Using Semantic Web Technologies to Facilitate XBRL-based Financial Data Comparability

Héctor Carretié[1], Beatriz Torvisco[1], Roberto García[2]

Universidad Rey Juan Carlos
Paseo Artilleros. 28032 Madrid, Spain
{hector.carretie, beatriz.torvisco}@urjc.es

Universitat de Lleida
Jaume II, 69. 25001 Lleida, Spain
roberto.garcia@udl.cat

Abstract. The XML Business Reporting Language (XBRL) is a standard for business and financial reporting. Many institutions are making available or requiring data in this format, e.g. the US SEC or the Spanish CNMV. However, XBRL data is loosely interconnected and it is difficult to mix and compare, especially when it is based on different accounting principles. Our contribution is based on converting XBRL reports into semantic data and then using Semantic Web technologies to formalise equivalences among terms from different accounting standards. This approach has been evaluated in a particular scenario and it is available online.

Keywords. Business, accounting, finance, interoperability, comparability, Semantic Web, ontology.

1 Introduction

There are many attempts to move existing data to the Semantic Web domain, especially relevant due to the amount of data being mapped are those around the Linked Data initiative [1]. The main motivation to do so is that usually this data is not offering its full potential because it is isolated, i.e. not connected to other external pieces of data that enrich them. It might even be the case that the data is loosely interconnected internally, because it lacks formal semantics. Most of the time this is due to the fact that the technological solutions used to publish that data do not make it easy to interconnect it internally and to other external data sources.

Business reporting is a domain where the need for a common data format for reports has already been identified. XBRL (eXtensible Business Reporting Language) is an XML language intended for modelling, exchanging and automatically processing business and financial information. XBRL is gaining a lot of momentum, especially thanks to the support of some regulators and government agencies worldwide. It is especially significant the importance of the XBRL program promoted by the U.S. Securities and Exchange Commission (SEC). Currently, all companies filing to the SEC are doing so using XBRL following the Government Information Transparency Act, which requires federal agencies to collect their data in a uniform, searchable format using XBRL thereby simplifying mandatory financial reporting for companies that receive federal funds.

However, despite the great success in the adoption of XBRL, we have observed some limitations in its support for cross analysis of financial information in XBRL tools and applications, as it is detailed in Section 2, that might threaten its usefulness. These limitations are not just among data based on different accounting principles, which are represented in XBRL using taxonomies. It even happens when comparing filings for different companies based on the same taxonomies or filings for the same company based on different versions of the taxonomies.

We argue that this limitation is inherited from the technologies underlying XBRL, especially XML. XML takes a document-oriented approach, where each document presents a tree structure. This makes it difficult for XML-based tools to provide functionalities that blur this separation into documents and that overcome the limitations of a tree structure when mashing-up data from different sources. Moreover, XBRL does not provide formal semantics that might help to integrate different taxonomies using logic reasoners.

In any case, the integration of XBRL data into comparable information is a strong requirement for the analysis of business and financial information at a global scale. This might increase the efficiency and effectiveness of the decision-making processes relying on this kind of information. For instance, bankruptcy prediction and other tasks related to the assessment of the solvency of a firm, a business sector or set of interrelated companies. Many have already pointed to this issue and propose Semantic Web technologies as a natural choice for XBRL data integration, cf. Section 2.

Despite these potential benefits, currently, financial and business data is being produced using XBRL and it seems that more and more XBRL data is going to be available in the future. XBRL is been promoted by regulators and government agencies like the US SEC, as it has been shown before, but also other bodies like the European Union or the Spanish Securities Commission (CNMV) [2].

Consequently, our opinion is that the best short-term approach to enjoy the benefits of Semantic Web technologies when working with financial data is not to propose and alternative language based on these technologies, but to apply methods to map existing XBRL to semantic metadata.

The rest of this paper is organised as follows. The next subsections introduce the structure of XBRL and Section 2 presents the related work. Then, in Section 3, the approach for generating semantic data from XBRL is presented. It is based on a transformation from XML data to RDF using the XBRL to RDF mapping, which is described in Section 3.1. Then, the second step is to map the XML Schemas that structure XBRL data to OWL ontologies using the XBRL Schema to OWL mapping detailed in Section 3.2.

The results of the previous mappings, as detailed in Section 4, are a set of OWL ontologies for the main XBRL taxonomies used by the US SEC and based on the US GAAP[1]. Based on these ontologies, it has been possible to map the XBRL instance documents sent to the US SEC since 2009 resulting in more than 100M triples availa-

---

[1] Generally Accepted Accounting Principles,
http://en.wikipedia.org/wiki/Generally_Accepted_Accounting_Principles_(United_States)

ble from the LOD Cloud as the Semantic XBRL dataset[2]. Some preliminary experiments have also been done with XBRL data based on the International Financial Reporting Standards (IFRS) and the Spanish PGC (Plan General Contable) accounting regulations.

Section 5 presents the main evaluations done so far. First of all, there are the results of a basic logical evaluation of the resulting ontologies. Then, we present a deeper evaluation of the overall approach through an scenario where comparability between two XBRL reports for the same company but based on different accounting principles is attained using Semantic Web technologies once they have been mapped to semantic data. Finally, Section 0 presents the conclusions and the future work.

