

Cross-lingual Querying and Comparison of Linked Financial and Business Data

Seán O’Riain¹, Barry Coughlan¹, Paul Buitelaar¹,
Thierry Declerck², Uli Krieger², Susan Marie-Thomas³

¹Digital Enterprise Research Institute, NUI Galway, Ireland
{sean.oriain, barry.coughlan, paul.buitelaar}@deri.org

²DFKI, Saarbrücken, Germany
{Declerck, krieger}@dfki.de

³SAP Research, Karlsruhe, Germany
susan.marie.thomas@sap.com

Abstract. Cross lingual querying of financial and business data from multi-lingual sources requires that inherent challenges posed by the diversity of financial concepts and languages used in different jurisdictions be addressed. Ontologies can be used to semantically align financial concepts and integrate financial facts with other company information from multilingual, semi-structured and unstructured Open Data sources. Availability as Linked Data then allows cross-lingual interrogation of the interlinked multi-lingual data set. This paper presents how the use of semantics and Linked Data enables the alignment and integration of business and financial facts provided by the different European Business Registers. The demonstrator allows business users to query multilingual data, perform comparisons, and review generated financial metrics.

Keywords: Multilingual, Ontology Alignment, XBRL, Financial Analytics

1 Business Motivation and Research Context

Business Registers throughout Europe are tasked with the collection, publishing exchange and provision of companies’ information and annual financial statements. Currently individual registers service specific information requests by returning populated pre-defined XBRL formatted templates with elements and content specific to the country of origin. Any further correlation or integrated analysis is left to the individual data requestor. To address this problem the Europe Business Registers Working Group (xEBR WG)¹ promotes registry information harmonization through the use of the eXtensible Business Reporting language (XBRL)² and the development of the xEBR Core Taxonomy, which was mapped to existing country-specific accounting taxonomies³, company identification taxonomies and extended with key financial ratios.

¹ <http://www.xbrleurope.org/working-groups/xebr-wg>

² <http://www.xbrl.org/>

³ Also referred to as Generally Accepted Accounting Principles or GAAP’s

This paper introduces the Business Intelligence Cross-lingual XBRL (BIXL) demonstrator, developed in cooperation with the xEBR WG as one of the Monnet Projects⁴ use cases. Building on xEBR WG efforts, BIXL enables information dependent activities such as financial analysis and general business investigation. Analysts are able to locate, access, and compare key data and financial figures from reports and text, regardless of the XBRL jurisdiction taxonomy used to mark up the original report or the language used. BIXL functionality was developed by combining:

- i) Alignment of xEBR Core Taxonomy mappings with selected XBRL country taxonomies and extraction of financial data from XBRL filings as RDF triples.
- ii) Extraction, annotation and alignment of other key business concepts (e.g. stock exchange name, number of employees, and activity sector) which are extracted from multilingual web semi-structured/unstructured text sources as RDF triples.
- iii) Integration of alignment results into the Monnet Financial Ontology (MFO) [1] and its instantiation as Linked Financial Data. Use of SPARQL queries to generate financial data summaries, metrics and comparison tables

The BIXL demonstrator builds on the notion of ontology localization [2] and emerges from previous attempts directed at financial term translation using statistical MT [3] and hybrid methods[4]. The demonstrators use of multilingual sources additionally progresses the availability and ability to deal with cross-jurisdictional financial Linked Data [5] as part of a move towards the Multilingual Semantic Web [6]. A recent monolingual commercial example offering a similar single entry point to company information, people, financial data, including ratios statement and disclosures, is 9W Search⁵, which targets mobile delivery of financial information.

A brief overview of the BIXL demonstrator architecture is first presented (Section 2) before a walk-through of the cross-lingual query and comparability functionality (Section 3).

2 Linked Data Integration and Processing Architecture

Figure 1 illustrates the 3 stage processing architecture and tools used to generate fact instances compatible with the Monnet Financial Ontology [1] (MFO). The data source layer outlines the XBRL jurisdictions and company information sources which are then converted (circular nodes)⁶ into MFO RDF compliant triples and axioms, as part of the MFO conversion process. OpenRDF Sesame was chosen as an RDF store due to its implementation of the SPARQL 1.1 syntax and the use of SPARQL property paths in simplifying tree structure navigation. The data layer is responsible for constructing and executing SPARQL queries against the RDF data store and marshalling of result sets into JSON Strings. Utility classes are available to convert result sets into a tree representation, needed later for traversal of the xEBR Core Taxonomy structure.

