Open Financial Data from the Macedonian Stock Exchange

Bojan Najdenov, Hristijan Pejchinoski, Kristina Cieva,
Milos Jovanovik, and Dimitar Trajanov

Faculty of Computer Science and Engineering, Ss. Cyril and Methodius University,
Skopje, Republic of Macedonia
{bojan.najdenov,milos.jovanovik,dimitar.trajanov}@finki.ukim.mk,
hristijan.pejcinoski@gmail.com, kristina_cieva@yahoo.com

Abstract. The concept of Open Data, which represents the idea that public data should be published in a machine-readable format, starts to take a significant role in modern society. Public data from various fields are being transformed in open data formats and published on systems which allow easier consumption from software agents and applications, as well as the users behind them. On the other hand, people in the business world are trying for a few decades now to establishing standards for financial accounting that govern the preparation of financial reports. Financial reporting has crucial significance for companies today, since it is a record of their work which is presented to their stakeholders and represents a starting point for future business decisions and strategies. In this paper, we use data from the Macedonian Stock Exchange and data from different web sites of Macedonian companies in order to create datasets of Open Financial Data relevant for our country, thus increasing the transparency and improving the data accessibility. We describe the process of transforming the data into 4 star Open Data, and present use-case scenarios which use data from our generated datasets and from the World Bank. The datasets are published and accessible via a SPARQL endpoint, and we demonstrate how a software application can make use of them.

Keywords: Finances, Open Data, Macedonian Stock Exchange, World Bank, RDF, Ontologies.

1   Introduction

The main idea that lies behind the concept of Open Data[1] is that public data should be free and available to everyone. We live in a world where information holds great value. Having the right information at the right time, in the right way, builds modern societies, drives technologies forward, develops businesses and even saves lives. The exponential growth of datasets about people, technological artifacts and organizations brought us in position where we have on disposal vast amounts of information ready

---

[1] http://okfn.org/opendata/

to be rearranged and shaped in order to create additional value [1]. This implies that the structured, machine-readable, open data that is free to access and interlink, is becoming the future of startup companies and business in general.

Linked Open Data[2] is a community effort to alleviate the problem of the lack of sufficiently interlinked datasets on the Web. Through this effort, a significant number of large-scale datasets have now been published in the LOD cloud[3], which is growing constantly [2]. As we see in Fig. 1, datasets from different fields are publicly available in Linked Open Data format, thanks to the contributors to the Linked Open Data community [3].

The concept of Linked Open Data provides us with a way to connect datasets stored on different locations, by using the Semantic Web standards such as RDF, OWL and SPARQL. By using the existing Web infrastructure, data from different data silos can be successfully interconnected and the Web can be used to decrease the barriers which occur during process of linking [4]. Linked Open Data enables better data analysis by simplifying the process of combining information sources. Datasets from various industry fields can then be used in different ways and by many entities, e.g. regulatory bodies and banks, when it comes to financial data [5].

Financial accounting is primarily oriented towards creating the financial reports for the companies' work. The process of creating the reports is dependent on a huge amount of datasets. Thus, this is a field that necessarily requires different approaches for representation, storage, querying and visualizing of the data. We find this issue of big importance for today's companies and economies which motivated us to work on a practical solution using data about Macedonian companies provided by the Macedonian Stock Exchange, the World Bank and the information and data they publish on their websites.
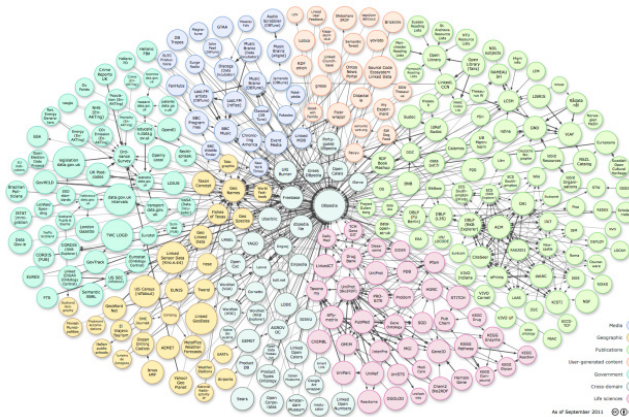

