MAPPING THE FUTURE OF FINANCIAL DATA: A TAXONOMY AND ONTOLOGICAL MODEL PERSPECTIVE TOWARDS STANDARDIZATION

Alessandro M.
Banca d'Italia

Abstract

The financial data services industry is complex, with numerous financial instruments and limited suppliers. The lack of a common language and standardized methodologies for taxonomy development poses challenges to the understanding and comparison of vendors' commercial offerings. This paper proposes a taxonomy and ontological model for the financial data services domain to address these challenges. The taxonomy organizes financial information services hierarchically, while the ontological model offers a conceptualization and specification of the domain for the development of applications based on semantic technologies. The taxonomy was developed empirically using three levels of reality and validated using closed card sorting and in-depth sessions with vendors of data. The ontological model is based on the OWL 2 ontology language and Description Logic and includes data vendors' classes and attributes that qualify the relationships between the ontology's objects. The proposed taxonomy and ontological model provide a common language and reference data standards for financial market services that capture the diversified, complex, and evolving nature of financial market data services. The taxonomy can facilitate strategic sourcing activities and improve spend analysis by providing a classification that is granular in terms of cost items and in line with the services used, regardless of the type of supplier and license agreement.

Keywords: financial data services, ontology, taxonomy, semantic technologies, spend analysis, strategic sourcing

INTRODUCTION

The Financial Data Services industry provides financial market data and related services, primarily real-time feeds, portfolio analytics, research, pricing, and valuation data, to financial institutions, traders and investors. Industry vendors aggregate data and content from stock exchange feed, broker and dealer desks, and regulatory filings to distribute financial news and business information to the investment community. They play a key role in the financial professional workflow and the demand for services is constantly growing. Over the past five years, according to some publicly available statistics, global spending on financial market data has grown significantly to reach $35 in 2021 and is currently dominated by a few large providers: Bloomberg, Refinitiv, S&P Global Market Intelligence, Moody's Analytics, FactSet, just to mention the main ones. They offer a kind of "one-stop-shop" digital platform that provides a wide range of financial data services. By contrast, in recent years the most successful providers in terms of growth rates are those whose core business services are providing proprietary data such as indices, financial benchmarks, evaluated pricing, and analytics.

Indeed, financial data services represent a complex market scenario that, besides the large number and variety of products and services, is mainly characterized by limited competition (with only a few suppliers to provide the related products), strong demand expansion (in particular, requests for regulatory requirements), and constantly rising prices (in sharp contrast with other products and services). In terms of variety, the range goes from plain, standard and individual services, and content, to complex data flows and sophisticated data processing platforms serving multiple business areas. It is also clear that the characteristics of the financial information industry reflect the increasing number and complexity of financial instruments that are traded on hundreds of different markets, leading to an extremely large and hard-to-define universe of data. The set of data can be referred to as two conceptual macro-categories: information about companies (corporate actions and events, valuation information, fundamental data including company performance, reference data on the entities themselves) and information about instruments (pricing data, volumes traded, reference data on the instruments).

Against this backdrop, the financial data services are usually analyzed and classified in commercial reports mainly according to the following classes: general business mix (data feed vs workstation); target segments (e.g., corporate, investment banking, investment management, retail wealth management); data coverage (real-time data, historical data, pricing, fundamental data, etc.). This perspective is not useful to support the user of financial market data in understanding and comparing vendors' commercial offerings.

Building on other initiatives concerning the development of applications based on new semantic technologies, the scope of this work is to investigate the possibility of creating a knowledge base that can support business applications based on semantic technologies. We argue that, despite a remarkable interest in the development of taxonomies and ontologies in the financial domain, little research work has been done on the financial data services domain. This paper is, therefore, an attempt to establish a common language for financial data services creating and making available reference data standards for financial market services capable of capturing the diversified, complex, and evolving nature of financial market data services.

The paper is organized as follows: section 2 provides the background and objectives of the proposed taxonomy; section 3 presents a comparison with other commercial classifications; section 4 describes the chosen methodological approach; section 5 describes some uses cases of the taxonomy; section 6 describes a possible prototype ontological model in the financial data services domain for the semantic web; section 7 is for discussion and conclusions.

