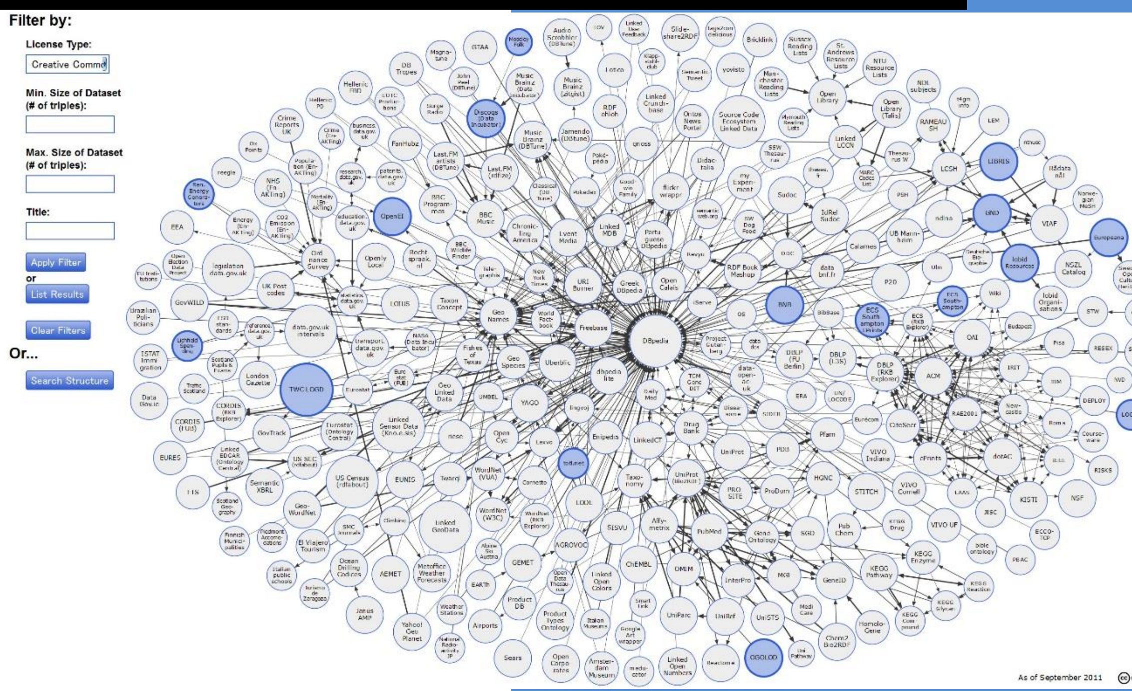
Linked Data adoption and application within financial business processes



Academic supervisor: Dr. Ir. Erwin Folmer
Second academic supervisor: Dr.Ir. Erwin Hofman

Company supervisor: Jasper Roes M.Sc.
Consultant Technical Sciences BIS

Author:

Desislava Kalcheva

Acknowledgements

I would like to express my gratitude to my university supervisor Erwin Folmer for introducing me to the concept of Linked Data, for the useful comments, remarks, ideas and engagement through the learning process of my master thesis. Furthermore, I would like to thank my second university supervisor Erwin Hofman for the much appreciated faith in my success, the creative ideas and unconditional support throughout the whole process. Also, I would like to show my appreciation to Jasper Roes for critically reviewing my whole work, for all improvement suggestions and for his active involvement and engagement. A big thanks to all interviewees, who willingly and honest shared their precious time and experiences with me and thereby enriched and verified the results of my research. Finally, I would like to thank my loved ones for the undying support in both, keeping me harmonious and helping me to keep everything together. Special thanks to my mom, my sister and my dad, who always stood by my side, believed in me and have always saw me as special. I will be forever grateful for your love.

Table of Contents

List of figures

List of tables:

1 Management summary.

Linked Data (hence LD) is a concept for efficient handling of data, which could be used for dealing with complex data and data structures. It operates on RDF triples, promotes keeping data at the source instead of copying it and allows for (semi-)automated reasoning. Therefore the interest towards applying it in different contexts and case studies increases, as well as the number of initiatives around it on a global scale rises accordingly. LD is framed in the context of the financial services domain as an enabler of a lot of opportunities for optimizing (financial) product rankings, as well as cross-country, cross-domain and cross-company benchmarks, without the need for special templates or enormous manual efforts. The LD adoption depends on its perceived ease of use and most heavily on the perceived usefulness. Several factors fall within this category, which could significantly influence it. Such factors are increasing the data re-use potential and facilitating (financial) product comparisons, recommendation applications, and product rankings. Next to that, legislative issues and the limited expressiveness of LD must be taken care of for the specific case. An equivalent of a “mandatory field” must be provided and relevancy of encountered resources argued and ensured. Furthermore, LD adoption can be enhanced by a higher availability of complementary products, such as ontologies and vocabularies, and increase in the perceived installed base. In the aspect of business reporting, LD has the potential to solve some of the main issues of XBRL and thereby help to gain the benefits, expected from its inauguration. Furthermore, solutions of these issues, based on LD are expected to be more dynamic than current considerations, such as enforcing a global standard or prohibiting private extensions of XBRL ontologies. Some of the main potentials for LD in this area are seen in solving semantic heterogeneity, enabling cross-country, cross-domain and cross-company benchmarks of business information without the need for extensive human manual efforts. Next to that, LD and XBRL are seen as very familiar to each other, possibly the next step in the system to system data exchange. Therefore, LD technology providers are advised to assure the availability of complementary products, such as ontologies and vocabularies for specific domains and across domains, and ensure maximal reuse of existing ontologies. Furthermore, an LD application on a larger basis can demonstrate the potential of LD for business processes and then in turn, increase the willingness among potential adopters. Such case can be initiated from the state, based on the financial statements submitted by companies. On the other hand, companies and other potential adopters are advised to consider LD solutions for their case studies. The overview of (dis-)advantages as demonstrated in chapter 10.4 allows for clear judgement over the advantageousness of a potential LD solution depending on the specific case study essentials. Next to that, it can be beneficial to follow current developments in the area of LD applications, to be able to recognize potential in the own field. Finally, it is recommended to ensure a LD advisory, when considering a change or optimization of business processes or even search for new business opportunities.

2 Problem statement

TNO has expertise and understanding on LD and of LD technologies. It consults clients on possible LD solutions within different domain. Furthermore, LD found application in different domains, such as electro technology, energy, aerospace etc. However, its adoption within the financial domain is still very limited. At the same time TNO considers to offer possible LD-based solutions to clients in the financial domain as well. Therefore an understanding of the advantages and disadvantages of LD application for the potential adopter within this domain is needed. Furthermore, it is relevant to identify factors that influence the adoption of LD. On the one side, this could increase the understanding of important factors that play a role in the LD adoption. On the other, it could create incentives among LD technology providers to positively influence those factors, thereby speeding up the process of adoption. Next to that, it could create more understanding of the LD capabilities among potential adopters.

Therefore an overview of advantages and disadvantages of LD application is dynamically modelled to enable application on different case studies and trigger a better informed decision on whether to initiate such within a certain domain and/or business scenario. Furthermore, a conceptual model on LD adoption is created to demonstrate the factors, which have influence on the process and can be possible levers on it. Factors that are specific to financial domain and relevant for LD adoption are included as well. The tangible research aim, goals and research questions will be elaborated on in the next chapter.

3 Research aim, goals and research questions.

3.1 Research aim and aggregated research structure.

As clarified in the previous chapter the necessity is given to research and illustrate (with a model) the adoption potential of LD for improving (financial) business processes and eventually even creating new business opportunities. Therefore the research aim is to showcase this potential by developing a conceptual model for the adoption of LD, apply it on an ongoing case within the financial industry and match it against experts' opinions to validate respectively adapt the findings.

Out of the research aim a main research question was formulated that covers all of the desired research areas. It reads: "What are the advantages and disadvantages towards the adoption of LD within the financial domain?".

The research structure consists therefore of three major parts, i.e. understanding of LD (for the financial domain in particular), construction of the model, whereas relevant design options will be taken into consideration and the key parameters and prerequisites will be included, demonstration with a case within the financial industry and matching against financial experts' opinions. An overview of the aforementioned research summaries and the corresponding disaggregated research goals can be found in the figure below.

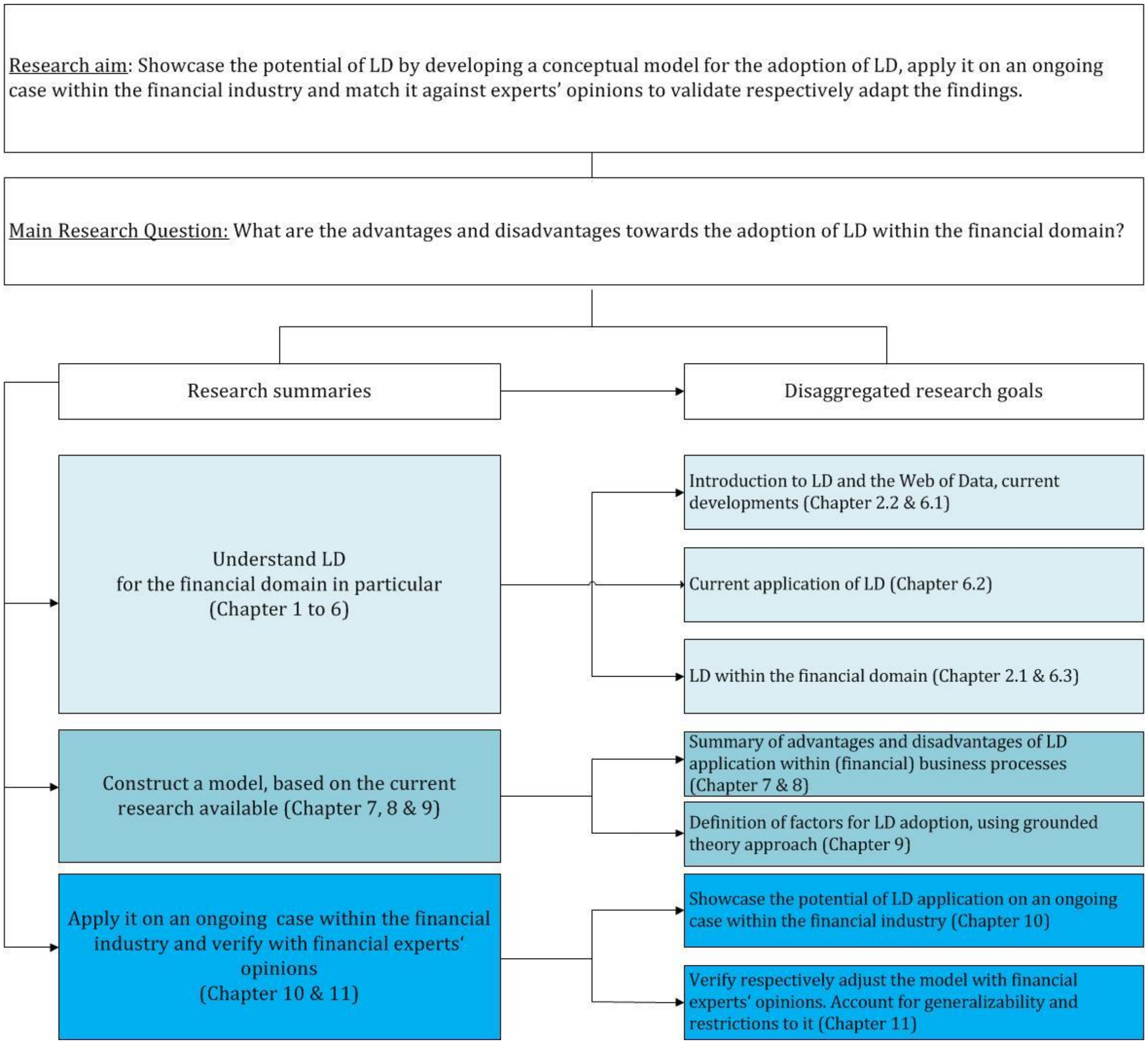


Figure 1: Research structure.

