Linked Data for Financial Reporting

Masamoto Goto, Bo Hu, Aisha Naseer,
Pierre-Yves Vandenbussche

Fujitsu
{masamoto.goto, bo.hu, aisha.naseer}@uk.fujitsu.com,
pierre-yves.vandenbussche@ie.fujitsu.com

Abstract. Recently, uncertainty has dominated global financial markets. Companies are struggling with their balance sheets while merges and acquisitions (M&A) happens on a more frequent basis. This calls for highly accurate corporate performance data, which serve as the ground truth for M&A decisions. Conventional financial reporting approach has drawn criticism from both investors and analysts, attribute to the inherent complexity, lack of timeliness, and inadequacy of scope.
In this paper, we try to address these shortcomings with a Linked Data approach. We showcase a system for enriching corporate financial reporting with other types of proprietary and public data, the latter are part of the Linked Open Data cloud or harvested from social media sites. Our motivation is to 1) source in a variety of data so as to portray a comprehensive image of a company and 2) render tedious financial figures through an intuitive and comparative perspective. We introduce the architecture of the Linked Financial Data for Reporting framework (*LFD*r) and outline the details of its implementation. We also propose methodologies for the evaluation of the whole framework.

1 Introduction

Financial reporting is the communication of financial information about an enterprise to the external world/public. Thus far, the usefulness of corporate financial reporting has been criticised in two aspects. On the one hand, current financial reporting framework was largely shaped during and immediately after the first *industrial revolution* in response to the emergence of corporate form, stock market, and the regulation of accounting and auditing practices [14]. Due to the intricate nature of financial instruments, the complexity of financial reporting is inevitable. Obscure legal terms designed to avoid stating responsibility in a black-white fashion have aggravated the magnitude of complexity. As a result, it becomes increasingly challenging for investors, who are not professionally educated/trained, to distill the messages conveyed in such reports [10]. Financial report tooling should, therefore, not only assist authoring, but also about facilitate comprehension. On the other hand, since the latest *technological revolution*, corporate structure has undergone fundamental changes that starts to render the conventional reporting approach less useful for modern companies. Some of such fundamental changes include the difference between market value and book value of company assets, the raise of offshore financial centres and offshore financing channels, far-reaching globalisation, etc. Financial reporting based on conventional auditing and accounting methodology starts to find itself struggling in faithfully reflecting the performance of companies, in particular the social media and e-commerce businesses whose true value can only be revealed using data other than balance sheets, profit/loss statements, and cash flow statements.

In this paper, built on our experience of Fujitsu Interstage[1], we propose a system enriching current financial reporting practice with linked data computing paradigm. The Linked Financial Data for Reporting framework (*LFD*r) addresses the shortcomings of the state of the art in terms of data timeliness, data completeness, and data consumption. Especially, in order to hide unnecessary complexity of financial report and present the data to ordinary users/investors in an easy to comprehend fashion, we compile data across multiple companies to offer performance comparison instead of isolated figures of individual ones. Meanwhile, we expanded the scope, from where data are solicited, by linking not only conventional information sources of financial reporting, but also "unusual" ones to construct performance summaries going beyond balance sheets.

2 Related Work

One important benefit of applying Linked Data principles to financial domain is the increased data interoperability across multiple financial systems and financial instruments [12]. The financial industry has long acknowledged the necessity of aligning different data providers [3]. This is particular evident in Financial Reporting area where international collaboration is already in place. For instance, US Securities and Exchange Commission (SEC) has mandated that by 2014 all financial entities should adopt the eXtensible Business Reporting Language (XBRL). XBRL is a family of XML-based global standards, enabling automated exchange of business information through machine-interpretable tags. XBRL taxonomies are constantly revised by a joint task force and promoted through non-profit XBRL International[2]. Apart from US SEC, major players of XBRL includes US FFIEC, Bank of Japan, Tokyo Stock Exchange, and many other European and Asian financial regulators.

