Distributed Data and Ontologies: An Integrated Semantic Web Architecture Enabling More Efficient Data Management

Oliver Browne
State Street Advanced Technology Centre, Cork University Business School, University College Cork, Ireland. E-mail: Oliver.Browne@ucc.ie

Philip O'Reilly*
State Street Advanced Technology Centre, Cork University Business School, University College Cork, Ireland. E-mail: Philip.OReilly@ucc.ie

Mark Hutchinson
State Street Advanced Technology Centre, Cork University Business School, University College Cork, Ireland. E-mail: M.Hutchinson@ucc.ie

Nenad B. Krdzavac
State Street Advanced Technology Centre, Cork University Business School, University College Cork, Ireland. E-mail: nenad.krdzavac@ucc.ie

Regulatory reporting across multiple jurisdictions is a significant cost for financial services organizations, due to a lack of systems integration (often with legacy systems) and no agreed industry data standards. This article describes the design and development of a novel ontology-based framework to illustrate how ontologies can interface with distributed data sources. The framework is then tested using a survey instrument and an integrated research model of user satisfaction and technology acceptance. A description is provided of extensions to an industry standard ontology, specifically the Financial Industry Business Ontology (FIBO), towards enabling greater data interchange. Our results reveal a significant reduction in manual processes, increase in data quality, and improved data aggregation by employing the framework. The research model reveals the range of factors that drive acceptance of the framework. Additional interview evidence reveals that the ontological framework also allows organizations to react to regulatory changes with much-improved timeframes and provides opportunities to test for data quality.

*Corresponding author

Received July 27, 2017; revised June 29, 2018; accepted September 15, 2018

 • Published online January 22, 2019 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.24144

Introduction

We are living in an era where technology is having a significant impact on financial services, characterized by highly distributed and heterogenic services. Yet the multitude of data storage systems means that efficient retrieval and querying of relevant data to answer business queries is problematic. Moreover, this problem is worsened by the lack of agreed data standards. Ontologies (Horrocks, 2008) are key ingredients for enabling the development of Semantic Web services that interpret heterogenic financial data.

Regulatory bodies, such as central banks, require institutions to deliver reports to illustrate that their activities comply with regulations. Many problems occur in the creation of these regulatory reports, with the lack of integrated data coupled with a lack of a standard data dictionary meaning that existing processes require significant manual intervention to create reports (Tripathy & Naik, 2014). The inflexibility of many implemented systems within financial services means that in many cases they can only be used to create one type of regulatory report and are often shelved once developed. Maintenance of the system, inflexibility to changes in accounting rules and regulations, and manual processes involved in validation of outputs are all reported issues (Chen & Sheldon, 1997). Furthermore, data provided to regulatory bodies are also open to interpretation. For example, current regulations often require regulator confirmation to identify treatment of many instruments. These challenges have spurred the need for the development of a standardized language across financial instruments and term sets, such that there is little room for interpretation and the regulator receives transparent and comparable data from all institutions (Bennett, 2013).

This article investigates the potential role that ontologies can have on both regulatory reporting initiatives and data management strategies. Financial Industry Business Ontology (FIBO) is an industry standard being developed by the Enterprise Data Management Council (EDM Council), an industry association, to standardize language across financial instruments and institutions. This research makes a number of contributions through developing and empirically validating an ontology-based architecture with significant business benefits. These include: the removal of requirements to replace Legacy Information Systems (LIS) to meet regulatory reporting requirements, more efficient and flexible data querying, and the dynamic reporting of data quality issues. Our research also extends the FIBO to incorporate previously unmapped equities and bonds.

A novel ontology-based Information Technology (IT) architecture was developed towards evaluating how ontologies can interface with distributed data sources. The viability of this data management framework was tested through the ingesting of industry data, provided by State Street Corporation. An application programming interface (API) was developed to enable users to interact with the system. The quality of the data management framework was evaluated using both a survey and a research model that captures user satisfaction and technology assessment by the reporting team, senior management within State Street Corporation, and by members of the EDM Council. This was also augmented with semistructured interviews. Multiple benefits were reported including; increased data quality, opportunities to automate manual processes, faster access to relevant data, and improved risk management. The research has significant implications for enterprise-level data management strategies. It provides empirical evidence of the potential of ontologies in relation to improved efficiencies regarding regulatory reporting and, in a broader context, an organization's data quality and approach to data management.

Theoretical Background

The global financial crisis has resulted in a large increase in regulatory reporting requirements within the financial services sector (Akhigbe, Martin, & Whyte, 2016). This sector has seen a myopic view of software investment since the 1970s (Arner, Barberis, & Buckley, 2015), with no system intending to survive longer than 15 years (Matei, 2012). Many of these systems remain in use long after the originally planned lifecycle and are often referred to as legacy information systems (LISs) (Bisbal, Lawless, Wu, & Grimson, 1999). The financial services sector, in particular, has traditionally chosen to maintain existing systems over developing new ones (Comella-Dorda, Wallnau, Seacord, & Robert, 2000). This has led to a data landscape characterized by a myriad of systems as organizations fear disruption to their core platforms. The existence of multiple systems with no standard data format or common definitions has led to poor quality control and duplicate data (Madnick & Zhu, 2006).

Investigations by regulatory bodies found that the financial industry had unacceptable levels of interdependency within the system itself (Ellis, Haldane, & Moshirian, 2014; Ye, Wang, Yan, Wang, & Miao, 2009). Systemic risk was difficult to quantify due to the poor quality and malleability of information within these financial institutions. These institutions were also found to have an unacceptable level of manual processes in the production of financial reporting data. Towards effecting change, the Bank for International Settlements issued the Basel Committee on Banking Supervision directive (BCBS) 239 (BIS, 2013), which aims to force organizations to automate reporting processes and reduce the dependency on manual input.

Several problems arise from the perspective of financial institutions with respect to the BCBS 239 directive. Legacy systems are a significant stumbling block to meeting these regulatory requirements, and the costs of resolving the issues with these systems are significant. In the past, three options have been presented when it comes to decisions about legacy systems: the system can be wrapped, maintained, or migrated (Bisbal et al., 1999). Yet theoretically there is a fourth option—adopting an ontological layer over the data using mapping (Baader, Calvanese, McGuinness, Nardi, & Patel-Schneider, 2003), which can avoid much of the costly methods that are being explored to enable compliance with BCBS 239. This approach involves mapping data from the originating systems to a standard ontology vocabulary. Many LISs suffer from a lack of documentation or original developers of the system (van Geet, Ebraert, & Demeyer, 2010). This process of mapping discovery has been identified as a difficulty where documentation for the system does not exist (Nallusamy, Ibrahim, & Naz, 2011; Noy, 2004). Yet once this mapping is complete, it can be extended over the ontology using extended relations and axioms.

Ontologies have seen widespread development and adoption in recent years, particularly in life science disciplines (Huang, Sherman, & Lempicki, 2009; Roy, Kucukural, & Zhang, 2011). However, applications in other industries have lagged. One potential explanation is the prevalence of low-quality data in organizations (Nagle, Redman, & Sammon, 2017). FIBO is largely at the implementation step of the building stage within the ontology lifecycle (Kayed, Hirzallah, Al Shalabi, & Najjar, 2008). FIBO has multiple modules containing aspects of the financial industry, including: securities, business processes, business entities, derivatives, and so forth. The current releases and utilized modules of FIBO are explored in the methodology section of this article. While many of these modules are released, many are still in development. We

DOI: 10.1002/asi

adopt many of the unpublished modules in this article. Our research develops and contributes to the standard. This is significant, as the collaborative standard depends on input from industry participants and researchers.

