# XBRL Taxonomies and OWL Ontologies for Investment Funds

Rubén Lara[1], Iván Cantador[2], and Pablo Castells[2]

[1] Tecnología, Información y Finanzas, Madrid, Spain
rlara@afi.es
[2] Universidad Autónoma de Madrid, Spain
{ivan.cantador, pablo.castells}@uam.es

**Abstract.** The analysis of investment funds information requires the availability of homogeneous information of the funds considered, which is usually generated and provisioned by different parties and in different formats. In this context, the gathering and integration of information from disparate, heterogeneous sources becomes a key task that can be considerably eased by the availability of explicit and shared information models. Furthermore, the analysis process leads to the generation of added-value information, whose consumption by other parties can also benefit from the existence of agreed information models. In this paper, we present our work on building explicit information models for investment funds in the Spanish market. An XBRL taxonomy of funds, and a translation process of XBRL taxonomies into OWL ontologies that has been applied to this taxonomy in order to obtain an OWL ontology of funds, are presented. The relative benefits of using OWL ontologies or XBRL taxonomies for the exchange and analysis of investment funds information are also discussed.

## 1 Introduction

Companies devote considerable efforts to the management of their information [2], requiring the integration of information from disparate and heterogeneous sources. In the financial field, a conceptually rich domain where information is complex, huge in volume and a highly valuable business product by itself [4], the exchange and integration of information for its posterior analysis is a key task for financial analysts. In particular, the analysis of investment funds requires the availability of homogeneous and consistent information, both up-to-date and historical, on the descriptive aspects (Net Asset Value -NAV-, commissions, etc.) of the funds subject of analysis.

Tecnología, Información y Finanzas (TIF), in cooperation with AFINet Global[1], acts in the Spanish market as a provider of analytical information. For providing this service, TIF continuously receives and aggregates information from the national stock markets, from firms managing investment funds, and from the national market supervisor (the CNMV)[2], covering all the investment funds currently commercialized in Spain with a

[1] http://www.grupoanalistas.com
[2] http://www.cnmv.es

J.F. Roddick et al. (Eds.): ER Workshops 2006, LNCS 4231, pp. 271–280, 2006.

10-years historical base (over 6000 investment funds at the time of writing). The information received includes descriptive aspects of a fund when it starts to be commercialized (entity commercializing the fund, investment policy, etc.), changes on any of these data, and the NAV of the fund at different points in time.

The different parties from which TIF receives information use heterogeneous information models and formats. This makes the reception, validation, and aggregation of information a difficult task, and requires ad-hoc validation procedures and a costly maintenance. As depicted in Figure 1, descriptive information about funds commercialized in the Spanish market is provided by the CNMV, and periodic information such as the NAV of a fund is provided by national stock markets (Madrid, Bilbao, Valencia and Barcelona) and by the firms managing the funds. This information is validated and transformed, leading to the creation of an aggregated and consistent information base that is ready for analysis. The analysis process leads to added-value information consumed by agents such as management firms, sellers, or investors.

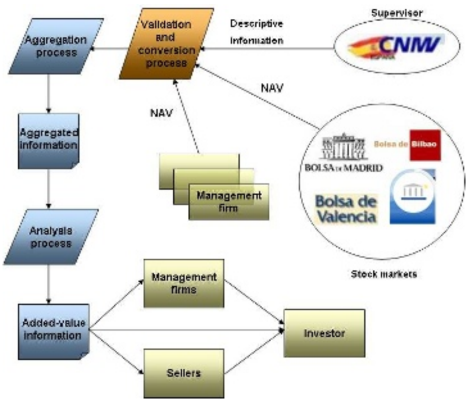


**Fig. 1.** Information life cycle

A gain in efficiency in the life cycle of Figure 1 can be achieved if the validation and conversion process, instead of dealing with heterogeneous information, would receive information according to a shared model so that ad-hoc processing can be avoided and maintenance needs are reduced. Furthermore, if the analytical, added-value information produced also follows an agreed model, the consumption of such information by different agents can be considerably eased.