1.1 XBRL

XBRL is based on two kinds of documents, instance documents and taxonomies. Instance documents report business facts and point to a set of taxonomies, which define the meaning of these facts, e.g. under what accounting principles they hold, what other facts they related to or what kind of things do they refer to.

1.1.1 Instances

More concretely, a XBRL instance document contains business Facts. An example of a Fact could be “sales in the last quarter”. If the Fact is simple valued, like “the long term debt is 350,000” whose value is just a number, it is called Item. If the Fact has a more complex value, like “for the *preferred stock*, the *preferred stock par value per share* is 0 and the *preferred stock shares authorized* is 2000”, it is called Tuple.

Items are represented in XBRL as a single XML element with the value as its content while Tuples are represented by XML elements containing nested Items or Tuples, i.e. subelements.

However, facts are not isolated entities and it is not enough to provide their values, it is also necessary to contextualize them. Consequently, four more entities are introduced in the XBRL model:

- **Context**: it defines the *entity* (e.g. company or individual) to which the fact applies, the *period* of time the fact is relevant and an optional *scenario*. The period of time can have zero length for instance and its value is based on ISO 8601 for date and time values. Scenarios provide further contextual information about the facts, such as whether the business values reported are actual, projected or budgeted. Contexts are referenced from Facts using the “contextRef” attribute, which specifies that the given Fact is valid for an *entity*, *period* and *scenario*.
- **Unit**: it defines a unit of measure, such as “USD” or “shares”. They are referenced from Facts using the “unitRef” attribute, which specifies that the numeric or fractional value of the Fact is based on that unit of measure. Complex units can also be defined, like “USD per share”. Currency units are based on ISO 4217.
- **Reference**: The kinds of facts under consideration are defined by taxonomies, which specify their meaning in the context of some accounting principles or pur-

---

[2] http://thedatahub.org/dataset/semantic-xbrl

pose, e.g. Facts relevant for banking and savings institutions. These kinds of facts are then used in instance documents in order to specify actual values for them. However, they are linked to their definition in the taxonomies, typically through schema references, in order to be able to retrieve their meaning.

- **Footnote**: it contains some additional support content and it is associated to Fact using XLink.

Table 1 shows part of an instance document from the EDGAR program that contains a Context element, which defines a company, a time period and the scenario "unaudited". Then, there is a Fact that holds in that context. The Fact references the Context and the value unit, while their content is the fact numeric value.

Roberto García Go

Eliminado: Table 1

Table 1. Context and facts examples from an EDGAR filing

```
…
<context id="From20080301-To20080530_EnterpriseSolutions_Unaudited">
    <entity>
      <identifier scheme="http://www.sec.gov/CIK">796343</identifier>
      <segment><adbe:EnterpriseSolutions /></segment>
    </entity>
    <period>
      <startDate>2008-03-01</startDate>
      <endDate>2008-05-30</endDate>
    </period>
    <scenario><adbe:Unaudited /></scenario>
</context>
…
<adbe:EnterpriseSolutionsRevenue decimals="-6"
contextRef="From20080301-To20080530_EnterpriseSolutions_Unaudited"
unitRef="USD">54400000</adbe:EnterpriseSolutionsRevenue>
…
```

1.1.2 Taxonomies

Taxonomies are the other kind of XBRL document. A taxonomy defines a hierarchy of concepts, basically kinds of Facts, and captures part of their intended meaning. In XBRL there is a set of base taxonomies that define the core concepts and other ones that extend them in order to particularize these concepts for concrete accounting principles, application domains, etc. Additionally, it is possible to extend existing taxonomies and accommodate them to particular needs.

Taxonomies are based on XML Schemas, which provide the taxonomy building primitives and the extension mechanisms. Moreover, there are also "linkbases", which allow establishing links beyond the taxonomy tree structure using XLink.

- **Schemas** define concepts that are instantiated as Items or Tuples, depending on their complexity, in the instance documents. They are based on XML Schema elements (xsd:element). A concept definition provides the fact name, whether it is a tuple or an item and its value data type (such as monetary, numeric or textual).
- **Linkbases** define links from concepts in a taxonomy to labels, pieces of content or other concepts. The XBRL specification defines five different kinds of linkbases.
    - **Label Linkbase**: set of links that provides human readable strings for concepts, potentially in multiple languages.

- **Reference Linkbase**: these links associate concepts with citations of some body of authoritative literature.
- **Calculation Linkbase**: these are links that associate a set of values of concepts in taxonomies with a mathematical calculation that must be checked for consistency, for instance that a set of concepts with percentage values sum up 100%.
- **Definition Linkbase**: it provides semantic relations between concepts like is-a, whole-part, etc.
- **Presentation Linkbase**: This linkbase associates concepts with other concepts so that the resulting relations can guide the creation of a user interface, rendering, or visualisation.

2 Related Work

The U.S Securities and Exchange Commission (SEC) offers some online tools that allow interacting with the data available in XBRL form. There is a tool called Interactive Financial Reports that allows viewing and charting companies financial information. It also provides some functionality that allows comparing different filings and different companies, though it is hard to use and prone to even the slightest differences between the compared filing facts, even when there is just a name change for facts from filings of the same company.