3.2 Research goals and questions.

After the relevant research field was framed, the main research question was broken down to main research activities. Seven research goals were formulated that deal with different distinctive parts of the main research problem. Those are formulated in a way that ensures they are self-sufficient in their research objectives but at the same time there is no overlap between different research goals. The first research goal deals with the conceptualization of LD and motivates the interest towards it. This research goal is covered by one research sub-question: "What are existing conceptualizations of LD and what justifies the interest towards it?". The second research goal covers the extent of current LD application of LD, measured on LOD research initiatives. This research goal is transferred into one research sub-question. It reads: "What is the current application of LD?". The second and third research goal represent a funnel from broad to narrow. Therefore the third research goal focuses on LD within the financial domain. The respective sub-question is: "What are specific cases of current LD application within the financial domain?". Within the next research goal a summary of (dis-)advantages is developed, based on analysis of current literature. It will answer the question: "What is the potential for LD in (financial) business processes?". In consequence the conceptual model for adoption of LD is designed, based on literature on LD, literature on the financial domain and adoption literature. The last two research goals aim at illustration of the LD potential using the developed conceptual model on an ongoing financial business case (Research goal 6) and verification respectively adjustment of the developed model using financial experts' opinions (Research goal 7). Sub-question 6 focuses on financial reporting and XBRL in particular: "What is the value added by the use and/or provision of LD in business reporting (standards) in the financial sector, illustrated with the developed conceptual model applied to the XBRL case?". The last sub-question reads: "What insights can be gained from financial experts' opinions?". The research goals and the thereof derived research sub-questions were carefully and thoughtfully formulated in order to be distinctive and in sum to answer the main research question completely. The main research question on the other hand was defined in a way to cover the complete area of interest expressed with the research aim. The research aim, the main research question, as well as the research goals and sub-questions were discussed with and approved by the TNO cooperation partners, assuring consonance of both, student's interests and TNO's research initiatives. A visualization of the research focus can be found in the figure below.

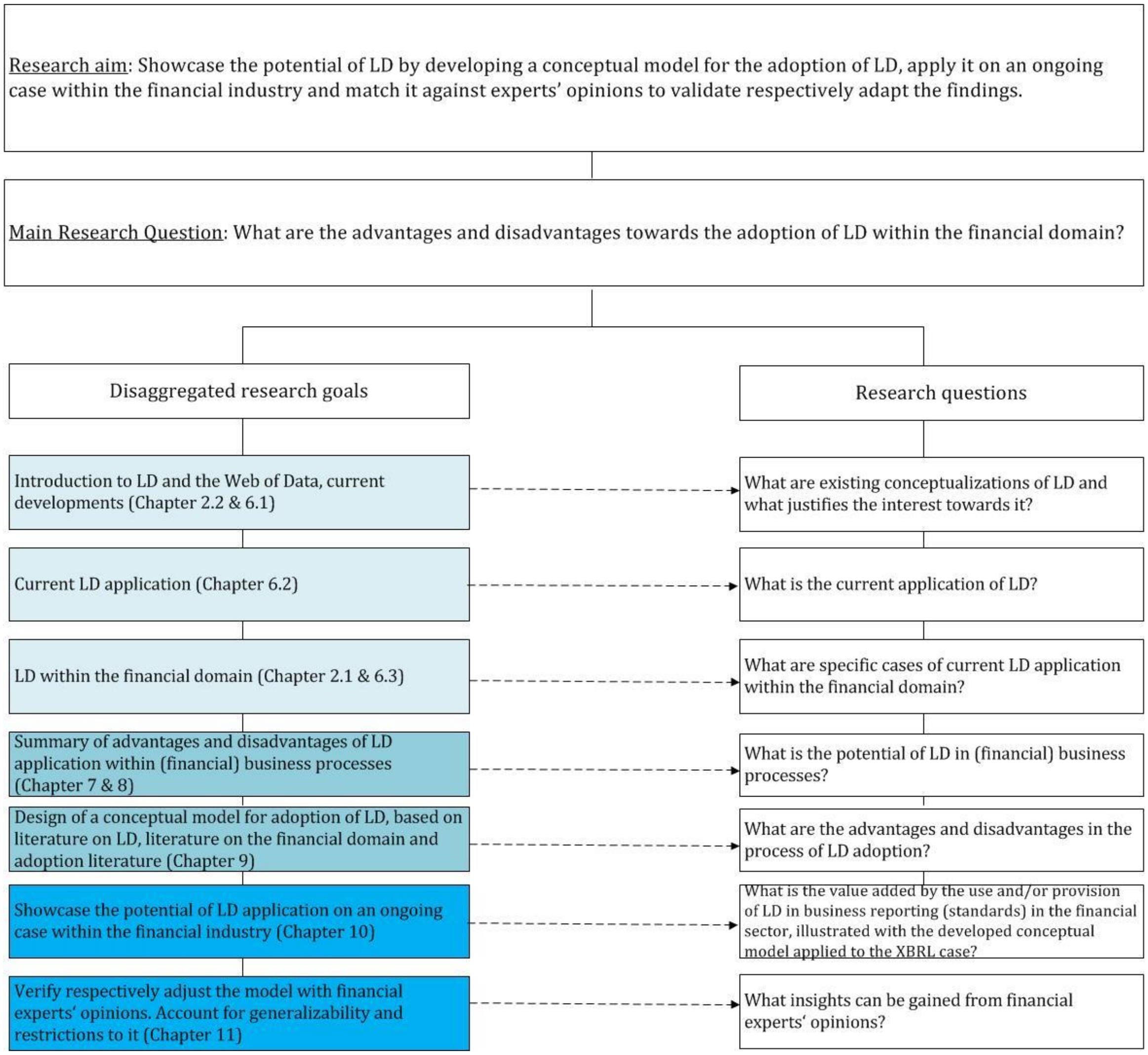


Figure 2: Research aim, main research question, research goals and research sub-questions. Visualization

In the next step the corresponding research methods were agreed upon and will be briefly introduced within the next section.

4 Research methods.

4.1 Systematic literature review.

4.1.1 The top journals and search engines selection.

In order to answer the first two research sub-questions a systematic literature review will be conducted. Even though such is more likely to be performed for PhDs and higher academic research purposes, the importance of the subject requires a guarantee that no important aspect will be left unconsidered and no issues will be considered with no (scientific) ground. The literature research will be conducted using Scopus and EBSCO. The keywords defined for this research, the used queries and selected journals and studies will be shown within this chapter.

In the search all top 25 CS/IS journals and top 25 Management journals were included. An overview of these journal was created by (Schwartz and Russo, 2004). However their coverage by the search engines changed significantly. Therefore the re-examination of the coverage by (Folmer, 2012) was taken over. In this re-examination the search engines INSPEC, ACM DB and Ei Compendex were excluded, since there is no published list of accessible journals (Schwartz and Russo, 2004). The top 25 CS/IS journals and the top 25 management journals, and their respective coverage by search engines as re-examined by (Folmer, 2012) are shown in Table 1 and 2 below.

	Top 25 CS/IS journals	*ISSN*	Scopus	Web of Science	Ingenta	ABI/Inform	EBSCO Business
1	MIS Quarterly: Management Information Systems	0276-7783	x	x		x	x
2	Communications of the ACM	0001-0782	x	x		x	x
3	Information Systems Research	1047-7047	x	x		x	x
4	Journal of Management Information Systems	0742-1222	x	x		x	x
5	Management Science	0025-1909	x	x		x	x
6	IEEE Transactions on Computer sciences (various)	e.g. 0018-9340	x	x			x
7	Harvard Business Review	0017-8012	x	x		x	x
8	Decision Sciences	0011-7315	x	x	x	x	x
9	Decision Support Systems	0167-9236	x	x	x	x	x
10	Information and Management	0378-7206	x	x	x	x	x
11	European Journal of Information Systems	0960-085X	x	x	x	x	x
12	MIT Sloan Management Review	1532-9194	x	x		x	x
13	ACM Transactions on Database Systems (various)	e.g. 0362-5915	x	x		x	x
14	Data Base	0095-0033					
15	Organization Science	1047-7039	x	x		x	x
16	Information Systems Journal	1350-1917	x	x	x	x	x
17	Academy of Management Journal	0001-4273	x	x		x	x
18	Communications of the AIS	1529-3181				x	x
19	IEEE Computer (Graphics and Applications) (various)	e.g. 0272-1716	x	x			x
20	Journal of Strategic Information Systems	0963-8687	x	x	x		x
21	Administrative Science Quarterly	0001-8392	x	x		x	x
22	Academy of Management Review	0363-7425	x	x		x	x
23	International Journal of E-Commerce	1086-4415	x	x		x	x
24	ACM Computing Surveys	0360-0300	x	x		x	x
25	Accounting, Management and Information Technologies	0959-8022	x		x		
25	Information and Organization	1471-7727	x		x		x
		Total	23	22	7	20	23

Table 1: Overview of the top 25 CS/IS journals and their coverage across search engines (Folmer, 2012).

	Top 25 International Business Journals	ISSN	Scopus	Web of Science	Ingenta	ABI/Inform	EBSCO Business
1	Journal of international business studies	0047-2506	x	x	x	x	x
2	Management international review	0938-8249	x	x		x	x
3	Journal of world business	1090-9516	x	x	x	x	x
4	International marketing review	0265-1335	x	x	x	x	x
5	Journal of international marketing	1069-031X	x	x		x	x
6	International business review	0969-5931	x	x	x		
7	International studies of management and organization	0020-8825				x	x
8	Journal of global marketing	0891-1762	x	x		x	x
9	International journal of research in marketing	0167-8116	x	x	x	x	x
10	Advances in international comparative management	0747-7929	x			x	
11	Advances in international marketing	1474-7979	x			x	x
12	Journal of international financial management and accounting	0954-1314	x		x	x	x
13	Multinational business review	1525-383X				x	x
14	Advances in international accounting	0897-3660	x			x	
15	International trade journal	0885-3908	x			x	x
16	International management	0020-7888	x			x	
17	The international journal of accounting	0020-7063			x	x	x
18	International journal of management	0813-0183				x	
19	Global finance journal	1044-0283	x		x	x	x
20	Journal of international management	1075-4253	x	x	x		
21	Thunderbird international business review	1096-4762	x			x	x
22	Journal of international consumer marketing	0896-1530	x			x	x
23	Advances in international banking and finance						
24	International journal of conflict management	1044-4068	x	x	x	x	x
25	International journal of finance	1076-9307	x	x		x	x
		Total	20	11	10	22	18

Table 2: Overview of the top 25 Management journals and their coverage across search engines.