Rather unintuitively, the successful story of XBRL does not naturally imply a better data integration in Financial Reporting domain. In practice, XBRL and the ecosystem built around it have effectively created bigger data silos, e.g. vendor lock-in of proprietary XBRL tools, exclusion of non-XBRL data, discouraged reference to non-XBRL data sets, and reluctance of modifying XBRL fundamentals. Linking XBRL compatible and incompatible data sets, therefore, is not a trivial task. It is our contention that Financial Reporting presents a typical case where a significant establishment lends itself to Linked Data initiative as both an opportunity and a challenge. Conceived before and been developed in parallel with the Semantic Web vision, XBRL enjoys a consensus that has already

[1] `http://www.fujitsu.com/global/services/software/interstage/`
[2] `http://www.xbrl.org`

been negotiated and implemented across the entire community and propelled by governing authorities of leading industrialised countries. It, therefore, can serve as a proving ground for new technologies. Current XBRL technology primarily focus collects financial data for reporting. As more XBRL-based financial data becomes available, it will need to effectively extract key financial facts and to link with other types of data [16]. Semantic technology and linked data are designed for such use.

On the other hand, established standards can be barriers, hindering the adoption of new technologies that are not fully compatible with existing ones. In recent years, the value of semantically enriched XBRL has been recognised [11] [7]. Leaving non-technical issues aside, this lack of large scale adoption can be attributed to several factors. Full-scale conversion from XML-based XBRL instance and taxonomy documents to genuine RDF graph is not straightforward. Naive conversion can lead to badly distorted RDF graphs, dearth of inter-connections, or knowledge loss [6]. Without properly populated RDF models, the advantage of semantic inferences cannot be fully appreciated. Meanwhile, XBRL leverages a large number of procedural knowledge, defining how financial figures are arrived. With its current capacity, RDF might find difficulties in modelling and reconstructing such knowledge. Finally, XBRL defines the right amount of data for a special purpose. Modifying or extending XBRL, therefore, can raise operating costs and incur doubts among the established XBRL community.

Having considered the above arguments, instead of fully "LODnising" XBRL-based financial reporting, we decided to take a less restrictive approach–centring our efforts around XBRL and keeping the integrity of XBRL for better community acceptance and low threshold for technology adoption. The high-level goal of *LFD*r is, therefore, to better integrate financial data with related information using LOD as an instrument.

3 Use Case and Requirements

Better financial data integration allows for more efficient ways of understanding the performance of a particular company by both professional analysts and amateur individual investors. In the following, we illustrate typical usage scenarios of *LFD*r concerning both types of potential users. The example was elicited from requirement studies conducted against EU and Japanese XBRL users.

Persona One: Joan is an experienced financial analyst working for Bank X. When logging into X's ICT system, Joan selects the list of companies she wants to check. The companies are based in different countries and thus their reports present different terminologies, such as Japanese EDINET[3], US GAAP [4], and International IFRS[4]. Joan finds her way through the reports using Legal Entity Identifier (LEI) which defines a universal standard identifier of any organisation involved in international financial transactions. With her domain expertise, Joan

[3] `http://info.edinet-fsa.go.jp`
[4] `http://www.ifrs.org/Pages/default.aspx`

is able to uniquely identify the borrower companies of X and linking multiple copies of financial reports. She extracts some key performance indicators (KPIs) from a company's quarterly and annually reports and produces charts that visualise the tendencies of the chosen KPIs. Upon finishing, Joan publishes her conclusion for the benefit of all X's institutional and individual investors.

Persona Two: Jane, an individual investor, accesses her investment portfolio using X's ICT system. Upon logging in, Jane is granted the access to all financial reports, KPIs, and analysis. After trudging through the available information, Jane becomes indecisive: whether continue holding her investment in company G or move it to F. Jane does not restrict herself to predefined KPIs. She browses for latest news and stock prices of G, information about G's subsidiary companies and main stakeholder. She also instructs the system to take into account public opinion from social media. Jane aggregates "new" KPIs with "conventional" ones using pre-defined operators. Jane juxtaposes comparable results from several companies for a comparative analysis.