With the multitude of disparate systems in the financial services sector, data are often stored in different relational databases, as well as nonrelational files, such as CSV. This structured and unstructured data requires innovative ways to confirm data quality and aggregation. Towards implementing a web service, this article aims to utilize the integration of XML/XSLT (Extensible Stylesheet Language Transformations) and Semantic Web technologies to implement services for financial reporting. We combine ontology-based data access (OBDA), a set of mathematical tools and implementation of these tools as software. With this method, a user can access data stored in a database by using concepts stored in ontologies (Ontop API) (Rodríguez-Muro, Kontchakov, & Zakharyaschev, 2013) and materialize nonrelational data into a triple store. This allows data users to query multiple databases and aggregate data automatically. By investigating the use of these technologies, we explore alternatives to complete system redesign. This is an important topic, as financial institutions are often averse to disrupting existing systems. Ontologies may allow new uses and improvement without significant upfront investment. Furthermore, we utilize open-source software to illustrate robust alternatives to proprietary software (Neumann, 1999), while also allowing us to modify the source code to extend software capabilities to meet our needs.

Methodology

Our methodology consists of two distinct steps. The first step was to design a data management framework to utilize the ontology standard over the source data. We refer to this data management framework as the Global Fund Reporting Ontology (GFRO), described in detail in the next subsection. The second step in our methodology is to empirically evaluate the improvements in implementing the GFRO system and the factors that affect its acceptance. This involves a survey instrument and a research model, augmented with interviews. The description of the survey, research model, and interviews is described in the second subsection below.

Data Management Framework

Towards enabling the evaluation of FIBO for financial reporting, we develop a data management framework (GFRO). The framework can be divided into four modules, as shown in Figure 1. These modules may be briefly described as follows: module A contains the source data, B contains FIBO and other enterprise-level ontologies with mappings to the source data, C enables the source data to be converted to a query enabled format using the ontologies, and finally, D imposes the logic and queries through an API. These are described in further detail below.

Module A contains a subset of structured and unstructured financial data relating to bonds and equities for evaluating and testing our proposed framework. These data were stored in relational databases, Excel, and CSV files. The data set was provided by State Street Corporation. Nonrelational data sources consist of data that are the result of calculation methods and formulas executed in Excel, using data from databases as input to same. For example, a column in an Excel file that denotes instrument holding value is calculated as a multiple of shares held multiplied by market value at a point in time. Values used in this calculation are obtained from relational databases.

Module B bridges Modules A and D by establishing T-mappings (Calvanese, Cogrel, Komla-ebri, Kontchakov, &

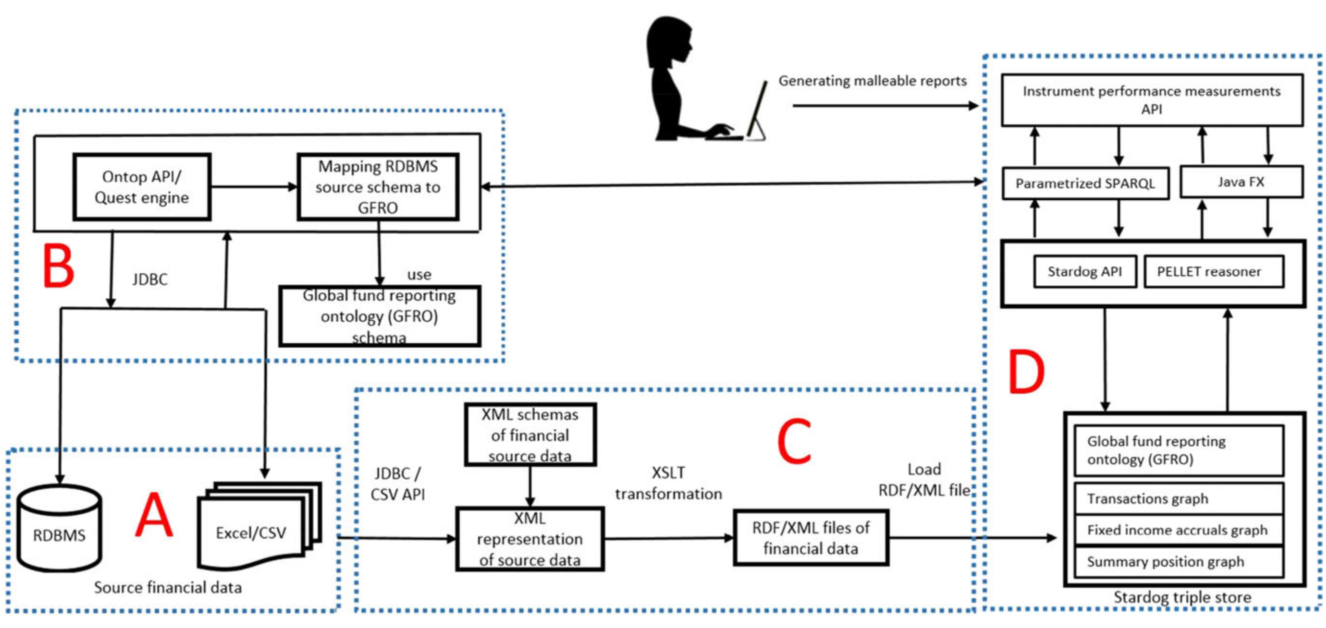

FIG. 1. Data management framework components. [Color figure can be viewed at wileyonlinelibrary.com]

DOI: 10.1002/asi

TABLE 1. List of FIBO modules and life-cycle stages.

FIBO Module	FIBO module abbreviation	FIBO release	Kayed et al. (2008) life cycle stage	Utilized in GFRO Y/N	Published Y/N
Business Entities	*be*	Yellow	Maintenance	Y	Y
Business Processes	*bp*	Red	Manipulation	Y	N
Corporate Action Events	*cae*	Red	Formalization	N	N
Collective Investment Vehicles	*civ*	Pink	Conceptualization	Y	N
Derivatives	*der*	Red	Implementation	Y	N
ETC - Other	*etc*	Yellow	Implementation	N	N
Financial Business and Commerce	*fbc*	Yellow	Manipulation	N	Y
Foundations	*fnd*	Yellow	Maintenance	Y	Y
Indices and Indicators	*ind*	Yellow	Maintenance	Y	Y
Loans	*loan*	Red	Formailzation	Y	N
Temporal Terms	*md*	Red	Conceptualization	Y	N
Securities	*sec*	Pink	Conceptualization	Y	N

Lanti, 2015) between the relational schema of source data and the GFRO dictionary. GFRO contains FIBO as the upper-level ontology (Asunción, Mariano, & Oscar, 2004) where common (industry standard) objects are stored. T-mappings relate data in source systems to the ontology terminology. Relational data necessary for financial reporting are queried over GFRO by using SPARQL. The process starts by sending a parameterized query to Module B. Before executing the SPARQL query, the Quest inference engine (Rodrıguez-Muro & Calvanese,) configures the environment by loading defined t-mappings and GFRO. The Quest engine is a core of the Ontop framework translating a SPARQL query submitted by an end user into the appropriate SQL query that is executed by a relational database engine (Calvanese et al., 2015). During execution of SPARQL queries, the Quest inference engine checks for data consistency (Rodríguez-Muro et al., 2013). The query result is delivered to instrument performance measurements that are part of Module D.