Based on the coverage and availability, the decision for usage of Scopus as a main search engine was made. Only the journals that weren't covered by Scopus, were searched using EBSCO. Three of the journals are not covered by any of the used search engines. Those are shown in the table above. 44 of the journals were covered by Scopus and 41 by EBSCO, whereas the three journals not covered by Scopus were included in EBSCO.

Furthermore since the analysis of the developed conceptual model will be carried out in the context of the financial (services) industry, the top 25 financial journal were included likewise. Their selection was made, using (Currie and Pandher, 2011) and includes the top 25 financial journals based on the ASA quality rank 2009. However the top 25 ASA importance rank 2009 delivered three additional journals that were included later on in the quality selection. This led to the selection of top 28 financial journals for this master thesis. These journals and their coverage within the chosen search engines was investigated. The results are shown in Table 3 below.

Rank	Title	ISSN	Scopus	Total Docs. (2014)	Total Docs. (5 years)
1	Journal of Finance	ISSN 15406261	x	60	671
2	Review of Financial Studies	ISSN 14657368	x	79	829
3	Journal of Financial Economics	ISSN 0304405X	x	105	2397
4	Journal of Financial and Quantitative Analysis	ISSN 00221090	x	47	746
5	Journal of Money, Credit and Banking	ISSN 15384616	x	72	439
6	Journal of Banking and Finance	ISSN 03784266	x	352	4348
7	Mathematical Finance	ISSN 14679965	x	44	331
8	Journal of Financial Intermediation	ISSN 10960473	x	32	319
9	Journal of Corporate Finance	ISSN 09291199	x	144	901
10	Financial Management	ISSN 00463892	x	22	546
11	Journal of Empirical Finance	ISSN 09275398	x	104	799
12	Journal of International Money and Finance	ISSN 02615606	x	155	1939
13	Journal of Financial Markets	ISSN 13864181	x	50	360
14	Financial Analysts Journal	ISSN 0015198X	x	22	936
15	Review of Finance	ISSN 15723097	x	42	301
16	Journal of Risk and Insurance	ISSN 15396975	x	52	393
17	Quantitative Finance	ISSN 14697696	x	167	965
18	Journal of Financial Research	ISSN 14756803	x	23	213
19	Journal of Portfolio Management	ISSN	x	28	980

		00954918			
20	Journal of Business Finance and Accounting	ISSN 0306686X	x	40	1139
21	Journal of Futures Markets	ISSN 10969934	x	68	517
22	Financial Review	ISSN 15406288	x	34	118
23	Finance and Stochastics	ISSN 14321122	x	1	9
24	Journal of Derivatives	ISSN 10741240	x	14	168
25	Journal of International Financial Markets, Institutions and Money	ISSN 10424431	x	106	755
26	Pacific Basin Finance Journal	ISSN 0927538X	x	78	719
27	Journal of Applied Corporate Finance	ISSN 1745662 2			
28	European Financial Management	ISSN 1468036 X	x	35	372

Table 3: Overview of the top 28 Financial journals (ASA importance and quality rank) and their coverage across search engines.

4.1.2 Keywords used for the search queries.

The keywords used for performing the literature search are of critical value for such research, since they ensure encounter of the most relevant studies and provide the basis for hypotheses and modelling. Therefore they were updated and adjusted iteratively to provide an appropriate number of relevant results, ultimately leading to the following results:

Keyword	Linked Data	Advantages	Finance	Semantics
Related		advantage	financial	semantic
		disadvantage	finances	
		advantageous		
		disadvantageous		

Table 4: Keyword and related words selection

Four keywords were selected and included in the search in a broader sense. The overview of the used keywords is shown in the table above. They secure proper coverage of the formulated main research question and research sub-question and ensure that no important aspect relevant for this research is left without consideration.

On the basis of the keywords and the related words as listed above, regular expressions were created for capturing different forms and spelling of words, using wildcards (Folmer, 2012). The used query expressions are shown in the table below.

Used query expressions
"linked data" AND *advantage* AND financ* AND *semantic
"linked data" AND *advantage* AND financ*
"linked data" AND *advantage* AND *semantic
"linked data" AND financ* AND *semantic
"linked data" AND *advantage*
"linked data" AND *semantic
advantage AND *semantic
financ* AND *semantic

Table 5: Used query expressions

4.1.3 Search process.

The search was performed within the selected search engines (Scopus & EBSCO), on title, abstract and keywords. Search within the top journals (top 25 CS/IS, top 25 Management and top 28 Financial journals) was conducted by means of the ISSN of those journals and is performed during January and February 2015. The combination of two keywords created a large but manageable selection of papers. A third keyword was added only to those searches that delivered more than 100 results.

4.1.4 Overview of selected studies.

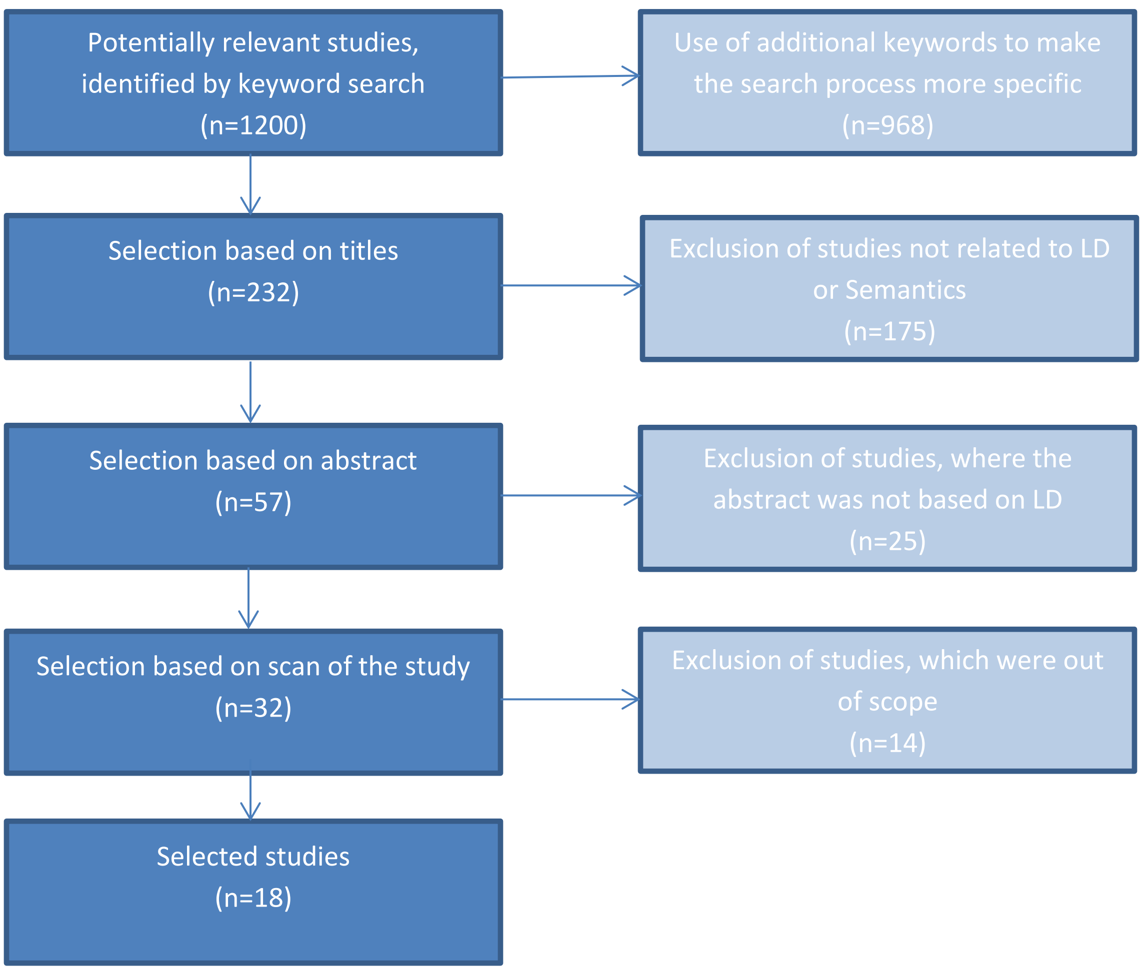


Figure 3: Quorum flowchart of selected studies.

In the following table, an overview of the selected studies according to this procedure, incl. the author(s), title, year publication is provided.

Authors	Title	Year	Publication title	Volume	Issue
Bizer, C., Heath, T., and Berners-Lee, T.	Linked Data - The Story So Far	2009	Int. J. Semantic Web Inf. Syst.		
Chan, S.W.K., Franklin, J.	A text-based decision support system for financial sequence prediction	2011	Decision Support Systems	52	1
Chowdhuri, R., Yoon, V.Y., Redmond, R.T., Etudo, U.O.	Ontology based integration of XBRL filings for financial decision making	2014	Decision Support Systems	68	
Chu, P.-C.	An object-oriented approach to modeling financial accounting systems	1992	Accounting, Management and Information Technologies	2	1
Currie, R.R., and Pandher, G.S.	Finance journal rankings and tiers: An Active Scholar Assessment methodology	2011	J. Bank. Finance	35	
Du, J., Zhou, L.	Improving financial data quality using ontologies	2012	Decision Support Systems	54	1
Hitzler, P., and Janowicz, K.	Linked Data, Big Data and the 4th Paradigm	2013	Semantic Web	4	
Hu, D., Yan, J., Zhao, J.L., Hua, Z.	Ontology-based scenario modeling and analysis for bank stress testing	2014	Decision Support Systems	63	
Janowicz, K.	The Role of Space and Time For Knowledge Organization on the Semantic Web	2010	Semantic Web	1	
Meijer, K., Frasincar, F., Hogenboom, F.	A semantic approach for extracting domain taxonomies from text	2014	Decision Support Systems	62	
Narock, T., Yoon, V., March, S.	A provenance-based approach to semantic web service description and discovery	2014	Decision Support Systems	64	

Nebot, V., Berlanga, R.	Building data warehouses with semantic web data	2012	Decision Support Systems	52	4
Nederstigt, L.J., Aanen, S.S., Vandic, D., Frasincar, F.	FLOPPIES: A Framework for Large-Scale Ontology Population of Product Information from Tabular Data in E-commerce Stores	2014	Decision Support Systems	59	1
Papavasileiou, V., Flouris, G., Fundulaki, I., Kotzinos, D., Christophides, V.	High-level change detection in RDF(S) KBs	2013	ACM Transactions on Database Systems	38	1
Roussinov, D., Zhao, J.L.	Automatic discovery of similarity relationships through Web mining	2003	Decision Support Systems	35	1
Steinfield, C.W., Markus, M.L., and Wigand, R.T.	Cooperative Advantage and Vertical Information System Standards: An Automotive Supply Chain Case Study	2011	IEEE		
Tummarello, G., Delbru, R., and Oren, E.	Sindice.com: Weaving the Open Linked Data	2007	ISWC'07/ASWC'07 Proceedings of the 6th international The semantic web and 2nd Asian conference on Asian semantic web conference		
Zhu, H., Wu, H.	Assessing the quality of large-scale data standards: A case of XBRL GAAP Taxonomy	2014	Decision Support Systems	59	1

Table 6: Overview of the selected studies

Additionally, a google search and google scholar search was performed, to enrich the study by more practical and most recent developments. Furthermore conferences for LD were visited and the productive talks delivered even more valuable insights from practitioners.