There is also the Financial Explorer, which presents company financial data through very informative diagrams but just from one company at a time, and the Executive Compensation tool. The later allows comparing just two facts, Public Market Capitalization and Revenue, across all filed companies.

Apart from the SEC tools, there are some other XBRL tools, most of them proprietary and with quite high licensing cost. Among them, the Fujitsu XBRL Tools[3] should be highlighted because they are one of the most popular tool sets and it is available for XBRL Consortium members and academic users. The tools comprise taxonomy and instance editors, viewers and validators.

The most powerful tool in this set, though still in beta and with many usability problems, is the Instance Dashboard. This application can consume multiple instance documents and, by specifying a base taxonomy, users can perform some comparison analysis, though limited to facts in a taxonomy that appears in all the filings.

As it can be noted from the previous analysis, the main limitation of XBRL tools is their limited support for cross analysis of financial information, not just among data based on different taxonomies, even when comparing filings for different companies based on the same taxonomies.

This limitation is inherited from the technologies underlying XBRL, especially from XML. XML takes a document-oriented approach, where each document presents a tree structure. This makes it difficult for XML-based tools to provide functionalities that blur this separation into documents and that overcome the limitations of a tree structure when mashing-up data from different sources.

---

[3] Fujitsu XBRL Tools, http://www.fujitsu.com/global/services/software/interstage/xbrltools/

Consequently, Semantic Web tools are being considered by people like Charles Hoffman, the father of XBRL: "*This field [W3C semantic standards] is rich with possibilities and stands as the next logical step in the natural progression of information technology to seek a higher value proposition*" [3].

This interest is materializing, and the combination of XBRL and the Semantic Web has been receiving some attention in different blogs [4,5], mailing lists and web groups[4]. The first attempts to combine both technologies focused on specific for some parts of XBRL. For instance, there is an ontology about financial information based on XBRL that is specific for investment funds [6] and, though it is generated using a generic XBRL taxonomy to OWL ontology algorithm, there is not and equivalent tool that maps generic XBRL instance data.

Another quite specific tool maps quarterly and semester accounting information submitted to the Spanish securities commission (CNMV) to RDF [2]. Both approaches are based on procedural code specially developed in order to extract specific patterns from the XBRL data. Consequently, they are difficult to scale to the whole XBRL specification and sensible to minimal changes in it.

More recent attempts have widened and generalised their scope. For instance, eTEN was an European Community programme providing funds to help make e-services available throughout the European Union. This programme ended in 2006. Within this programme there was the WINS project: Web-based Intelligence for common-interest fiscal Networked Services.

WINS provides a Web-based Business Intelligence (BI) Service to public and private Financial Institutions by integrating BI products and knowledge discovery tools to produce new financial knowledge on companies from information gathered through interoperable information services. Within the WINS context, Declerk and Krieger [7] pointed out some limitations encountered in the XBRL schema documents mainly due to the lack of reasoning support over XML-based data. They proposed the "ontologization" or process to translate XBRL taxonomies into OWL to overcome these limitations.

The "ontologization" starts from the WINS information extraction (IE) task, which gathers financial facts from PDF files and converts them into XBRL documents. From these document, the process continues based on a hand-made translation of XBRL facts into OWL ontologies that then helps classifying the facts into higher-level concepts like Balance Sheet or Statement of Income. However, the ontologies are not exploited beyond this point in order to facilitate the comparability of the financial facts across different accounting standards.

Another example of mapping from XBRL to Semantic Web technologies is OpenLink XBRL Sponge, which maps generic XBRL instance data to RDF [8]. However, in this case, there is not and associated mapping from the taxonomies instance data is based on to ontology languages. Therefore, it is not easy to facilitate the comparability of the financial facts by working at the conceptual level provided by the ontologies.

Bao et al. [9] do consider the comparability issue and they point out the tremendous human cognitive effort that must be done when comparing financial data written

---

[4] XBRL Ontology Specification Group,
http://groups.google.com/group/xbrl-ontology-specification-group

in XBRL. Their proposal is to overcome this problem by defining the *logic model* of XBRL reports using the Web ontologies language OWL to design ontologies that capture the meaning of the reports beyond just their structure. They transform concepts into classes and arcroles into properties. However, the possibilities of the logic models generated are not put into practice in comparability scenarios that involve different accounting regulations.

Finally, latest approaches start to focus on comparability and attempt to profit from Semantic Technologies and Linked Data principles to attain it [10]. For instance, the XBRL European Business Registry (xEBR) is an XBRL Europe project to create a list of concepts, which are common across the various European business registries. The concepts encompass basic financial data as well as company profiles. However, this Project is still limited by the fact that there is no common regulation for Business Registries in Europe. Therefore, many Registries in Europe have built their own set of taxonomies.

Our proposal, as detailed in the next sections, focuses on facilitating comparability at the semantic level, where it is easier to establish the equivalences among financial facts independently of the particular taxonomies and associated accounting standards they come from. In order to do that, we propose an approach that, instead of directly processing XBRL data, takes profit from the fact that it is expressed using XML and specified using XML Schemas. The instance XML documents are translated into RDF that models the financial facts and refers to the concepts modelled in ontologies generated from the schemas. From this point, it is now possible to establish equivalences that facilitate comparability at the ontology level use inference to benefit from this knowledge at the instance level.