In mapping between relational database schemas and ontologies in Ontop (Calvanese et al., 2015), end users should specify how to construct from the concrete values in the data sources the (abstract) objects that populate virtual ABox with financial data; an ABox is an assertions component containing facts associated with a set of ontological terms allowing inference functionality (Kontchakov, Rodriguez-Muro, & Zakharyaschev, 2013). Kontchakov et al. (2013) show that Ontop provides high query scalability over huge amounts of data where end users benefit from not needing to know where and how financial data are stored. Furthermore, they demonstrate that completeness in OBDA is guaranteed by the utilization of inference engines and rewriting query techniques.

FIBO is available in multiple Semantic Web languages, of which we chose OWL DL to adopt. FIBO is a conceptualization of the financial domain and as such is a software artifact in knowledge representation. Therefore, our selection of OWL DL satisfies a need to create a vocabulary and generate inferences (Almeida, 2013). On the other side, materialization offers query and inference on financial data with respect to ontologies developed by very expressive ontology languages such as OWL DL (McGuinness & Harmelen, 2004).

Within Module C, nonrelational data are converted into XML files that conform to the XML schema and later transformed into Resource Description Framework (RDF) graphs by using XSLT. RDF is a data model for objects and relations between them (subject, predicate, and object), which can be represented in an XML syntax (McGuinness & Harmelen, 2004). R2RML, a language for expressing customized mappings from relational databases to RDF (Das, Sundara, & Cyganiak, 2012), is not used within the framework, as XSLT-based transformations showed better performance and flexibility. XSLT allows complex financial mathematical formulae to be implemented by using financial rules like yield calculations. In the XSLT implementation we use GFRO classes and properties, so the resulting RDF graphs should be consistent with respect to the GFRO schema. To generate RDF graphs, in XSLT implementation, the values of XML tags are used in order to generate appropriate RDF nodes and relations between them. The root tag in all XML files is denoted with <*root*>. For example, the attribute value of <*fund_id* > *(Fund identifier)* tag is mapped into objects (URIs) that are instances of the *fibo-civ-fun-civ: Fund* class, where *fibo* identifies the ontology and *civ-fun-civ* identifies the submodules in which the tagged class (Fund) is contained. A list of FIBO modules utilized and their lifecycle stage can be seen in Table 1. Each <*asset_id* > *(Asset Identifier)* is mapped to a URI that is an instance of a subclass of the *fibo-fbc-fi-fi:FinancialInstrument* class but under different conditions. For example, to transform the value of <*asset_id* > tag into URI that is an instance of *fibo-der-fx-spots:FxSpotContract*, a conjunction of values of set parameters (*investment type*, *fund*, and *asset identifier)* are used as conditions during this transformation and to uniquely identify each instrument. These graphs are subsequently loaded into the Stardog triple store. A triple store is a specially designed graph database used for storing RDF data.

Module D provides a set of Java interfaces and their implementations to be used by end users. The Instrument Performance Measurements (IPM) API consists of two modules: a performance dashboard layer, and regulatory templates layer. The performance dashboard layer is a set of Java interfaces and classes that implement fund distributions; for example, geographical distribution instruments, and sector distributions of instruments. This layer uses the Java FX (Java, 2018) library to generate reports of the fund-reporting template as a dynamic dashboard.

TABLE 2. Survey respondent's demographics.

	Number	Percent
Organizational level:		
Analyst	17	33
First-level Supervisor	11	21
Middle Management	15	29
Senior Management	9	17
Functional Area:		
Accounting	26	50
Finance	2	4
Information Systems	10	19
Other	5	10
Research and Development	9	17

Geographical distribution of instruments is calculated as the value of a funds instrument represented by each country. The same template is used to calculate industry sectors' distribution. The regulatory template includes instruments held within a fund and their performance over the previous reporting period including changes to holding amounts, yields, and prices. It is delivered to a user as an automatically populated Excel sheet. The template is provided by the Central Bank of Ireland (CBI). To populate this sheet, the IPM API implements business logic for calculating the yield of bond instruments and populate the sheet with results of yield calculations. To calculate yield, we use data that represent coupon rate and current market price in line with the guidelines for the CBI Money Market and Investment Funds (MMIF) report (CBI, 2016). The coupon rate is the interest rate payable on the bond. The current market price is the most recently recorded price available in the source data for the instrument itself. For variable coupons, the most recent rate is used. Data used in the IPM layer are queried from databases and graphs loaded into the Stardog triple store. The IPM layer communicates with data sources over a layer that implements parameterized SPARQL queries.

We combine the ODBA (Module B), and materialization of source data snapshots as graphs in Stardog triple store (Modules C and D). This approach is taken, as fund level reporting is more appropriately implemented using parameterized SPARQL queries over the GFRO dictionary that does not have expressivity beyond OWL QL (Kontchakov et al., 2013; Rodrıguez-Muro & Calvanese, 2012). An inference engine, embedded in Ontop (Rodrıguez-Muro & Calvanese, 2012), is able to provide query-answering inference for consistent reporting. To provide fund-level reporting, we do not implement axioms and rules on the top of GFRO that would require a reasoner for more expressive language than OWL QL (Rodrıguez-Muro & Calvanese, 2012).

To implement instrument-level reporting, we implement axioms and Semantic Web rules on the top of the GFRO ontology using OWL DL (McGuinness & Harmelen, 2004), and apply the Pellet inference engine in order to complete knowledge; that is, infer all missing relations between financial data sorted as an RDF graph in Stardog triple store.

Research Model to Empirically Validate the Framework

To assess the success of the GFRO implementation, we initially develop a survey instrument, utilizing Wixom and Todd's (2015) model and approach, to gather data on user satisfaction and acceptance of the new framework. This approach

TABLE 3. Selected multi-item constructs results.

Construct	Standard loading	Mean	Standard deviation
Completeness			
Provides me with all the information I need	0.96	5.44	1.37
Accuracy			
The GFRO produces correct information	0.88	5.73	1.15
Information quality			
In general, GFRO provides high-quality information	0.89	5.85	1.20
Accessibility			
GFRO makes information easy to access	0.95	5.79	1.43
Flexibility			
GFRO can be adapted to meet a variety of needs.	0.88	5.83	1.12
Integration			
GFRO effectively integrates data from different areas of the company	0.96	6.14	1.09
Timeliness			
It takes too long for to GFRO to respond to my requests. (RC)	0.98	5.10	1.48
GFRO provides information in a timely fashion.	0.98	5.33	1.31
Information satisfaction			
I am very satisfied with the information I receive from GFRO	0.94	5.60	1.17
Ease of use			
GFRO is easy to use.	0.98	5.50	1.22
Usefulness			
GFRO allows me to get my work done more quickly.	0.98	5.89	1.31
Using GFRO enhances my effectiveness on the job.	0.97	5.64	1.33
Attitude			
My attitude towards using GFRO is favorable	0.97	6.12	1.14

Note: Scale items are based on a seven point Likert-type scale (1 = "strongjy disagree," 7 = "strong]y agree"). $p < .05$.