## THE TAXONOMY: BACKGROUND, DESCRIPTION, AND OBJECTIVE

In general, the classification of objects or items helps researchers and practitioners understand and analyze complex domains. As already noted in the relevant literature (Nickerson et al., 2010), complexity reduction and the identification of similarities and differences among objects are major advantages provided by taxonomies. Taxonomies help structure and organize knowledge, grouping objects from a distinct domain based on common characteristics and explaining the relationships among these characteristics. Moreover, there are several problems to be solved before the methodologies can be considered mature and can be applied with concrete prospects for implementation.

Concerning financial services taxonomies, we have found that there are no shared methodologies for taxonomy development, but it is possible to identify some classifications used in different areas of

analysis. Finding and comparing financial information services and their vendors' standard methods has always been an extremely complex task. The commercial offer is heterogeneous, and vendors sometimes have a strong position in particular market segments that, in some cases, could influence the users' choice of market data and financial information. In this sense, it may be extremely useful to be able to place each service offered within a shared classification to better understand its content and, by comparing it with other available alternatives, make a more considered choice.

Within the domain of interest, therefore, classes and subclasses have been defined based on the primary service offered, which can satisfy one or more needs of the purchaser at the same time. According to this approach, since the stated objective of the taxonomy is to organize the financial information services currently available on the market according to a hierarchical structure, there is an awareness that sometimes the identified individual services may partially overlap in terms of the granular content of the services. For example, we see in Figure 1 how financial platform services can be considered as an aggregate of services that can sometimes be purchased separately from other vendors.

FIGURE 1 EXAMPLE OF OVERLAPPING IN THE FINANCIAL DATA SERVICES COMMERCIAL OFFER
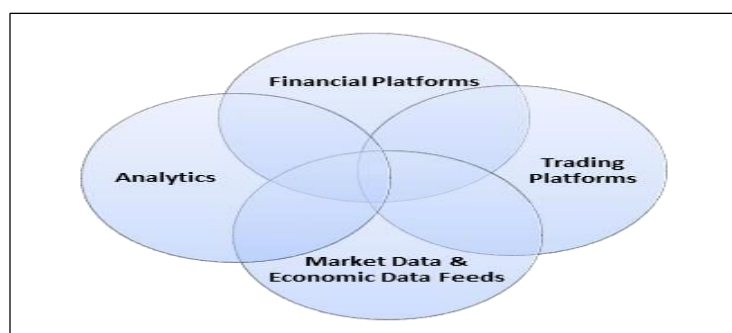
On the other hand, it must be clear that this taxonomy, built according to an empirical rather than a theoretical-based approach, is intended to represent the total range of services offered for the declared domain and not to classify the individual elements that combine to make up the various products and services useful to financial market operators. According to these premises, it is also evident that, given the extreme dynamism of the financial services industry and the role that incoming players (e.g. FinTechs) will play in this market, the classes identified in the proposed taxonomy are susceptible to updating concerning the changes.

Other Available Classifications

To achieve a common understanding of a product domain classification, it is crucial to define standard product classification schemes. Over the past years, considerable effort has been made to develop both private e-Procurement and public procurement classification to enhance the coverage of domains, and enrich the semantic and formal precision (Leukel & Maniatopoulos, 2005). The main classifications available are:

•   the United Nations Standard Products and Services Code (UNSPSC), which provides an open, global multi-sector standard for the classification of products and services;
•   the Global Product Classification (GPC), the chosen GS1 standard mandatory classification system for the Global Data Synchronization Network (GDSN);

• eCl@ss, an ISO/IEC-compliant industry standard, forming a worldwide reference-data standard for the classification and unambiguous description of products and services;

• the Common Procurement Vocabulary (CPV), which is the only classification system that has to be used for the publication of public procurement notices in the EU.

The discussion will focus on the CPV concerning its role in European public procurement and take into account the ongoing debate on the possibility of revising it and linking it more closely to other private sector classifications.

The CPV is a single classification system for public procurement in the EU and consists of some 9.500 codes structured in a five-level tree hierarchy. The purpose of the CPV is to help bidders to identify relevant tender notices thus fostering cross-border procurement. The rationale behind the CPV is to increase competition and ensure a higher level of transparency. If relevant publications can be identified more easily, this will result in more bids and increase competition between bidders and this, in turn, could eventually lead to better value for money in public procurement. As shown in Figure 2, the CPV is part of an integrated system that allows the comparability of statistics produced in different statistical domains according to a coherent and consistent classification structure for products based on a set of internationally agreed concepts, definitions, principles and classification rules (Eurostat, 2008; European Commission, 2008).

