Porting the xEBR Taxonomy to a Linked Open Data compliant Format*

Thierry Declerck[1] and Dagmar Gromann[2]

[1] DFKI GmbH, Stuhlsatzenhausweg 3,
D-66123, Saarbrücken, Deutschland,
declerck@dfki.de,
https://www.dfki.de
[2] Artificial Intelligence Research Institute (IIIA-CSIC), Campus de la UAB,
E-08193 Bellaterra, Spain,
dgromann@iiia.csic.es,
http://iiia-csic.es

Abstract. The XBRL Europe Business Registers (xEBR) working group has developed a Core Reference Taxonomy over the last years. This work represents a milestone in the support of conceptual interoperability concerning the information on company identification and financial statements across European legislations and languages. Our main contribution is to port the current data of this taxonomy, which is available both as an Excel table and in the standard XML format of XBRL, to a Linked Data compliant format in order to make the taxonomy interoperable also at the semantic level. This paper describes the current version of the ontological model resulting from this transformation and explains some of our design decisions.

Keywords: xEBR Taxonomy, Linked Open Data, Multilingualism

1 Introduction

Transparent financial reporting is key to preventing corporate fraud and allowing for stakeholders to compare business results of different corporations over a certain reporting period. However, business reports summarize several types of information, such as financial statements and risk and performance reports, in various formats for different jurisdictions and in specific natural languages. To make this variety of information comparable, the eXtensible Business Reporting Language (XBRL)[1] has been introduced, a freely available and global business reporting standard encoded in an XML-based language. This XML schema allows for the tagging of specific items of company reports, which can then be

* We dedicate this paper to the memory of our colleague, Dr. Hans-Ulrich Krieger, who recently passed away.

[1] https://www.xbrl.org. See also [5].

automatically processed, such as being automatically compared in a spreadsheet. Even though the focus of this XML-based format is on syntactic rather than semantic correctness, its adoption by businesses has already shown to lead to an increased business value as proven empirically (e.g. [8]). We believe that this increase in value can be further tapped into by a semantically expressive version of XBRL, especially for the European case, in which each member state defines specific rules for filing business reports under its own jurisdiction, which is reflected in its local XBRL taxonomy by means of different name-spaces and tag names. Some tags only exist in specific local XBRL taxonomies.

To render those local XBRL taxonomies interoperable across borders, the XBRL Europe Business Registers (xEBR) working group has developed the so-called xEBR Core Reference Taxonomy[2]. This taxonomy serves as an interlingua between XBRL jurisdictions that represents shared information and assigns them canonical names as well as further information, such as key financial ratios and company identifiers. Several local XBRL taxonomies have already been aligned with this centralized xEBR format, however, this alignment process is tedious and its verification is difficult [13]. Furthermore, the XML-based xEBR format solves primarily the syntactic interoperability problem, but does not provide a semantically grounded representation across languages and jurisdictions [12].

Porting the xEBR Core Reference Taxonomy to a Linked Open Data (LOD)[3] compliant format enables the use of semantic technologies. Thereby, continuously updated financial information, such as stock exchange data, can be aligned with more static data from XBRL reports (see e.g. [3]). Semantic technologies allow for automated consistency checking of represented knowledge, which is beneficial for the verification and generation of financial summaries. Thereby, the comparison of companies reporting to different jurisdictions can be facilitated. The LOD-compliant model we propose in this paper assigns in a declarative way multilingual information to each concept, which eases its use by human analysts across countries. By analysts we refer to all stakeholders of a company, including shareholders, government, and also the general public.

This paper first sets the proposed work in relation to previous approaches before we discuss the xEBR taxonomy in more detail. We then describe the details of the transformation process from an Excel version of the taxonomy to a machine-processable ontology in Section 4.

2 Related Work

The proposed approach addresses the problem of formalizing an existing financial reporting taxonomy for the purpose of increasing the semantic interoperability of its associated local taxonomies and to leverage the benefits of Semantic Web technologies. To this end, we address two important problems that have been

[2] See `http://www.xbrleurope.org/working-groups/xebr-wg/xebr-taxonomy` for more details.