TABLE 4. Reliability of instruments and correlations of latent variables.

	α	ACCE	ACCU	ATTI	COMP	EASE	FLEX	FORM	INFQ	INFS	INTE	RELI	SYSQ	SYSS	TIME	USEF
Accessibility	0.94	**0.94**														
Accuracy	0.91	0.52	**0.92**													
Attitude	0.86	0.76	0.45	**0.88**												
Completeness	0.91	0.70	0.66	0.56	**0.92**											
Ease of use	0.95	0.67	0.52	0.71	0.52	**0.96**										
Flexibility	0.93	0.68	0.60	0.64	0.60	0.70	**0.93**									
Format	0.89	0.72	0.46	0.74	0.70	0.60	0.56	**0.91**								
Info. quality	0.97	0.79	0.72	0.76	0.70	0.78	0.77	0.72	**0.97**							
Information sat.	0.97	0.74	0.75	0.64	0.82	0.53	0.65	0.70	0.75	**0.98**						
Integration	0.98	0.79	0.68	0.70	0.63	0.73	0.74	0.59	0.84	0.72	**0.98**					
Reliability	0.88	0.66	0.50	0.67	0.40	0.61	0.59	0.50	0.57	0.59	0.66	**0.90**				
System quality	0.97	0.81	0.69	0.77	0.63	0.80	0.76	0.73	0.90	0.74	0.85	0.74	**0.97**			
System sat.	0.94	0.70	0.67	0.77	0.62	0.73	0.63	0.74	0.79	0.69	0.72	0.62	0.76	**0.97**		
Timeliness	0.91	0.56	0.32	0.60	0.43	0.43	0.40	0.59	0.50	0.52	0.38	0.65	0.60	0.50	**0.92**	
Usefulness	0.94	0.81	0.71	0.74	0.64	0.76	0.72	0.67	0.91	0.71	0.86	0.60	0.89	0.78	0.51	**0.94**

Note: α measures the reliability of the instruments with all above the minimum threshold of 0.70. Diagonal elements in bold are the square root of the AVE, which should not exceed the intercorrelations for each construct.

allows us to assess system and information characteristics and the causal effects they have on ease of use and attitude. We adopt a mixed methods approach by augmenting this with qualitative analysis, as the team using the framework provides a relatively small sample size for quantitative analysis and we wish to avoid small sample bias (Dennis & Garfield, 2003; Venkatesh, Brown, & Bala, 2013).

A survey instrument is prepared following Wixom and Todd (2005), adapted to meet the GFRO framework. The adaptation of the survey instrument involved a review process by academics and practitioners. Following this process, we exclude the measurement properties of intention and currency. These measurements were deemed surplus to the scope of the research, as the API would become the

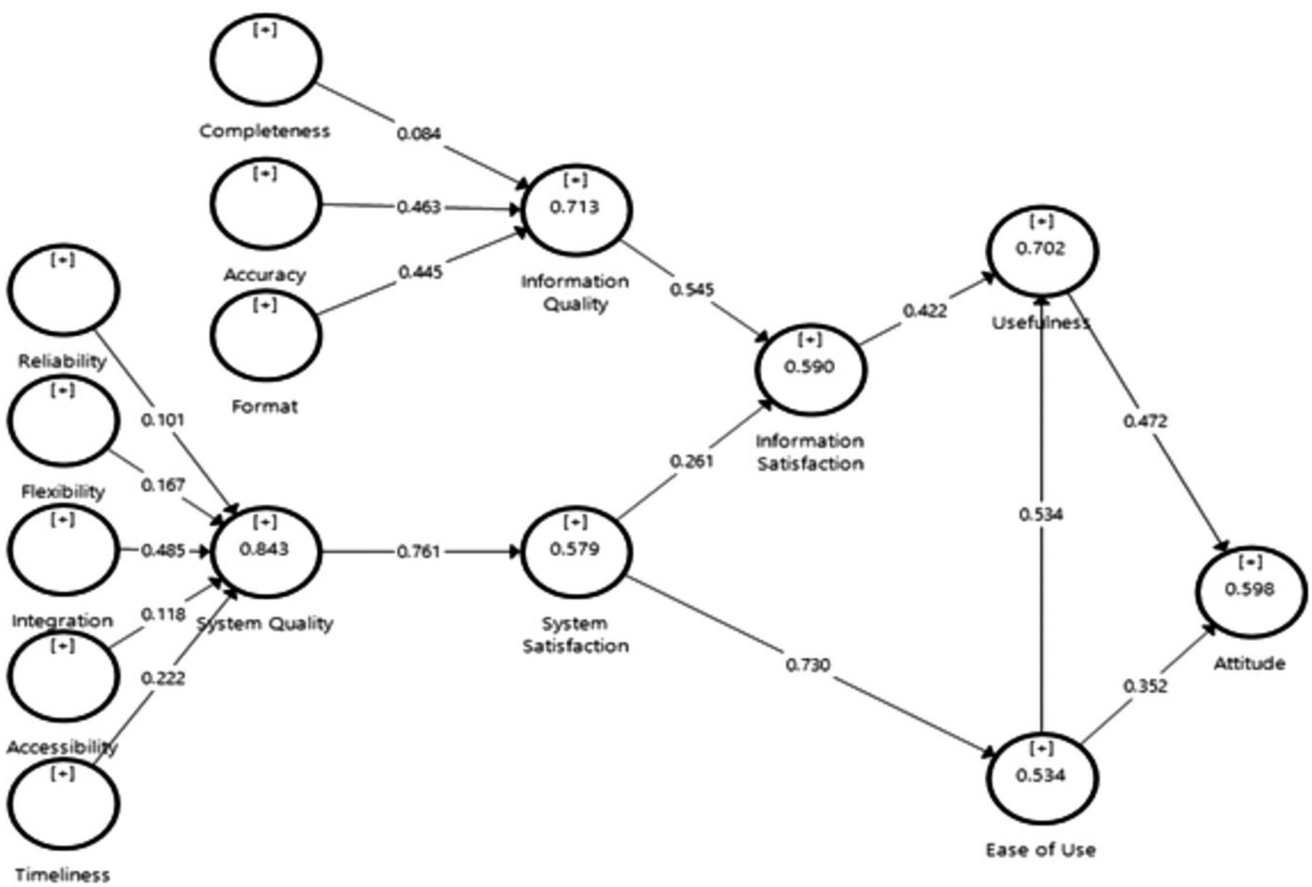

FIG. 2. Wixom and Todd (2005) structural model results.

DOI: 10.1002/asi

TABLE 5. Pellet reasoner identified bond anomalies.

Asset identifier	Accrual basis 1	Accrual basis 2	Accrual basis 3
1101010B	30/360	Actual/actual	—
0201010A	Actual/Actual	30/360	—
0420101B	Actual/Actual	30/360	—
0503010D	30/360	Actual/actual	30/365

sole method of producing regulatory reports and thus its use would not be optional. The data provided are related to static data; therefore, currency is irrelevant. All other structural model measurement properties were included. The survey instrument was confidential; no identifying personal information was gathered. Demographic survey questions were also removed, in line with privacy policies in place within the organization. Within the relatively small sample of respondents, the combination of demographic question answers would have negated the anonymity of the respondent.