FIGURE 2 INTERNATIONAL SYSTEM OF ECONOMIC AND PRODUCT CLASSIFICATIONS
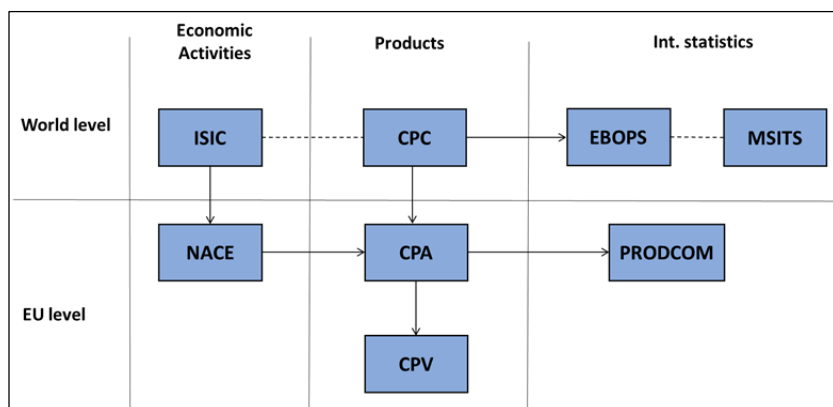
A recent report commissioned by the European Commission (Cosinex, 2018) showed several inadequacies in the functioning of the classification. Among these, from a buy-side perspective, it is worth mentioning that contracting authorities usually do not use CPV internally to describe their needs or to project or structure them. Only when the tender documents are finalized and publication of the notification is pending, the CPV becomes relevant. This means that CPVs, at least in some areas, do not correspond to the real structure of the reference markets and are therefore unable to capture their peculiarities.

A sample analysis was carried out on the Tenders Electronic Daily (TED) database – the online version of the "Supplement to the Official Journal" of the EU, dedicated to the European public procurement – by searching for some of the most important financial platforms based products showed a low frequency of notices containing such types of services, presumably because the main users of these services are private entities that do not apply European procurement rules. Nevertheless, an empirical

study was carried out on the limited information set available to identify the main codes used. As shown in Table 2, this category of services is classified mainly into three divisions, presumably depending on a subjective choice by the contracting authority.

As emerging from these results, the available classification, although including very detailed categories of works or supplies, is largely inadequate to grasp the specificities of the financial information services market.

In the private sector, in addition to the classifications proposed and adopted to promote e-procurement, it is possible to find non-standard classifications developed by consulting firms that analyze the financial information services market. This kind of study have a two-fold objective: to help market players to position themselves about their competitors and to facilitate strategic sourcing activities by data users. By carrying out a comparative analysis of the different classifications adopted by consulting firms and within the most relevant market data User Groups, several categories were identified (e.g., terminals, exchanges/brokers data, fundamental & reference data, indexes). These kinds of commercial classifications seem to reflect more closely the structure of the financial information services industry and have been taken into account to test the Taxonomy.

TABLE 1 CPV MAIN CODES INVOLVED IN PROCUREMENTS OF FINANCIAL INFORMATION SERVICES

| Division	Group	Class	Category
48000000-8 Software package and information systems	48400000-2 Business transaction and personal business software package	48410000-5 Investment management and tax preparation software package	48411000-2 Investment management software package
		48440000-4 Financial analysis and accounting software package	48441000-1 Financial analysis software package
			48442000-8 Financial systems software package
	48800000-6 Information systems and servers	48810000-9 Information systems	48812000-3 Financial information systems
66000000-0 Financial and insurance services	66100000-1 Banking and investment services	66140000-3 Portfolio management services	
		66150000-6 Financial markets administration services	66151000-3 Financial market operational services
79000000-4 Business services: law, marketing, consulting, recruitment, printing and security	79900000-3 Miscellaneous business and business-related services	79980000-7 Subscription services	

DESIGN AND DEVELOPMENT

The development of a taxonomy is a complex activity that requires complete knowledge of the reference domain and implies a significant effort of conceptualization. A corporate taxonomy, on the other hand, allows a greater degree of freedom in the definition of concepts and is applied with specific reference to the identified business case. From an operational perspective, the taxonomy has been developed following some consequential steps that have allowed arriving at the prototype version (see Figure 3).