3 Approach

The proposed approach is based on the transfer of existing XBRL taxonomies and instance data to Semantic Web technologies. This transfer is based on the XML Semantics Reuse methodology [11,12] and the XML Schema to OWL and XML to RDF tools implemented in the ReDeFer project[5].

This methodology combines an XML Schema to web ontology transformation, XSD2OWL, with a transparent translation from XML to RDF, XML2RDF. The ontologies generated by XSD2OWL are used during the XML to RDF step to generate semantic metadata that takes into account the XML Schema intended meaning.

This approach differs from other attempts to move metadata from the XML domain to the Semantic Web. Some of them just model the XML tree using the RDF primitives [13]. Others concentrate on modelling the knowledge implicit in XML languages definitions, i.e. DTDs or the XML Schemas, using web ontology languages [14,15]. Finally, there are attempts to encode XML semantics integrating RDF into XML documents [16,17].

However, none of them facilitate an extensive transfer of XML metadata to the Semantic Web in a general and transparent way. Their main problem is that the XML

---

[5] ReDeFer project, http://rhizomik.net/redefer

Schema implicit semantics are not made explicit when XML metadata instantiating this schemas is translated. This is so because the RDF data produced from XML instance data looses its links to the XML Schemas that structure them and model the relations among different XML entities.

These relations among different XML entities are what carry the XML Schema implicit semantics. They capture part of the meaning intended by the schema developer that, though XML Schema does not provide a way to encode semantics, is recorded in the way XML Schema constructs are used. For instance, by modelling that element "father" is a *subtitutionGroup* for element "parent", it is possible to interpret that "parent" is more general than "father" and that "father" can appear where "parent" appears. More details about the implicit semantics of XML Schema constructs as compared to OWL ones are provided in Section 3.2.

Therefore, the previous transformations from XML to RDF do not take profit from the meaning encoded in XML Schemas and produce RDF metadata almost as semantics-blind as the original XML. Or, on the other hand, they capture this semantics but they use additional ad-hoc semantic constructs that produce less transparent metadata.

3.1 XML2RDF

The XML to RDF transformation follows a structure-mapping approach [13] and tries to represent the XML metadata structure, i.e. a tree, using RDF. The RDF model is based on the graph so it is easy to model a tree using it. Moreover, we do not need to worry about the loss of semantics produced by structure-mapping. We formalised the underlying semantics into the corresponding ontologies and we will attach them to RDF metadata using the instantiation relation *rdf:type*.

The structure-mapping is based on translating XML metadata instances to RDF that instantiates the corresponding constructs in OWL. The more basic translation is from *xsd:elements* and *xsd:attributes* to *rdf:Properties* (*owl:ObjectProperties* for node to node and *owl:DatatypeProperties* for node to value relations).

Values are kept during the translation as simple types and RDF blank nodes are introduced in the RDF model in order to serve as the source and destination for properties. They will remain blank until they are enriched with semantic information.

The resulting RDF graph model contains all that we can obtain from the XML tree. It is already semantically enriched thanks to the *rdf:type* relation that connects each RDF property to the *owl:ObjectProperty* or *owl:DatatypeProperty* it instantiates. It can be enriched further if the blank nodes are related to the *owl:Class* that defines the package of properties and associated restrictions they contain, i.e. the corresponding *xsd:complexType*. This semantic decoration of the graph is formalised using *rdf:type* relations from blank nodes to the corresponding OWL classes.

At this point we have obtained a semantically enabled representation of the input metadata, a representation that makes the meaning intended by the XML and XML Schema modelers explicit from a computer point of view. The instantiation relations can now be used to apply OWL semantics to metadata. Therefore, the semantics derived from further enrichments of the ontologies, e.g. integration links between different ontologies or semantic rules, are automatically propagated to instance metadata thanks to inference.

Focusing on XBRL data, what we get by applying this triplification process of the corresponding XML data is summarised in Fig. 1. This figure shows the XBRL core concepts as they are modeled in the resulting RDF data. The report is modelled as an instance of the class “ReportType” and facts are modelled as instances of “FactType”.

In fact, if a direct modelling of the underlying XML tree was performed, facts should be modelled as RDF Properties because they correspond to XML elements. However, in order to make the resulting RDF data more usable as it is more intuitive to view a fact as class instance than as a relation one, we have introduce a modification in the basic XML2RDF algorithm as it is detailed in the next subsection.

Then, continuing from the “FactType” instance, there are relations to the actual value of the financial fact modelled using rdf:value and two properties stating the decimals and unit used for that value. There is also a property linking the fact to its context, which details the involved entity, the time period and the scenario.