4.2 Design science.

For the following research sub-questions design science will be applied in developing a conceptual model about the advantages and disadvantages associated with LD within (financial) business processes. It will be designed, based on insights and analysis of current research available and then iteratively adjusted using a case study within the financial industry/standards and insights from interviews with financial experts.

The preliminary outline in terms of reporting format and language, lay-out and APA-style conventions can be found in the appendix. Same applies to the time horizon and corresponding planned activities.

5 Introduction to the master thesis assignment.

For this master thesis LD is of interest. The improvements in (financial) business processes that a LD application could trigger and the new business opportunities that could come along with it, are elaborated on in the next chapters. Moreover the disadvantages and difficulties associated with such application, are opposed. From this a conceptual model for the adoption of LD is developed. It captures the potential of LD in terms of both, advantages and disadvantages and eases decision makers in the trade-off whether to apply it or not, depending on their specific business case and/or domain. The model is developed based on the current research on LD available. After it is being conceptualized, it is applied on an ongoing case within the financial industry, i.e. XBRL. Furthermore, it is validated by interviews with financial experts to prove practicability and improve its applicability, as well as capabilities to solve present problems and challenges. Next to that semi-structured interviews with LD professionals were conducted to enrich the understanding of LD and its relevancy for practitioners. Furthermore, they ensured proper usage of terms and understanding of the context as a whole. The corresponding research model is shown below.

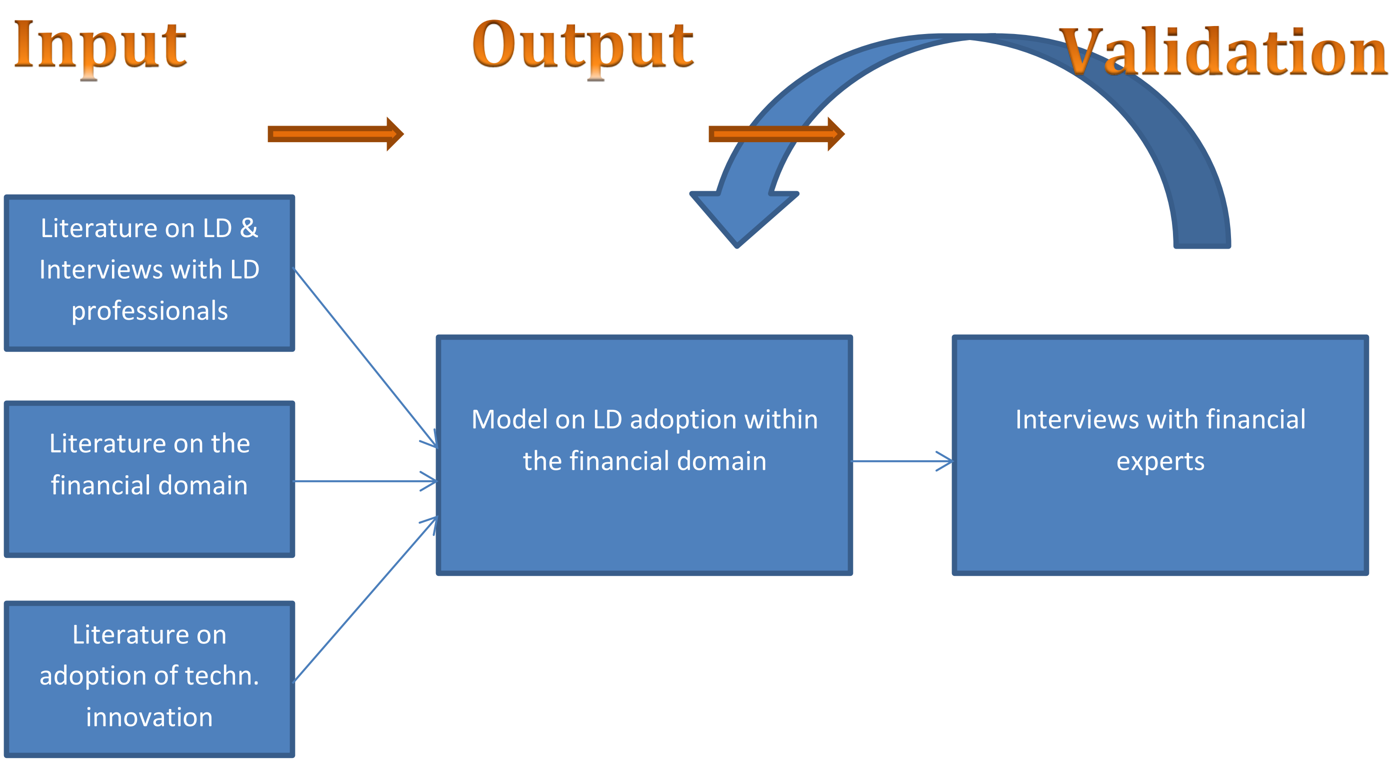


Figure 4: Research model.

5.1 Introduction to the financial domain.

SIC[1] codes are industry classifications. They are four-digit numerical codes, assigned by the US government to business establishments. The purpose of SIC codes is to "identify the primary business of the establishment" and facilitate the "collection, presentation and analysis of data" (SICCODE.com, 2015b). This classification covers all economic activities. The economy is thereby divided into 11 divisions, consisting of 83 two-digit major groups, further divided into industry groups and finally disaggregated into industries. The financial (services) domain is a part of SIC 60-67. The division includes establishments from finance, insurance and real estate, whereas finance incorporates "depository institutions, non-depository credit institutions, holding (but not predominantly operating) companies, other investment companies, brokers and dealers in securities and commodity contracts, and security and commodity exchanges" (SICCODE.com, 2015a). Moreover, monetary authorities that perform monetary control belong to the Finance sector. Furthermore, those establishments engage in financial transactions, i.e. transactions that involve the "creation, liquidation, or change in ownership of financial assets" and/or facilitate those transactions (SICCODE.com, 2015a). Those include three types of activities – raising of funds by taking deposits and/or issuing securities, pooling risk by underwriting insurance and annuities, and providing services to facilitate or support financial intermediation, insurance and employee benefit programs.

More importantly, financial industries are "extensive users of electronic means" (SICCODE.com, 2015a). Those are used for verification of financial balances, authorization of transactions, transfer of funds, notifications of individual transactions for banks or credit card issuers, providing daily summaries etc. (SICCODE.com, 2015a). The selected case study in chapter 10 shows this in the context of electronic communication of business and financial data.

The financial industry had undergo many changes. In earlier days institutions were the main players. However currently more and more individual or non-institutional investors are drawn to the financial markets (Chowdhuri et al., 2014). Furthermore they mainly rely on publicly available data sources when making investment decisions and not on financial advisors' consultations (Ervin, 2004). Therefore the accuracy of traditional data sources and online financial data is essential and yet questionable (Du and Zhou, 2012).

The purpose of financial accounting is to "collect, process, and report information related to financial transactions." (Chu, 1992) The presented financial picture, including the firm's financial condition, i.e. assets and liabilities, periodic operational results, transactions with customers and creditors etc. must be objective, consistent and reliable (Chu, 1992). The

[1] Standard Industrial Classification

decision of the US SEC[2] for all tier-1 public companies to report their financial statements in XBRL[3] aimed at a solution for this (SEV.GOV, 2013). It was expected that the XBRL adoption will improve accessibility, transparency and efficiency of the dissemination of financial data (Pinsker and Li, 2008). Its benefits are expected to extend to economic stakeholders, both inside and outside of the organization (Baldwin et al., 2006). However XBRL still poses cognitive and technical challenges to investors (Debreceny et al., 2010). LD might offer solutions to (some of) the issues, associated with XBRL.

5.2 Introduction to Linked Data.

The concept of LD is somewhat related to two other concepts, namely Open Data and Big Data. The representation of these relations looks as follows:

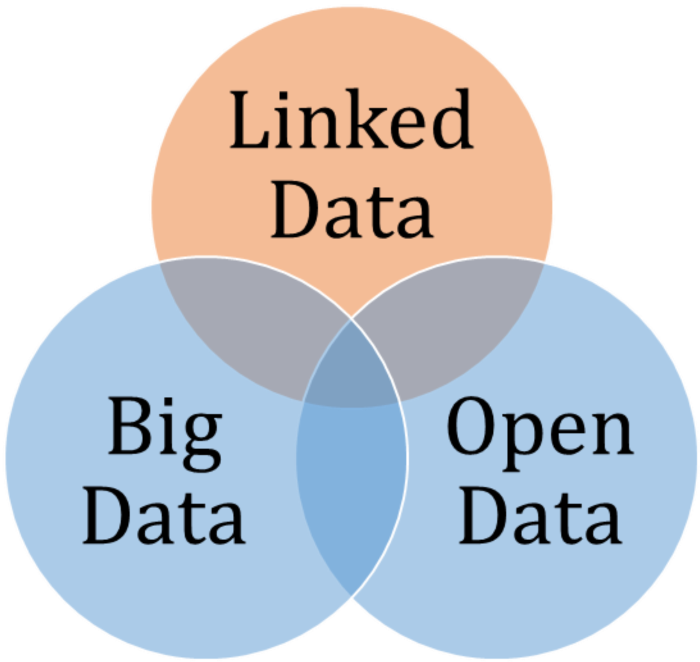


Figure 5: The views on data (Folmer and Krukkert, 2015)

It shows the main view points on data and their interrelation. Big data is mainly characterized by the three V's: Volume, Velocity and Variety. The focus of this research work will be on LD. Open data will be somewhat touched upon with Linked Open data, hence LOD. LOD is simply LD associated with no cost of acquiring/ usage due to the presence of an open license, its open format and machine readability. LD is "a data publication methodology that utilizes the semantic web to make data publicly accessible on the Web" (Narock et al., 2014). It enables combination of data from multiple sources that can be queried over by exposing RDF databases on the Internet (Narock et al., 2014).

[2] The United States Securities and Exchange Commission
[3] eXtensible Business Reporting Language

Resource description framework (hence RDF) is "a framework for representing information in the Web" (Klyne and Carroll, 2014). LD works with triples – a subject, a predicate and an object. An RDF graph is a set of such triples. This is shown in the figure below.