The CNMV is considering the definition of XBRL [1] taxonomies for modelling some descriptive aspects of investment funds. However, these taxonomies would not include analytical information. Furthermore, OWL [3], the W3C recommendation for the definition of ontologies [9], has not been considered so far as an alternative for defining shared information models for investment funds. In this setting, we have worked on: a) an XBRL taxonomy that includes descriptive *and* analytical information of funds, and that can serve as a basis for possible future developments led by the CNMV or for their extension, and b) on the evaluation of OWL as an alternative to XBRL.

The paper is structured as follows: XBRL is introduced in Section 2, as well as the XBRL taxonomy created. Section 3 presents the process designed for translating XBRL taxonomies into OWL ontologies. Finally, Section 4 discusses the advantages and disadvantages of using XBRL or OWL for modelling investment funds information and provides some concluding remarks.

## 2 An XBRL Taxonomy of Investment Funds

In this section, we introduce XBRL and present the XBRL taxonomy developed for the modelling of descriptive and analytic aspects of investment funds.

### 2.1 XBRL in a Nutshell

XBRL builds on top of XML, XML Schema and XLink. **Taxonomies**, which provide the elements that will be used to describe information, and **instances**, which provide the real content of the elements defined, are the main ingredients of XBRL.

An **XBRL taxonomy** is constituted by an XML Schema and the XLink linkbases contained in or referenced by that schema, and it can be part of a set of related taxononomies called a Discoverable Taxonomy Set (DTS) [1]. The XML Schema in a taxonomy defines reporting concepts, which are given a name and a type, as XML Schema element definitions. For example, the *NAV* concept of a fund would typically have a *monetary* type. Additional constraints on how concepts can be used (e.g. instant/duration period, debit/credit balance) are documented by other XBRL attributes on the XML Schema element definitions.

Linkbases are collections of XLink extended links, which provide further information about the meaning of the concepts by expressing relationships between concepts (inter-concept relationships) and by associating concepts to their documentation. Taxonomies make use of five types of XLink linkbases, namely: definition, calculation, presentation, label, and reference linkbases.

Definition links describe relations among concepts in a taxonomy, such as general-special relations, that provide information on what an element actually. Calculation linkbases provide information on how some elements are calculated in terms of some other elements. Presentation linkbases contain relations such as parent-child that are exclusively used for presentation purposes.

Label links define labels in natural language for concepts in the taxonomy; the same concept can have labels in different languages. Reference links point to legal or other type of documentation that explains the meaning of a given taxonomy element. As it can be seen, both label and reference links do not define relations among elements but document elements in a taxonomy.

Actual facts are reported as **XBRL instances**, which organize reporting information using two main elements: XBRL items and XBRL tuples. *Items* are defined as extensions of primitive data types (String, Integer, Boolean, etc.), and they represent atomic information elements of an XBRL instance. However, some facts are dependent on each other and they must be grouped for a proper and complete understanding. For instance, in reporting information of a fund, each deposit entity name has to be properly associated to a correct deposit entity identifier. Such sets of facts (entity name, entity identifier) are called *tuples*. Tuples may contain both items and other tuples.

In addition to the actual values of a fact, XBRL instances provide contextual information necessary for interpreting such values e.g "NAV is 50 *today*" through the use of XBRL context elements. Furthermore, for numeric facts, XBRL instances can document measurement units e.g. "NAV is $50" by using XBRL unit elements.

### 2.2 An XBRL Taxonomy of Investment Funds

The lack of explicit and shared models for exchanging information in the investment funds market and the promotion and increasing adoption of XBRL by Spanish regulators and supervisors e.g. Bank of Spain and CNMV led us to consider XBRL as a candidate language for creating an explicit information model. We started by evaluating and reviewing the information model used by TIF in order to define a revised model that could meet the needs of different agents in the market. We counted with the cooperation of Analistas Financieros Internacionales[3], a leading company in the analysis of the Spanish financial market, and Gestifonsa[4], a funds management firm.

The revised model has been described using XBRL. The reuse of existing taxonomies (IPP[5], DGI[6], and ES-BE-FS[7]) has been evaluated, and it has been concluded that the DGI taxonomy can be reused for the description of the entities that commercialize or manage a given fund. Figure 2 shows the DTS of the taxonomy built, where *dgi-lc-es-2005-03-10.xsd* contains the information elements of the imported DGI taxonomy in Spanish and its respective linkbases, and *dgi-lc-int-2005-03-10.xsd* contains the international elements of the DGI taxonomy. The information elements of the taxonomy created have been divided into the following groups:

- Descriptive information: models all the descriptive aspects of a fund, such as the name of the fund, the entity managing the fund, etc.
- Relevant facts information: models relevant facts about a given fund, such as changes in its investment policy.
- Periodic descriptive values: models descriptive information periodically updated, such as the NAV of the fund or the number of unit holders.
- Analytic information: models the analytic values associated to a fund, such as performance measures, the rating of the fund in its category, etc.