[3] See `http://linkeddata.org/` for more details.

tackled by other approaches before: a) alignment of xEBR and XBRL, and b) formalization of accounting concepts.

2.1 xEBR and XBRL Alignments

To overcome the multiplicity in jurisdiction and natural language of reporting standards, several approaches to align individual XBRL taxonomies have been proposed. Spohr et al. [12] suggest a supervised ranking support vector machine algorithm to perform multilingual ontology alignment. The system developed by Spohr et al [12] is utilized as a baseline for a logic-based alignment approach proposed by Thomas et al [13]. While there is a whole range of multilingual and crosslingual ontology alignment approaches (see [14] for an overview), the two described here are particularly relevant since they are also interested in explicating the implicit semantics of XBRL taxonomies. However, the focus of this paper is on the translation of an existing xEBR taxonomy to a LOD-compliant format rather than formalizing alignments.

2.2 Formalizing Accounting Concepts

In the literature on formalizing accounting concepts the two major approaches are directly specifying accounting principles in ontologies and transforming existing taxonomies to a Semantic Web format. In the first category, Bai et al. [1] propose an ontology-based extension of XBRL to allow for its use in a financial service matching framework. Thomas et al. [13] propose a logic-based approach to explicate shared semantics that are implicit across XBRL taxonomies with the objective to align those shared concepts across taxonomies from different jurisdictions. Krahel [7] suggests the use of a formalization process of accounting principles to detect inconsistencies in existing standards.

Approaches to porting XBRL taxonomies to ontological formats [4, 2] and to Linked Open Data [6, 10] are vital to linking XBRL data to other data on the Web. In fact, a whole range of XBRL-related ontologies has been published, an overview of which can be found in O'Rian [9]. While we base our research on lessons learned from those previous approaches, most of them represent a very literal conversion of the XBRL taxonomies with little semantic addition [13]. More recent approaches suggested to port XBRL to multidimensional data models in order to ensure the correctness of XBRL formulas [11]. While the discussion of format and standard in this work is very insightful, little is added to the semantic explication of accounting principles and definitions in XBRL. Instead of a literal conversion, we focus on a semantic explication of the principles represented in xEBR to unleash its full potential in an LOD-compliant format. This also means closely analysing all existing definitions and representing them in explicit semantics. In a second step this ontologization of xEBR facilitates alignments across resources, however, this is not the focus of the current publication.

3 The xEBR Core Reference Taxonomy

The xEBR Core Reference Taxonomy defines common concepts for financial statements and company identification, across a certain number of legislations and countries[4]. This way, matching tables have been proposed between the xEBR taxonomy and local taxonomies, covering the legislations of Belgium, Germany, France, Italy, The Netherlands, Spain and the United Kingdom. The xEBR taxonomy also describes equivalences to codes of the BACH database[5].

Figure 1 below gives a very partial view on the xEBR taxonomy data encoded in Excel[6].

