Semantic Approach to Financial Data Integration for Enabling New Insights

Emmanuel Asimadi[1,2], Stephan Reiff-Marganiec[1], Brian Donnelly[2], Josef Baker[2] and Daren Fang[2]

[1] Department of Informatics, University of Leicester, Leicester, UK
[2] Synapse Information Ltd, Birmingham, UK
{ea251, srm13}@le.ac.uk

Abstract. Financial regulators around the world are following in the footsteps of the US SEC by mandating businesses to share their financial information in an XML based business reporting standard called XBRL. Businesses are periodically reporting on their finances, hence there is a wealth of financial data waiting to be explored. The structural complexities in the XBRL format and the spread of data across many files pose a hurdle in exploiting the data. This paper presents a semantic approach to integrate, process and query the financial information embedded in the XBRL to allow for new insights into the financial ecosystem.

Keywords: Semantic Web; Data integration; Advanced Queries; Financial Applications.

1 Introduction

eXtensible Business Reporting Language (XBRL) is an XML-based business and financial reporting standard attributed to Charles Hoffman's work in 1998 investigating the use of XML for financial reporting. XBRL aims at providing a common vocabulary, a flexible and self-describing data structure that makes reporting domain assumptions explicit in a way that supports automated processing. Considering a typical financial statement (Fig. 1), the presentation of financial facts implicitly conveys meaning available only to a human reader and not a computer – to a computer this highly informative report is nothing but text. A Financial report contains information about a business Entity (a resource that can be further described e.g. by its country of registration, incorporation date and industry classification). XBRL provides a means of capturing these: Concepts, Labels and Facts. The same Concept might have different names e.g. Revenue or Turnover, but they mean the same thing in financial practice. Labels provide multiple lexical representation of the same concept and thus support multiple language presentation of the data. Financial Facts are the actual data communicated by the report against the identified Concepts. Facts may correspond to a period or represent a measure at an instance in time. Figure 1 captures Facts corresponding to 2014 and 2013.

The goal of XBRL is to facilitate information exchange and generate value along the entire data supply chain from business report production through to its consumption and analysis, thus leading to greater efficiency, cost savings, improved accuracy and reliability. Although XBRL is often considered very complex, its value proposition is immediately obvious when compared with the paper/document (pdf/word) based reports which it replaces. Unlike paper/document based reports, XBRL provides well-defined annotation and access to data in financial reports making automated processing possible.

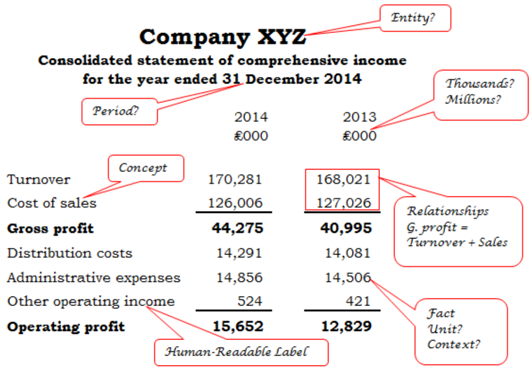

Fig. 1. Implicit Metadata in Financial Report [1, 2]

Regulatory requirements have been the primary driver for the uptake of XBRL around the world with Japan among the earliest adopters in 2005 [1]. In 2009 the Security Exchanges Commission[1] required Public and Foreign Private companies reporting against U.S. General Accepted Accounting Principles and International Financial Reporting Standards (IFRS) – in the case of foreign private companies – to submit their filings in XBRL. According to SEC's press release [2] this requirement was not only to enable investors to better analyze financial information but also "assist automation of regulatory filings and business information processing" thus achieving greater efficiency, accuracy, usability and importantly reduced cost. In 2011 Her Majesty's Revenue and Customs (HRMC) in the UK mandated all companies to submit their company tax returns in Inline XBRL (iXBRL). iXBRL is XBRL tagged data is presented in human-readable HTML format, allowing the single document to be accessible to both humans and machines. XBRL filings to HMRC are not made public. UK's Company House (company register) on the other hand allows voluntary submission of accounts and company information. That notwithstanding, the number of filings has almost doubled from a little over a million in 2012 to almost two million in 2015 [3]. Company House publishes iXBRL files and the volume of iXBRL published presents a treasure drove of data which is standardized and more accessible.

While XBRL achieves significant annotation and standardization of financial reporting, actually integrating and analyzing data stored in XML-based XBRL remains complex. Hence, despite all of these efforts that are leading to great availability of XBRL reports, the true value of XBRL data is yet to be exploited: storing and querying and

1 https://www.sec.gov

integrating data across filings is hard due to limitations of the underlying XML document data structure (for a start each financial report is in one or more separate XML files) and integration with other data sources (such as company or geographical information) is not readily available. Furthermore, there are significant issues focused on accessing the information in convenient tools.