FIGURE 3 DEVELOPMENT PHASES OF THE TAXONOMY
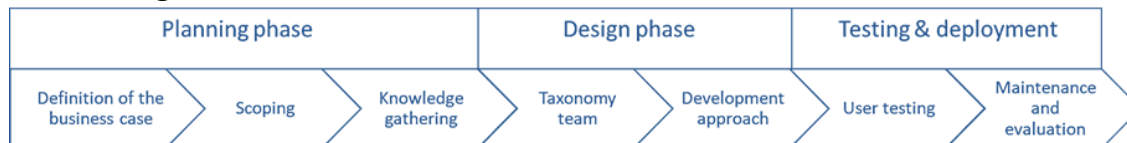
Regarding the development strategies for identifying concepts, it is useful to refer to classification methods in social science (Bailey, 1984). In this context, a distinction is often made between three levels of reality: the conceptual level, which starts from purely conceptual premises, sometimes hypothetical or imaginary; the empirical level, preferable when empirical cases have an important descriptive value; the operational level, which is a combination of the two previous approaches (see Figure 4).

FIGURE 4 LEVELS OF REALITY FOR CONCEPT IDENTIFICATION
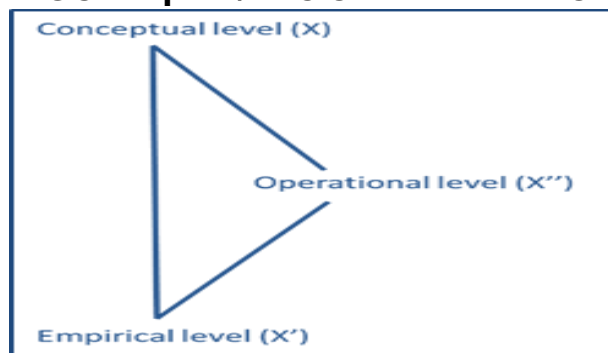
Moving on to a more operational level, a bottom-up approach may result in a high level of detail but, on the other hand, can make it difficult to identify commonalities between related concepts and increase the risk of inconsistencies. A top-down approach, conversely, assures a better level of detail but at the cost of choosing and imposing arbitrarily high levels of detail; this, in turn, can lead to poor stability. All this considered, we started looking for the most general and the most particular concepts as key concepts, but then decided to focus on the most important ones that were used to complete the hierarchy by generalization and specialization (middle-out approach). Then, the identification of the key concepts and relationships in the domain of interest tried to focus the attention on the concepts as such, rather than on mere words representing them.

Consistently with the methodological approach valid for the development of ontologies, we have taken into account the fundamental principles of clarity (the taxonomy effectively communicates the intended meaning of defined terms, the defined terms minimize ambiguity, and examples are provided to understand definitions); coherence: (it is possible to perform inferences that are consistent with the descriptions or definitions); conciseness (there are no unnecessary or useless definitions; redundancies between definitions do not exist); adaptability (the taxonomy does not need a continuous adaptation that calls into question its overall structure; the taxonomy anticipates its uses

and offers a conceptual foundation for anticipated tasks). The validation of the taxonomy structure has been performed in two different ways:

1. Closed card sorting, where participants are provided with a predetermined set of categories/classes that are already labeled and they have to place the items into these categories. This kind of exercise helps disclose the degree to which the participants agree with the predetermined categorization. To do that, in-depth sessions were organized with some of the main vendors of data and economic and financial news. In particular, they were asked to place the main services and products within a grid built taking into account the concepts, relationships, and attributes identified during the process of construction of the taxonomy. This exercise not only confirmed the overall structure of the taxonomy but also excluded the presence of arbitrary or ad hoc dimensions and characteristics that would have affected the conceptual validity of the artifact;

2. Competency questions, already identified in the taxonomy design phase, have been proposed to the users of the services. These questions assure the targeted value of the structure is achieved and indicate when the taxonomy development is sufficiently complete. In other words, this step aims to ensure that the results are accurate, sufficient, and have the right level of granularity, which is identified by the subject matter expert.