⁴ <http://www.monnet-project.eu>

⁵ <http://9wsearch.com/>

⁶ Available for download from <https://github.com/monnetproject/rdffconverters>

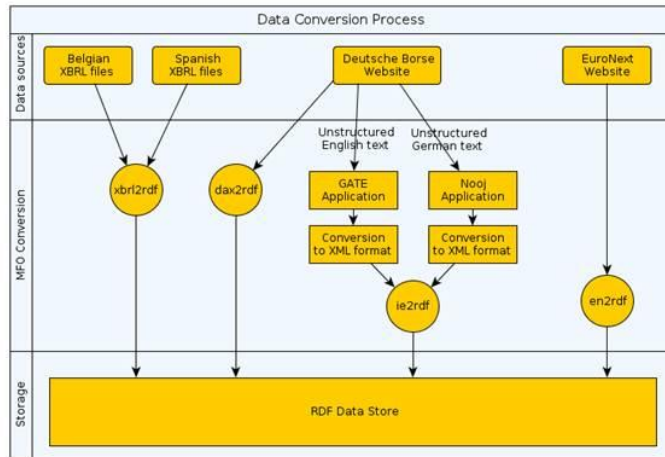


Figure 1 Multi-lingual Financial and Business Data Integration, adapted from [5]

3 Cross-lingual Querying and Comparison of Linked Financial and Business Data Demonstrator

The BIXL’s web interface allows business information querying from XBRL and stock exchange sources. Users can query, navigate and compare financial filing and company profile information across multiple languages and financial jurisdictions. Users are initially presented with a search form (Figure 2) to narrow selection based on data source (i.e. the type of query), the particular companies under investigation and the applicable time period.

Figure 2 Business Information Source Selection Filter

Available query types are Company Profiles, Financial Filing Summaries/Metrics or a combination of both. Companies are selected using the auto-complete search box and company name, identifier (or ISIN number if unavailable), and the number of available filings/company profile snapshots in the RDF store, displayed. Users can

select from the project supported languages of English, Spanish, German or Dutch, also used to specify the language for cross-lingual querying and interface labels. Search forms display results in an easy to access tabularized comparable format. The 'Company Profile' view shown in Figure 3, provides a side-by-side comparison of key data for two selected companies. Temporal specific snapshots for each company can be selected via the drop-down boxes. Displayed information extracted from unstructured text is underlined with dashes. When the user hovers over a fact extracted from unstructured text, a tool-tip displays the original text and outlines where the fact was extracted from, here in German.

Select Snapshot:	Deutsche Börse - 2012-06-06	Deutsche Börse - 2012-03-01
Name	Bayerische Motoren Werke Aktiengesellschaft BMW AG St BMW Group	DEUTZ AG Deutz AG
Stock Exchange	Deutsche Börse	Deutsche Börse
ISIN	DE0005190003	DE0006305006
Total Capital Stock	€655,566,568	€308,978,241
City	80788 München	51149 Köln
Country	Germany	Germany
Telephone	+ 49 (0)89 382-0	+49 (0)221 822-0
Employees	95500	3839
Internet	http://www.bmwgroup.com	http://www.deutz.com
End of Business Y	Im Geschäftsjahr 2010 erzielte die BMW Group einen weltweiten Absatz von 1,46 Millionen Automobilen und über 110.000 Motorrädern. Das Ergebnis vor Steuern belief sich auf rund 4,8 Mrd. Euro, der Umsatz auf 60,5 Milliarden Euro. Zum 31. Dezember 2010 beschäftigte das Unternehmen weltweit rund <u>95.500 Mitarbeiterinnen</u> und Mitarbeiter.	

Figure 3 Company Profiles

The 'Financial Filings Summary' view shown in Figure 4, provides a comparative view of financial data extracted from XBRL filings.