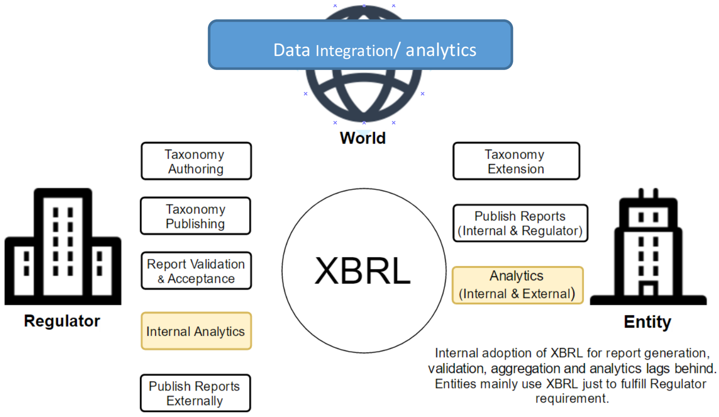

Fig. 2. XBRL in the real world

For example, an investor might be asking for geographic centres of rapid growth in a specific technology domain or a tax investigator might be interested in directors that are acting in numerous companies worldwide – while the information for this is available in XBRL filings and other data sources, extracting it is time consuming detective work. Taking advantage of semantic approaches our work integrates and advance queries for financial data and other external data sources to answer exactly such queries in practical ways and seamlessly integrated in the tools that they are using normally, namely spreadsheets. This paper focuses on a real world application semantic technology to exploiting XBRL data with a focus on the UK.

The specific novelty of this work is (1) a practical approach for using XBRL in a wider semantic context, which (2) ensures traceability between semantic financial data and XBRL reports and (3) allows queries spanning reports and XBRL standards.

The rest of this paper is organized as follows: Section 2 introduces background and related work, Section 3 details our approach, both high level as well as the details. Sections 4 considers some initial evaluation and discusses results with section 5 summarizing the paper and looking at next steps.

2 Background and Related Work

XBRL is an XML-based business information exchange format owned and freely licensed by XBRL International Inc. (XII). XII defines XBRL as “a language for the electronic communication of business and financial data which is revolutionizing business reporting around the world. It provides major benefits in the preparation, analysis and communication of business information. It offers cost savings, greater efficiency and improved accuracy and reliability to all those involved in supplying and using financial data” [5].

XBRL provides some guarantees for accuracy of an individual financial report, at least at the syntactic level. This is achieved by using instance validation against the reference XBRL taxonomy ensuring that data being exchanged obey datatype and predefined business. Hence, syntactic errors like wrong datatype, omission of required facts, accuracy of derived facts (for example ensuring that *Fixed Asset = Tangible Asset + Intangible Assets + Investments*) are detected (but not avoided).

Secondly, reports submitted in XBRL can be easily repurposed to serve new reporting requirement eliminating the need to ‘re-key’ the data, as is the case for paper-based approaches. For instance, the facts in a statutory accounts report can be re-used in a tax report by using appropriate presentation

Thirdly XBRL significantly benefits the analysis process by providing much needed unified structure and context (at least as long as the same taxonomy is used). It significantly enables comparative analysis of the financial information from a large number of entities and deriving financial ratios useful for gauging the performance of the entities. Figure 2 illustrates the XBRL use-case with major participants being business entities, analysts interested in financial data and regulators like the Securities and Exchange Commission (SEC) of the US, Her Majesty’s Revenue & Customs (HMRC) of UK who are mandating the use of XBRL to facilitate information exchange.

In a typical workflow, regulators or authorized bodies author taxonomies, e.g. based on the General Accepted Accounting Principles (US-GAAP, UK-GAAP) or International Financial Reporting Standard (IFRS). These taxonomies provide the ‘dictionary’ and the business rules against which instances (reports) are generated. Technically, XBRL taxonomies allow for extensions, where entities can add their own concepts and rules to the standard taxonomy to cater for their unique needs. This is discouraged by regulators as it compromises comparability of reports [6].

Entities publish XBRL reports to their website or to fulfill regulatory requirements; these are automatically validated and accepted by the regulatory body who processes and analyses it internally to manage industry. A less applied use-case for the Entity is the adoption of XBRL within the organization. The expectation is to have subsidiaries within the business exchange information using XBRL thus taking advantage of validation, aggregation and other promises of XBRL. Many organizations view it as a burden and thus only attach XBRL to tail end of their report generation process [7], just to meet regulator’s requirement to submit reports in XBRL.

2.1 Technical Overview of XBRL

XBRL is driven by XML technology. The specification comprises to main components the Instance document and a Taxonomy Set.
The Instance Document is essentially the financial report that contains the facts about the business entity. The Instance references a Taxonomy Set which can be considered as a dictionary of terms and provides further meaning to the concepts used in the Instance. This separation means information exchange requires only the transmission of

the Instance - any destination with the referenced taxonomy can interpret and consume the Instance. Figure 3 captures some high-level components of XBRL.