The graphical representation in Figure 5, generated using the open-source ontology editor Protégé (Musen, 2015), shows the results of these refinements and depicts the overall structure of the proposed taxonomy. The definitions adopted do not necessarily reflect those predominantly used by financial operators to identify the types of data and/or instruments used. This is because, as already clarified, the chosen perspective tries to reflect, as much as possible, the commercial offer of the main suppliers from a procurement perspective (e.g., identification of all possible suppliers of that particular service, carrying out market surveys on particular segments).

FIGURE 5 OVERALL STRUCTURE OF THE TAXONOMY
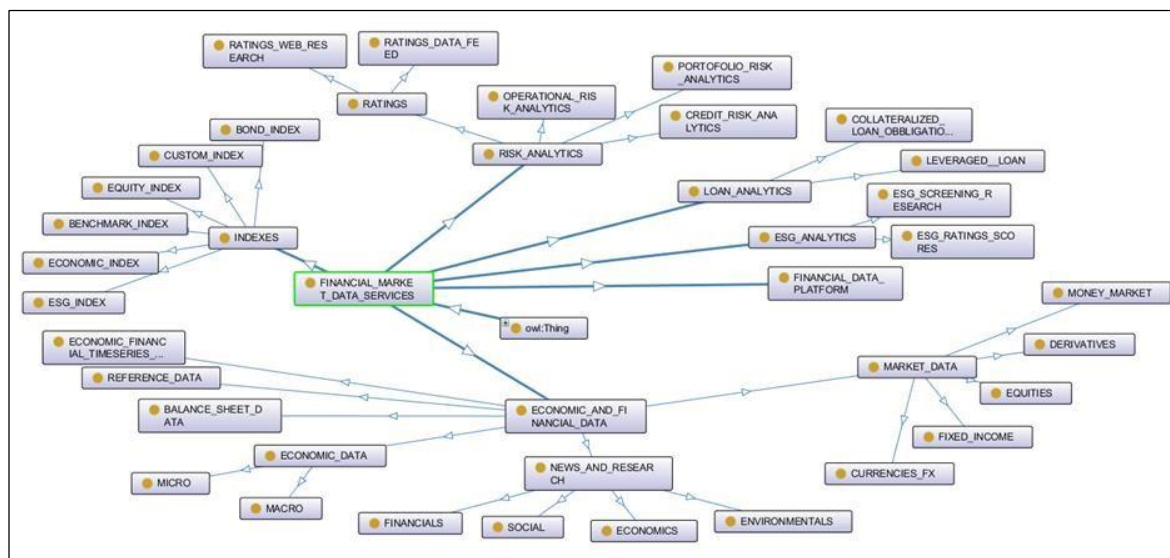
Source: author' elaboration (OntoGraf plugin for Protègè)

To unambiguously define the relevant domain, it can be useful to recall the Industry Classification Benchmark (FTSE, 2019), which is a globally utilized standard for the categorization and comparison

of companies by industry and sector. According to this classification, financial data providers are companies that "provide financial decision support tools for investment institutions (including financial database operators and index data providers)". In this context, the "financial market data services" domain covers the overall offer of such services provided by different types of companies (e.g., data vendors, exchanges, brokers, index providers, etc.) to provide financial decision support for investment decisions and financial markets analysis. Under this domain, six meta-classes ("economic and financial data", indexes", "ESG analytics, "financial data platform", and "loan analytics") and twenty sub-classes are identified in a hierarchical structure. Finally, to better describe the internal structure of concepts, several intrinsic and extrinsic attributes are considered.

TAXONOMY USE CASES

Procurement Activities

The market for financial services and financial information is large, complex, fragmented, and oligopolistic in many segments. The search for a solution that meets the requirements of the buyer is difficult due to the lack of complete references on the offer. The main search modes used include web search engines, specialized press, references from other companies, user groups/associations, and advertising suppliers. Matching supply and demand is complicated by the indeterminacy of the requirements and the lack of transparency in the description of the products.

From the demand side, financial operators often express the requirements of service ambiguously: in general, the trader tends to maintain the suite of software products and platforms that he knows and is not often willing to bear the learning costs of a new product. The operating mechanisms are consolidated around a product configuration and they become a further constraint to change. On the other hand, the configuration of an operating station is a complex set of hardware, software, information, and additional services. There is often a noticeable "lock-in effect" that binds services to each other. The offer, on its side, does not allow you to easily find the product or service you need, for several reasons: it is difficult to find the candidate suppliers and it is not easy to identify the features required within highly articulated, and often bundled, offers that also include unnecessary elements. Finally, the pricing mechanism is very complex and it is difficult to evaluate the convenience of one offer compared to another. The risks for buyers are the use of sub-optimal products and services, high purchase and use costs, and the creation of lock-in situations.