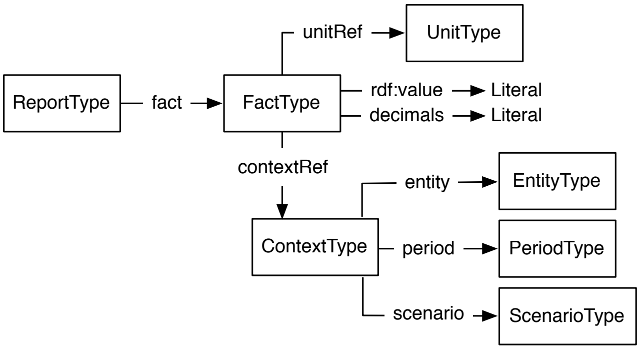


Fig. 1. RDF model for the core XBRL concepts generated using XML2RDF and XSD2OWL (boxes correspond to classes and arrows to properties having them as domain/ranges)

3.2 XBRL Schema to OWL Mapping

The XML Schema to OWL transformation is responsible for capturing the schema implicit semantics, which is determined by the combination of XML Schema constructs. The transformation is based on translating these constructs to the OWL ones that best capture their intended meaning. These translations are detailed in Table 2.

Roberto García Go

Eliminado: Table 2

The XML Schema to OWL transformation is quite transparent and captures a great part XML Schema semantics. The same names used for XML constructs are used for OWL ones, although in the new namespace defined for the ontology. XSD and OWL constructs names are identical; this usually produces uppercase-named OWL properties because the corresponding element name is uppercase, although this is not the usual convention in OWL. Therefore, XBRL Schema to OWL produces OWL ontologies that make explicit the semantics of the corresponding XBRL taxonomies.

The only caveats are the implicit order conveyed by *xsd:sequence* and the exclusivity of *xsd:choice*. For the first problem, *owl:intersectionOf* does not retain its operands order, there is no clear solution that retains the great level of transparency that has been achieved. The use of RDF Lists might impose order but introduces ad-hoc constructs not present in the original metadata.

Table 2. XBRL Schema to OWL translations for the XML Schema constructs

XML Schema	OWL	Mapping motivation
element[@substitutionGroup= "xbrli:item"]	owl:Class	Facts, though elements, are mapped to classes
element | attribute	rdf:Property owl:DatatypeProperty owl:ObjectProperty	Named relation between nodes or nodes and values
element@substitutionGroup="xbrli:item"	rdfs:subClassOf	The corresponding element is mapped to a owl:Class rdfs:subClassOf xbrli:item
element@substitutionGroup	rdfs:subPropertyOf	Relation can appear in place of a more general one
element@type	rdfs:range	The relation range kind
complexType|group |attributeGroup	owl:Class	Relations and contextual restrictions package
complexType//element	owl:Restriction	Contextualised restriction of a relation
extension@base | restriction@base	rdfs:subClassOf	Package concretises the base package
@maxOccurs	owl:maxCardinality	Restrict the number of occurrences of a relation
@minOccurs	owl:minCardinality	
sequence	owl:intersectionOf	Combination of relations in a context
choice	owl:unionOf	

Moreover, as it has been demonstrated in the Semantic Web community, the element ordering does not contribute much from a semantic and knowledge representation point of view [18] in most cases and when it is a requirement it is more convenient to explicitly represent it using some sort of order attribute or property. For the second problem, *owl:unionOf* is an inclusive union, the solution is to use the disjointness OWL construct, *owl:disjointWith*, between all union operands in order to make it exclusive.

4 Results

First of all, we have generated an ontological infrastructure for the XBRL core, currently XBRL 2.1. It is composed by the ontologies resulting from mapping the XBRL core XML Schemas using the XBRL Schema to OWL mapping: XBRL Instance, XBRL Linkbase, XBRL XL and XBRL XLink. These ontologies have been adapted to accommodate the changes introduced by XBRL to RDF that make the output semantic data more usable, basically by making facts classes and no longer properties.

Apart from the previous schemas, the following schemas have been also mapped in order to be able to map the XBRL data submitted to the US SEC.

From US GAAP (Generally Accepted Accounting Principles) the schemas, and corresponding ontologies, are: Primary Terms Elements (USFR-PTE),Primary Terms Relationships (USFR-PTR), Financial Services Terms Elements (USFR-FSTE), Financial Services Terms Relationships (USFR-FSTR) and Investment Management Terms Relationships (USFR-IME). For specific industries: Banking and Savings Institutions (US-GAAP-BASI), Commercial and Industrial (US-GAAP-CI), Insurance (US-GAAP-INS) and Investment Management (US-GAAP-IM).

There are also some non-GAAP schemas that have been also mapped to OWL ontologies: Accountants Report (USFR-AR), Management Discussion and Analysis (USFR-MDA), Management Report (USFR-MR) and SEC Certifications (USFR-SECCERT).

The same approach has been followed to map the IFRS taxonomies and the ones used by the Spanish securities commission (CNMV). Most of the previous ontologies are available from the BizOntos Business Ontologies web page[6] and the semantic data for all the processed filings can be queried and browsed from the Semantic XBRL site[7]. Currently, more than 25 thousand filings have been processed from the US SEC, plus some from the CNMV. The combination of all these filings once mapped to RDF amounts slightly more than 100 million triples. At this step, it is possible to take profit from semantic web technologies in order to improve the interconnectedness of the dataset by means of semantics-enabled data integration.

5 Evaluation

The proposed approach has been evaluated using two input XBRL reports for the same company but based on different accounting principles, and consequently different taxonomies. The input data is from Telefonica S.A., one of the reports was submitted to the Spanish CNMV and the other to the US SEC[8], more specifically the consolidated Balance Sheet for the years 2009 and 2008.