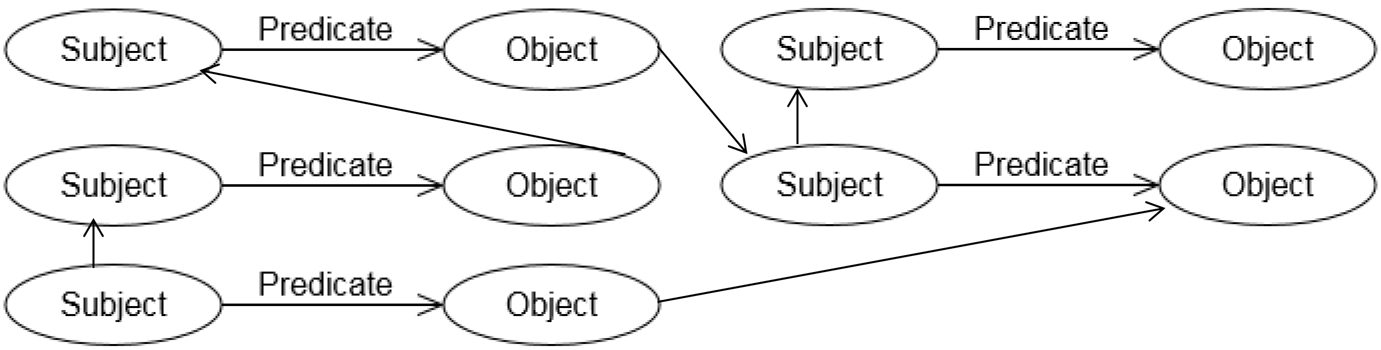


Figure 6: An RDF triple (Klyne and Carroll, 2014).

SPARQL on the other hand, is an RDF query language, i.e. it facilitates manipulating and querying RDF graphs on the Web or in an RDF store (Keio and Beihang, 2013).

LD promotes four main statements:

- The Open world assumption – it poses that for a complete picture information from different/additional sources is needed.
- Anybody can say Anything about Any topic (AAA principle), which also can be extended if space and time are being added (Hitzler and Janowicz, 2013) to AAAAA principle. Every party that has information at its disposal is able to share it and/or make it available. Thereby different perspectives come to play. This leads to a higher level of
- Variety – information from different sources (public, private, governmental) is available in different forms (illustrations, text documents, web pages, databases etc.) and can be connected ("linked") using the LD principles. Therefore one is able to
- Say more, do more, play more.[4] LD allows for assigning context to data, which enables digital processing and interpretation. This in turn improves business processes and possibly leads to new insights.

Although the research on LD is advanced, the practical implementations within the financial domain appear to be still limited in our opinion. We believe that this is due to its novelty and the lack of clear showcase of its potential for business, in terms of both,

[4] "Play more" refers not to actual playing with the data but to higher potential and brighter scope of data application and reuse, which ultimately could lead to improving business processes and even creating new business opportunities.

advantages and disadvantages. This research gap should be closed with the creation of a model to capture the LD potential.

For this purpose a cooperation with TNO was established, due to its expertise in terms of LD and furthermore the presence of interest in researching its potential within current (financial) business processes. TNO – innovation for life, Netherlands Organization for Applied Scientific Research is an organization for independent research that "connects people and knowledge" (TNO Innovation for life, retrieved from https://www.tno.nl/en/about-tno/mission-and-strategy/). It focuses on five closely related themes "with a prominent place in the national and European innovation agenda". These are:

- Industry
- Healthy living
- Defense, safety & security
- Urbanization
- Energy (TNO Innovation for life, retrieved from https://www.tno.nl/en/focus-area/)

This master thesis assignment is seen as theme overarching since LD find/could find application within each of the five themes. However the business study and expert's opinions used for validation of the model findings will be within the financial industry. The exact content of this study assignment will be shown in detail within chapter 3.

5.3 Introduction to adoption and the most imposed adoption models and theories

The adoption of LD technology can be considered as adoption of technological (IT) innovation, since it has the "potential to reduce information overload and to enable semantic integration" (Joo, 2011) through its capabilities, such as semantics and machine-processability (Antoniou and Harmelen, 2008). Adoption is stage two within the stage model of technology diffusion. It consist of initiation, adoption and acceptance, adaptation, routinization, and infusion (Zmud and Apple, 1989). In the following, relevant models will be mentioned and briefly introduced, starting with the Technology Acceptance Model (hence TAM).

The basic concept that underlies user adoption models looks as follows:

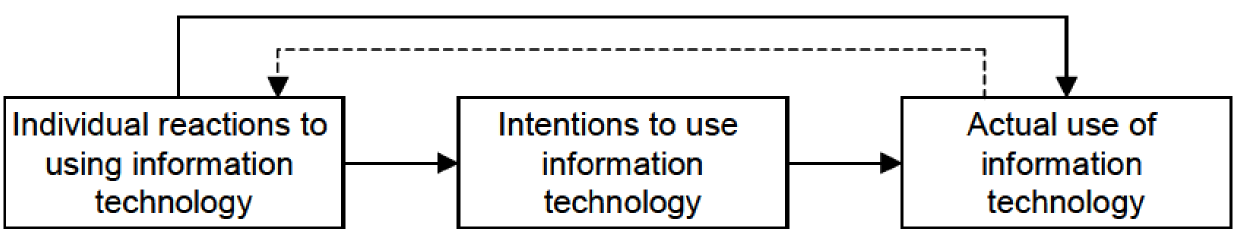


Figure 7: Basic concept underlying user acceptance models (Venkatesh et al., 2003).

Some streams of research focus on individual acceptance of technology and focuses on using intention as a dependent variable (Davis et al., 1989). Other aim at implementation success at the organizational level (Leonard-Barton and Deschamps, 1988). Goodhue (1995) focuses on technology fit. However, all those streams, incl. the TAM, are based on the concept, described in the figure above, considering individual reactions to using the technological innovation, intentions to use it, as well as the actual usage.

Within this model a potential user's overall attitude towards using a system is assumed to directly influence the actual usage of this system. In turn, this attitude is a function of perceived usefulness and perceived ease of use. On the other hand, these two beliefs are directly influenced by design features (Davis Jr, 1986). The latter have no direct effect on attitude or behaviour according to this model. Their only effect is thus only indirect through perceived usefulness and perceived ease of use. This is shown in the figure below.

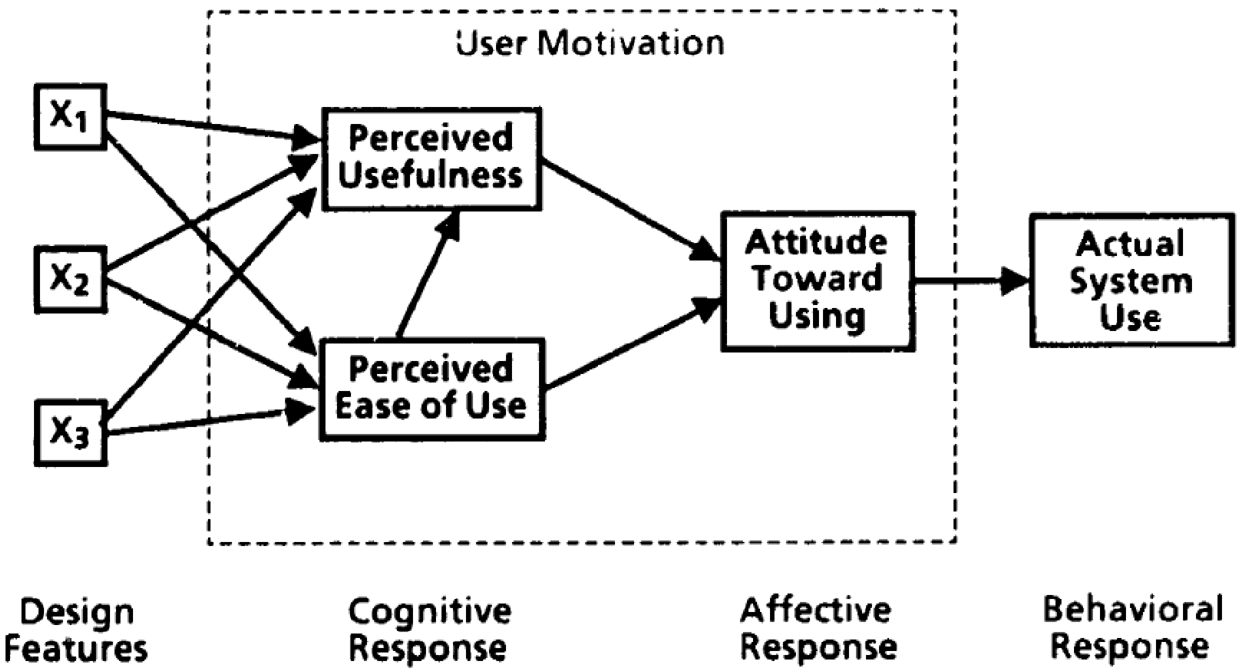


Figure 8: Technology Acceptance Model (TAM) (Fishbein and Ajzen, 1975).

The hypothesized relationships within the model are linear. Hereby use is defined as a "repeated, multiple-act behavioural criterion" (Fishbein and Ajzen, 1975, p. 353) and refers to the actual usage of the system in the context of one's job. It is specific in regards to target (the given system), action (the actual usage) and context (one's job), and non-specific in regards to time frame. The attitude, on the other hand, is defined with respect to the degree of evaluative affect (Fishbein and Ajzen, 1975) that one associates with using the target in the specified context. Furthermore, perceived usefulness is specified as "the degree to which an individual believes that using a particular system would enhance his or her job performance", whereas perceived ease of use represents "the degree to which an individual believes that using a particular system would be free of physical and mental effort" (Ajzen and Fishbein, 1977). Multi-item measurement scales for those two variables are developed and validated to enable "better measures for predicting and explaining use" (Davis, 1989). For a complete overview of the measures, please refer to the appendix.

The TAM is founded on the Fishbein model, however there are major differences between the two models, which will be briefly explained in the following. Within the Fishbein model the summed belief-evaluation term is treated as one independent variable. Therefore the regression coefficient shows the overall effect of beliefs on attitude. This could, however, substantially distort the outcomes (Davis Jr, 1986). The TAM, on the other hand, treats each belief separately in the regression equation. Therefore by analysing the coefficients, relative influence of the different beliefs on attitude toward using can be compared. Furthermore, no evaluation term is employed in the TAM, whereas each belief is weighted by its corresponding evaluation term in the Fishbein model (Davis Jr, 1986). Relationships between beliefs are not explicitly specified in the latter and instead assumed to have equal weights and summed together regardless of existing relationships. TAM, on the other hand, poses a causal relationship between perceived ease of use and perceived usefulness. Another difference lies in the elicitation procedure of identifying the salient beliefs. Within the TAM it is a qualitative free-response procedure, whereas Fishbein focuses on specific subjects and questions about the specific system that is being analysed. The gathered beliefs information is then measured, using single-item measurement scales (Fishbein and Ajzen, 1975) and multi-item measurement scales within the TAM (Ajzen and Fishbein, 1977). The variable behavioural intention is omitted in the TAM, since it reflects on a decision, formed through factors that may change over a significant period of time (Davis Jr, 1986).