LABEL (en) & HIERARCHY	NAME SPACE	ELEMENT NAME	ABSTRACT / TYPE
COMPANY BALANCE SHEET, HORIZONTAL LAYOUT [REPORT]	xebr	CompanyBalanceSheetHorizontalLayoutReport	role
Assets [Presentation]	xebr	AssetsPresentation	abstract
Subscribed capital unpaid [Presentation]	xebr	SubscribedCapitalUnpaidPresentation	abstract
Subscribed capital unpaid [Total]	xebr	SubscribedCapitalUnpaidTotal	monetary
Subscribed capital unpaid: Of which there has been called (if not in current assets)	xebr	SubscribedCapitalUnpaid:OfWhichThereHasBeenCalledIfNotInCurrentAssets	monetary
Formation expenses [Presentation]	xebr	FormationExpensesPresentation	abstract
Formation expenses [Total]	xebr	FormationExpensesTotal	monetary
Fixed assets [Presentation]	xebr	FixedAssetsPresentation	abstract
Intangible assets [Presentation]	xebr	IntangibleAssetsPresentation	abstract
Costs of development	xebr	CostsOfDevelopment	monetary
Concessions, patents, licences, trade marks and similar rights and assets	xebr	ConcessionsPatentsLicencesTradeMarksAndSimilarRightsAndAssets	monetary
Goodwill	xebr	Goodwill	monetary
Payments on account	xebr	PaymentsOnAccount	monetary
Intangible assets [Total]	bach	IntangibleAssetsTotal	monetary
Tangible assets [Presentation]	xebr	TangibleAssetsPresentation	abstract
Property, plant, and equipment [Presentation]	xebr	PropertyPlantAndEquipmentPresentation	abstract
Land and buildings	xebr	LandAndBuildings	monetary
Plant and machinery	xebr	PlantAndMachinery	monetary
Property, plant, and equipment [Total]	xebr	PropertyPlantAndEquipmentTotal	monetary
Other fixtures, fittings and tools	xebr	OtherFixturesFittingsAndTools	monetary
Payments on account and tangible assets in the course of construction	xebr	PaymentsOnAccountAndTangibleAssetsInTheCourseOfConstruction	monetary
Tangible assets [Total]	bach	TangibleAssetsTotal	monetary
Financial assets [Presentation]	xebr	FinancialAssetsPresentation	abstract
Shares in affiliated undertakings	bach	SharesInAffiliatedUndertakings	monetary
Loans to affiliated undertakings	xebr	LoansToAffiliatedUndertakings	monetary

Fig. 1. A screen shot of the Excel version of xEBR – preliminary version 8

In this partial view of the Excel sheet encoding the xEBR taxonomy, the reader can see that the labels/terms used in the taxonomy have various (recurrent) endings that are carrying a specific meaning. In the second row of the first column, for example, the term used is: "COMPANY BALANCE SHEET, HORIZONTAL LAYOUT [REPORT]", one of the eight different instantiations of the type"role" (the various types the xEBR elements can belong to are listed in the right column of the displayed partial Excel sheet). We observe that the word used in square brackets has a special meaning: we are dealing with a "report" element, which does not come as a surprise since XBRL is precisely about

[4] Details on the xEBR taxonomy are given in `http://www.xbrleurope.org/working-groups/xebr-wg/xebr-taxonomy`.

[5] BACH stands for the "Bank for the Accounts of Companies Harmonized". See `https://www.bach.banque-france.fr/?lang=en` for more details.

[6] But the reader can download the full Excel table of a former version of the taxonomy at `http://www.xbrleurope.org/working-groups/xebr-wg/xebr-taxonomy`

business reporting. The third row is introduced by the term "Assets [Presentation]". This one is stating that "Assets" is a presentation element, which as such is not associated with a specific value, but which is relevant for describing the (conceptual) structure of a report.

As a first comment, we can state here that, contrary to the impression given by the hierarchical organisation displayed in the first column of the Excel sheet, the relation between a "report" element and a "presentation" element is not a sub-class hierarchy but much more of a part-of relation. The presentation element "Assets" is a (probably necessary) part of a XBRL report, to be included in its specific "COMPANY BALANCE SHEET, HORIZONTAL LAYOUT" section. While we can assume that there is a sub-class relation between the presentation element "Assets" and the presentation element introduced in the fourth row of the Excel sheet: "Subscribed capital unpaid [Presentation]"; our first interpretation is that "Subscribed capital unpaid" is an "Asset". We think that this distinction between part-of and sub-class relations between reporting elements listed in this hierarchical Excel structure is important.

We observe also that the information about the type of a term is included in its typographical realisation. An exception for this are the terms that represent single financial monetary values ("Plant and machinery" vs. "Property, plant, and equipment [Total]", where "Total" clearly marks the associated type "monetary"). This inclusion of the type name in the term is needed in order to differentiate terms that would have the same typographical realisation, such as "Property, plant, and equipment [Presentation]" vs. "Property, plant, and equipment [Total]". We believe that it is much better to avoid this strategy of using this kind of typographical conventions for naming concepts, but instead to use interpretation independent codes, which mark only the structural relations "A." and "A.1" followed by "A.1.2" or the like, as done in many taxonomies, and to include the terms inside the objects associated with those codes in the form of "labels". Actually, standard XBRL practices include such labels in their XML serializations, but keep the used labels for naming the concepts just by leaving out the blank spaces and using a camel case notation (as can be seen in the third column "Element Name") of the Excel sheet in Figure 1. We advocate for replacing this convention by using more standard codes, in the form of alpha-numeric sequences.