Joan's and Jane's use (in particular the latter) of X's ICT system broaches some unique requirements that differentiate our use case from other financial applications:

1. Semantic and syntactical discrepancies abound among individual reports, even with authoring support.
2. Cross report analysis lacks tooling.
3. Analysis is largely single faceted whereas financial applications become increasingly multiple faceted.
4. Financial reports with release and audit latency fail to give timely results.
5. The performance of one company cannot be easily judged in isolation.

4 The *LFD*r Platform

*LFD*r implements the envisaged Bank X's ICT system. In this section, we report the main data model and system architecture.

4.1 A model to capture facts about companies

Where our main concern is to build a common analytic and visualisation layer over several heterogeneous data sources, we formulated the hypothesis that RDF technology and Linked Data bridge the gaps among separately maintained data silos. In order to observe the integrity of XBRL-annotated data/reports, an ontology was defined.

The model showed in Figure 1 follows best practices in ontology engineering [15] and reuses as much as possible existing vocabularies. Each fact captured in the report is transformed into an Observation of a Concept concerning a Company. This domain model uses the well-known n-ary relation design pattern [9] and is able to handle rich numerical statements. An exemplar instantiation of

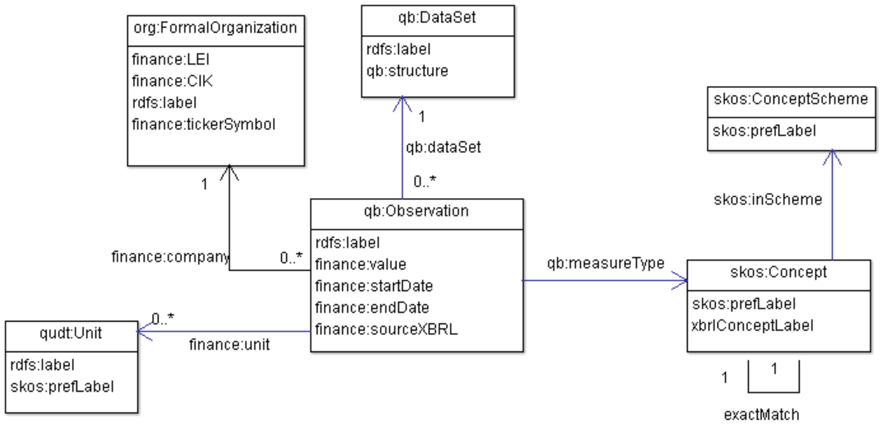

Fig. 1. UML model for the *LFD*r ontology

the domain model is the observation of "Yahoo! Inc."'s "Gross Profit" between "2009-04-01 and 2009-06-30" at a value of "860,444,000 USD". Yet another example is the observation of the "Number of Employees" of "Apple Inc." in "2012", which has a total value of "72,800" excluding subsidiaries.

4.2 System Architecture

The architecture of *LFD*r is depicted in Figure 2, which gives an overview of the data sources, essential components, and the two types of end users (i.e. Joan and Jane respectively). Key *LFD*r components include modules for (1) importing and pre-processing data, (2) storing and querying the data, and (3) providing innovative services/interfaces to end users.

Data preparation module Depending on the types, data ushered into *LFD*r go through different preparation routes. For RDF-ready data, e.g. LOD data sets, a dump can be downloaded and bulk-loaded into the system. Sources such as *US SEC* corporate ownership and *US Census* offer SPARQL end-points through `www.rdfabout.com`, which can be incorporated on-demand into the system. For instance, historical stock price is retrieved using queries with start and end dates.

For (semi-)structured data that are not available in RDF format, pre-processing is necessary. There is a long-lasting debate on the treatment of tabular data. We again adopted an on-demand approach: the original data schemata are observed where LEI/CIK systems and DBpedia based data reconciliation is utilised to align database and ontology instances.