To counter the critical issues described above, the buyer requires organizational measures and market research tools reports. The creation of financial services market specialized teams, in charge of carrying out all procurement activities for the company, allows the synergic exploitation of the skills (financial, legal, marketing) necessary for this task.

The networks of buyers who exchange information, regularly, about the offer and commercial policies of the vendors are a powerful means of sharing data and news. Through networking mechanisms, it is possible to discover new services that can solve problems at low costs and with better performance. The networks of buyers also allow the creation of purchasing groups that can prove effective in contrasting the aggressive policies of monopolistic suppliers.

One European user group, among the others, is particularly relevant: the Information Providers User Group (IPUG). The IPUG is a non-profit organization, established in 1989 to represent the current and future interests of its member firms. It is now the principal organization in the UK representing users of market data services on a technical, administrative, and strategic level. IPUG has developed strong working relationships with the major real-time information service vendors, benchmark suppliers, and

pricing and fundamental service providers. IPUG acts on behalf of its membership to focus these suppliers on generic issues affecting all users. IPUG is recognized by these vendors as the users' legitimate voice and is often consulted and asked to contribute to supplier policy decisions that affect the membership. In line with its commitment to represent new industry trends, IPUG continually seeks to monitor the technology and business process developments that affect the industry.

Finally, yet importantly, the use of a taxonomy allows for better target market research, because it allows you to direct the search in homogeneous clusters (classes) where fungibility can be found and makes it possible to use expert research systems (machine learning) by providing standardized definitions of classes of services. Moreover, it simplifies the exchange of info with other buyers (shared sector studies) and facilitates compliance with regulatory procurement rules (e.g., "Public Procurement Code").

AI Systems for Searching Financial Data Providers and Solutions

As already highlighted, the financial data providers' market is extremely complex and is characterized by low competition. In this scenario, it is extremely difficult to source small financial services providers, since they only own very little market shares and are therefore often excluded from the market data procurement process. In the past, Procurement Automation (aka eProcurement) mainly focused on using ERP management tools to record and examine previous buying decisions and expenditure data. In recent years, machine learning and artificial intelligence have been applied to procurement workflows, introducing computation of external or third party unstructured data to achieve a higher level of market knowledge and decision automation. This new kind of procurement is often referred to as AI Procurement or Digital Procurement. Most of the time, this information is text-based, i.e., collections of several documents from multiple data sources (social networks, blogs, forums, etc.).

In the last years, many powerful machine learning models have been published and released to the community like BERT (Devlin, 2018) and USE (Cer et al., 2018), achieving state-of-art results for many NLP tasks over this kind of information. By the end of the day, however, the final user does not feel comfortable with unstructured data. Hence, the above models need to be used to display clear information, ready to be used by humans. Among all the NLP tasks available, the "Named Entity Disambiguation and Linking" task, aims to automatically match information against knowledge bases containing structured data. Between all the existing Knowledge Bases, it is mandatory to cite Google Knowledge Graph (Singhal, 2012) and Wikidata (Vrandečić, 2014). However, neither the former nor the latter, provide a classification system, of the entities they are composed of, that suits the needs of the financial world. Indeed, Google Knowledge Graph uses a finite and standardized vocabulary for types defined by "schema.org" (Guha et al., 2016), which does not provide any detailed categorization for the complex scenario of financial services. On the other hand, Wikidata classifies each entity utilizing the "instance of" property (P31). However, since every one of its entities can be used as a value for this property, we have millions of potential classes.

The above situation enhances the need to build a Knowledge Base for financial institutions and players, in which entities are classified using a specific financial data services taxonomy, built by business experts who deeply know the relevant market. Any candidate taxonomy for this role, as well as done by schema.org, must be defined with a severe versioning system that clearly states a finite number of entries per version. Moreover, it should be standardized; i.e. accepted and verified by a representative team of the most important financial players, becoming, therefore, a common language for financial services.