Fig. 1. The LOD Cloud, as of September 2011

---

2 http://linkeddata.org/
3 http://lod-cloud.net/

2   Related Work

Numerous projects exist which have a major target to either publish financial or corporate data in Open Data formats, or enable their annotation with the technologies of the Semantic Web, in order to leverage their value. The World Bank, as one of the most important financial institutions on a global level, puts great effort in many projects which result with creating Open Data. Other significant projects in this area are the Financial Industry Business Ontology (FIBO), the Open Corporates project and the Financial Report Ontology.

The World Bank aims towards decreasing extreme poverty in the world, through proving financial and technical assistance to developing countries. The financial support the developing countries receive is in form of low-interest loans, credits and grants, or investments in various areas like healthcare, education, infrastructure, resource management etc. The World Bank, as a global institution, supports the ideas behind the Open Data concept, and therefore shares its public data freely on their website[4].

In [6], the authors introduce an interesting project which aims towards designing new methods for extraction of data and, based on that, developing a prototype for extracting financial information from the semi-structured text. They believe that in the financial world numbers are often one main target, but they are meaningless without any semantic meta-data describing what kind of information they represent.

The Financial Industry Business Ontology[5] (FIBO) is an initiative to define and describe terms and rules for financial data. Its goal is building a representation of the information about financial instruments, market data, business entities, etc. along with the relationships between them.

Open Corporates[6] is one of the largest open databases of companies in the world, having information about 63 million companies from around the globe. They publish the data in XML, RDF or JSON format and it can be downloaded from their website. They believe that basic corporate information about all the companies in the world should be brought together in one place, making it easier to access, use and connect with other data.

The Financial Report Ontology[7] is a project developed with the idea of providing an ontology that would describe the financial reports as concepts, as well as their individual entries. The ontology aims to assist companies in the process of creating annotated financial reports.

3   Macedonian Open Financial Data

3.1   Public Data from the Macedonian Stock Exchange

The Macedonian Stock Exchange (MSE)[8] is the only financial institution in Macedonia that is authorized to organize, execute and regulate the trading of

---

[4] http://data.worldbank.org/

[5] http://www.omg.org/hot-topics/fibo.htm

[6] http://opencorporates.com/

[7] http://financialreportontology.wikispaces.com/

[8] http://www.mse.mk/en/

securities. It was established in 1995 as a joint stock company and the first trading occurred in March, 1996. The main purpose of MSE is to provide security and efficiency in the organized trading of securities in Macedonia.

MSE is comprised of two market segments: Official Market and Regular Market. The stock market indices are MBI10 (Macedonian Blue Chip Index), which includes the stocks of the 10 most traded companies, MBID (Macedonian Stock Exchange Index of publicly held companies), which includes the stocks of the publicly held companies listed on MSE and OMB (Bond Index), which includes issued bonds listed on MSE.

MSE publishes most of its data on their website, either as PDF files or in HTML tables. Among all of the published data, like stock prices, different indices, information about growth trends on securities, etc., our main topic of interest are the financial reports which MSE member companies publish. We gathered the financial report data from the MSE website, converted it and stored it in CSV format. We did the same process for gathering and storing the company data, which we obtained from individual companies websites.

3.2   Open Data from the World Bank

As we already noted, the World Bank published data from its projects on their website. Parts of these data are the financial data, which allow us to see what global funds the World Bank manages, visualize them or build models over them.