The reason for identifying these four distinct groups of information (being the root of each group an XBRL tuple) is that the information they contain has a different nature, the sources providing the information are different, and the periodicity with which each group of information is produced is diverse. Besides the information elements created, presentation, label, calculation and reference linkbases have been defined.

3 http://www.afi.es
4 http://www.cajacaminos.es/
5 http://www.xbrl.org.es/informacion/ipp.html
6 http://www.xbrl.org.es/informacion/dgi.html
7 http://www.xbrl.org.es/informacion/es_be_fs.html

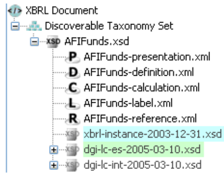


**Fig. 2.** DTS of investment funds

Definition links have not been used as: a) the use of links of type *requires-child* is not recommended in [10], b) there are no equivalent elements in the taxonomy, so links of type *essence-alias* have not been used, c) no use was found for links of type *general-special*, and d) there are no similar tuples for which a link of type *similar-tuples* makes sense. The latest version of the taxonomy can be found at *http://www.tifbrewery.com/tifBrewery/resources/XBRLTaxonomies.zip*.

## 3 Translating XBRL Taxonomies into OWL Ontologies

We have identified OWL as a potential alternative to the use of XBRL which might present some features that are of practical interest in the investment funds market. In order to evaluate the use of OWL ontologies we have developed a generic translation process of XBRL taxonomies into OWL ontologies so that existing and future taxonomies can be easily converted into OWL ontologies.

### 3.1 Description of the Translation Process

XBRL taxonomies provide explicit and shared information models and, thus, they are very similar to ontologies except that they do not have a formal semantics for all the aspects of the model. Similarly, XBRL instances can be seen as ontology instances and expressed as such. Therefore, we have designed a translation process of XBRL taxonomies into OWL ontologies, and of XBRL instances into OWL instances. In the following, we will restrict ourselves to the translation of taxonomies into ontologies.

An automatic translator has been implemented based on the process that will be presented. It has been tested by translating not only our funds taxonomy but also other XBRL taxonomies, including the DGI, IFRS-GP[8], ES-BE-FS and IPP taxonomies. The latest version of the ontologies obtained can be found at *http://www.tifbrewery.com/resources/OWLOntologiesv2.zip*.

As XBRL is an XML based technology, the first step in the translation process is to parse the XML elements. Using JDOM[9], an XML parsing module obtains the XML elements in the XBRL taxonomies, instances, and links to be translated. The translation steps described below are then applied to the obtained elements, resulting in a Jena[10]

[8] http://xbrl.iasb.org/int/fr/ifrs/gp/2005-05-15
[9] http://www.jdom.org
[10] http://jena.sourceforge.net

model that corresponds to the OWL ontologies and instances derived from the XBRL taxonomy and instances given, which is saved to text files.

The correspondence between the upper level XBRL elements and the OWL classes generated in summarized in Table 3.1. In the following, we describe the steps for the automatic translation of XBRL elements, following this correspondence. For the sake of simplicity, we will refer to the DGI taxonomy in the explanations. The translation process for other taxonomies is analogous.

**Table 1.** Summary of parsed taxonomy element translations

| Parsed taxonomy element | Root OWL class | Direct OWL subclasses |
|---|---|---|
| XML complex types | DGI_ComplexType | A subclass for each complex type |
| XBRL Tuples<br>XBRL items | DGI_Element | DGI_Tuple<br>DGI_Item |
| XLink links | DGI_Link | DGI_LabelLink<br>DGI_PresentationLink<br>DGI_CalculationLink |
| XBRL Contexts | Context<br>(range of properties<br>is subclass of<br>ContextElement) | Subclasses of ContextElement:<br>ContextEntity<br>ContextEntityElement (Identifier)<br>ContextPeriod<br>ContextScenario |
| XBRL units | Unit<br>(range of properties<br>is subclass of<br>UnitElement) | Subclass of UnitElement:<br>UnitMeasure |

*1. Declaration of a root OWL class Element* from which complex (tuples) and simple (items) information parts of the taxonomy will inherit, named DGI_Element for the DGI taxonomy. This class has associated a property *xbrl_id*, corresponding to the XBRL attribute *id* common to all XBRL elements.