Instance Document. This is the actual financial report The Instance contains facts of the report and its context, e.g. the period it corresponds to, the business entity it relates to, whether it represents actuals, budget, audited or forecasted data and most importantly the concepts whose value is captured by the Facts, such as Asset, Profit among others.

Taxonomy Set. This is a collection of documents that make up the taxonomy, which acts as a dictionary extending the meaning of concepts used in the instance and their relations. It comprises the schema (.xsd) and linkbases (.xml).

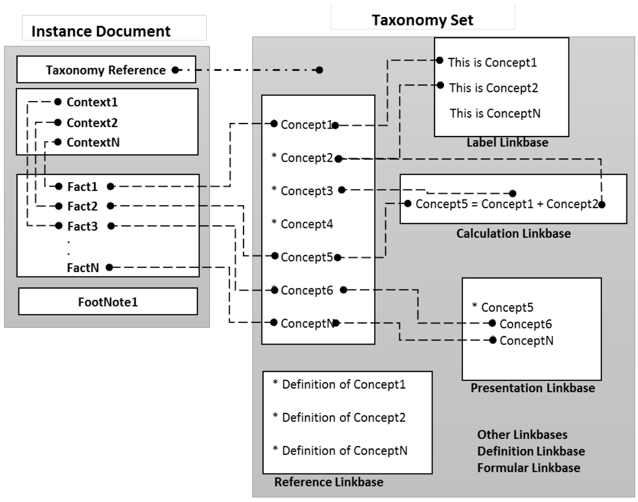

Fig. 3. Schematic representation of XBRL data structure [7]

Schema. The schema lists all concepts in the taxonomy and provides typing information which are used to validate the instance.

Linkbases (LBs). The Labels LB defines human readable labels for concepts in the taxonomy including multi-language support; the Calculation LB captures mathematical relationships between concepts; the Presentation LB captures the hierarchy and order for presenting concepts in reports; the Reference LB provides an authoritative reference to definitions of the concepts and the Definition LB defines other relations between concepts and is particularly useful for hypercube representation of data. Finally, the Formula LB supports more advanced business rules for Instance validation [8].

2.2 Semantics and XBRL

Ontologies are a means of representing knowledge in the form of graphs (classes and relationships between them) so that computers can reason about them. Relational databases due allow storage of data but do not permit for the integrated reasoning which makes it possible to infer new knowledge from what is explicitly stated. Additionally, the semantic web comes with an equally expressive query language (SPARQL) that allows us to answer complex questions about domain of interest. The basic unit of knowledge representation in the semantic web are triples *Subject-Predicate-Object*, which can be encoded in RDF.

Based on our understanding of the requirements of XBRL and our target queries we opted for a more purposeful and efficient transformation of XBRL into a semantic model based on OWL/RDF as compared to the more dynamic approach of [9] parts of which it admits to be counter intuitive and suggests future improvements. Our approach allows us to avoid propagating the limitation of XBRL into our Semantic model. This is particularly important because XBRL was designed with the intent of annotating reports and not necessarily to be efficiently queried semantically.

Others, such as [10] have proposed generic models for converting XML into ontologies. The generic models typically have the drawback that the generated ontology matches the structure of the XML, so if one takes separate XML standards the resulting ontologies will not match and hence integration of the data remains challenging.

In deciding on which financial ontology to apply we found Financial Reporting Ontology (FRO) very rich, capturing the meaning and representing the domain knowledge. [11] also defines an ontology, which again attempts to convert the structure faithfully and completely, leading to an in practical terms unnecessarily complex ontology. This however did not meet our need for simplicity and efficient query requirements hence the need to create a core ontology to hold the financial data and integrate with other sources.

3 Our Approach

Our Semantic Model borrows ideas from the XBRL Abstract Model, which attempts to define an XBRL data model lifting the level of abstraction from the XML syntax [11]. Thus, the abstract model carries the implicit meaning of a financial report devoid of constraints of representing it using XML and we can envision a time when UK Company House (UK's company register) will publish financial data using ontologies like ours to make the data more accessible. This future vision is supported by Company House already publishing company profile information in RDF and making it available as linked data service with a SPARQL endpoint[2].

Also important to our approach is the deliberate choice to benefit from retaining XBRL but clearly separating its function as a syntactic layer in line with the semantic layered cake. This syntactic layer allows to then build semantic (using RDF), query and inference layers as is the case for typical semantic applications.

This approach leads to a two-fold benefit 1) it allows for continued use of XBRL for what syntactic validation and 2) frees up the semantic layer to focus on serving queries that provide new insights. To this end, the model assumes syntactic correctness of the underlying XBRL data and thus focuses on integration to external data and delivery of queries. Figure 4 shows how XBRL and Semantic Web co-exist with the former providing the much needed annotation and syntactic validation needed for the latter. As

[2] *http://business.data.gov.uk/companies/app/explore/sparql.html*

a side effect this also provides traceability as the link between the semantic model and the original data are maintained, a matter that can be of high relevance for many financial applications.