Data storage module Due to performance consideration, *LFD*r (on-demand) caches data used in company performance analysis. Given the sheer size of the

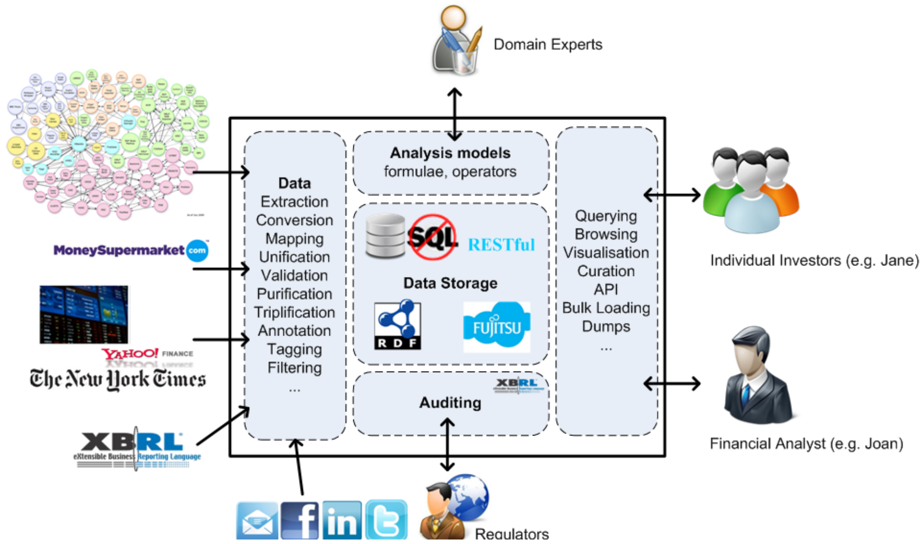

Fig. 2. *LFD*r system architecture

data, a distributed ordered Key Value Store (KVS) with enhanced range query and data locality is used in the underlying RDF storage [8]. Triples are stored as keys of the key-value pairs while the value part is reserved for triple metadata, e.g. provenance trace, access control policy, caching, etc. Data replication and fault tolerance is achieved as follows. For an RDF triple, denoted as $\langle s, p, o\rangle$, two replicates are created in the forms of $\langle p, o, s\rangle$ and $\langle o, s, p\rangle$, accommodating range queries over any combinations of triple elements. The list of triples are segmented where data segment locality (in terms of which data server) are decided based on co-occurrence patterns of triples in relation with other segments.

User and programming interactions *LFD*r interfaces with both human users and applications. The web-based user interface is illustrated in Figure 3. For end users, comparative results of the selected companies are centred on the web interface where users can decide whether to drill down into individual KPIs or roll-up to acquire an overview. Interaction with the data storage and analytic modules is through a web service interface fully compliant with RESTful principles [5]. Programmable interface of *LFD*r is materialised by Fujitsu RESTful implementation that standardises the modelling, exposure, and reference of data items in the RDF data repository.

5 Exploring Linked Data

XBRL aims to have a global and standard format while LOD facilitates the mapping between different XBRL taxonomies and provides the benefit of mashing-up financial reports with other types of data concerning the subject companies.