Surveys were carried out with the team responsible for the processes that the GFRO automates as well as other individuals to whom the application was demonstrated using workshops and presentations. Survey instruments were distributed in both paper-based and digital formats. Paper-based responses were gathered by the researchers and amalgamated with digital responses for the overall model. Construct items were included in a random order and measured on a seven-point Likert-type scale (1 = "strongly disagree," 7 = "strongly agree"). Results of the survey instrument were then coded and analyzed using the Smart PLS software package (Lowry & Gaskin, 2014).

We use the methods of Wixom and Todd (2005) to analyze the quantitative data from the surveys. This approach allows us to assess improvements upon existing processes for reporting financial data. Wixom and Todd's (2005) research has verified the relationships between object-based and behavioral beliefs with attitudes. Therefore, allowing for the combination of the user satisfaction and technology acceptance models. Using this approach, we can assess system and information characteristics and the causal effects they have on ease of use and attitude. The research model was tested using partial least squares (PLS), which is highly suited for complex predictive models (Chin, 1998) and appropriate to a large number of constructs. Smart PLS 3 (Lowry & Gaskin, 2014) was used for the analysis of the paths within the structural model.

A number of semistructured interviews with senior figures in the organization were also completed. In total, there were 26 demonstrations of the framework and associated interviews. The duration of interviews was 30–45 min. Interview participants ranged in background including: IT, accounting, finance, and client relations. Participants also varied widely in organizational seniority from vice-president, senior vice-president, to C-suite executives. This exposure to a broad cross-section of employees allowed us to gain insights into the operational benefits of GFRO as well as the potential organizational benefits. Candidates were selected based on two criteria: their role in future development the GFRO framework, and their involvement in similar projects in the past within the organization.

Results

This section explores the impact that the GFRO framework had on the regulatory reporting data quality, timeliness, and integration of distributed data using the mixed methods empirical evidence outlined in the methodology. We begin with the results of the survey before providing estimates of the structural model. We then discuss findings from the semistructured interviews before providing a short discussion of the FIBO extensions, resulting from this research.

Survey Results

Empirical testing of the improvements provided by the GRFO framework is measured using an integrated research model approach as previously outlined. A survey instrument is constructed following Wixom and Todd (2005) and instrument items tested for discriminant validity. There were 52 completed responses to the survey (Table 2), including nine members of senior management.

These benefits of the ontological framework are evidenced by the results of the survey, with key constructs reported in Table 3. Construct items were measured on a seven-point Likert-type scale (1 = "strongly disagree," 7 = "strongly agree"). Respondents were particularly in agreement with the ability of the system to integrate data effectively from disparate sources (6.14) as well as the accuracy of the information produced (5.73). Respondents also agreed that the GFRO framework also improved employee's effectiveness (5.64) and allowing tasks to be completed more quickly (5.89). There was also significant agreement on the ability of the framework to provide information in a timely fashion (5.33) and provide high-quality information (5.85). Overall, there was significant support for the adoption of the framework and its overall improvement in current processes (6.12). This evidence is supported by the interview feedback provided by participants.

Research Model Results

The results of the reliability of instrument items can be seen in Table 4, column 1, with correlations reported in the other columns and the square root of the average variance extracted (AVE) in the diagonal. All reliability measures exceed 0.85, indicating strong internal consistency (Nunnally, 1978), above

DOI: 10.1002/asi

TABLE 6. Selected extensions to FIBO standard.

	Description logics concept inclusion axioms
Object properties	
hasAccruedInterestMoneyAmount	$\exists$ hasAccruedInterestMoneyAmount.T $\sqsubseteq$ debt-pricing-yields:AccruedInterestAmount
	T $\sqsubseteq$ $\forall$ hasAccruedInterestMoneyAmount.currency-amount:MoneyAmount
hasTradeBuy	$\exists$ hasTradeBuy.T $\sqsubseteq$ fibo-fbc-fi-fi:FinancialInstrument
	T $\sqsubseteq$ $\forall$ hasTradeBuy. global-fünd-reporting:TradeBuy
hasTradeSell	$\exists$ hasTradeSell.T $\sqsubseteq$ fibo-fbc-fi-fi:FinancialInstrument
	T $\sqsubseteq$ $\forall$ hasTradeSell. global-fund-reporting:TradeSell
hasCostAmount	T $\sqsubseteq$ $\forall$ hasCostAmount.global-fund-reporting:CostAmount
hasCostAmountCurrency	$\exists$ hasCostAmountCurrency.T $\sqsubseteq$ global-fünd-reporting:CostAmount
	T $\sqsubseteq$ $\forall$ hasCostAmountCurrency. fibo-fnd-acc-cur:Currency
hasGainOrLoss	T $\sqsubseteq$ $\forall$ hasGainOrLoss. global-fund-reporting:GainOrLoss
hasGainOrLossCurrency	$\exists$ hasGainOrLossCurrency.T $\sqsubseteq$ global-fund-reporting:GainOrLoss
	T $\sqsubseteq$ $\forall$ hasGainOrLossCurrency. fibo-fnd-acc-cur:Currency
Datatype property	
hasCostCalculationMethod	$\exists$ hasCostCalculationMethod. DatatypeLiteral $\sqsubseteq$ fibo-fbc-fi-fi:FinancialInstrument
	T $\sqsubseteq$ $\forall$ hasCostCalculationMethod. Datatypestring
hasFairValueMethodCode	$\exists$ hasFairValueMethodCode. DatatypeLiteral $\sqsubseteq$ fibo-fbc-fi-fi:FinancialInstrument
	T $\sqsubseteq$ $\forall$ hasFairValueMethodCode. Datatypestring
hasGainOrLossValue	$\exists$ hasGainOrLossValue. DatatypeLiteral $\sqsubseteq$ global-fund-reporting:GainOrLoss
	T $\sqsubseteq$ $\forall$ hasGainOrLossValue. Datatypedecimal
Class	
RealEstateInvestmentTrust	global-fund-reporting:RealEstateInvestmentTrust $\sqsubseteq$ global-fund-reporting:IssuedShare

the recommended level of 0.70. Some variable intercorrelations were relatively high, ranging from 0.32–0.90. This, however, can be partially explained by the relatively small sample used for testing. Furthermore, the AVE indicates convergent validity, as the AVE values of each construct do not exceed the interconstruct correlations (Chin, 1998) and exceed 0.50 (Fornell & Larcker, 1981). Data were then tested for multicollinearity using the variance inflation factor, where all constructs fell below the 5.0 level (Hair, Anderson, Tatham, & Black, 1995).

The results of the structural model (Figure 2) estimate the path coefficients between dependent and independent variables. The R^2 also indicates the variance explained by the independent variables. All paths specified in the model are statistically significant ($p < .05$). This indicates a strong overall fit of the model with system quality and user satisfaction for the GFRO framework.