Furthermore, the TAM has been extended by consideration of consumer perceptions of network externalities. It is claimed and empirically tested that the existence of direct and indirect network externalities will positively influence the consumer purchase intentions and perceptions of an innovation's usefulness and ease of use (Song et al., 2009). This justifies the efforts of management of network externalities on adoption. Stimulating the size of the installed base and the development of a variety of complementary products can increase purchase intention and in turn, the perceived usefulness and ease of use. Given the existence of network externalities, the perceived value of an innovation depends on the number of consumers, adopted the innovation, e.g. the telephone. Furthermore, an increasing number of consumers can have an indirect impact on the perceived innovation's value. It can stimulate the availability of complementary products and services, thereby leading to a higher perceived value of an innovation, e.g. a DVD player's value may depend on the number of available DVD discs. Therefore the perceived installed base and the perceived availability have direct influence on the perceived ease of use, as well on the perceived usefulness. There are multiple reasons for this. Higher volume of installed base positively reflects on the volume of innovation-related word-of-mouth communication that a potential adopter experiences. This then in turn, directly increases the perceived usefulness and ease of use, by providing knowledge about the innovation (Arndt, 1967). Furthermore the size of installed base influences the probability of acquiring information about innovation usefulness and ease of use by observing prior adopters and potential borrowing to try it (Rogers, 1983). Next to this, the size of installed base signals about the ease of use of an innovation (Song et al., 2009). Furthermore, it creates a perception of a strong user support network, since it is associated with higher number of existing users, available to support a new adopter (Redmond, 1991). Moreover, it is hypothesized that network externalities also have a direct influence on the purchase intent. This is how the TAM is extended (Song et al., 2009) and it looks as follows:

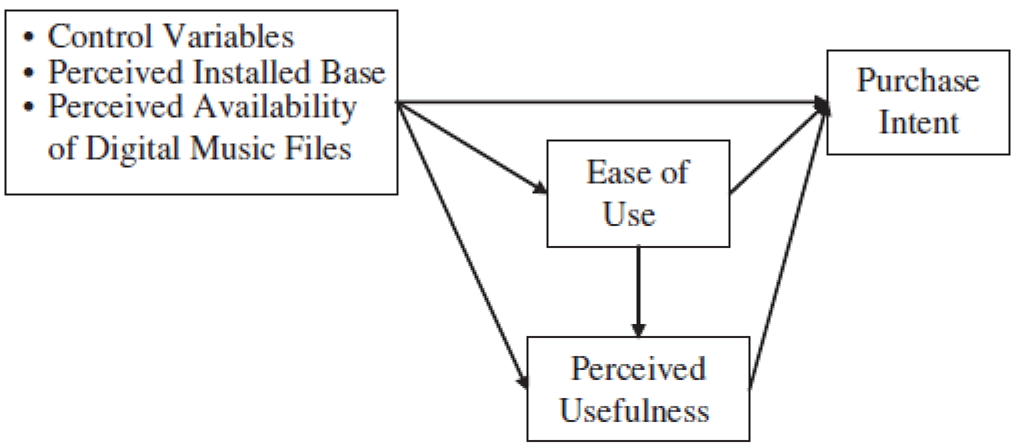


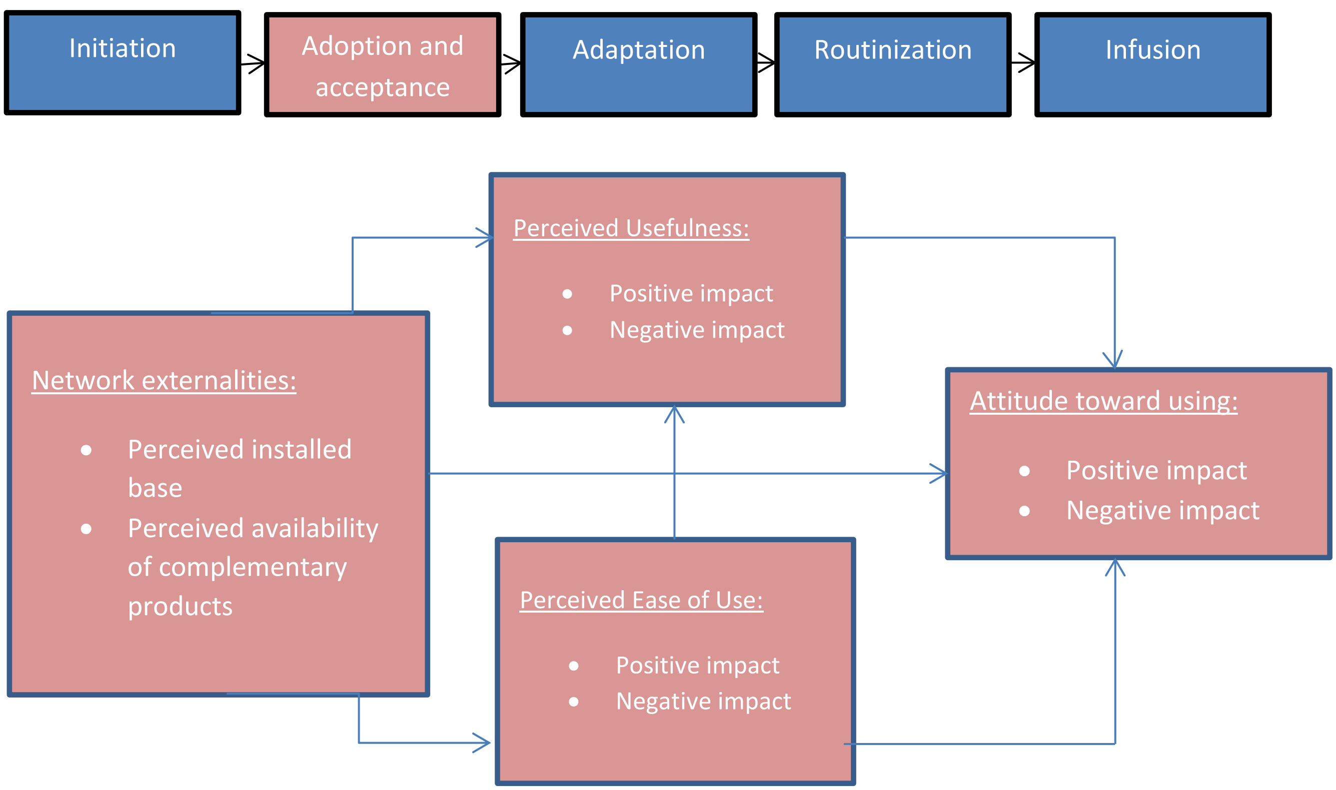
Figure 9: Extension of the TAM model by network externalities (Song et al., 2009).

Many studies researched the robustness of TAM. Adams, Nelson, and Todd (1992), Davis and Venkatesh (1996), Hendrickson, Massey, and Cronan (1993), Sagars and Grover (1993) and Szajna (1994 and 1996) worked on the computer-related usage of the TAM. Additional explanatory variables are examined, such as age (Morris and Venkatesh, 2002), gender (Gefen and Straub, 1997) and culture (Phillips, Calantone and Lee, 1994 and Straub, Keil and Brenner, 1997). The most recent interest of researchers has been moved to the adoption of home computers (Brown and Venkatesh, 2005 and Venkatesh and Brown, 2001) and mobile communication services (Cheong and Park, 2005, Lu, Yao and Yu, 2005, Nysveen, Pederson and Thorbjornsen, 2005 and Wu and Wang, 2005), focusing on the TAM usefulness in this regard. The focus of this work will be on the adoption of LD.

The TAM is related to the work of Holak and Lehmann (1990), who claim that relative advantage and compatibility have direct impact on purchase intention. This is then extended with three more factors by Rogers (2003). These are: complexity, observability and trialability, leading to the five determinants of consumer adoption. Relative advantage is described as the degree, to which an innovation is perceived as better than the previous idea by a particular group of users. It is measured in terms that matter to those users, i.e. economic advantage, social prestige, convenience or satisfaction (Robinson, 2009). It is therefore related to perceived usefulness, whereas compatibility is related to ease of use. However, relative advantage fails to capture the difference between performance benefits and cost savings of an innovation (Song et al., 2009). Therefore, the variations in relative advantage values can only limited be interpreted. Given this, perceived usefulness appears to be more suitable to be used in further analysis within this research. The next determinant, observability, refers to observable results, which lower uncertainty.

Roger's diffusion of innovations is radically different to most other theories of change because it proclaims change not in people but in the innovations (Robinson, 2009). Another main insight from it is the importance of peer-to-peer conversation and peer networks for innovation's adoption. Next to that, innovation suppliers need to understand the needs of the different user segments, i.e. innovators, early adopters, early majority, late majority and laggards, as the adoption process varies per segment (Rogers, 2003).

The proposed model for LD adoption within this work is based on the TAM model (Fishbein and Ajzen, 2005) and includes the extension by Song et al. (2009). It looks conceptually as follows:



6 Results from the systematic literature review.

6.1 Linked Data and the Web of Data (Web 3.0).

Nowadays the amount of digital data grows over more than 50% per year. Therefore any means to structure this data, becomes increasingly relevant (IDC, 2010). Next to that knowledge management and decision-making tasks rely on this data (Meijer et al., 2014). This justifies the necessity of a research work that discovers and illustrates the potential of LD, in terms of advantages and disadvantages, in improving current (financial) business processes and eventually even creating new business opportunities.

LD is a "set of best practices for publishing and connecting structured data on the Web" (Bizer et al., 2009) and its essence consists of semantics and standards. The semantics capture the meaning of the data. The standards allow for interpretation because they imply how the meaning and relations should be set in order to enable digital exchange and processing (Folmer and Verdonk, 2014). This set of best practices has been adopted by a high and constantly increasing number of data providers allowing for the creation of a global data space that contains billions of assertions or the so called Web of Data (Bizer et al., 2009). More elaboration on the number of datasets, triples and its growth rate will be provided in the next section.

The Web of data, also referred to as the Semantic Web or the Web 3.0, is a "global information space", in which not only the documents but also the data itself is linked (Bizer et al., 2009). It was conceived in 2001 (Berners-Lee et al., 2001). It is a large knowledge-base of sources that delivers (references) information as RDF files or through SPARQL endpoints. The idea behind it is to add machine-understandable, semantic annotation to web-published contents. That way they can be retrieved and effectively processed by humans and machines in a variety of tasks. This is done by attaching semantics to resources, from very simple to very complex annotations depending on the requirements, by using semantic web technologies. They enable a new dimension to data integration by providing a common terminology, standard format for resources (RDF/(S)[5] and OWL[6]), semantically linked data (Nebot and Berlanga, 2012). The result is more and more (semi)-structured data and knowledge resources, published on the Web, all together creating the Web of Data (Bizer et al., 2009). The main difference between the Web 2.0 and the new Web of data, or Web 3.0 is thus that Web 2.0 mashups work against a fixed set of data sources, whereas the LD applications operate on top of an "unbound, global data space." This then enables answers from new data sources on the Web, expressed in the AAA principle, "Anybody can say Anything about Any topic". Furthermore this can be extended if space and time are added (Hitzler and Janowicz, 2013) to AAAAA. Thus every party that has information at its disposal is able to share it and/or make it available with LD. Thereby different perspectives come into play. However from businesses' perspective this may not

[5] RDF/(S): http://www.w3.org/TR/RDF-concepts/
[6] OWL: http://www.w3.org/TR/owl-features

always be beneficial, as they do not want to "play" but to use relevant data from reliable sources. This will be further elaborated on within the disadvantages section.