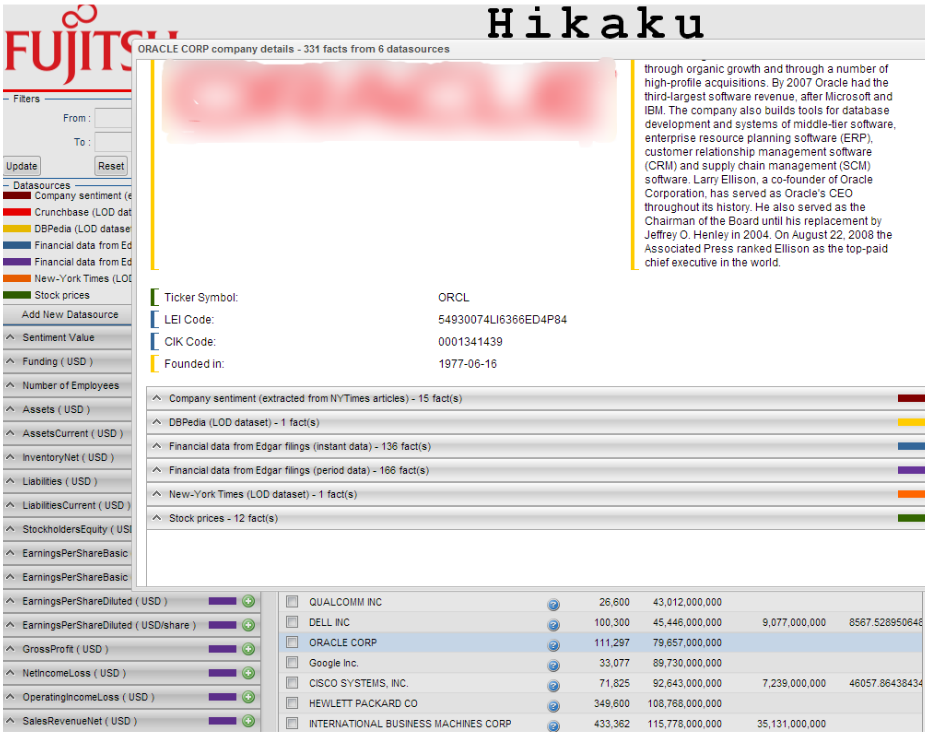

Fig. 3. *LFDr* UI: mash-up multiple data sources

5.1 Complementing Official Reports

Numeric figures in financial reports can be unintentionally and/or deliberately manipulated to present a false and misleading image of a company (*c.f.* the recent scandalous acquisition of Autonomy by Hewlett-Packard [2]). Even though such incidents cannot be entirely avoided, incorporating other sources of data can promote informed decision making and minimise potential risks and mistakes due to a lack of transparency. Typical public data that can compliment financial reports include stock market data, digitised mass media coverage, mailing list/online bulletin board systems (BBS) and the emerging social media. The use of public data is based on the following observation:

1. Official financial reports are normally published quarterly (*aka* 10Q) and yearly (*aka* 10K). They normally lag behind media coverage of major events concerning the subject company.
2. Official financial reports tend to be summarising over a long period of time and may not reflect the stock price fluctuation at given time points in that period.
3. More and more users or customers start to share their opinions about a product, a service or a company in channels other than customer services. Though the dynamics are not fully understood, we witnessed boycott/promotional activities campaigned through social media that have strongly affected a company's performance in real-world.

Choosing the most appropriate data sets can not only impinge the scope and accuracy but also system performance in terms of query execution time and memory consumption. The data sets being considered by *LFD*r can be grouped into several categories.

LOD data sets When comparing financial performance, one needs to cover a wide spectrum of aspects of corporate entities. Even though data sets published on LOD cloud may not explicitly bear a "finance" label in their titles, they can be of great assistance in discovering relevant information which is otherwise hard to access. The following LOD data sets are chosen at this stage.

- *DBPedia* is used for general company data such as logo and location, as well as KPIs such as the number of employees, revenue assets, equity, net income, etc. We also compose company's subsidiaries out of DBpedia data. The quality of data varies. Hence, data collected from DBpedia are cross-validated and complimented with those from other sources.
- *Linked Crunch Base* is a free database of technology companies, people, and investors. From CrunchBase, we retrieve such data as funding, competitors, company acquisitions, main people in charge, and products. It allows us to identify similar and comparable company profiles. For instance, companies with similar size, products and competitors can be grouped together and recommended for performance comparison.

– *Linked New-York Times*, as of 13 January 2010, has published approximately 10,000 subject headings as linked open data[5]. It compliments company profiles compiled from the above two sources.