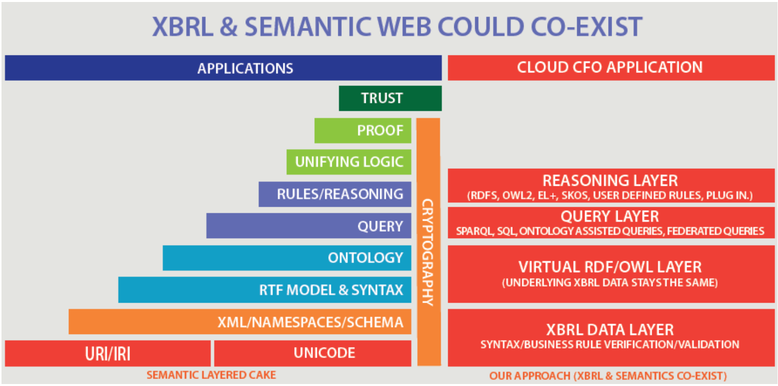

Fig. 4. Our approach in the context of Semantic Layered Cake.

3.1 The Model

The model is designed to be intuitive, efficient to query as well as have minimal memory footprint. As such it does not follow the usual attempt of idealistic semantic modelling of trying to completely capture all possibilities of a domain, but rather being focused to the range of target queries. XBRL components are translated into the semantic model. This type of approach is not unusual for modern NoSQL databases like DynamoDB, Cassandra, HBase and the like where underlying data models are built based on target queries for efficiency.

Figure 5 depicts the high-level overview of the resulting data model[3]. It highlights the main classes in the model and the relations between them. In brief, the *FinancialReport* references a *Taxonomy* which provides further meaning for the *Concepts* used in the report. The *FinancialReport* contains *Facts*. A *Fact* derives additional context information from nodes around it such as period, unit, concept, footnote. In translating from XBRL to RDF we omitted the notion of contexts in XBRL as it does not benefit the semantic model but introduces additional triples to the model. XBRL Extensions that are in the Linkbase will automatically be integrated using the approach.

To guarantee the model's agility, the translation from XBRL to RDF is stateless; this implies new requirements on the model lead to additional triples and not the creation of a new model. To facilitate traceability, the key URIs are generated to be the same as in the linked stores for additional information that is not part of the XBRL documents. For example the URI for the entity in the model is "*http://business.data.gov.uk/id/company/02050399*", which is the same as Company House's URI for company. This way,

3 The ontology and generated data samples are provided online http://download.synapseinformation.com/semantic_xbrl/index.html.

we avoid relying on inference to integrate external RDF stores thus making queries quicker.

However, following good semantic design principles, the model employs a number of ontology design patterns [2], some of which are discussed below.

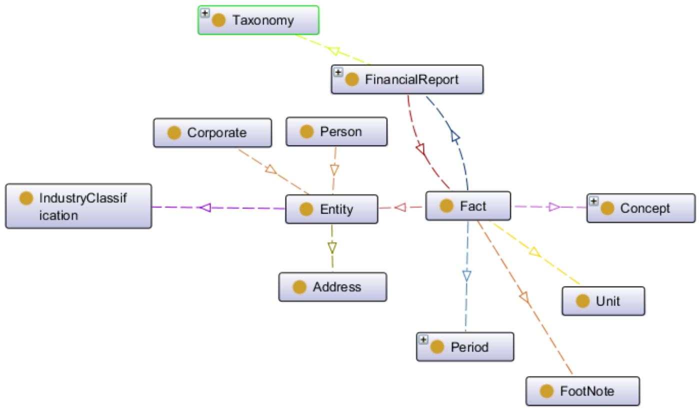

Fig. 5. XBRL Semantic Model

Persons with Significant Contribution

UK Company House publishes the data on persons with significant control in json format. This dataset contains information about people with significant control of businesses namely: the nature of control, their name, date of birth, nationality, address, country of residence among others. The compressed zipped snapshots are made available periodically by Company House.

In order to avoid creating multiple instances of the same individual, who for example manages multiple companies, we adopt the *context slices* design pattern [13] as captured in Fig. 7. Thus for an individual who is significant controller of multiple companies, we create that individual only once. Any other occurrence of this individual is treated as a projection of the primary individual to which we associate a context and then attach additional information that is valid only within that context. This allows us to avoid data duplication and supports more complex queries like finding information that is valid for an individual only within a particular context. A typical query could be "*given a context instance:c1 that hasPeriod instance:date1 find all other relations that hold for the primary individual instance:SignificantController1*".