The direct and indirect effects of reliability (0.101), flexibility (0.167), integration (0.485), accessibility (0.118), and timeliness (0.222) were all significant determinants of system quality and account for 84% of the variance in that measure. These measurements indicate a strong relationship between ratings of system quality and its dependent variables. This is further evidenced by multi-item construct results (Table 4). Respondents rated highly the timeliness of the (5.10), integration (6.14), and flexibility (5.83) of the GFRO framework, thus supporting a high perception of system quality (5.65).

Furthermore, completeness (0.084), accuracy (0.463), and format (0.445) account for 71% of the variance in information quality. The strong relationship between the dependent and independent variables can also be seen in Table 4. Respondents rated the completeness of the information provided by the system (5.44) and the accuracy of information (5.73), thus explaining strong support in information quality (5.85).

System satisfaction (0.261) and information quality (0.545) had a significant impact on information satisfaction (5.60), explaining 59% of the variance in this measure. Furthermore, system satisfaction (0.73) had a significant influence on ease of use, with a rating of 5.50. Information satisfaction (0.422) and ease of use (0.534) proved significant to perceived usefulness, explaining 70% of the variance in this measure and a rating of 5.89.

The implemented data management framework demonstrates the effectiveness of ontologies within the data management of an organization and several benefits are evidenced by using this framework over the current processes of preparing and submitting regulatory reports. In particular, our survey results show that GFRO allows for improved data quality, reduced time to completion, and enhanced integration of data from disparate sources.

Semistructured Interview Results

Next, we provide further evidence on the effectiveness of the data management framework revealed in the series of semistructured interviews.

A significant contribution of this research reported in the interviews is the identification of data quality issues using the Pellet inference engine. The framework detects inconsistencies in financial reports and data using the imposed logic of the ontology (Noy, 2004). Graphs loaded into Stardog triple store are accurate but not complete. This provides completeness of inference: the reasoner infers all necessary and sufficient relationships among graph nodes. If a clash appears, then the Pellet inference engine flags inconsistencies and provides explanations.

DOI: 10.1002/asi

TABLE 7. Selected regulatory property extensions.

Object properties	Description logics concept inclusion axioms
hasMMIFYield	$\exists$ hasMMIFYield. T $\sqsubseteq$ bonds-common: Bond
	T $\sqsubseteq \forall$ hasMMIFYield. global-fund-reporting:MMIFYield
hasMMIFCouponType	$\exists$ hasMMIFCouponType. DatatypeLiteral $\sqsubseteq$ bonds-common: Bond
	T $\sqsubseteq \forall$ hasMMIFCouponType. Datatypeinteger
hasMMIFIdentifierCode	$\exists$ hasMMIFIdentifierCode. DatatypeLiteral $\sqsubseteq$ fibo-red-sec-securities-secuiities-identification-individuals:ISINIdentifier
	T $\sqsubseteq \forall$ hasMMIFIdentifierCode. Datatypeinteger
hasMMIFInstrumentType	T $\sqsubseteq \forall$ hasMMIFInstrumentType. Datatypeinteger
hasMMIFOrlginalMaturlty	$\exists$ hasMMIFOriginalMaturity. DatatypeLiteral $\sqsubseteq$ bonds-common: Bond
	T $\sqsubseteq \forall$ hasMMIFOriginalMaturity. Datatypeinteger
hasMMIFQuoteType	T $\sqsubseteq \forall$ hasMMIFQuoteType. Datatypeinteger

The business benefits associated with the inference engine are evidenced in the comments of State Street's Vice President for Consultancy Services, who states that "This functionality is of huge value as it enables us to identify issues with our data at a very early stage, prior to any reports being generated for submission to the regulator." Furthermore, the framework has significant implications for regulatory reporting. BCBS 239 could require that a robust taxonomy and metadata for the qualification and automated collation of regulatory reporting data is in place. This validated framework can play a significant role in efficiently enabling the organization to meet their regulatory obligations, through the efficiencies that the system enables.

Table 5 provides examples of bonds for which the securities have a different day count basis in different funds. This example is important in a business context, as bonds have a stated accrual day count basis in the initial contract. An accrual is the amount of interest earned to date on a bond. If this differs across funds, it is an anomaly because both cannot be true. Rows with anonymized asset identifiers are a list of bond identifiers that violate accrual basis axioms implemented on the top of the *DayCountBasis* FIBO class.

From a data quality perspective, the inference capabilities of the Pellet reasoner provide significant data quality insights. In addition to the example outlined in Table 5, additional data anomalies were identified. These included: (i) Equities with updated prices in one fund and not in the other, and (ii) Bonds that do not contain an individual par value in monetary terms, only as a %, meaning no par value or quantity could be inferred. The ability of the framework to dynamically identify these issues is a significant benefit and one that isn't possible with a traditional database/data warehouse approach.

From a risk management perspective, the ability to efficiently retrieve data from an LIS, infer missing data, and flag data quality issues mean that the proposed framework has significant value. The ability to efficiently utilize the data from LISs effectively negates the requirement to migrate to a centralized data warehouse. This is reiterated by State Street's Vice President for Consultancy Services, who states that "it removes the need and huge costs associated with migrating to a centralised data warehouse."

The representation of the data is standard in its presentation; however, the method of retrieval and querying is unique in its flexibility. Databases find it difficult to query from a data endpoint and must be queried more generally and data collated and validated (Rodríguez-Muro et al., 2013). This approach allows us to be flexible over querying by using shared characteristics over reasoning, removing the need for collation of spreadsheets and manual processes. From a business perspective, several significant tangible advantages were identified during the evaluation process. The system was evaluated by 26 senior managers from State Street and regulators from the Central Bank of Ireland in a series of workshops. The business impact of the proposed framework is illustrated in the comments of State Street's Chief Scientist, who states that "In total, it is estimated that 200-person hours a quarter would normally be spent preparing an MMIF report. Some of the aggregations that used to take 3–4 hours individually can now be accomplished in 3–4 seconds and the quality of the reporting is dramatically improved."

It was unanimously agreed by all interviewees that this is a much more efficient way to access data from multiple systems. This fact is reiterated in the comments of the Assistant Vice President in the Consulting Services Group, who stated that "this approach is so much more efficient than the 'as is' approach which has a significant manual aspect and is extremely time-consuming."

From an industry standards and data exchange perspective, the proposed extensions to the FIBO classes have significant business value. The Managing Director of the EDM Council characterized the research efforts as having "Major implications for the financial industry" and the associated data standardization efforts was very much welcomed in feedback received from the Irish regulatory authority. Specifically, the R&D efforts mean that it is now easier for FIBO to be adopted by industry, facilitating data standards and interorganizational data exchange across the financial services industry. State Street's Chief Scientist states that "such standardization efforts have significant potential in facilitating greater levels of data interchange and better risk management."

FIBO Extensions

We also make two contributions to FIBO as a standard that are planned to be adopted in subsequent releases of the ontology. First, we contribute to the conceptualization of unpublished modules extracting and categorizing terms in a conceptual model. We also contribute to the formalization of the unpublished modules by reducing ambiguity in existing and duplicated terms as well as implementing conceptually present relationships and hierarchies. Furthermore, we contribute to the manipulation stage of published modules by implementing and testing SPARQL queries over the modules.