Social media Nowadays a lot of valuable information concerning a company's public image and subjective assessment is available from popular social networks. We incorporate this type of data as well (focusing primarily on Twitter). Automatic data/knowledge extraction from social media is challenging, due to the noisy and dynamic nature of social media data [1]. Many companies offer sentiment analysis services with various licensing costs. Sentiment analysis research is beyond the scope of *LFD*r, where we take advantage of off-the-shelf offerings. Initially, we take into account only feeds from pre-defined sets of key accounts consisting both influencers and followers in order to assess the relative amount of useful content we can get. Generic and simple sentiment analysis tools (*c.f.* tools surveyed in [13]) are used with the assumption that the sheer size and diversity of data can offset intrinsic inaccuracy of simple solutions.

Mess media Mess media coverage provides more up-to-date information of a company and in many occasions leads/misleads the market on a wide and profound basis, e.g. causing stock to brief plunge or rise. *LFD*r reflects this through sentiment analysis of new articles. *NY Times* and *Yahoo! Finance* APIs are used to gether news articles and commentaries of a specific company. Sentiment scores are then computed with off-the-shelf tools/services and accumulated to reflect a company's mass media image.

5.2 Finding the Subsidiaries

Though required by transparency regulations, public companies are still coy in revealing their full corporate structure. It is much worse in case of private companies who are not bound by law to disclose their subsidiaries. How to capture a comprehensive picture summarising the performance of not only the main entity but also its subsidiaries, therefore, becomes a challenge. There are company databases, most of which are commercial ones with high licensing cost or maintained by university libraries with access limited to library members only. Consuming such data and linking them to other LOD data sets might become problematic, as their availability and accessibility cannot be guaranteed. We experimented several sources of uncovering the corporate structure.

– DBpedia defines two overlapping properties, namely `parentCompany_of` and `parent_of`, indicating the relationships between parent and subsidiaries. For instance, DBpedia defines "Fujitsu Ten", "Fujitsu Consulting India", and "Amdahl Corporation" as child companies of "Fujitsu Limited". Values of these two properties are consolidated and denoted as $\mathcal{S}_{dbp1}$.

[5] `http://data.nytimes.com/`

– DBpedia also reveals parent-child relationships in comment/abstract with a natural language pattern "*(A| |'is') a subsidiary of P*". Exhaustively checking the abstract/comment of all DBpedia organisational entities is not feasible. We narrow down the scope to DBpedia *categories* related to the name of the target parent company. For instance, when collecting subsidiaries of Fujitsu, we query the categories containing the word "Fujitsu" under the broader "DBpedia:*Categories named after conglomerate companies*". We then query all the listed entities in that category, retrieve the abstract/comments, and scan for the pattern. This gives us new subsidiaries such as "Fujitsu Technology Solution", "HAL Computer System", and "Nifty Corporation" which are saved in $\mathcal{S}_{dbp2}$.
– For IT companies, CrunchBase can be another useful source. Very rarely CrunchBase gives direct subsidiary relations. More often, such information can be extracted from the content of "Milestones" section using natural language pattern "*P acquired A*". Again, for "Fujitsu", we are able to extract "PFU Limited" and "KAZ Group" as Fujitsu subsidiaries (denoted as $\mathcal{S}_{cb}$).

We are aware that this may not compute a complete subsidiary list, for which human curated commercial corporate databases are always a better option. Alternatively, one can analyse company websites, annual reports, or Wikipedia pages to extract a complete list. It is our contention that before such resources are made LOD ready and/or open to general public, the present approach offers the most cost-effective solution.

6 Discussions and Conclusion

In this paper, we introduced the *LFD*r framework aiming at data integration and performance comparison of corporate financial data. From the development and deployment of *LFD*r, we learnt interesting lessons (on both technical and socio-technical fronts) to share with the community.

6.1 Incorporating external data

On a more technical front, when developing *LFD*r, we made several design decisions in favour of XBRL or LOD. These are highlighted in this section.