Many different financial datasets can be found on World Bank's website[9] in various different formats, such as CSV, JSON, PDF, RDF, RSS, XLS, XLSX and XML. Some of their datasets can be accessed via the public SPARQL endpoint which the World Bank provides[10], as part of their Linked Data project. The dataset that we are interested in contains data on commitments against contracts that were reviewed by the Bank before they were awarded (prior-reviewed Bank-funded contracts) under IDA/IBRD[11] investment projects and related Trust Funds. We downloaded this dataset in RDF format and linked its data with data published by MSE and Macedonian companies. The procedure will be described in details.

4   Ontologies for the Datasets

4.1   Ontology for the World Bank Dataset

We loaded the dataset from the World Bank data store into a local Virtuoso Universal Server[12] instance, as an RDF graph. Since all the entries in the dataset refer to a loan awarded to a company by the World Bank, a single entry in the dataset can be considered as a resource which provides all the details related to a specific loan.

---

[9] https://finances.worldbank.org/all-datasets
[10] http://worldbank.270a.info/sparql
[11] http://data.worldbank.org/indicator/DT.DOD.MWBG.CD
[12] http://virtuoso.openlinksw.com/

4.2   Corporate Registry Ontologies

As we already mentioned, Open Corporates holds a large publicly available dataset of information about companies as legal entities, for all around the world. Unfortunately, they do not hold any information about Macedonian companies, and therefore we cannot use their datasets in the context of Macedonian financial data.

However, we did analyze their data and the ontologies they use for semantic annotation, so we decided to reuse those ontologies and annotate our data in a similar manner. Another motivation for this was the similarity between the structures of the dataset from Open Corporate had with the data we were able to collect for Macedonian companies. The ontologies we use in describing the companies as legal entities are listed in Table 1.

Table 1. The ontologies we reused for Macedonian company data

Prefix	URI
foaf	http://xmlns.com/foaf/0.1/
vCard	http://www.w3.org/2006/vcard/ns#
adms	http://www.w3.org/ns/adms#
rov	http://www.w3.org/ns/regorg#
skos	http://www.w3.org/2004/02/skos/core#

We use the rov:RegisteredOrganization class in order to represent a legal entity or organization which is legally registered, i.e. a company that we have data about. The rest of the DataType properties we use to describe a Registered Organization can be found in Table 2.

Table 2. The DataType properties we use

Property	Description
rov:legalName	The legal name of the company.
rov:registration	The registration is a fundamental relationship between a legal entity and the authority with which it is registered and that confers legal status upon it. rov:registration is a sub property of adms:identifier which has a range of adms:Identifier.
vCard:extended-address	The address of the object.
vCard:hasTelephone	To specify the telephone number for telephony communication with the object.
skos:notation	Refined name of a company.
foaf:homepage	A homepage for some company. Every value of this property is a foaf:Document.
rdfs:label	Information about the basic activities of a company.

4.3   Financial Report Ontology

Every member of the Macedonian Stock Exchange provides annual financial reports which are the balance sheet, income statement, statement of cash flows and the statement of retained earnings. Our focus in this paper is the balance sheet of the companies in particular, which requires an ontology to be provided so that we could semantically annotate that data.

For this purpose we decided to reuse the Financial Report Ontology which, as we already described, defines the basic financial report terms.

In the ontology we find the class Fundamental Accounting Concept, which represents one full financial report. Its properties are divided into five groups: General Information properties, Balance Sheet, Income Statement, Statement of Comprehensive Income and Cash Flow Statement properties. For our local reports we will use only General Information, Balance Sheet and Income Statement properties.

Table 3. The properties in the CFRL ontology

Property	Description
cfrl:hasReport	This property connects a company i.e. instance of RegisteredOrganization class, with its financial report.
cfrl:hasLoan	This property points to the World Bank loans that are made by that company.

4.4   Corporate Financial Reports and Loans Ontology

In order to be able to successfully complete the annotation and linking process between the datasets, we developed the Corporate Financial Reports and Loans Ontology (CFRL). In it, we introduce two object properties: "hasReport" and "hasLoan". Their main role is to provide means of interlinking the datasets. The description of these two properties can be found in Table 3.