The motivation is that Telefonica is one of the few Spanish corporations that files their financial statements to the Spanish securities commission (CNMV) in XBRL format and also to the American Securities Exchange Commission (US SEC). The 2009 period was the last period available in the CNMV and SEC websites, at the time of the elaboration of the present evaluation.

The elaboration of the financial statements for the CNMV has been done under the Spanish GAAP regulations[9], i.e. Plan General de Contabilidad, issued in 2007 and based on IFRS. Meanwhile, financial information filled to the US SEC was elaborated under the IFRS, following SEC's provisions for foreign corporations.

Therefore, it could be expected that both XBRL financial reports would be the same or at least quite similar. However, as the Table 4 shows, there are some differences mainly due to different levels of disaggregation. The totals for assets or liabilities coincide but not the figures contained under these main sections.

---

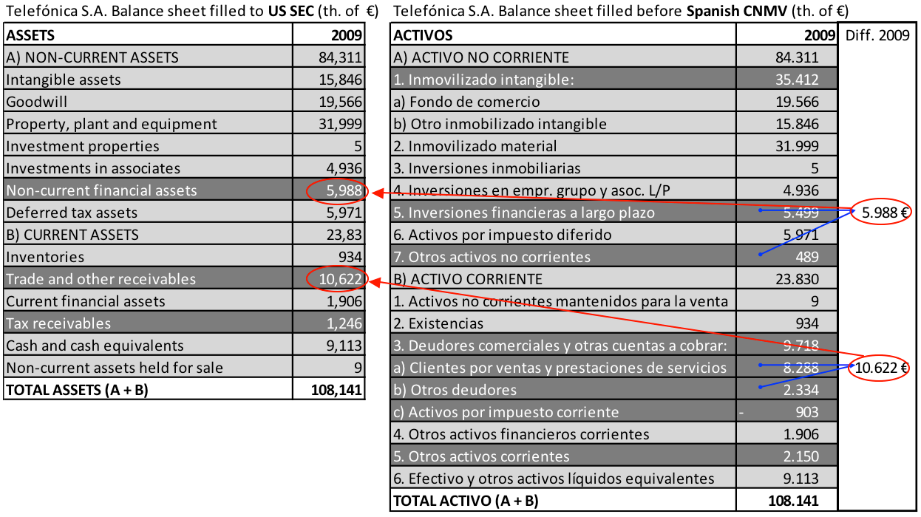
[6] BizOntos, http://rhizomik.net/ontologies/bizontos

[7] SemanticXBRL, http://rhizomik.net/semanticxbrl

[8] Telefonica's report to the CNMV is available from http://www.cnmv.es/ipps/default.aspx and the one sent to SEC is available from http://www.sec.gov/Archives/edgar/data/814052/000095010310000881/dp16939_20f.htm

[9] Models recently modified by Ministerial Oder JUS/1698/2011 of June 13, approving the model for presentation at the Mercantile Registry of the consolidated financial statements

Fig. 2 highlights the accounts where quantity differences are found. For instance, in the 2009 balance sheet for the SEC (on the left), “Non-current financial assets” amounts 5,988 millions of euros, meanwhile in balance sheet for the CNMV (on the right) “Inversiones financieras a largo plazo” (long-term financial investments) amounts 5,499 millions and “Otros activos no Corrientes” (Other non-current assets) amounts 489 millions. Both accounts sum up 5,988 millions, so the sum of “Inversiones financieras a largo plazo” and “Otros activos no Corrientes”, two terms specific to CNMV taxonomies, is equivalent to the IFRS term “Non-current financial assets”.



Telefónica S.A. Balance sheet filled to **US SEC** (th. of €)

ASSETS	**2009**
A) NON-CURRENT ASSETS	84,311
Intangible assets	15,846
Goodwill	19,566
Property, plant and equipment	31,999
Investment properties	5
Investments in associates	4,936
Non-current financial assets	5,988
Deferred tax assets	5,971
B) CURRENT ASSETS	23,83
Inventories	934
Trade and other receivables	10,622
Current financial assets	1,906
Tax receivables	1,246
Cash and cash equivalents	9,113
Non-current assets held for sale	9
TOTAL ASSETS (A + B)	**108,141**

Telefónica S.A. Balance sheet filled before **Spanish CNMV** (th. of €)

ACTIVOS	**2009**	Diff. 2009
A) ACTIVO NO CORRIENTE	84.311	
1. Inmovilizado intangible:	35.412	
a) Fondo de comercio	19.566	
b) Otro inmobilizado intangible	15.846	
2. Inmovilizado material	31.999	
3. Inversiones inmobiliarias	5	
4. Inversiones en empr. grupo y asoc. L/P	4.936	
5. Inversiones financieras a largo plazo	5.499	5.988 €
6. Activos por impuesto diferido	5.971	
7. Otros activos no corrientes	489	
B) ACTIVO CORRIENTE	23.830	
1. Activos no corrientes mantenidos para la venta	9	
2. Existencias	934	
3. Deudores comerciales y otras cuentas a cobrar:	9.718	
a) Clientes por ventas y prestaciones de servicios	8.288	10.622 €
b) Otros deudores	2.334	
c) Activos por impuesto corriente	- 903	
4. Otros activos financieros corrientes	1.906	
5. Otros activos corrientes	2.150	
6. Efectivo y otros activos líquidos equivalentes	9.113	
TOTAL ACTIVO (A + B)	**108.141**	

Fig. 2. Assets section Telefonica’s Balance Sheet filled to the US SEC

Other equivalences requiring the addition of different account are also highlighted in Fig. 2 and marked using dark grey. The accounts marked with light grey have direct equivalences between the taxonomies used by the US SEC and the CNMV. For instance, “Intangible assets” in the US SEC document is equivalent to “Otro inmobilizado intangible”.