Annotation versus conversion There is a long-standing debate over the treatment of non-RDF data sources. "Conversionists" advocate a full translation into RDF for seamless integration while "annotationists" support "let tables/mess be tables/mess" philosophy. Our experience show better acceptance and lower induction curve of the latter approach. Even though "openness" and "transparency" have been widely advocated and are considered the foremost principle of LOD, we recognise the necessity of observing the integrity and independency of certain data sources and the technical challenges of full RDF translations from certain data formats. In *LFD*r, on-demand annotation is widely used for non-RDF data sources. It sustains the original data ownership, simplify data safety and data provenance issues, and allows incremental data collection for better system performance. In the meantime, it ensures the minimum necessary exposure for facilitating LOD applications and thus encourage LOD take-up.

Resource reference In *LFD*r, public data are aligned with proprietary data. The boundaries between them should be well-defined. Depending on the application and data sources, one of the following two approaches will be taken:

URI concatenation If the foreign data are crawled and harvested into local repository, the foreign URI can be attached to the end of local identifier. For instance, data about Hewlett-Packard retrieved from NY Times is appended to the local base URL as:

```
http://foo.bar/resolve/http://data.nytimes.com/N73356871238605776012
```

Following this approach, we keep up-to-date with foreign data sources and more importantly do not assume any ownership of the data.

Internalising with equivalence If semantic equivalence is confirmed and if local assertions about external resources are necessary, collapsed equivalence should be applied. We encourage exploiting `owl:sameAs` for confirmed equivalence and `skos:broader`/`skos:narrower` and/or `skos:related` in case of local assertion. For instance, the follows illustrate the equivalence between a local instance and a foreign instance and the case when local instance is more restrictive.

```
<http://foo.bar/organisation/xxx>
  owl:sameAs <http://data.nytimes.com/N73356871238605776012>;
<http://foo.bar/organisation/yyy>
  skos:narrower <http://data.nytimes.com/N73356871238605776012>;
  skos:related <http://data.nytimes.com/N73356871238605776012>;
  local:produce <http://foo.bar/product/y>.
```

The potential risk of leveraging full-fledged semantic equivalence is two-fold: i) internalising external resources assumes an ownership of the latter and thus leads to unexpected provenance and privacy issues; ii) further updates of the internalised resources can create inconsistence between local and remote assertions that violate the equivalent constraints.

6.2 Preliminary Evaluation

The evaluation of *LFD*r is and will be focusing on both quantitative and qualitative aspects. Quantitative evaluation provides a numerical means for feasibility and comparative studies of the technologies employed in *LFD*r. This includes tools for data/knowledge extraction, sentiment analysis, semantic alignment, etc. We also resort to the qualitative evaluation to assess the general applicability and industry-relevance of the platform. Participatory development philosophy is applied to ensure early engagement of target users. Multiple formative usability studies and one summative study were/will be conducted at critical milestones ensuring a continuous integration of user feedback into the prototypical system, leading to better alignment with domain requirements, quick roll-out to early adopters, and shortened product cycle.

The first formative study was performed in XBRL26[6]. In general, the concept of LOD-enhanced financial reporting is well received by the XBRL community. *LFD*r demonstrated functionalities that cannot be achieved by either LOD or XBRL if used alone, leading to very positive and interesting feedback, such as "how can I publish my reports to LOD cloud?" and "has LOD been standardised by any industries?". Technical improvements focused on two leading issues that indicate the potential directions of system enhancement. Firstly, when incorporating LOD sources, one needs to ensure the trustworthiness of data. At this moment, data sets are reviewed and handpicked individually. When more data sources are deemed necessary, an automatic data authentication method needs to be conceived. Moreover, the authentication results should be reflected quantitatively in the KPI calculation and made transparent to users consuming the KPIs. Secondly, as the financial world seems to increasingly revolve around LEI, generating automatic alignment between LOD (more specifically DBpedia entities) and LEI should provide better value for both communities. In the first prototype, this alignment is based on string distance between LEI labels/prepherial data and DBpedia entity properties. More sophisticated and accurate alignment should be used to relief human experts.