Other design patterns applied in this model include the *part-of* design pattern which enables modelling XBRL concept hierarchies from the Presentation Linkbase, such as for example *CurrentAssets* being part of *Asset*.

3.2 Integration

Recall that the goal of our solution is not only to make XBRL data more accessible by exposing it to the expressive power of Semantic Web query languages, but also to enable integration to other data stores that together will enhance the value of the financial data. In this first iteration we integrate financial data derived from XBRL to 1) Company Profile found in Company House Linked Data Service, 2) Location data found in Ordinance Survey, 3) Significant person's information and 4) Industry classifications which are also accessible from the company profile ontology. These are by no means the only datasets that can be integrated; other high value dataset include stocks and other linked datasets depending on use-case.

Linkage to external data stores is through federated queries and we facilitate this by making our internal URIs match that of the external store. This way, we avoid relying on inference and *sameAs* axioms making the queries more efficient. The ideal solution would be to publish the translated XBRL data to the linked data cloud making it available to the larger linked data ecosystem. This would form the basis of using the expressive power of semantic web to curate (make filings comparable, embed domain knowledge and business rules) and integrate with other sources for valuable queries.

For example, the company profile provides information about the entity, i.e. its registered ID, legal name, address, and date of incorporation among others. In addition to making this data available in csv/json format, Company House publishes this data as Linked Data Service [14] with a SPARQL endpoint to which we connect. Thus benefitting from Company House's rich ontology and live data directly in our queries.

This linked data service connects further to Ordinance Survey data (a location RDF dataset for UK), the UK Standard Industry Classification (a taxonomy of industries) and makes use of a number of well-established vocabularies to annotate the company profile. These vocabularies include SKOS, registered organisation vocabulary, Dublin Core, and vcard among others. Additionally, it supports efficient text search on company name using indexed datatypes on `SKOS:prefLabel` and `:legalName`.

Company House's linked data service fulfils relatively complex queries like "*list all technology (SIC) companies in a District (Ordinance Survey) in UK*". The processing speed and use-cases of this ontology meets our requirement hence we choose to connect to it rather than replicate functionality. Thus, we benefit from Company House's logic and latest data when we need it – a classical benefit of linked data.

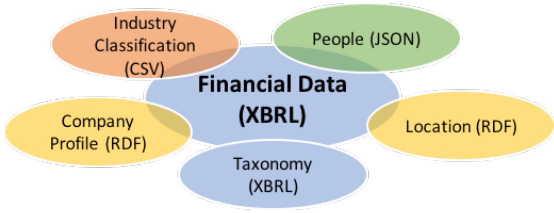

Fig. 6. Data sources integrated and their formats

3.3 Implementation & Technology Stack

Our deployment is implemented in the Oracle XML DB with XBRL and Semantic graph extensions. Oracle serves this use-case well by providing the end-to-end infrastructure for storing XML-based XBRL to performing inference on semantic graphs (note that these parts are not usually connected in Oracle). That being said, any other open source RDF-store could serve our need from the technical perspective, but using a standard technology reassures customers in the financial sector.

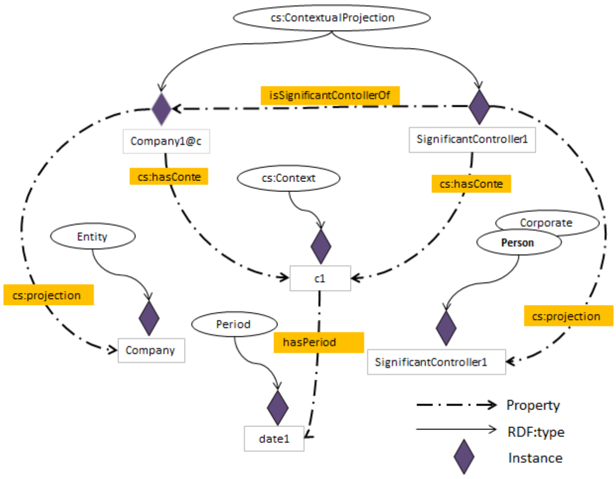

Fig. 7. Modelling Persons with Significant Contribution using Context Slices Design Pattern [13]

The main workflow in is the translation of XBRL to RDF based on our ontology. To begin with, ETL processes collect and transform iXBRL filings from Company House to RDF. We rely on the RDB to RDF Mapping Language (R2RML) to achieve the transformation. R2RML [15] allows to express mappings from relational data to RDF. R2RML processors either offer virtual SPARQL endpoints for querying underlying relational data or provide RDF dumps based on R2RML mappings. Oracle allows 'virtual' RDF views through R2RML mappings. This has the advantage of retaining the connection between the underlying XBRL data and the RDF data model. Data in the RDF model is live and changes in the XBRL layer are immediately reflected in the RDF layer. Figure 8 presents a simplified view of this transformation. Alternatively, generic XML processors could be used to transform XBRL into the target semantic model. [9] proposes XML2RDF to transform XML to RDF based on ontology derived from the schema using XSD2OWL mappings.