5   Linking the Datasets

Before we begin explaining the process of interlinking the datasets, we must state that our goal is to interlink the data from our corporate registry dataset, i.e. the data we gathered from various websites of different companies, with the data we acquired from the World Bank about loans that companies were awarded and also with the financial reports data we got from the Macedonian Stock Exchange. Conceptually, the linking we wish to achieve is shown in Fig. 2.

Fig. 2. Linking the datasets

5.1   Mapping the Data from CSV to RDF

The next step of our work is mapping and transforming datasets from CSV files to RDF and to do that, we use the Virtuoso Universal Server, which provides mechanisms for data transformation management and querying using SPARQL.

The technical process of mapping and transforming the data from CSV to RDF was done using the R2RML mapping language[13], as described in [7], for the corporate registry dataset and the financial reports dataset, respectively. The dataset about loan details originating from the World Bank was already in RDF format, so we imported it directly in the Virtuoso Universal Server instance, as an RDF graph.

5.2   Interlinking the RDF Datasets

Having transformed all the datasets into RDF graphs in Virtuoso, our next step was to interlink the data, as shown in Fig. 2. For that purpose of we created two properties in our CFRL ontology: "cfrl:hasReport" and "cfrl:hasLoan".

The "cfrl:hasReport" property links a company with its published financial reports. That means, we connect a "RegisteredOgranization" entity with its financial report i.e. "FundamentalAccountingConcept", by matching the values of the company name. For this purpose we use the "skos:notation" property of a "RegisteredOgranization" entity, and the "fac:EntityRegistrantName" property of a "FundamentalAccountingConcept" entity.

The property "cfrl:hasLoan" interlinks a "RegisteredOgranization" entity with its loan entities. Similarly to the previous property, we create the connection by matching the names of the correspondent companies. For this purpose we use the "rov:legalName" of an "RegisteredOrganization" entity, and the "worldbank:supplier" property of a loan entry.

These interlinking processes were done using SPARQL queries over the datasets.

The resulting linked data that we generated, can be accessed through a public SPARQL endpoint[14].

---

[13] http://www.w3.org/TR/r2rml/
[14] http://linkeddata.finki.ukim.mk/sparql

6   Use-Cases

The main purpose of using interlinked Open Data datasets is the ability to increase the value and usability of the separate datasets, by providing advances use-case scenarios. We are going to describe two of the many possible scenarios.

6.1   Displaying Information from the World Bank

We demonstrate the use of the "hasLoan" property to retrieve information about a company which obtained a loan from the World Bank, or to be more precise, the dates when the company signed contracts for getting loans with the World Bank, the total contract amount (USD) and which sector was the loan dedicated to. For the purpose of the demonstration, we show the top 5 loans and their details. The SPARQL query is the following:

```
prefix cfrl: <http://linkeddata.finki.ukim.mk/lod/ontology/cfrl#>
prefix worldbank: <http://finances.worldbank.org/resource/>
prefix rov: <http://www.w3.org/ns/regorg#>

SELECT ?s ?csd ?tca ?ms WHERE {
   ?company rov:legalName ?s .
   ?s cfrl:hasLoan ?l .
   ?l worldbank:contract_signing_date ?csd ;
      worldbank:supplier_contract_amount_usd ?tca ;
      worlbank:major_sector ?ms .
} ORDER BY DESC (?tca) LIMIT 5
```

The result of the executed query at our Virtuoso SPARQL endpoint, are shown in Table 4.