LD is not relational data, SQL etc. but graph data. The graphs are decentralized, which means that there isn't a single knowledge-base of statements but anyone can contribute with statements to the information space of the Web of Data. Shared identifiers (URIs) and shared terms allow for merging of these statements and therefore providing useful services to human and software clients, (Tummarello et al., 2007) This enables new types of applications, which were not possible until now with the Web 2.0. Semantic search is one of the first applications that makes use of and exploits the Web of Data (Nebot and Berlanga, 2012). Search and browsing over RDF data are developed (Cheng et al., 2008) and this enhances conventional information retrieval, since it enables search services centered on entities, relations and knowledge. These new types of applications enabled by the usage of graph (RDF) data can be classified into three categories.

- LD Browsers

Traditional Web browsers allow for navigation between HTML pages by following hypertext links. LD browsers however enable browsing within and between data sources by following links expressed as RDF triples. For example an user can browse the description of the new Hobbit movie "The Hobbit: The Battle of the Five Armies" (2014) and then discover that its country of origin is New Zealand/USA. He or she then can follow the link to the country and find more information about the country and possibly a list with actors born there. This navigation is done following RDF links rather than HTML links (Bizer et al., 2009).
A particular example of a LD browser is the Disco Semantic Web browser[7]. The encountered documents are browse-able resources themselves and the user can follow them to decide which of them are to be used for his or her specific query (Tummarello et al., 2007). The Marbles LD Browser will be shown on an example in the following (Bizer, 2003).

- LD Search engines and Indexes

LD search engines crawl LD from the Web by following RDF links. That way users can query aggregated data from a variety of sources. Two basic categories of these are to be distinguished, namely human-oriented search engines, such as Falcons and application-oriented indexes, such as Watson[8] (Bizer et al., 2009).

- Domain-specific applications

Domain-specific applications are of main interest for this thesis, since application within the (financial) business domain is being researched. Current domain-specific applications are Revyu for film reviews, based on LD, as well as DBpedia Mobile, which is a location-aware LD browser for exploring cities. As indicated before LD applications merge information from different sources. For example DBpedia Mobile connects information

[7] Disco Semantic Web browser: http://www4.wiwiss.fu-berlin.de/rdf_browser/
[8] For more information and technical details, please refer to (Bizer et al., 2009)

from DBpedia, reviews from Revyu and related photos from the Flickr photo-sharing API (Bizer et al., 2009). The biomedical domain is as well active, if not one of the most active, in effort to export database semantics to data representation formats (RDF), following open standards that explicitly state the content semantics. For a long time ontologies and languages, such as RDF and OWL are being used there (Nebot and Berlanga, 2012). An increasing number of scientific communities use Semantic Web ontologies to share and interpret data within and across research domains (Papavassiliou et al., 2009). Many initiatives characterize these efforts, such as Bio2RDF (Belleau et al., 2008) or Linked Life Data[9]. Some Semantic Web developments have already been adopted on a large-scale use, such as the introduction of schema.org[10]. It is a semantic vocabulary, proposed by the four major search engines Bing, Google, Yahoo! and Yandex. It is not domain-specific and not an ontology of "everything" but rather a high-level vocabulary on popular Web concepts (Nederstigt et al., 2014). Next to that Google has its own development, i.e. the Knowledge Graph[11]. It is a project that enhances search results with appropriate semantic metadata. Those movements are often attributed to the concept of Linked data (Nederstigt et al., 2014).

Linked data operates on the so called "decentralized publication model" (Tummarello et al., 2007). Taking advantage of the new data sources appearing on the Web due to the AAA principle, does not require changes in the application code due to LD. Each application that supports LD can thus automatically make use of them. This contributes connecting different data depositories currently available on the Web into a single global information space. This is arguably the main benefit of LD from users' perspective and one of the main building blocks of the Semantic Web. Nevertheless there lies also one of the key challenges for the LD principles. Indeed the user has integrated access to data from various distributed and heterogeneous data sources, however he/she has not selected them explicitly (Bizer et al., 2009). The challenges that come along with this will be shown within the disadvantages section.

Let's first see how this looks like on an existing example. In *Figure 4: The Marbles LD browser* (Bizer, 2003) below it will be shown how information from different sources is combined when browsing in Marbles (a LD browser) data about Tim Berners-Lee. The colored dots show the different data sources. Here we see about 7 different data sources, which are integrated into a single view to provide the user with as much as possible perspectives and facets of his/her query. The proposed sources are all listed below in order to secure proper selection by the user.

[9] Linked Life Data: http://linkedlifedata.com

[10] Bing, Google, Yahoo! and Yandex, schema.org: http://bit.ly/196asth2013.

[11] Google Inc., Knowledge Graph: http://bit.ly/18BtaMI2012.

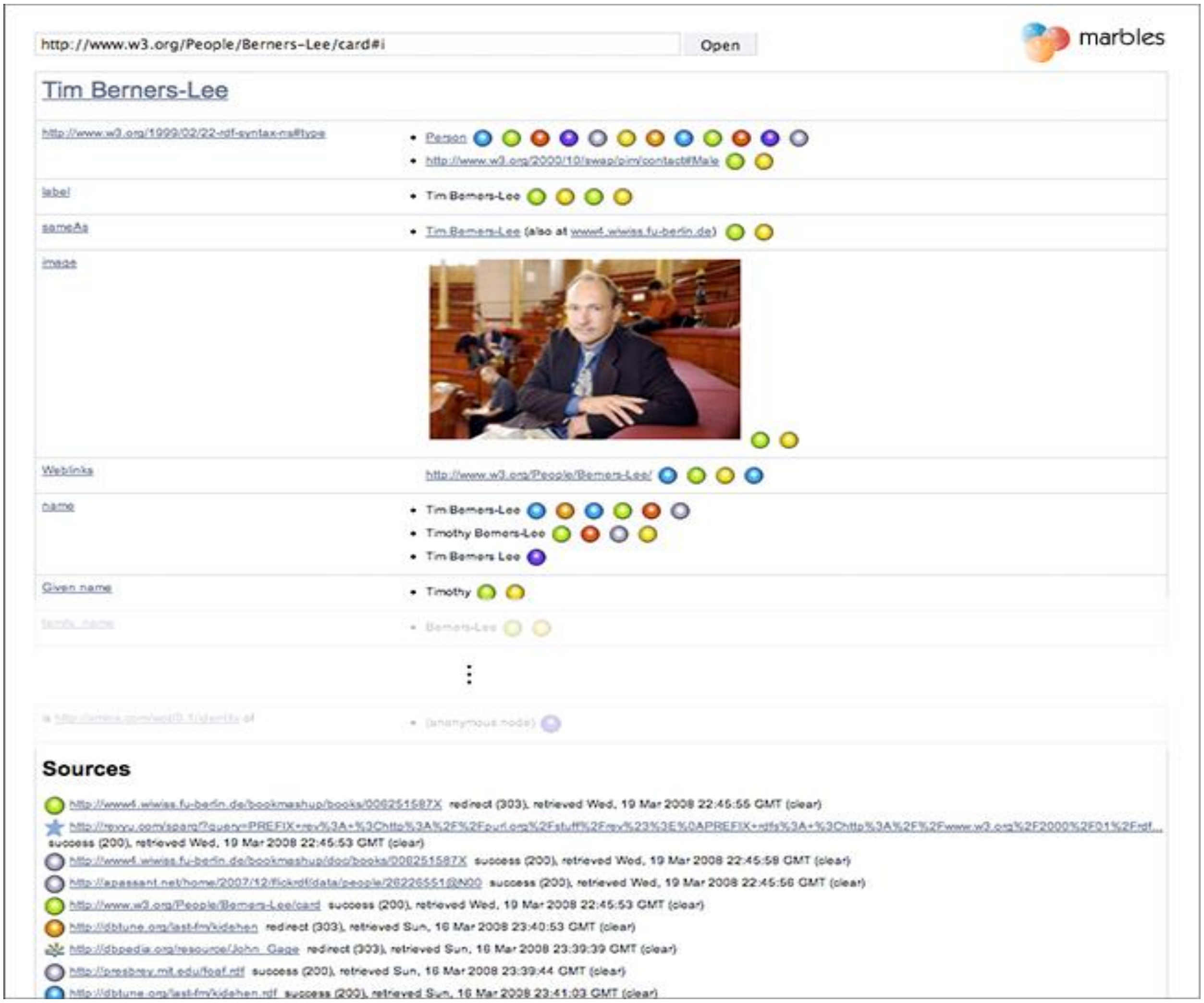


Figure 10: The Marbles LD Browser (Bizer, 2003)

Another of the main reasons for considering the application of LD within business processes is its machine readability. The original and main goal of the Semantic Web as formulated by (Berners-Lee, 2000, pp.191): "is putting data on the Web in a form that machines can naturally understand, or converting it into that form. This creates what I call a Semantic Web – a web of data that can be processed directly or indirectly by machines". LD is a means to achieve this goal. The LD principles can lower the data reuse barrier, as well as ease its integration and application (Bizer et al., 2009).

- LD is thus data published on the Web, machine readable, with explicitly defined meaning, typed linkages to other external data sets and can be linked to from external data sets (Bizer et al., 2009). There are four main rules to publish data as LD, also called the "LD design principles". Those are:
- Use URIs as names for things
- Use HTTP URIs so that people can look up those names

- When someone looks up an URI, provide useful information, using the standards (RDF, SPARQL)
- Include links to other URIs, so that they can discover more things (Berners-Lee, 2006). LD works with RDF (triples) as basic data representation language to prevent syntactic issues. Furthermore it uses vocabularies and ontologies created in formally well-defined language such as OWL to vanish interoperability issues (Hitzler and Janowicz, 2013). Ontology or taxonomy is "a concept hierarchy in which the broader-narrower relations between concepts are stored" (Meijer et al., 2014). It is defined as "an explicit specification of conceptualization" (Gruber, 1993). With "conceptualization" it is meant that ontologies are an abstract model of a given domain of knowledge, e.g. an ontology can be an abstract model of the reality of the financial domain. "Explicit specification" in this definition refers to the specified concepts, their attributes and the defined relationships between them. They are based on classes and instances, where classes model the domain structure and instances belong to classes, modelling the "ground level" objects. In the case of ontologies, two main processes can be distinguished – ontology building and ontology mapping (Ding and Foo, 2002). Ontology building can be done manually, but since it is a very time- and resource-intensive process, semi- and fully-automated approaches have been developed. Furthermore an ontology can be created bottom-up, top-down or using a hybrid approach. The bottom-up usually starts with text documents and moves from specification to generalization. Top-down would start with top level concepts and move from generalization to specification. A hybrid would start with the most important concepts and move to generalization and specification. Ontology mapping likewise is very crucial, since more and more ontologies are created and their re-use becomes essential (Du and Zhou, 2012). Ontology mapping can be done using an one-to-one approach, ontology clustering but also it can also be done as creation as a single-shared ontology. Ontologies are useful not only for information search but also for classification and navigation through data (Berners-Lee et al., 2001). Moreover within a specific domain they play an important role in improving information consistency, reusability, systems interoperability and knowledge sharing. Next to that they can improve information organization, management and understanding (Fensel and Brodie, 2003). Therefore they can be used in decision support systems. However due to the heterogeneity of used vocabularies and ontologies, the context plays a very important role. The context is "largely determined by space and time" (Janowicz, 2010). This will be further elaborated on within the advantages and disadvantages sections.