Simplified extract from XBRL Instance (.xml)

Report_ID	Entity	XBRL_Concept	Fact_Value	Units	Period
1001	7626170	TangibleFixedAssets	2752	GBP	instant_31-MAY-15
1001	7626170	CashBankInHand	7674	GBP	instant_31-MAY-15

R2RML

Triples

Subject	Predicate	Object
http://www.companieshouse.gov.uk#07626170	:submits	http://example.com/1001
http://example.com/1001	:hasFact	:TangibleFixedAssets
http://example.com/1001	:hasFact	:CashBankInHand
:TangibleFixedAssets	:hasValue	2752
:TangibleFixedAssets	:hasUnits	GBP

Fig. 8. Transforming XBRL into RDF triples

The ontology resulting from these transformations forms the basis for further inference and advanced semantic queries. We apply SPARQL CONSTRUCT and inference to curate and derive other calculated ratios. This curation is required because of variations in the way entities tag their accounts even in the same taxonomy and also to cater for the need to make accounts accounts submitted in different taxonomies comparable. Using forward chaining, entailments (inferences) are computed and indexed at the time of creating the model in the ORACLE DB, rather than handling these a query time allowing queries that rely on inference to run fast as well. SPARQL queries against this data are more expressive and intuitive compared to XML or SQL approaches and are discussed in the ensuing section.

3.4 Advanced Queries

With the ontology populated, we have the base model/data to derive additional information which include financial ratios for assessing financial performance of the company (e.g. Liquidity, Return on Assets and Return on Equity etc). With the expressive power of semantic languages we rely on the CONSTRUCT function to compute ratios within specific contexts. Taking this further, we build new abstractions of companies that are fast growing, highly profitable, high leverage and low liquidity among others. These classes in themselves fulfil complex queries e.g. "Technology Businesses that have low liquidity" and can also be put together to fulfil even more complex queries such as finding suitable acquisition targets (as illustrated in section B.2).

To begin with, we compute the derived ratios from the primary data using SPARQL's CONSTRUCT. This illustrates the expressivity and generally more intuitive nature of SPARQL compared to SQL or XQuery for querying the XML-based XBRL documents. In the construct captured in Listing 1 we bind early in the query the required subset of data to make the query more efficient and extract primary facts with the same context (Period, Entity among others) and then proceed to use them in computing the derived types e.g. working capital, liquidity, profitability, leverage among other financials. Listing 1 shows a sample construct query for Working Capital and Current Ratio. With a working knowledge of SPARQL it is easy to see the intuitiveness in making sure that the facts used to derive the financial ratios have the same context and subsequently assigning this same context to the computed financial ratio.

We apply inference to derive concepts/classes that facilitate the analysis of the financial report and the status of the company. The notions defined by Forbes [16] are used in determining suitable acquisition targets are inferred.

Fast growing – this is measured by the three-year compound annual growth rate of sales. Companies with faster growth rates are more likely to be acquired.

High profitability – is the ratio of Earnings Before Interest, Tax, depreciation and amortization (EBITDA) to Sales. Private companies with much higher profitability are more likely targets for acquisition.

High leverage – is the ratio of debt to EBITDA. Private companies with higher than average leverage are more likely to be acquisition targets.

Low liquidity – measured by the ratio of current asset to current liabilities. This is an indication of much money the entity has, to cater for its short-term needs. Acquisition targets have lower levels of liquidity.

Listing 1. Low Liquidity Company Inference Example

```
Antecedent
   (?fact a ex:FinancialRatio)
   (?fact rdfs:label "CurrentRatio"@en)
   (?fact ex:hasEntity ?entity)
   (?fact ex:hasValue ?value)
   (?fact ex:hasPeriod ?period)
   (?period a ex:CurrentPeriod)
Filter
   (value < 2)
Consequence
   (?entity a ex:LowLiquidityCompany)
```

4 Evaluation & Discussion

The methods reported in this paper form parts of an emerging software product, and hence the two items at the forefront of our evaluation were feasibility in terms of functionality and feasibility in terms of performance and storage. A number of complex and valuable queries become easily enabled with the transformed XBRL taking advantage of constructs, inference and federated queries (connecting to external data). Two of these are illustrated below as evaluation.

Benchmarking Queries. Such queries allow the analyst to compare businesses that a similar along some attributes example location and industry. One such query is: "*Find businesses in the same industry and district (location) as mine with similar Financial Asset/Profit*".

In the SPARQL query fulfilling this question, we connect to external data on company house's linked data service to obtain SIC Code (industry classification) and the District (locality) of the company of interest. We then proceed to extract companies with same industry and district for benchmarking. This external data can be merged with internal data.