The instance document from Telefonica filed to the Spanish CNMV includes terms specific to the Spanish terminology, defined in the “ipp-gen” namespace in the XBRL instance documents, but with an equivalent term in IFRS. Other terms reuse the international standard, and thus are in the “ifrs-gen” namespace, but in some cases they do not coincide with the terms specified in the IFRS taxonomy. Finally, some elements are specific to the CNMV, e.g. “ipp-gen:TotalActivoNiif”.

Both, numerical and terminological differences, dramatically decrease the comparability of the two consolidated balance sheets. However, it is possible to establish equivalences at the conceptual level, and arithmetic operations among them when there is not a direct equivalence. This is easily achievable thanks to Semantic Web technologies once the involved taxonomies have been mapped to OWL and the corresponding instance documents to RDF. The next section presents some examples.

5.1.1 Mappings between Spanish PGC and IFRS

Table 3 shows some of the semantic mappings generated for the Telefonica scenario between the ontologies corresponding to the IFRS taxonomies and thus for the CNMV taxonomies.

Table 3. Mappings between Spanish PGC and IFRS

<table>
<tr><th>Spanish CNMV
(PGC taxonomies)</th><th>US SEC
(IFRS taxonomies)</th><th>Semantic Mappings</th></tr>
<tr><td>ipp-gen:
ActivoNoCorrienteNiif
84.311 €</td><td>ifrs:
NoncurrentAssets
84.311 €</td><td>ipp-gen:ActivoNoCorrienteNiif
owl:equivalentClass
ifrs:NoncurrentAssets</td></tr>
<tr><td>ifrs-gp:
TradeAndOtherReceivablesNetCurrent =
ipp-gen:
ClientesVentasPrestacionesServicios +
ipp-gen:
OtrosDeudores
8.288€ + 2.334€</td><td>ifrs:
TradeAndOtherCurrentReceivables
10.622€</td><td>ifrs-gp:TradeAndOtherReceivablesNetCurrent
owl:equivalentClass
ifrs:TradeAndOtherCurrentReceivables<pre>
CONSTRUCT {
  [] a ifrs-gp:
     TradeAndOtherReceivablesNetCurrent;
  xbrli:contextRef ?context;
  xbrli:unitRef ?unit;
  xbrli:decimals ?decimals;
  rdf:value ?value. }
WHERE {
  ?cvps a ipp-gen:
     ClientesVentasPrestacionesServicios;
  xbrli:contextRef ?context;
  xbrli:unitRef ?unit;
  xbrli:decimals ?decimals;
  rdf:value ?cvps-value.

  ?od a ipp-gen:OtrosDeudores;
  xbrli:contextRef ?context;
  xbrli:unitRef ?unit;
  xbrli:decimals ?decimals;
  rdf:value ?od-value.
  BIND(?cvps-value+?od-value AS ?value) }
</pre></td></tr>
</table>

The approach is to model the accounts determined to be equivalent, because the are in the same part of the balance sheet and correspond to the same quantity, as equivalent at the ontology level using the *equivalentClass*[10] OWL construct. When the relation is more complex than a simple equivalence, for instance when the value for a term in one vocabulary is the sum of more than one value in other vocabularies, then the approach is to use a *Construct*[11] SPARQL query that computes the combined value, for instance the sum, and creates the computed fact.

The complete set of mappings is available from an online demo[12], where they are also put into practice using a Semantic Web repository that includes an inference an inference and a SPARQL engine that “execute” these mappings. For the demo, just the CNMV XBRL document is loaded into the repository and the mappings are used to generate most of the IFRS version of the assets part of the balance sheet using the semantic mappings.

---

[10] OWL Equivalent Class, http://www.w3.org/TR/owl-ref/#equivalentClass-def

[11] SPARQL Construct, http://www.w3.org/TR/sparql11-query/#construct

[12] SemanticXBRL Demo, http://rhizomik.net/semanticxbrl-demo/

6 Conclusions and Future Work

As it has been shown, it is possible to map the XML data for XBRL filings in order to generate semantic data that keeps all the original information and structure. This mapping also includes the involved XML Schemas that structure the XML data. These schemas are mapped to Web ontologies, which make all the semantics implicit in the original XML Schemas explicit and available when querying semantic data.

Moreover, it is also possible to take profit from Web ontology primitives in order to semantically integrate different filings following different XML Schemas, i.e. XBRL taxonomies. Once mapped to ontology concepts and relations, the XBRL contexts, facts and other resources defined for different filings can be related as more specific, more general or equivalent.

This approach has been put into practice in the context of the US SEC's XBRL program. It has been possible to apply the previous XML to RDF and XML Schema to Web ontology mappings to filings sent to the US SEC and some from the Spanish CNMV. More than 100 million triples have been obtained, which are structured by the ontologies generated from the corresponding taxonomies.