Table 4. Results from the SPARQL query

Supplier	Contract signing date	Total contract amount	Major sector
Granit	Mar 26, 2009	$9,802,524.00	Transportation
Granit	Dec 04, 2009	$6,197,108.00	Transportation
Granit	Dec 04,2009	$5,323,028.00	Transportation
Granit	Mar 26, 2009	$4,519,095.00	Transportation
Granit	Dec 04,2009	$3,785,761.00	Transportation

6.2   Displaying Information from the Financial Reports

In this section we show how the "hasReport" property that we defined in our CFRL ontology, can be used to provide additional information about companies. One such scenario would be to retrieve information about the top 5 companies by the profit they

made in the year of 2012, in Macedonian Denars (MKD). For that purpose we can use the following SPARQL query:

```
prefix cfrl: <http://linkeddata.finki.ukim.mk/lod/ontology/cfrl#>
prefix fac:
<http://www.xbrlsite.com/2013/FinancialReportOntology/Prototype04/FundamentalAc
countingConcepts.xml#>
prefix rov: <http://www.w3.org/ns/regorg#/>

SELECT ?name ?profit ?period WHERE {
   ?cmp cfrl:hasReport ?rep ; rov:legalName ?name .
   ?rep fac:GrossProfit ?profit ; fac:FiscalPeriod ?period .
   FILTER (?period = 2012)
} ORDER BY ?profit LIMIT 5
```

The result of this query, showing the name of such companies and the profit they made in the year of 2012, can be seen in Table 5.

Table 5. Results from the SPARQL query

Name	Profit (MKD)	Period
ALKALOID AD SKOPJE	3,291,423	2012
Stopanska Banka AD Skopje	2,376,477	2012
Tikvesh AD Skopje	339,049	2012
GD GRANIT AD - Skopje	291,238	2012
Vitaminka AD Prilep	102,378	2012

7   Conclusion and Future Work

Data, information and knowledge management are key activities in modern economies and considerable efforts and resources are devoted for research in these areas, by different organizations in the world. Having data structured and interlinked provides a whole new area of opportunities for data usage and management. This provides huge benefits in the information dissemination processes and provides mechanisms so that information can be shared easily between bank divisions, institutions and distributed to all stakeholders.

In this paper we gave an overview of the process of transforming the one-star and two-star data about companies into four-star Open Data and connected it with a dataset from the World Bank. We also provided use-case scenarios which gave examples of how our local data and how the data from the World Bank can be used in order to provide information which is not available when the datasets are isolated. With this, we hope our work contributes to the goals of the Open Data Initiative[15] in Macedonia.

---

[15] http://opendata.gov.mk/

In the future, we plan to continue our work in these fields, increase the amount of datasets, connect our data with other remote resources and transform these datasets further to five-star data, interlinked with financial data published on the LOD cloud. This would improve the quality of the use-cases we provide and also create new opportunities for development of creative applications and analysis. We hope our work serves as a motivation to companies, financial institutions, organizations around the world, to recognize the benefits of open financial data and publish their public data on the Web in raw and machine-readable format.

References

1. Cardoso, J., Pedrinaci, C., Leidig, T., Rupino, P., De Leenheer, P.: Open semantic service networks. In: International Symposium on Services Science (ISSS), Leipzig, Germany (2012)
2. Möller, K., Hausenblas, M., Cyganiak, R., Handschuh, S., Grimnes, G.: Learning from Linked Open Data Usage: Patterns & Metrics. In: Web Science Conference (WSC) (2010)
3. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems (IJSWIS), 1–22 (2009)
4. Kundra, V.: Digital Fuel of the 21st Century: Innovation through Open Data and the Network Effect. Joan Shorenstein Center on the Press, Politics and Public Policy, Harvard College (2012)
5. Radzimski, M., Sánchez-Cervantes, J.L., Rodríguez-González, A., Gómez-Berbís, J.M., García-Crespo, A.: FLORA –Publishing Unstructured Financial Information in the Linked Open Data Cloud. In: First International Workshop on Finance and Economics on the Semantic Web (FEOSW) (2012)
6. Bjoraa, E.: Ontology guided financial knowledge extraction from semi-structured information sources. Master Thesis in Information and Communication Technology, Agder University Colledge, Grimstad (May 2003)
7. Jovanovik, M., Najdenov, B., Trajanov, D.: Linked Open Drug Data from the Health Insurance Fund of Macedonia. In: 10th International Conference for Informatics and Information Technology (2013)