Several class and relationship contributions were also contributed to FIBO that are anticipated to be adopted in a later release of the standard. In total, 62 extensions to FIBO were provided following the implementation of the GFRO framework. A sample of the contributions are outlined in Table 6, described in a logical language (Baader et al., 2003; Calegari & Sanchez, 2009). These include datatype properties, object properties, and classes. One major contribution to FIBO is the class for Real Estate Investment Trust (REIT). This investment type is a share-like instrument for property investment funds and is particularly important in ontologies for exploration of contagion effects.

Finally, the implemented framework allows data in traditional relational databases and unstructured data sets to be extended using the imposed logic of the ontology. We utilize this ability using a set of classes (Table 7) that compose part of a regulatory ontology for the purposes of preparing a report submitted to the Central Bank of Ireland. These classes do not exist explicitly within the data set provided. However, we can use the Pellet reasoner to infer that these classes should logically exist within the data. This inference replaces the need to manually aggregate the data and provides significant time and cost savings as well as allowing the LISs used to prepare the data to remain unchanged. The MMIF properties are specific to the required report; however, multiple regulatory ontologies could be imposed on a single set of standard data as required. State Street's Vice President for Consultancy Services confirms that this as a key contribution of the GFRO framework in stating "we spend so much time using macros and manual efforts for various regulatory bodies, as they require similar but slightly different reports. This can be automated and affirmed using the ontology-based approach."

Discussion

Data management has received considerable attention in the information systems discipline in recent years. Abbasi, Sarker, Chiang, and Lindner (2016) note that a key issue for organizations relates to how they manage their data. Furthermore, the authors note that today's data environment is characterized by an array of data quality and credibility concerns. The requirement that business data must be accurate remains a critical issue, with fitness for use being a key criterion (Baesens, Bapna, Marsden, & Vanthienen, 2016).

Hicks (2017) notes that ontologies are an understudied area in information systems. Hence, this research makes a significant contribution, extending existing research through illustrating and validating the potential of ontologies as a plausible mechanism by which to integrate both structured and unstructured data from multiple diverse sources, including legacy information systems. Therefore, this research has significant implications for theory. It illustrates that the inference engines within triple store databases can play a significant role relating to data quality and advances theory that data quality issues can be identified and dynamically captured via technology. Furthermore, it advances theory relating to the importance of context in data quality by empirically illustrating the power of technology in capturing data inconsistencies. These inconsistencies and quality issues serve as a key barrier to the automated mapping of financial data. While an accuracy of 93% has been achieved in automated mapping (Rodríguez-García et al., 2014), a regulatory report would require complete accuracy. These levels have been achieved in other fields (Jean-Mary, Shironoshita, & Kabuka, 2009) and may increase adoption of ontologies in financial services.

The presented framework can serve as an IT infrastructure blueprint for organizations across multiple sectors to utilize existing systems, both legacy and emerging, to efficiently search their data based upon characteristics and to dynamically identify data quality issues. We also present an alternative approach to data warehouses, lakes, and Hadoop systems. These options often store the data from source systems without any alteration, maintaining the problem of garbage in–garbage out. Many companies who have been early adopters of such mass data storage have realized the need for a master data management strategy and are building data dictionaries, which implementing an ontology could bypass.

Conclusion

Today, there is much discussion regarding the potential of technology for improved regulatory reporting. However, this is very much an emerging domain, with very few examples of technology being successfully applied. Furthermore, there are no empirical examples of ontologies being applied in an operational context in financial services. Therefore, this research makes a number of significant contributions.

First, the use case illustrates that through the implementation of ontologies, organizations can maintain existing distributed data sources incorporating both structured and unstructured data sources, thereby providing a very real and tangible alternative to the need to pursue a strategy of both costly and time-consuming migration to centralized data warehouses. Second, our research empirically validates the approach using a survey instrument, an integrated model of technology assessment and user satisfaction, and a series of semistructured interviews. Third, it extends the state of the art through illustrating the application of semantic technologies for financial services through the extension of bonds and equities in FIBO, extending FIBO to incorporate 62 extensions, encompassing a combination of classes and properties. These extensions are formally being adopted into

DOI: 10.1002/asi

a version of FIBO to be published in the near future. Finally, it contributes through extending and validating the power of inference engines for identifying and flagging data quality issues, thereby empirically validating their potential for addressing data quality issues.

Nevertheless, there are limitations to the current research that should be addressed in future research efforts. The framework was tested on a limited source of real financial data. There is a decrease in scalability in the case of enabling both Pellet and inference engines embedded into Ontop API. The framework does not use R2RML for mapping and transforming of nonrelational and relational data into Stardog triple store. XSLT-based transformations showed better performances and flexibility than R2RML because it allows complex financial mathematical formula to be implemented by using financial rules like yield calculations. The main disadvantage of using XSLT is the maintenance of transformations because the current implementation contains more than 10,000 lines of source code. The framework does not offer real-time financial reporting and does not offer the end users the ability to generate consistent financial reports based on streaming data, as implemented in Calbimonte, Oscar, and Gray (2010).

Towards addressing this limitation, the researchers would suggest that the framework be evaluated with ongoing live "Big Data" feeds from multiple sources to determine its reliability and scalability in such circumstances. There are also opportunities to evaluate the approach using data sets from multiple organizations to determine its usefulness in such circumstances.

References

Abbasi, A., Sarker, S., Chiang, R.H.L. & Lindner, C.H. (2016). Big data research in information systems: Toward an inclusive research agenda. Journal of the Association of Information Systems, 17(2), I–XXXII.

Akhigbe, A., Martin, A.D. & Whyte, A.M. (2016). Dodd-Frank and risk in the financial services industry. Review of Quantitative Finance and Accounting, 47(2), 395–415.

Almeida, M.B. (2013). Revisiting ontologies: A necessary clarification. Journal of the American Society for Information Science and Technology, 64(8), 1682–1693.

Arner, D.W., Barberis, J. & Buckley, R.P. (2015). The evolution of FinTech: A new post-crisis paradigm? UNSW, Sydney, NSW 2052, Australia, UNSW Law.

Asunción, G.-P., Mariano, F.-L. & Oscar, C. (2004). Ontological Engineering: with examples from the areas of knowledge management, e-Commerce and the Semantic Web. (Advanced information and knowledge processing) Automatic Classification of Histological Images and Histological Knowledge Modelling of the Human Cardiovascular System View project Grid Database View project. London: Springer Retrieved from https://www.researchgate.net/publication/271589401.

Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D. & Patel-Schneider, P.F. (2003). The description logic handbook: Theory, implementation, and applications. Cambridge, UK: Cambridge University Press. Retrieved from https://dl.acm.org/citation.cfm?id=885746.

Baesens, B., Bapna, R., Marsden, J. & Vanthienen, J. (2016). Transformational issues of big data and analytics in networked business. Management Information Systems Quarterly, 40(4), 807–818.

Bank for International Settlements. (2013). Principles for effective risk data aggregation and risk reporting, (January 2013), 28. https://doi.org/92-9197-138-3.

Bennett, M. (2013). The financial industry business ontology: Best practice for big data. Journal of Banking Regulation, 14(3–4), 255–268. https://doi.org/10.1057/jbr.2013.13.

Bisbal, J., Lawless, D., Wu, B. & Grimson, J. (1999). Legacy information systems: Issues and directions. IEEE Software, 16(5), 103–111.