4 xEBR to Ontology

Our mapping from xEBR to a LOD-compliant format concerns first the hierarchical organisation of the taxonomy, then the relations between the concepts listed in the taxonomy and their types: "role", "abstract", "monetary", "tuple", "string", "boolean", or "uri", whereby "role" is subdivided (in this version of the taxonomy) into eight classes (like "CompanyBalanceSheetHorizontalLayoutReport", "CompanyHistoryReport", etc.). And we also foresee two sub classes for the monetary role: "total" value and "single" item value. We finally address the relations between the multilingual labels. In this paper we do not focus on the

information related to company identification, so that we will not expand on its associated types “tuple”, “string”, “boolean” and “uri”.

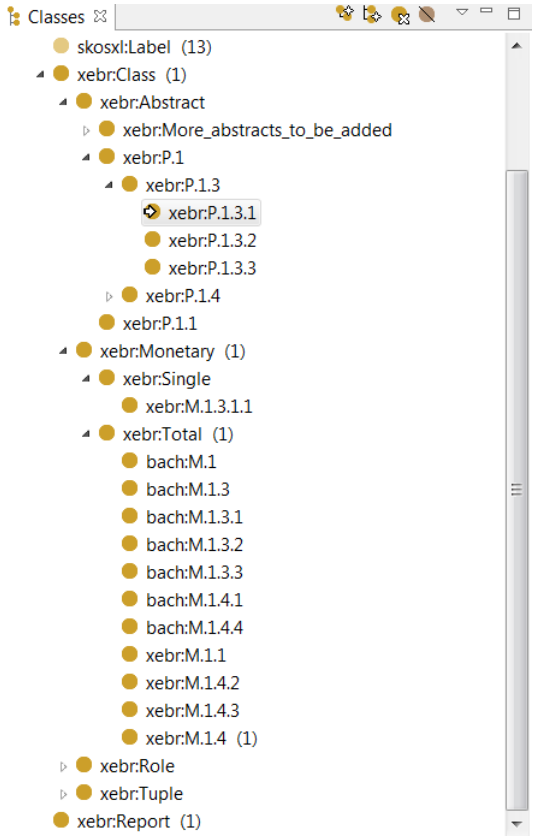

Fig. 2. A screen shot showing partially the class hierarchy of the xEBR Ontology

In Figure 2 we display a partial view of the hierarchy of concepts we derived from the xEBR taxonomy in an OWL and RDF environment. The reader can see that we organize the xEBR concepts as sub-classes of the 4 main types: “abstract”, “monetary”, “role” and “tuple”. The reader can also observe that we are using alpha-numeric codes for encoding the original xEBR taxonomy elements to be seen in the third column of the Excel sheet partially displayed in Figure 2. The labels of the original xEBR elements are now represented by the annotation property rdfs:label, as it is shown in Figure 3.

In the following we just present some examples of the way we ported elements of xEBR to an OWL/RDF(S) ontology. The displayed examples are in the so-called Turtle notation.

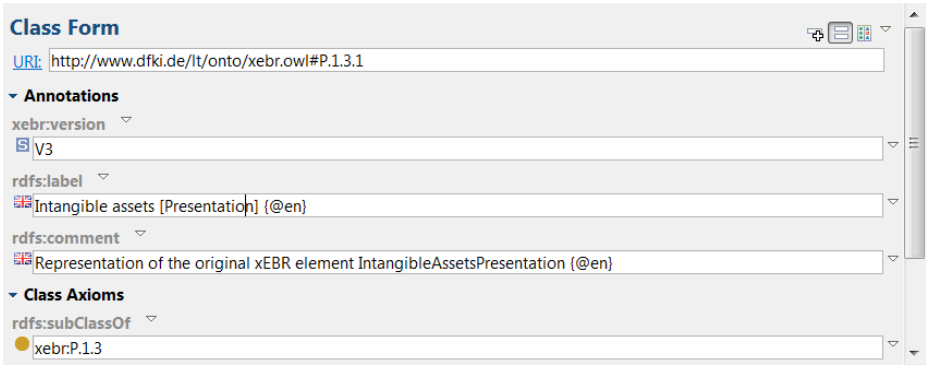

Fig. 3. A screen shot showing the encoding of the IntangibleAssetsPresentation element