*2. Declaration of DGI_Tuple and DGI_Item subclasses of DGI_Element. XBRL tuples and items correspond to OWL subclasses of DGI_Tuple and DGI_Item, respectively. The attributes of XBRL Item are translated into the OWL properties:* xbrl_balance*, with possible values "credit" and "debit";* xbrl_periodType*, with possible values "instant" and "duration";* xbrl_contextRef*, whose range is the OWL class Context (step 11); and* xbrl_unitRef*, whose range is the OWL class Unit (step 12).*

*3. Declaration of a root OWL class DGI_ComplexType*. XML complex types are translated into classes that inherit from DGI_ComplexType, having OWL properties: *xml_name* to store the name of the complex type, *xbrl_periodType*, with possible values "instant" and "duration", and *xbrl_contextRef*, whose range is the Context class.

*4. Syntactic translation of XML complex types into OWL subclasses of DGI_Complex Type*. The names of the obtained subclasses are those stored in the XML attribute *name* of the complex type elements. Each subclass of DGI_ComplexType has a property whose name is the concatenation of the complex type name and the word "value", and whose type is the primitive data type associated to the complex type (xsd:string, xsd:integer, xsd:boolean, etc.). Additionally, they contain those properties defined in the primitive XBRL data types (xbri:stringItemType, xbrli:integerItemType, xbrli:boolean ItemType, etc.). For example, in the DGI taxonomy, the class AddressFormatCodeItem-Type has the property *length* with a fixed value of 2, indicating that the possible values of the data type can only have 2 characters.

*5. Syntactic translation of XBRL Items into OWL subclasses of DGI_Item*. The names of the obtained subclasses are those stored in the XML attribute "name" of the item elements. Each subclass of DGI_Item has a property for storing the value of the item, and whose range is the type of the XBRL item.

*6. Record XBRL Tuples as OWL subclasses of DGI_Tuple*. Initially, they are created empty, and their properties are added in step 7. The reason is that tuple properties will reference other tuples, which might be not yet created and which will have to exist in the OWL model that is being built.

*7. Syntactic translation of the XBRL tuple attributes into OWL object properties*. The attributes of the tuples are added to the subclasses of DGI_Tuple as OWL object properties. These properties will have as range a class associated to a complex type of step 4, a class created in step 5 or a class recorded in step 6.

*8. Declaration of a root OWL class DGI_Link*. Its instances, which correspond to the XLink links of the XBRL taxonomies, contain the properties: *xlink_from*, created for the translation of the XLink attribute *from*, stores the origin element of the link; *xlink_to*, created for the translation of the XLink attribute *to*, indicates the destination element of the link; *xlink_role*, created for the translation of the XLink attribute *role*, indicates the role assigned to the link: "label", "calculation", etc.

*9. Declaration of OWL subclasses of DGI_Link*. Subclasses of DGI_Link are built for each type of link: DGI_LabelLink, DGI_PresentationLink, DGI_CalculationLink, DGI_ReferenceLink, and DGI_DefinitionLinks.

*10. Syntactic translation of XBRL linkbases into instances of the corresponding subclasses of DGI_Link*. Links in XBRL linkbases are translated into OWL instances of the different subclasses of DGI_Link (for reasons of space, only the translation of label, presentation and calculation linkbases is presented):

- Label links are translated into OWL instances of DGI_LabelLink. In addition to the common link properties (*from*, *to*, *role*), label links have properties: *xbrl_label*, obtained from the translation of the XBRL attribute *label* and used to store the text of the label, and *xml_lang*, obtained from the translation of the XML attribute *lang* and used to indicate the language of the label.
- Presentation links are translated into instances of DGI_PresentationLink. Besides common link properties, presentation links have properties: *xbrl_order*, from the translation of the attribute *order* and used to store the relative position of the destination element within the presentation of the origin element, and *xbrl_preferedLabel*, obtained from the translation of *preferedLabel*.
- Calculation links are translated into OWL instances of DGI_CalculationLink. Additionally to common link properties, calculation links have properties: *xbrl_order*, obtained from the translation of the XBRL attribute *order* and used to store the relative position of the destination element value within the calculation of the origin element value, and *xbrl_weight*, obtained from the translation of the XBRL attribute

*weight* and used to store the weight of the destination value within the calculation of the origin element value.