Target Acquisiton Queries. A typical investor query might be to find good targets for acquisition. Using Forbes definition of interesting acquisition targets: "*Private companies are more likely to become acquisition targets if they are large, fast growing, and have high profitability, high leverage, and low liquidity*" [16].

Figure 9 shows a SPARQL query that fulfills this question. We make use of the inferred classes defined earlier to answer the complex question of finding acquisition targets. This specific example queries private companies in the information technology industry, located in a specific district (Birmingham) that meet acquisition target description i.e. low liquidity, high profitability, high leverage among others. This query uses our in-house ontology for computation and inference involving financial data and federates to company house for information on company profile and then to ordinance survey for location information. To optimize this query Oracle provides a push down option which allows variables of local query to be bound before being dispatched to external source. The impact of this directed query is significant as response time of your test query dropped to under 1 second from 30 seconds. Without this option the external query runs first and its outcome (possibly large) are combined with internal query.

In summary, the SPARQL queries shown demonstrate the expressive nature of our approach and its ability to fulfil more complex queries relatively easily and very intuitively when compared to relational or XML based approaches.

In terms of performance and scalability evaluation, we have populated our database with 10000 company reports. We can typically process a batch of 1000 reports into the database in under 30 minutes (this allows to process more than the reports filed in a day in a batch mode overnight as the process is unsupervised). Initial results for running queries against this dataset are very promising with most queries run in under 1 second thanks to the optimization of the ontology for querying. Also, running the queries against different sized data sets showed that the query time does not significantly increase, so performance on that account seems unproblematic. It should be noted that to gain the same insight in a manual way is almost infeasible as an accountant would need to study the reports to derive the answers, a job that takes many hours and is costly.

Storage of data is a slightly different issue; as we are retaining the XBRL filings in addition to the RDF tuples we require twice the storage space for the data, so with every filing added the storage need increases. However, as traceability is required and the data obviously needing to be stored for querying this is unavoidable. Making more use of linked data could address this concern, however it introduces a stronger reliance on third parties' live data services, which for a commercial product is less desirable.

```
Select ?company ?legalname ?financialreport ?sic ?orgtype ?district
  where {
    ?company a    ex:LowLiquidityCompany,ex:FastGrowingCompany,
                   ex:HighProfitableCompany,ex:HighLeverageCompany;
             ex:submits ?financialreport.

    SERVICE <http://business.data.gov.uk/companies/query>  {

      ?company  rov:orgActivity ?sic;
                rov:orgType  cat:private-limited-company;
                rov:orgType  ?orgtype;
                rov:legalName ?legalname;
                terms:registeredAddress ?address.
      ?sic skos:prefLabel "Other information technology…"@en.

      ?address postcode:postcode ?postcode.
      ?postcode postcode:district ?d.
       ?d    skos:prefLabel   "Birmingham"@en.
     }

   }',

     SEM_Models('XBRL_INST_MODEL'),
     SEM_Rulebases('OWLPRIME','xbrl_inst_rb'), null, null, null, ' SERVICE_JPDWN=T '
));
```

Fig. 9. Acquisition target SPARQL query

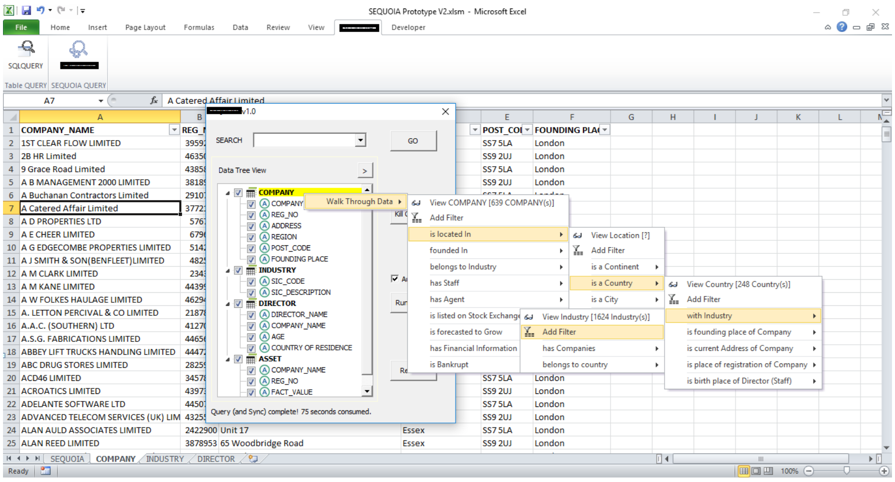

Fig. 10. The prototype spreadsheet based breadcrumb query interface

5 Summary & Future Work

We have presented an approach for integrating financial filings based on XBRL in a semantically enhanced way, which allows to run complex queries using easy to understand, standard semantic web techniques to gain new insights into financial markets. True to Synapse's philosophy of enabling advanced functionality in user-friendly and familiar environment of most accountant and financial analysts, the proposed interface is embedded in the familiar spreadsheet (Excel) environment of Synapse's Cloud CFO product (Fig. 10). The intuitive UI allows users to navigate natural questions from the main concepts that exist in the domain. The exploratory nature of the UI enables users to view intermediary results while working to fine tune their questions about the data. The ability to break the query and view intermediate results and then start off again from any other concept allows us to support complex queries.