The first example states that "AssetsPresentation" (encoded in our ontology as P.1) is an owl:Class and as such a subClass of "Abstract". The second example repeats the same type of encoding, this time for the subClass "FixedAssetsPresentation" (encoded as P.1.3 – showing the hierarchical relation to "AssetsPresentation" by adding a"." to"P.1"). The subClass property is transitive, so that we do not have to state the partOf relation of P1.3 to R.1.

```
<http://www.dfki.de/lt/onto/xebr.owl#P.1>
  rdf:type owl:Class ;
  xebr:isPartOf <http://www.dfki.de/lt/onto/xebr.owl#R.1> ;
  xebr:version "V1" ;
  rdfs:comment "Representation of the xEBR presentation element
   AssetsPresentation"@en ;
  rdfs:label "Assets [Presentation]"@en ;
  rdfs:subClassOf xebr:Abstract ;
.

<http://www.dfki.de/lt/onto/xebr.owl#P.1.3>
  rdf:type owl:Class ;
  xebr:version "V1" ;
  rdfs:comment "Representation of the xEBR presentation element
   FixedAssetsPresentation"@en ;
  rdfs:label "Fixed assets [Presentation]"@en ;
  rdfs:subClassOf <http://www.dfki.de/lt/onto/xebr.owl#P.1> ;
.
```

The partOf relation of "P.1" to the role "Report" in the first example above is marked by the line"xebr:isPartOf `<http://www.dfki.de/lt/onto/xebr.owl#R.1>`". With R.1 we encode the original xEBR report element "CompanyBalanceSheetHorizontalLayoutReport", as displayed just below:

```
<http://www.dfki.de/lt/onto/xebr.owl#R.1>
  rdf:type owl:Class ;
```

```
  xebr:version "V1" ;
  rdfs:comment "Representation of the xEBR report element
   CompanyBalanceSheetHorizontalLayoutReport"@en ;
  rdfs:label "COMPANY BALANCE SHEET, HORIZONTAL LAYOUT [REPORT]"@en ;
  rdfs:subClassOf xebr:Role ;
.
```

The next example introduces another case of a "part-of" relation. The class "FixedAssetsTotal", which is of type "Total" (itself of type "Monetary", and therefore we encode it prefixing it with a "M": M.1.3) is introduced as an element that is part of the P.1.3 presentation element (it is not a subClass of this element, as can be seen also in Figure 2, in which the Monetary has its own class hierarchy, something new compared to the original xEBR taxonomy).

```
<http://www.bach.banque-france.fr#M.1.3>
  rdf:type owl:Class ;
  xebr:isMonetaryPartOf <http://www.dfki.de/lt/onto/xebr.owl#P.1.3> ;
  rdfs:comment "An element of type Monetary representing a total"@en ;
  rdfs:label "Fixed Assets [Total]@en" ;
  rdfs:subClassOf xebr:Total ;
.
```