PROTOTYPE ONTOLOGICAL MODEL

In recent years, the development of ontologies – which can be defined as an explicit formal specification of the terms in the domain and relations among them (Gruber, 1993) – has gained attention, and many disciplines, including social sciences, now develop and use standardized ontologies to share and annotate information (Guarino et al., 2009). Within the financial industry, it is worth mentioning the project called Financial Industry Business Ontology (FIBO), which proposes a set of formal models for financial industry concepts (Bennet, 2013). The main objective of this 'artifact' is to solve long-standing reconciliation problems in the field of data management by using the principles of the semantic web.

Even though there is no unique and correct way to develop an ontology, it is possible to identify the main steps to follow, namely: determine the "domain" and "purpose" of ontology; identify the key concepts of the phenomenon to describe; organize concepts into "classes" and "hierarchies" between classes (i.e. define a taxonomy); define class "properties" and "constraints" (lawful values). Finally, it is necessary to create "instances" and assign "values" to properties for all instances created. The best solution depends on the business case you follow but, considering that an ontology is a model of reality, the concepts in the ontology must reflect this reality or, in other words, should be close to objects (physical or logical) and relationships in the domain of interest (Noy & McGuinness, 2001).

Looking at the general construct that an ontology applied to the domain of financial information services can have, both the conceptualization work and the subsequent specification conducted for the design of the taxonomy are certainly reusable.

However, several attributes needed to qualify the domain were deliberately not considered in the design of the class taxonomy, as they would have made it too complex to manage. Now, to establish a simplified ontology schema, it is necessary to introduce a set of new elements (see Figure 6) such as the data vendors' class and a series of attributes that qualify the relationships between the objects (instances) of the ontology.

FIGURE 6 SIMPLIFIED ONTOLOGY SCHEMA
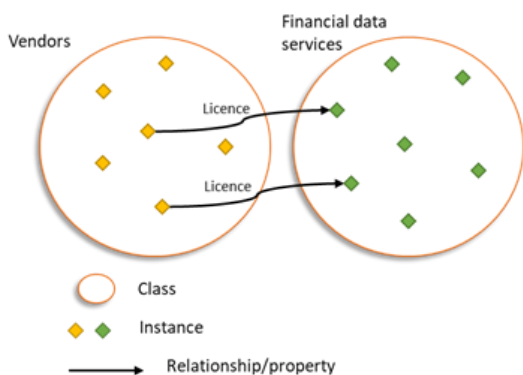
To provide an idea of what an ontology might be in the domain under discussion, we have chosen to use the syntax of the OWL 2 ontology language (W3C®, 2012) and at the same time indicate the relevant logic notation (Description Logic – DL). Table 3 shows the main class axioms used in most of the ontologies available today. The aim is to provide an idea of how abstract concepts related to financial information services can be represented in a knowledge base.

TABLE 2 CORRESPONDENCE BETWEEN OWL SYNTAX AND DLS FOR THE PROTOTYPE ONTOLOGY

Type	OWL syntax	Logic notation	Example
Class Expression Axioms	SubClassOf (a:EquityIndex a:Indexes)	$A \subseteq B$	Each *Equity Index* is an *Index*
	DisjointClasses (a:Risk Analytics a:Loan Analytics)	$A \cap B = \perp$ *or* $A \subseteq \square\, B$	Nothing can be both a *Risk Analytics Service* and a *Loan Analytics Service*
	IntersectionOf (a:FinancialDataPlatform a:EconomicAndFinancialData)	$A \cap B$	*Financial Data Platform* and Economic *And Financial Data* have *Market Data* in common
	SubClassOf (ObjectSomeValuesFrom(a:IsDisplayOnly a:MarketData) a: Financial Data Platform)	$A \subseteq \square\, P.B$	If some object *Is Display Only*, then this object is a *Financial Data Platform*.
	ClassAssertion(a:Equity Market Data a:EoDEquityPrice)	$A \in B$	The individual *a:EoD Equity Price* can be used to represent a particular *Equity Market Data*

The prototype version of the ontology should be accurately evaluated and debugged by using it in applications or problem-solving methods by discussing it with experts, or both. As a result, we will almost certainly need to revise the initial ontology; this process of iterative design will likely continue through the entire lifecycle of the ontology.