Calbimonte, J.-P., Oscar, C. & Gray, A. (2010). Ontology-based access to streaming data sources. In 7th Extended Semantic Web Conference ESWC2010 (Vol. 6496, pp. 2–3). LNCS(PART 1).

Calegari, S. & Sanchez, E. (2009). Object-fuzzy concept network: An enrichment of ontologies in semantic information retrieval. Journal of the American Society for Information Science and Technology, 3(2), 80–90.

Calvanese, D., Cogrel, B., Komla-ebri, S., Kontchakov, R. & Lanti, D. (2015). Ontop : Answering SPARQL queries over relational databases. Semantic Web Journal, 8(3), 471–487.

CBI. (2016). Resident Investment Funds Return (MMIF) Notes on Compilation. Dublin, Ireland: Central Bank of Ireland.

Chen, H.-M. & Sheldon, P.J. (1997). Destination information systems: Design issues and directions. Journal of Management Information Systems, 14(2), 151–176.

Chin, W.W. (1998). Commentary: Issues and opinion on structural equation modeling. MIS Quarterly, 22(1), 1. Retrieved from https://doi.org/Editorial.

Comella-Dorda, S., Wallnau, K., Seacord, R.C. & Robert, J. (2000). A survey of legacy system modernization approaches (Vol. 4, pp. 1–30). Pittsburgh, Pennsylvania: Carnegie Mellon University. Retrieved from https://doi.org/10.1080/02783190009554032.

Das, S., Sundara, S. & Cyganiak, R. (2012). R2RML: RDB to RDF Mapping Language. Cambridge, Massachusetts: W3C. Retrieved May 27, 2018, from https://www.w3.org/TR/r2rml/.

Dennis, A.R. & Garfield, M.J. (2003). The adoption and use of GSS in project teams: Toward more participative processes and outcomes. MIS Quarterly, 27(2), 289–323.

Ellis, L., Haldane, A., & Moshirian, F. (2014). Systemic risk, governance and global financial stability. Journal of Banking and Finance, 45(1), 175–181.

Fornell, C., & Larcker, D.F. (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18(1), 39–50.

Hair, J.F., Anderson, R.E., Tatham, R.L., & Black, W.C. (1995). Multivariate data analysis with readings. Prentice Hall. Retrieved from https://dl.acm.org/citation.cfm?id=207590.

Hicks, C. (2017). An ontological approach to misinformation: Quickly finding relevant information. In Proceedings of the 50th Hawaii International Conference on System Sciences (pp. 942–949). Retrieved from https://scholarspace.manoa.hawaii.edu/bitstream/10125/41263/1/-paper0114.pdf.

Horrocks, I. (2008). Ontologies and the Semantic Web. Communications of the ACM, 51(12), 58–67.

Huang, D.W., Sherman, B.T. & Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols, 4(1), 44–57.

Java. (2018). JavaFX official documentation. Retrieved from https://docs.oracle.com/javase/8/javase-clienttechnologies.htm.

Jean-Mary, Y.R., Shironoshita, E.P. & Kabuka, M.R. (2009). Ontology matching with semantic verification. Journal of Web Semantics, 7(3), 235–251.

Kayed, A., Hirzallah, N., Shalabi, L.A.A. & Najjar, M. (2008). Building ontological relationships: A new approach. Journal of the American Society for Information Science and Technology, 59(11), 1801–1809.

Kontchakov, R., Rodríguez-Muro, M. & Zakharyaschev, M. (2013). Ontology-based data access with databases: A short course. In Reasoning Web: Semantic Technologies for Intelligent Data Access (pp. 194–229). Retrieved from https://doi.org/10.1007/978-3-642-39784-4_1.

Lowry, P. B., & Gaskin, J. (2014). Partial least squares (PLS) structural equation modeling (SEM) for building and testing behavioral causal theory:

DOI: 10.1002/asi

When to choose it and how to use it. IEEE Transactions on Professional Communication, 57(2), 123–146. https://doi.org/10.1109/TPC.2014.2312452.

Madnick, S. & Zhu, H. (2006). Improving data quality through effective use of data semantics. Data & Knowledge Engineering, 59(2), 460–475.

Matei, C.M. (2012). Modernization solution for legacy banking system using an open architecture. Informatica Economica, 16(2), 92–102.

McGuinness, D.L. & van Harmelen, F. (2004). Owl web ontology language overview. W3C Recommendation 10 February 2004 (pp. 1–12). Retrieved from https://doi.org/10.1145/1295289.1295290.

Nagle, T., Redman, T.C. & Sammon, D. (2017). Only 3% of companies' data meets basic quality standards. Harvard Business Review Retrieved from https://hbr.org/2017/09/only-3-of-companies-data-meets-basic-quality-standards.

Nallusamy, S., Ibrahim, S. & Naz, M. (2011). A software redocumentation process using ontology-based approach in software maintenance. International Journal of Information and Electronics Engineering, 1(2), 133–139.

Neumann, P.G. (1999). Inside risks: Robust open-source software. Communications of the ACM, 42(2), 128.

Noy, N. (2004). Semantic integration: A survey of ontology-based approaches. SIGMOD Record, 33(4), 65–70.

Nunnally, J.C. (1978). Psychometric theory. New York: McGraw-Hill.

Ontop Framework. Retrieved from http://ontop.inf.unibz.it. [Online; last accessed 08-April-2018].

Rodríguez-García, M.Á., Rodríguez-González, A., Colomo-Palacios, R., Valencia-García, R., Gómez-Berbís, J.M. & García-Sánchez, F. (2014). Using data crawlers and Semantic Web to build financial XBRL data generators: The SONAR extension approach. The Scientific World Journal, 2014, 1–18. Retrieved from https://doi.org/10.1155/2014/506740.

Rodrıguez-Muro, M. & Calvanese, D. (2012). Quest, an OWL 2 QL reasoner for ontology-based data access. In Owled 2012.

Rodríguez-Muro, M., Kontchakov, R. & Zakharyaschev, M. (2013). Ontology-based data access: Ontop of databases. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Vol. 8218, pp. 558–573). Berlin, Heidelberg: Springer Berlin Heidelberg.

Roy, A., Kucukural, A. & Zhang, Y. (2011). Function prediction. Nature Protocols, 5(4), 725–738.

Tripathy, P. & Naik, K. (2014). Legacy information systems. In Software evolution and maintenance (pp. 187–222). Hoboken, NJ: Wiley.

van Geet, J., Ebraert, P. & Demeyer, S. (2010). Redocumentation of a legacy banking system: An experience report. In Proceedings of the Joint ERCIM Workshop on Software Evolution (EVOL) and International Workshop on Principles of Software Evolution (IWPSE) (pp. 33–41). Retrieved from https://doi.org/10.1145/1862372.1862382.

Venkatesh, V., Brown, S.A. & Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. MIS Quarterly, 37(1), 21–54.

Wixom, B.H., & Todd, P.A. (2005). A theoretical integration of user satisfaction and technology acceptance. Information Systems Research, 16(1), 85–102. https://doi.org/10.1287/isre.1050.0042.

Ye, K., Wang, S., Yan, J., Wang, H. & Miao, B. (2009). Ontologies for crisis contagion management in financial institutions. Journal of Information Science, 35(5), 548–562.