The following example is also about an element that is of type "Monetary", but it is not representing a "Total" figure, therefore we introduced a "calculation" property (very similar to what is foreseen in XBRL). This property is used to relate the element "CostsOfDevelopment" (M.1.3.1.1) to the "Total" element it contributes to: "IntangibleAssetsTotal" (M.1.3.1), which itself is a part of P.1.3.1 ("IntangibleAssetsPresentation").

```
<http://www.dfki.de/lt/onto/xebr.owl#M.1.3.1.1>
  rdf:type owl:Class ;
  xebr:isCalculationElementOf <http://www.bach.banque-france.fr#M.1.3.1> ;
  rdfs:comment "Element of the class Monetary that
   has a single value" ;
  rdfs:label "Costs of development@en" ;
  rdfs:subClassOf xebr:Single ;
.

<http://www.bach.banque-france.fr#M.1.3.1>
  rdf:type owl:Class ;
  xebr:isMonetaryPartOf <http://www.dfki.de/lt/onto/xebr.owl#P.1.3.1> ;
  rdfs:comment "An element of type Monetary representing a
   total"@en ;
  rdfs:label "Intangible assets [Total]@en" ;
  rdfs:subClassOf xebr:Total ;
.

<http://www.dfki.de/lt/onto/xebr.owl#P.1.3.1>
  rdf:type owl:Class ;
```

```
  xebr:version "V3" ;
  rdfs:comment "Representation of the original xEBR element
   IntangibleAssetsPresentation"@en ;
  rdfs:label "Intangible assets [Presentation]"@en ;
  rdfs:subClassOf <http://www.dfki.de/lt/onto/xebr.owl#P.1.3> ;
.
```

The following lists a few properties that we have created so far.

```
xebr:isCalculationElementOf
  rdf:type owl:FunctionalProperty ;
  rdf:type owl:ObjectProperty ;
  rdfs:comment "This property marks an element as a part of a
   total Monetary item" ;
  rdfs:domain xebr:Monetary ;
  rdfs:label "Is Calculation Element Of"@en ;
  rdfs:range xebr:Total ;
.

xebr:isMonetaryPartOf
  rdf:type owl:FunctionalProperty ;
  rdf:type owl:ObjectProperty ;
  rdfs:comment "This functions marks the \"monetary\" parts of a
   presentation part of a report" ;
  rdfs:domain xebr:Monetary ;
  rdfs:label "Is Monetary Part Of"@en ;
  rdfs:range xebr:Abstract ;
.

xebr:isPartOf
  rdf:type owl:FunctionalProperty ;
  rdf:type owl:ObjectProperty ;
  rdfs:comment "This functions marks the part of xEBR document/Report
      in which the items should be included" ;
  rdfs:domain xebr:Abstract ;
  rdfs:label "Is Part Of"@en ;
  rdfs:range xebr:Role ;
.
```

5 Representing Multilingual Information

One of the major assets of XBRL, and even more xEBR, is the fact that conceptual information is equipped with (multilingual) labels, representing the way the concepts are linguistically realised in the corresponding national legislation, with the expectation that companies will use a very similar terminology in their reports. Just to name an example, the Belgian XBRL taxonomy comes with labels in four languages (English and the three national languages: Dutch, French and German). The xEBR initiative supports the specification of multilingual

equivalents within the context of the identified shared concepts across various legislations.

While the XBRL/xEBR support of multilingualism is definitely a major achievement, we see that this information can be expressed only in a somewhat cumbersome and redundant manner in the used XML code. Also the so-called "link base" organisation of the label data[7] is not able to express terminological generalisations and relations between the used terms. Additionally, we would like to use a representation language that is compatible with the one we used for the representation of the concepts, their types and their relations for encoding the multilingual terms (or XBRL/xEBB labels). We opted therefore for (at least in a first step) SKOS-XL. SKOS has been designed as an OWL and RDF(S) compatible vocabulary for representing light-weight ontologies, like thesauri and terminologies. SKOS-XL is an extension of SKOS that allows to handle the classical rdfs:label as an object, upgrading thus the content of some OWL annotation properties to an autonomous element that can be manipulated from several ontologies[8]. And it seems that this is exactly what we need for representing the XBRL/xEBR (multilingual) labels, as we illustrate with a few examples below.

