a critical mass of business experts, bank supervisors, technology providers and consultants to agree on taxonomies.

Still, XBRL will face similar problems to the ones semantic Web researchers are already working on, such as taxonomy mapping and alignment, semi-automatic extraction of instances from different data sources, etc. Therefore, it might be worthwhile for the semantic Web community to explore the XBRL effort and to combine strengths.

4 Conclusions

The economic and financial domain is a particularly interesting domain for the application of novel techniques that could improve current information integration processes. TIF has experienced the application of such technologies to its business, realizing that the process of creating ontologies has hardly started in industry. On the other hand, XBRL is succeeding at attracting a critical mass of business actors, which is enabling the creation of agreed taxonomies that will be used in the near future, promoted by e.g. bank supervisors. A brief study of both lines of research can easily show that XBRL has been running closer to industry, while the semantic Web community has partly forgotten the pre-requisite to have a semantic Web: having ontologies. Therefore, we expect semantic Web research to be slightly redirected to work closer with industry and with other efforts such as XBRL that are working closer to the needs of technology users. Semantic technologies, if there are no ontologies providing the semantics to work with, will be of little use.

References

1. eXtensible Business Reporting Language. Technical report, XBRL International. <http://www.xbrl.org>.
2. Pablo Castells, Borja Foncillas, and Rubén Lara. Semantic web technologies for economic and financial information management. In *ESWS 2004*, Heraklion, Greece, May 2004.