*11. Syntactic translation of XBRL contextRef elements*. In order to translate XBRL contexts, a new ontology has been created, which will be imported by the ontologies resulting from the translation of XBRL taxonomies. This ontology contains a main class Context. The Context class has the following properties: a) *xbrl_id*, of type xsd:ID, for the translation of the XBRL attribute *id* to identify each context, b) *xbrl_entity*, of type ContextEntity, defined for the translation of *entity*, c) *xbrl_period*, of type ContextPeriod, defined for the translation of *period*, and d) *xbrl_scenario*, of type OWL Thing, and defined for the translation of *scenario*. Other classes such as ContextPeriod (with subclasses ContextForeverPeriod, ContextInstantPeriod, and ContextStartEndPeriod), ContextEntityElement and ContextScenario are defined corresponding to the types of values that define an XBRL context.

*12. Syntactic translation of XBRL unitRef elements*. For the translation of units defined in an XBRL taxonomy, an independent OWL ontology has been created. This ontology will be imported by ontologies resulting from the translation process. Its main class is Unit, which has a property *xbrl_unitMeasure* of type UnitMeasure and whose content is the definition of the associated unit. The UnitMeasure class, used to define the units added in a given context, does not have properties. Its subclasses distinguish the different types of units: Divide for units defined by means of a ratio (with properties *xbrl_unitNumerator* and *xbrl_unitDenominator*), and Measure for simple units (with property *xbrl_measure*).

Besides the order of steps presented above, the hierarchy and relationships between elements within a taxonomy, and the relationships among different taxonomies, will define their translation order.

## 4 Discussion

The translation process presented in the previous sections helps to identify similarities and differences between XBRL and OWL, described below.

*XBRL items and tuples.* There is a correspondence between XBRL items and tuples and OWL classes. Items correspond to classes that only have one value (besides information such as the period, context, etc.), and tuples correspond to classes with object properties that store the constituent parts of the tuple. In this sense, items and tuples can be naturally represented by OWL classes.

*XBRL contexts and units.* An important feature of XBRL is the possibility of associating contexts and units to elements. This can be done in OWL by creating ontologies for contexts and units, as presented in the previous subsection, and by including appropriate object properties in OWL classes representing XBRL items and tuples.

*Reference and label links.* They can be represented in OWL by creating appropriate classes and instances, as done in our translation. As these links are intended for documentation, no formal semantics is associated to them. Furthermore, no application of a possible formal semantics for this type of links is envisioned.

*Definition links.* Definition links can be represented by creating instances of the classes introduced in the previous subsection. Special attention deserves the representation of *general-special* definition links which, even though they are currently translated into instances of definition link classes, naturally correspond to subclass relations in ontologies. However, existing taxonomies e.g. IPP, DGI, or IFRS-GP hardly make use of general-special definition links. A reason for this is that this type of links is not exploited by current XBRL tools to infer additional information, as this kind of relation does not currently have a formal semantics. We believe that the formalization of subclass relations can be of interest in practical applications, and that general-special definition links could be given formal semantics by using OWL.

*Calculation links.* Calculation links can be represented in the way outlined in the previous section. However, these links have a formal, mathematical semantics in XBRL, while in OWL this semantics is not supported. Therefore, we believe that for OWL ontologies to be adopted in the financial domain, where mathematical relations are highly relevant for data validation, building of mathematical support on top of OWL would be required.

*Presentation links.* Presentation links can be represented as described by our translation process. However, unless OWL visualization tools are adapted to take into account presentation information, they will be meaningless.