6.3 Future Work

The crux of our future work lies in the improvement of current prototype to reflect feedbacks from the first formative study and extended quantitative studies of employed technologies in the financial domain. More evaluations have been scheduled. With XBRL community already becoming the initial adopter, reach-out to other financial communities can be facilitated through the XBRL "channel".

Future plans also include an improved *LFD*r's financial dashboard with features such as 1) time series analysis (e.g. "Fujitsu's performance since the latest tsunami disaster.", "is Fujitsu performing better this year comparing with other Japanese companies?"), 2) a data set explorer and quality-checker (e.g. "FT.com with a quality score of 0.8 and a trust score of 0.75"), and 3) user-defined KPI validation (e.g. "combining sentiment score and total number of employees does not make sense.").

[6] `http://conference.xbrl.org/`

Acknowledgements

The present work is a joint efforts of Digital Enterprise Research Institute (DERI) of Ireland, Fujitsu Ireland, and Fujitsu Laboratories Ltd.

References

1. G. Barbier and H. Liu. Data mining in social media. In C. C. Aggarwal, editor, *Social Network Data Analytics*, pages 327–352. 2011.
2. BBC. Hewlett-packard accountants sued over autonomy purchase, 2012. accessed on 04 March 2013, `http://www.bbc.co.uk/news/business-20538392`.
3. D. Burdick, M. A. Hernández, H. Ho, G. Koutrika, R. Krishnamurthy, L. Popa, I. Stanoi, S. Vaithyanathan, and S. R. Das. Extracting, linking and integrating data from public sources: A financial case study. *IEEE Data Eng. Bull.*, 34(3):60–67, 2011.
4. P. Delaney and F. A. S. Board. *Generally Accepted Accounting Principles*. J. Wiley & Sons, 1986.
5. R. T. Fielding. *REST: Architectural Styles and the Design of Network-based Software Architectures*. Doctoral dissertation, University of California, Irvine, 2000.
6. R. García and R. Gil. Triplificating and linking xbrl financial data. In *Proceedings of the 6th International Conference on Semantic Systems*, pages 3:1–3:8. ACM, 2010.
7. A. Harth, B. Kämpgen, and S. O'Riain. Financial information observation system (FIOS), 2012. http://fios.ontologycentral.com.
8. B. Hu, N. Carvalho, L. Laera, and T. Matsutsuka. Towards big linked data: a large-scale, distributed semantic data storage. In *Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services*, IIWAS '12, pages 167–176, New York, NY, USA, 2012. ACM.
9. N. Noy, A. Rector, P. Hayes, and C. Welty. Defining n-ary relations on the semantic web. *W3C Working Group Note*, 12:4, 2006.
10. T. Omberg, S. Sakr, S. Sethi, B. Murrell, and A. Popken. *Key factors shaping financial reporting: The decade ahead*. Deloitte, 2011.
11. S. O'Riain, E. Curry, and A. Harth. Xbrl and open data for global financial ecosystems: A linked data approach. *International Journal of Accounting Information Systems*, 13(2):141 – 162, 2012.
12. S. O'Riain, A. Harth, and E. Curry. Linked Data Driven Information Systems as an Enabler for Integrating Financial Data. In A. Yap, editor, *Information Systems for Global Financial Markets: Emerging Developments and Effects*, chapter 10, pages 239–270. IGI Global, 2012.
13. B. Pang and L. Lee. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, Jan. 2008.
14. B. Ryan, R. W. Scapens, and M. Theobald. Thomson London, 2002.
15. Y. Sure, S. Staab, and R. Studer. Ontology engineering methodology. *Handbook on Ontologies*, pages 135–152, 2009.
16. J. Zaino. Linking xbrl to rdf: The road to extracting financial data for business value, September 2012.