CONCLUSIONS

Despite a remarkable interest in the development of taxonomies and ontologies in the financial domain, little research work has been done on the financial data services domain, as this area of interest seems to be still confined to analysis and discussion in the various information providers' user groups or consulting firms specializing in market data analysis. The taxonomy, and the related prototype ontological model, presented in this paper is a first attempt to address the complex issue of financial information services categorization. The variability and complexity of the financial instruments, the exponential growth of the economic and financial data, and the consequent complexity, and sometimes opacity, of the commercial offer, make it a challenging task.

While being aware that much work remains to be done, considering the peculiarity and complexity of the domain of reference, some initial objectives have been achieved. In particular, the taxonomy is currently used in the following business activities:

•   spend analysis, utilizing a classification that is both granular in terms of cost items and more in line with the content of the services used, regardless of the type of supplier and the related license agreement;

•   demand management, using a common metric and language for the identification of services in the context of market data management;

• strategic sourcing, where the continuous process based on a data-driven approach is enhanced with a greater awareness on the part of both the purchasing department and the business units capable of producing positive externalities on the cost side (i.e. getting the best service or product at the best possible price).

The activity of developing an ontology proved to be a rather complex activity that will need further investigation, especially in terms of defining class instances and properties. Nevertheless, through this work, we wanted to demonstrate how it is possible to apply the conceptual metrics used in the definition of ontologies to the domain of financial information services. This may have interesting implications soon in terms of the exploitation of large amounts of data (big data) related to financial information services, thus paving the way for the development of business applications based on semantic technologies.

ACKNOWLEDGMENT

REFERENCES

Bailey, K.D. (1994). *Typologies and Taxonomies - An Introduction to Classification Techniques*. Los Angeles, SAGE Publications.

Bennet, M. (2013). The financial industry business ontology: Best practice for big data. *Journal of Banking Regulation*, *14*(3/4), 255–268.

Cer, D., Yang, Y., Kong, S.Y., Hua, N., Limtiaco, N., John, R.S., . . . Kurzweil, R. (2018). *Universal sentence encoder*. arXiv preprint arXiv:1803.11175

Cosinex. (2017). *Revision of CPV – Final Report. Consultancy Services for Common Procurement Vocabulary expert group*. Ref. Ares(2018)773667 - 09/02/2018. European Commission.

Devlin, J., Chang, M.W., Lee, K., & Toutanova, K. (2018). *Bert: Pre-training of deep bidirectional transformers for language understanding*. arXiv preprint arXiv:1810.04805

European Commission. (2008). *Public Procurement in the European Union. Guide to the Common Procurement Vocabulary*.

Eurostat. (2008). *NACE Rev. 2 Statistical classification of economic activities in the European Community*. Methodologies and working papers.

FTSE Russell. (2019). *Industry Classification Benchmark*.

Gruber, T.R. (1993). *Towards Principles for the Design of Ontologies Used for Knowledge Sharing*. Stanford Knowledge Systems Laboratory.

Guarino, N., Oberle, D., & Staab, S. (2009). What is an ontology? In *Handbook on ontologies* (pp. 1–17). Springer, Berlin, Heidelberg.

Guha, R.V., Brickley, D., & Macbeth, S. (2016). Schema. org: Evolution of structured data on the web. *Communications of the ACM*, *59*(2), 44–51.

Leukel, J., & Maniatopoulos, G. (2005). A Comparative Analysis of Product Classification in Public vs. Private e-Procurement. *The Electronic Journal of e-Government*, *3*(4), 201–21271.

Musen, M.A. (2015). The Protégé project: A look back and a look forward. AI Matters. *Association of Computing Machinery Specific Interest Group in Artificial Intelligence*, *1*(4). DOI:10.1145/2557001.25757003

Nickerson, R.C., Muntermann, J., & Varshney, U. (2010). Taxonomy Development in Information Systems: A Literature Survey and Problem Statement. *AMCIS 2010 Proceedings*, paper 125.

Noy, N.F., & McGuinness, D.L. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880.

Singhal, A. (2012). Introducing the Knowledge Graph: things, not strings. *Official Google Blog*. Retrieved from http://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-thingsnot.html

Vrandečić, D., & Krötzsch, M. (2014). Wikidata: A free collaborative knowledgebase. *Communications of the ACM*, *57*(10), 78–85. W3C®. (2012). *OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax*. Retrieved from https://www.w3.org/TR/owl2-syntax/