The approach covers the full circle from integration of company filings into a unified database to querying the combined data. The approach importantly retains traceability of the source of information while allowing to enhance the reports with additional data from other sources to get even wider insights. Initial results show that good performance can be achieved while the required functionality is fully achieved.

As immediate future work we are analyzing scalability further and are adding user friendly interfaces, which can be used by any financial advisor, to the approach. We are also considering integration of further data sources into the mix to provide a yet wider network of data. Future work will also integrate more dynamically to the UI making the UI completely driven by the underlying semantic graph by dynamically generating menus from the graph. This loose coupling will enable the interface to work with any appropriately labelled underlying graph allowing businesses to ask complex questions to their internal and external data while retaining the ability to analyze and visualize the results from familiar spreadsheet environments.

Based on the product, we will work towards financial data being published as linked-data using our ontology. This will form the basis for further curating the XBRL based financial data and integration of other sources providing new technical challenges and business opportunities alike.

Acknowledgement: This work is partly funded by InnovateUK KTP009972.

References

1. “Who else uses XBRL? | XBRL.” [Online]. Available: https://www.xbrl.org/the-standard/why/who-else-uses-xbrl/. [Accessed: 24-Feb-2017].
2. “Interactive Data to improve financial reporting”. [Online] Securities and Exchange Commission. Available: https://www.sec.gov/rules/final/2009/33-9002.pdf [Accessed: 8-05-2017].
3. CoreFiling, “CoreFiling: Company Filing Data Search.” [Online]. Available: http://companies.corefiling.com/search. [Accessed: 24-Feb-2017].
4. R. Debreceny, C. Felden, B. Ochocki, M. Piechocki, and M. Piechocki, *XBRL for Interactive Data: Engineering the Information Value Chain*, 1st ed. Springer Publishing Company, Incorporated, 2009.
5. D. Valentinetti and M. A. Rea, “Critical Reflection on XBRL: A " Customisable Standard " for Financial Reporting?,” *Int. J. Account. Financ. Report.*, 3(2), 2162.
6. C. Hoffman and L. Watson, *XBRL For Dummies*. For Dummies, 2009.
7. V. Morilla, B. Ochocki, G. Shuetrim, M. Goto, R. Hommes, and H. Wallis, “XBRL Formula Overview 1.0,” *XBRL International Inc*, 2011. [Online]. Available: https://www.xbrl.org/wgn/xbrl-formula-overview/pwd-2011-12-21/xbrl-formula-overview-wgn-pwd-2011-12-21.html. [Accessed: 07-Feb-2017].
8. H. Carretié, B. Torvisco, and R. García, “Using semantic web technologies to facilitate XBRL-based financial data comparability,” *CEUR Workshop Proc.*, vol. 862, pp. 16–30, 2012.
9. Klein, M.C.A.: Interpreting XML Documents via an RDF Schema Ontology. Proceedings of the 13th International Workshop on Database and Expert Systems Applications, DEXA’02. pp. 889–894. IEEE Computer Society, Washington, DC, USA (2002).
10. M. Spiess. An ontology modelling perspective on business reporting. Information Systems 35 (4), pp 404-416. 2010.
11. XBRL International Inc., “XBRL Abstract Model 2.0,” *XBRL International Inc*, 2012. [Online]. Available: http://www.xbrl.org/Specification/abstractmodel-primary/PWD-2012-06-06/abstractmodel-primary-pwd-2012-06-06.html. [Accessed: 07-Feb-2017].
12. C. Welty, “Ontology Design Patterns:Context Slices.” [Online]. Available: http://ontologydesignpatterns.org/wiki/Submissions:Context_Slices. [Accessed: 18-Jan-2017].
13. Companies House, “Companies House – Linked Data Service.” [Online]. Available: http://business.data.gov.uk/companies/docs/getting-started-with-query.html. [Accessed: 07-Feb-2017].
14. S. Das, S. Sundara, and R. Cyganiak, “R2RML: RDB to RDF Mapping Language,” *W3C*, 2012. [Online]. Available: https://www.w3.org/TR/r2rml/. [Accessed: 09-Feb-2017].
15. “6 Key Financial Indicators Of Attractive Acquisition Targets,” *Forbes*, 2016.
16. H. Averkamp, “Financial Ratios - Balance Sheet | AccountingCoach.” [Online]. Available: http://www.accountingcoach.com/financial-ratios/explanation/2. [Accessed: 27-Feb-2017]