All labels of xEBR taxonomy are then introduced as an instance of the skosxl:Label class. We can then explicitly describe relations between those labels, so for example that "Crediti Verso Soci Versamenti Ancora Dovuti" can be considered an Italian equivalent to "Subscribed capital unpaid [Presentation]". Both labels are then marked as prefLabel for the concept "Subscribed Capital Unpaid Presentation". It is compliant with the guidelines of SKOS-XL that a concept has more than one object marked as prefLabel, if each object represents a term in a distinct language. We foresee the use of skosxl:altLabel for terminological variants one can observe, for instance, in the actual reports generated by companies.

```
xebr:L_SubscribedCapitalUnpaidPresentation
  rdf:type skosxl:Label ;
  rdfs:comment "English PrefLabel for P.1.1
   SubscribedCapitalUnpaidPresentationSubscribe"@en ;
  rdfs:label "Subscribed Capital Unpaid Presentation"@en ;
  skosxl:isTermTranslationOf xebr:L_CreditiVersoSociVersamentiAncoraDovuti ;
  skosxl:literalForm
   "Subscribed capital unpaid [Presentation]"@en ;
.

xebr:L_CreditiVersoSociVersamentiAncoraDovuti
  rdf:type skosxl:Label ;
  rdfs:comment "Italian Pref Label for P.1.1.
   SubscribedCapitalUnpaidPresentation" ;
  rdfs:label "Crediti Verso Soci Versamenti Ancora Dovuti"@it ;
  skosxl:isTermTranslationOf xebr:L_SubscribedCapitalUnpaidPresentation ;
```

[7] See `https://www.xbrl.org/2003/xbrl-linkbase-2003-12-31.xsd` for more details.

[8] See `https://www.w3.org/TR/skos-reference/skos-xl.html` for more details.

```
  skosxl:literalForm
   "Crediti Verso Soci Versamenti Ancora Dovuti"@it ;
.

<http://www.dfki.de/lt/onto/xebr.owl#P.1.1>
  rdf:type owl:Class ;
  xebr:isPartOf <http://www.dfki.de/lt/onto/xebr.owl#R.1> ;
  xebr:version "V5" ;
  rdfs:label "Subscribed capital unpaid [Presentation]"@en ;
  rdfs:subClassOf xebr:Abstract ;
  skosxl:prefLabel xebr:L_SubscribedCapitalUnpaidPresentation ;
.
```

Using SKOS-XL for modelling the (multilingual) labels, we are thus in the position of establishing generalisations and marking explicitly relations between the various labels within a legislation and the terminological variants of such concepts as used in concrete reports and other data sources, as exemplified in the next section.

6 Multilingual Short- and Long-Term Data Example

Shared semantic concepts between up-to-date information on, e.g. stock exchanges, and the XBRL reported information can provide a very powerful picture on the finances of a corporation. Those shared concepts can be utilized in order to align those two types of information. It has been shown before that structural as well as value-based similarities can be utilized to align a company's financial data across languages [3]. Concentrating on various reports in various languages for one company for a specific year allows for the additional use of a simple heuristics in order to detect multilingual term correspondences: the financial positions associated with terms have the same values. While this heuristic alone is too simplistic and would lead to faulty results, combined with a semantic similarity measure applied to its labels it can be a powerful tool to align short-term and long-term reporting information.

German	**Figure**	**English**	**Source**
Kurzfristige Vermögenswerte	24.861	Short-term assets	BASF
Kurzfristige Vermögen	24.861		Bundesanzeiger
Umlaufvermögen	24.861	Total Current Assets	DAX

Table 1. Example of heuristic-based alignment of information on one specific company across sources

As shown in Table 1, the figure-based aggregation of multilingual information is capable of uncovering terminological variants in German and English. Based on this heuristic, i.e., the identical figure shown for this one company's position, we learn that "Kurzfristige Vermögenswerte" is equivalent in meaning to

"Kurzfristige Vermögen" as well as "Umlaufvermögen". Since two of the sources provided in Table 1 are bilingual, we can learn two English translations of those German terms, namely "Short-term assets" and "Total Current Assets". While similarity measures might perform well on the first two German terms, their capability of identifying a similarity to "Umlaufvermögen" might be limited. Here the multilingual aspect fosters alignment, since most (not purely letter-based) similarity measures provide higher values for the two English terms. Mediated by the corresponding xEBR concepts – "xebr:CurrentAssetsTotal" – these German and English term variants can then be automatically linked to other languages, such as the xEBR Spanish label "activo corriente" and the associated value can be compared to values reported in the long-term financial report. Linking the xEBR label to the English DAX terms allows for enriching the ontology with terminological variation that can be useful for detecting terms in actual company reports. However, the method for this linking should be elaborated on by, for instance, using distributional semantic models.