*Open-World Assumption (OWA) vs Closed-World Assumption (CWA).* The semantics of OWL is based on classical First-Order Logic (FOL) [8], and the OWA is made i.e. information is not assumed to be false if it cannot be proven to be true. However, in an industrial setting the CWA is widely made e.g. in relational databases. In fact, XBRL users are expected to intuitively make the CWA when, for example, querying for particular information of an investment fund. Due to his background, an average user would most likely see natural a "no" answer to que question "Is the investment fund $myFund$ classified in category $myCategory$?" if, from available information, the investment fund is not classified under this category. Locally closing the world using an epistemic operator for OWL could be a solution to this problem [7]. In adition, OWL does not define constraints but restrictions, as explained in [6]. However, for validation purposes we believe that the use of constraints is required.

Summarizing, the major advantage we see from the use of OWL is its formal semantics, which can be exploited for the automatic classification of funds if general-special relations are used and represented as OWL subclass (or subsumption) relations. As implicit subsumption relations can be automatically inferred using Description Logics reasoners [11], customers or analysts can e.g. formally define the characteristics of funds they are interested in and appropriate funds will automatically and precisely be found. In particular, we are investigating the application of formal semantics to personalization in the reception of information in the investment funds market and to the automated classification of funds. For this purpose, we plan to analyze subsumption relations present in current taxonomies but not explicitly declared. However, the Open-World semantics of OWL and the use of restrictions instead of constraints can hamper the use of OWL for querying investment funds information and for validating information reported. This problem will be further investigated in the future. Additionally, we believe extensions of OWL to incorporate and validate mathematical relations in the style of XBRL are necessary for the use of OWL ontologies in the financial domain.

Our conclusion is, thus, that extensions to OWL are required in order to fulfill all the requirements of financial information reporting, and that while its semantics can be appropriate e.g. for investment funds classification, it might be problematic for e.g. validation purposes. We believe that the XBRL community has accomplish a remarkable success in the definition of agreed, shared models; the existence of these models is actually good news for the semantic Web community, which seeks similar goals. However, XBRL can be improved in the direction of adding formal semantics to it and, thus, benefit from the work done by the semantic Web community. Similarly, the semantic Web community can identify paths for improvement and development of OWL by studying the increasing adoption of XBRL in the financial domain, proposing extensions and modifications targeting at this domain.

Future work will concentrate on evaluating alternative languages for the formal description of investment funds, especially the use of the WSML family of languages [5], which provides a basic interoperability layer and extensions in the direction of Descriptions Logics and in the direction of Logic Programming.

**Acknowledgements.** This work has been partially funded by the PROFIT programme of the Spanish Ministry of Industry, Tourism and Commerce (FIT-340000-2005-256).

## References

1. eXtensible Business Reporting Language. Technical report, XBRL International.
2. V. Alexiev, M. Breu, J. de Bruijn, D. Fensel, R. Lara, and H. Lausen. *Information Integration with Ontologies : Experiences from an Industrial Showcase*. Wiley, 2005.
3. S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, and L. A. Stein. OWL Web Ontology Language Reference. Technical report, 2004.
4. P. Castells, B. Foncillas, and R. Lara. Semantic web technologies for economic and financial information management. In *ESWS 2004*, Heraklion, Greece, May 2004.
5. J. de Bruijn, H. Lausen, R. Krummenacher, A. Polleres, L. Predoiu, M. Kifer, and D. Fensel. The Web Service Modeling Language WSML. Technical report, WSML, 2005.
6. J. de Bruijn, A. Polleres, R. Lara, and D. Fensel. OWL DL vs. OWL Flight: Conceptual Modeling and Reasoning for the semantic Web. In *WWW conference*, 2005.
7. F. M. Donini, M. Lenzerini, D. Nardi, W. Nutt, and A. Schaerf. An epistemic operator for Description Logics. *Artificial Intelligence*, 100(1-2):225–274, 1998.
8. M. Fitting. *First order logic and automated theorem proving*. Springer, 2nd edition, 1996.
9. T. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, 1993.
10. W. Hamscher, M. Goodhand, C. Hoffman, B. Homer, J. MacDonald, G. Shuetrim, and H. Wallis. Financial reporting taxonomies architecture 1.0. Technical report, XBRL, 2006.
11. D. Nardi, F. Baader, D. Calvanese, D. L. McGuinness, and P. F. Patel-Schneider (edts.), editors. *The Description Logic Handbook*. Cambridge, January 2003.