Moreover, the benefits of the approach have been validated in a real scenario where it is possible to generate an XBRL report following the IFRS taxonomies starting from one based on the Spanish CNMV taxonomies using semantic mappings established at the ontology level.

Future work focuses on, once we establish more semantic mappings at the conceptual level that can be reused to map instance documents for different companies, obtaining financial statement analysis ratios, taking profit from the semantic data already available.

For instance, to compute the debt ratio (equivalent to total liabilities / total assets), and current ratio (equivalent to total liabilities / total assets) by analysing the balance sheets, or the Return on Sales (ROS, equivalent to net income / sales revenue). From these ratios and the semantic mapping, we will be able to create a ranking showing the best-positioned international companies for each ratio mixing the data they submit to different regulators.

Acknowledgements

The work described in this paper has been partially supported by Spanish Ministry of Science and Innovation through the Open Platform for Multichannel Content Distribution Management (OMediaDis) research project (TIN2008-06228).

References

[1] Bizer, C., Heath, T., Berners-Lee, T.: Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems. 5, 1–22 (2009).

[2] Núñez, S., de Andrés, J., Gayo, J.E., Ordoñez, P.: A Semantic Based Collaborative System for the Interoperability of XBRL Accounting Information. Emerging Technologies and Information Systems for the Knowledge Society. pp. 593–599. Springer, Berlin/Heidelberg, DE (2008).

[3] Hoffman, C.: Financial Reporting Using XBRL: IFRS and US GAAP Edition. Lulu.com (2006).

[4] Raggett, D. XBRL and RDF. In: Dave Raggett's Blog, (2008). Available from http://people.w3.org/~dsr/blog/?p=8

[5] DuCharme, B. Changing my mind about XBRL again. In: Bob DuCharme's weblog, bobdc.blog, (2008). Available from http://www.snee.com/bobdc.blog/2008/08/changing_my_mind_about_xbrl_ag.html

[6] Lara, R., Cantador, I., Castells, P.: Semantic Web Technologies For The Financial Domain. In: Cardoso, J. and Lytras, M. (eds.) The Semantic Web: Real-World Applications from Industry. pp. 41–74. Springer, New York, NY, USA (2008).

[7] Declerck, T., Krieger, H.: Translating XBRL into Description Logic: an approach using Protege, Sesame and OWL. In: Abramowicz, W. and Mayr, H.C. (eds.) Proceedings of the 9th International Conference on Business Information Systems, BIS'06. pp. 455–467. GI, Bonn, DE (2006).

[8] Erling, O., Mikhailov, I. RDF Support in the Virtuoso DBMS. In: Pellegrini, T., Auer, S., Tochtermann, K. and Schaffert, S. (eds.) Networked Knowledge - Networked Media. pp. 7-24. Springer, Berlin/Heidelberg, DE (2008).

[9] Bao, J., Rong, G., Li, X., Ding, L.: Representing Financial Reports on the Semantic Web: A Faithful Translation from XBRL to OWL. In: Dean, M., Hall, J., Rotolo, A., and Tabet, S. (eds.) Semantic Web Rules. pp. 144–152. Springer, Berlin/Heidelberg, DE (2010).

[10] O'Riain, S., Curry, E., Harth, A.: XBRL and open data for global financial ecosystems: A linked data approach. International Journal of Accounting Information Systems. In Press (2012).

[11] García, R.: Chapter 7: XML Semantics Reuse. In: García, R. A Semantic Web Approach to Digital Rights Management. VDM Verlag, Saarbrücken, Germany (2010).

[12] García, R., Gil, R.: Linking XBRL Financial Data. In: Wood, D. (ed.) Linking Enterprise Data. pp. 103–125. Springer, New York, NY, USA (2010).

[13] Klein, M.C.A.: Interpreting XML Documents via an RDF Schema Ontology. Proceedings of the 13th International Workshop on Database and Expert Systems Applications, DEXA'02. pp. 889–894. IEEE Computer Society, Washington, DC, USA (2002).

[14] Amann, B., Beeri, C., Fundulaki, I., Scholl, M.: Ontology-Based Integration of XML Web Resources. Proceedings of the 1st International Semantic Web Conference, ISWC 2002. pp. 117–131. Berlin/Heidelberg: Springer (2002).

[15] Cruz, I., Xiao, H., Hsu, F.: An Ontology-based Framework for XML Semantic Integration. Eighth International Database Engineering and Applications Symposium, IDEAS'04. pp. 217–226. IEEE Computer Society, Washington, DC, USA (2004).

[16] Lakshmanan, L., Sadri, F.: Interoperability on XML Data. Proceedings of the 2nd International Semantic Web Conference, ICSW'03. pp. 146–163. Springer, Berlin/Heidelberg, DE (2003).

[17] Patel-Schneider, P.F., Simeon, J.: The Yin/Yang web: XML syntax and RDF semantics. Proceedings of the 11th International World Wide Web Conference, WWW'02. pp. 443–453. ACM Press (2002).

[18] Berners-Lee, T. Why RDF model is different from the XML model. W3C Dessign Issues, (1998). Available from http://www.w3.org/DesignIssues/RDF-XML.html