7 Conclusion

Representing the xEBR taxonomy as an OWL ontology enables the use of semantic technologies. However, this benefit can only be fully leveraged if the implicit semantics of the taxonomy are explicated in the ontology, an approach that we present in this paper. One aspect is also that we are able to improve the modularity of the original taxonomy, as can be seen in the fact that monetary elements have now their own class hierarchy. And we can mark explicitly the relations between labels by the use of specialized properties. We also show how this formalization of an XML-based format (or in case of xEBR, an Excel based format) can foster the alignment of short-term reporting information to long-term XBRL reporting by means of a brief example of stock exchange data.

One of the major benefits of this representation format is the ability to automatically check the consistency of data redundancy and generate financial summaries - short- and long-term. Using xEBR as an ontology allows for the comparison of companies across jurisdiction and across languages. If stakeholders from different countries wish to obtain detailed information on their investment in their native language, the proposed multilingual representation of financial information can be very handy for human analysts.

Acknowledgement

We would like to thank the XBRL Europe Business Registers Working Group (xEBR), and more specially Thomas Verdin, for their exciting work on the taxonomy and for providing us with the latest version of the Core Reference Taxonomy. We also thank Hans-Ulrich Krieger for his ground-breaking and inspiring work on the ontologisation of XBRL. Our thank goes also to the anonymous reviewers for their helpful comments on our submission.

References

[1] Bai, L., Koveos, P., Liu, M.: Applying an ontology-augmenting xbrl model to accounting information system for business integration. Asia-Pacific Journal of Accounting & Economics pp. 1–23 (2016)

[2] Bao, J., Rong, G., Li, X., Ding, L.: Representing financial reports on the semantic web. In: International Workshop on Rules and Rule Markup Languages for the Semantic Web. pp. 144–152. Springer (2010)

[3] Declerck, T., Gromann, D.: Extraction of multilingual term variants in the business reporting domain. In: Proceedings of CHAT 2012: The 2nd Workshop on the Creation; Harmonization and Application of Terminology Resources; Co-located with TKE 2012; June 22; 2012; Madrid; Spain. pp. 41–46. No. 072, Linköping University Electronic Press (2012)

[4] Declerck, T., Krieger, H.U.: Translating xbrl into description logic. an approach using protege, sesame & owl. In: BIS. pp. 455–467 (2006)

[5] Fourny, G.: The XBRL Book – Simple, precise, technical. Ghislain Fourny (1 2017), the first two chapters can be downloaded at https://ghislainfourny.github.io/the-xbrl-book/download/the-xbrl-book.pdf

[6] García, R., Gil, R.: Publishing xbrl as linked open data. In: CEUR Workshop Proceedings. vol. 538. Citeseer (2009)

[7] Krahel, J.P.: On the formalization of accounting standards. Ph.D. thesis, Rutgers University-Graduate School-Newark (2012)

[8] Liu, C., Luo, X.R., Wang, F.L.: An empirical investigation on the impact of xbrl adoption on information asymmetry: Evidence from europe. Decision Support Systems 93, 42–50 (2017)

[9] O'Riain, S.: Semantic Paths in Business Filings Analysis. Ph.D. thesis, National University of Ireland, Galway (2012)

[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 13(2), 141–162 (2012)

[11] Santos, I., Castro, E., Velasco, M.: Xbrl formula specification in the multidimensional data model. Information Systems 57, 20–37 (2016)

[12] Spohr, D., Hollink, L., Cimiano, P.: A machine learning approach to multilingual and cross-lingual ontology matching. In: International Semantic Web Conference. pp. 665–680. Springer (2011)

[13] Thomas, S.M., Wu, X., Ma, Y., O'Riain, S.: Semantically assisted XBRL-taxonomy alignment across languages. In: Buitelaar, P., Cimiano, P. (eds.) Towards the Multilingual Semantic Web, pp. 277–293. Springer (2014)

[14] Trojahn, C., Fu, B., Zamazal, O., Ritze, D.: State-of-the-art in multilingual and cross-lingual ontology matching. In: Towards the Multilingual Semantic Web, pp. 119–135. Springer (2014)