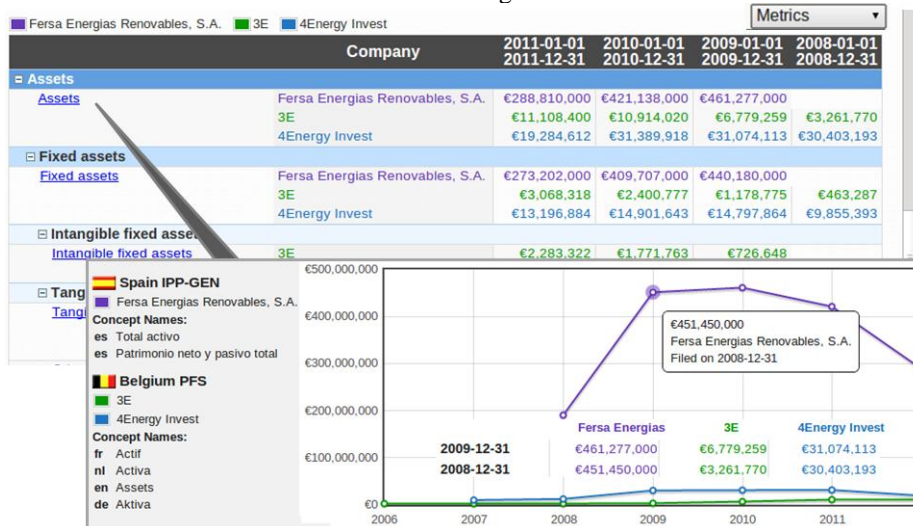


Figure 4 Financial Filings Summary Comparison

The data, displayed in a tree table corresponds to the hierarchical structure of the xEBR Core Taxonomy. To assist users manage the level of information displayed, abstract concept nodes (aggregated financial figures) and their child nodes (individual financial figures) can be collapsed or expanded. Concepts found in the XBRL filings with corresponding instances are not shown with the displayed data. Colour coding is added to assist easy visual identification and comparison. Selecting the hyper-linked financial concept name (here “Assets”), provides, a detailed plot of values over time across the companies considered by the query. Source taxonomies provenance used to generate the xEBR data is also listed along with equivalent concept names in available project languages. Generated Financial Metrics (Figure 5) can be viewed by selecting “Metrics” from the drop down combo box in Figure 4.

Company		2011-12-31	2010-12-31	2009-12-31	2008-12-31
Net Result	Fersa Energias Renovables, S.A.	€-126,381,000	€-35,597,000	€140,000	€-26,039,000
	3E	€-2,611,535	€-689,938	€272,839	€308,894
	4 Energy Invest	€-11,899,155	€72,521	€508,227	€-1,593,721
Own Funds	Fersa Energias Renovables, S.A.	€227,562,000	€353,079,000	€389,891,000	€394,424,000
	3E	€2,697,900	€5,312,626	€1,001,501	€733,978
	4 Energy Invest	€18,598,559	€30,497,714	€30,425,193	€29,916,966
Current Ration (Measurement of Liquidity)	3E	1.54	3.36	1.01	1.14
	4 Energy Invest	9.10	21.96	35.38	42.26
Net Working Capital	3E	€2,811,779	€5,982,852	€80,286	€349,683
	4 Energy Invest	€5,418,387	€15,737,484	€15,816,176	€20,061,573

Figure 5 Metrics Summary View

The Metrics view displays both Business Registry specified financial metrics (e.g. Own Funds, Net Results) and generally applied investigative accounting ratios (e.g. Current Ration, Net Working Capital) calculated from the balance sheet of the XBRL filings aligned to the xEBR taxonomies. Here again the tabulated view allows quick side-by-side comparison of company figures or analysis ratios. The **BIXL demonstrator** is available from <http://monnet01.sindice.net:8080/financial>.

Acknowledgements. Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Lion-2), and the EU FP7 MONNET project under Grant Agreement No. 248458.

References

1. Krieger, H.K., Declerck, T., Nedunchezian, A.K. *MFO - The Federated Financial Ontology for the MONNET Project*. in *Proceedings of the 4th International Conference on Knowledge Engineering and Ontology Development*. 2012. Barcelona, Spain.
2. Cimiano, P., Montiel-Ponsoda, E., Buitelaar, P., Espinoza, M., Gomez-Peresz, A., *A note on ontology localization*. *Appl. Ontol.*, 2010. **5**(2): p. 127-137.
3. Arcan, M., Federmann, C., Buitelaar, P., *Experiments with Term Translation in 24th International Conference on Computational Linguistics*, 2012: Mumbai, India.
4. Wunner, T., Buitelaar, P., O’Riain, S. *Semantic, Terminological and Linguistic Interpretation of XBRL*. in *Workshop on Reuse and Adaptation of Ontologies and Terminologies*, EKAW, 2010. Lisbon, Portugal.
5. O’Riain, S., Edward, E., Harth, A., *XBRL and open data for global financial ecosystems: A linked data approach*. *IJAIS*, 2012. **13**(2): p. 141-162.
6. Buitelaar, P., Choi, K.S., Cimiano, P., Hovy, E. *The Multilingual Semantic Web (Dagstuhl Seminar 12362) Dagstuhl Reports*. 2013. Germany: 2/9, Schloss Dagstuhl--Leibniz-Zentrum fuer Informatik.