6.2 Open linked data and its magnitude.

The size of LD is continuously growing in number and participants. However it is difficult to judge about its exact magnitude and/or growth rate, since only a part of it is made available with linked open data, hence LOD. However the size of LOD and its growth rate can somewhat give an indication of the growing importance of LD. This is best illustrated with the so called LOD cloud. It consists of 1014 data sets in total, in eight different domains and the linkage relationships between them (Schmachtenberg et al., 2014). It is shown in the figure below.

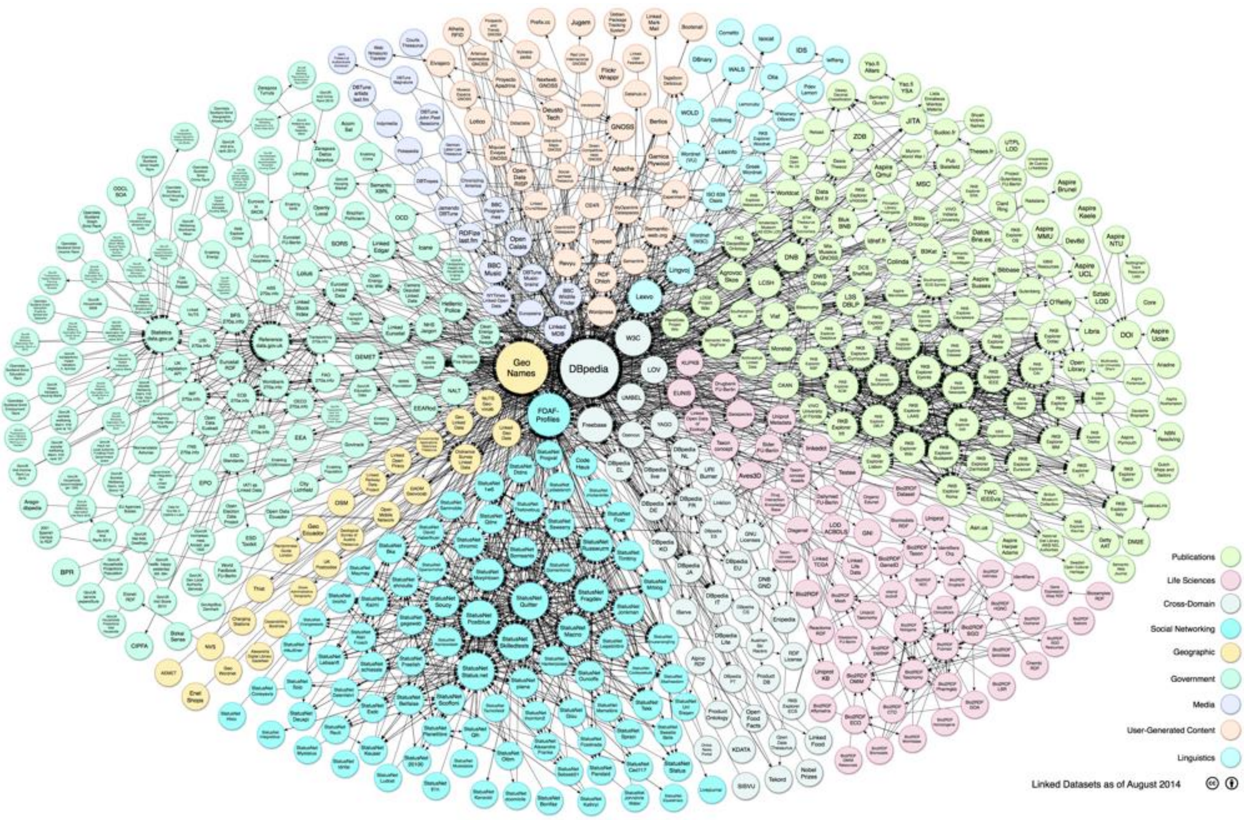


Figure 11: The Linking Open Data Cloud (Schmachtenberg et al., 2014a)

The arcs in figure 2 indicate existing links between items in the two corresponding data sets. Heavier arcs correspond to higher number of linkage relationships between two data sets. The size of the bubbles shows the relative number of RDF triples within a data set, while the colors represent the topical domain. The three biggest domains currently, as shown in the picture, are the Social Web, Government and Publications. The complete overview with number of datasets and a percentage of the total number datasets per topical domain is given in *table 7: Datasets by topical domain* (Schmachtenberg et al., 2014).

Datasets by topical domain.		
Topic	**Datasets**	**%**
Social web	520	51.28%
Government	183	18.05%
Publications	96	9.47%
Life sciences	83	8.19%
User-generated content	48	4.73%
Cross-domain	41	4.04%
Media	22	2.17%
Geographic	21	2.07%
Total	1014	

Table 7: Datasets by topical domain (Schmachtenberg et al., 2014)

As the reader may notice the Financial (services) domain is not yet present in the overview of datasets per topical domain. This may indicate that LD (LOD) did not yet find application within this domain. However it may also be interpreted in a way that the usage of LD in the financial domain is not done in the form of open data but moreover using LD instead of LOD. Furthermore each of the topical domains has financial aspects involved, in form of accounting, innvestments, fincancing ect.

The size of the Web of Data, in the representation of LOD is further estimated based on the number of RDF triples and the linkages between them by the community in the ESW wiki. In 2009 it consists of 4.7 billion RDF triples, interlinked by around 142 million RDF links (Bizer et al., 2009). The last data available[12] is from 2010, where the number of triples increased up to four times its previous size for about an year, thereby consisting of 19.6 billion RDF triples (W3C, 2010). This is a very up forth growing pace. However it cannot be assumed that this corresponds directly with a continuous growth rate of the LOD cloud, since it is based on one year's observation. Furthermore, this does not imply a direct correspondence with the growth rate of LD, since only LOD is accounted for. A part of this rapid growth could be for instance accustomed by the creation or converting into LOD of a

[12] As of December 2014.

large dataset in 2010. Furthermore, the growth in number of triples in 2010 could be an exceptional case, thus not an indicator of a constant growth rate. However, it does illustrate that more and more LOD is published and/or more data is converted into LD according to the LD principles mentioned before. This is a certain clue about the growing in importance of LD. The focus of this work lies, however, on LD (and not LOD) within the financial domain.

6.3 Linked data within the financial domain.

Financial decision-making applications can be divided into five categories: stock forecasting, portfolio management, bankruptcy prediction, foreign exchange market and fraud detection (Zhang and Zhou, 2004). All of them depend on the collection of data. Poor data quality can have enormous negative social and economic impact (Wang and Strong, 1996). Poor data quality can also have a variety of issues, the primary of which according to Bansal et al. (1993) are data consistency, data accuracy and data integrity. They are especially applicable to financial data, since it is highly time-variant, non-linear and noisy (Bansal et al., 1993). The noise of financial data includes dynamic noise and observation noise. The dynamic noise is related to distortions in the obtained information, for instance due to changes of the factors after measurement occurred. Observation noise, on the other hand, corresponds with the accuracy of the measurement and includes distortions during the measurement Therefore two lines of distortion can occur – one that results from differences between the real-world system and the view inferred from the information system and another one from differences between the real-world system and the observed view of the real-world system (Wand and Wang, 1996). Furthermore data quality concerns both, objective, i.e. intrinsic to the data and contextual aspects, i.e. vary across users and tasks. The use of LD approaches and ontologies can improve the quality of online financial data and support decision-making in finance and in other domains, in which data is spread across multiple sources with overlap but complementary in content, even if they are different in data format and naming schemes. Du and Zhou (2012) develop an ontology to grasp this potential. The proposed framework has two dimensions – completeness, unambiguity, correctness and meaningfulness[13] being the one dimension, based on data quality problems, and concept and instance, based on the abstraction level in ontologies (Du and Zhou, 2012). This results in six categories of data quality problems: terminological ambiguity, conceptual inaccuracy, missing data, unreliable data, inconsistent representation and incomplete domain. They are shown in the table below.

[13] Meaningfulness is left without consideration outside the discussed framework, since meaningfulness of the financial data to the prospect users is expected, since the financial domain is a well-established domain.

	Completeness	Unambiguity	Correctness
Concept	Incomplete domain	Terminological ambiguity, inconsistent representation	Conceptual inaccuracy
Instance	Missing data	Inconsistent representation	Unreliable data

Table 8: Classification schema of data quality problems (Du and Zhou, 2012).

On the concept level “Incomplete domain” refers to lack of complete knowledge of a certain domain (Wand and Wang, 1996). For instance, financial firms have no inventory but this is not the same as having inventory on the balance sheet equal 0. “Terminological ambiguity” would resolve for instance following problem: Yahoo!Finance uses accounts payable in terms of taxes payable and notes payable, which differs from the definitions within Google Finance and MSN Money Central. Google Finance considers only the taxes payable, whereas MSN Money Central’s definition includes taxes payable, notes payable and reserves. Therefore the developed ontology can help in cases where the same term is used for different concepts. Furthermore it can deal with situations, where different terms are used but refer to the same concept, i.e. sales in Compustat implies the same meaning as revenue in Google Finance. “Conceptual inaccuracy” corresponds with the lack of precision in concept definitions, e.g. revenue on the Income Statement is the sum of revenue of goods and revenue of service. Therefore a revenue in Google Finance can be interpreted as revenue of goods, revenue of service or a combination of both, when sufficient verification information is missing.
Furthermore on an instance level, data can be missing. Again it is challenging to determine whether the data is actually missing or the corresponding position in the financial statement equals zero. Inconsistent representation of financial data across different sources might be for instance that the same data is represented in units of millions or thousands of USD (Du and Zhou, 2012).

The created ontology addresses all of the aforementioned issues. Furthermore it is tested with empirical data and proven to improve the performance of financial decision-making (Du and Zhou, 2012). The framework consists of three components: Financial ontology (FinO), online financial data resources and financial decision-making. It is shown in the figure below.

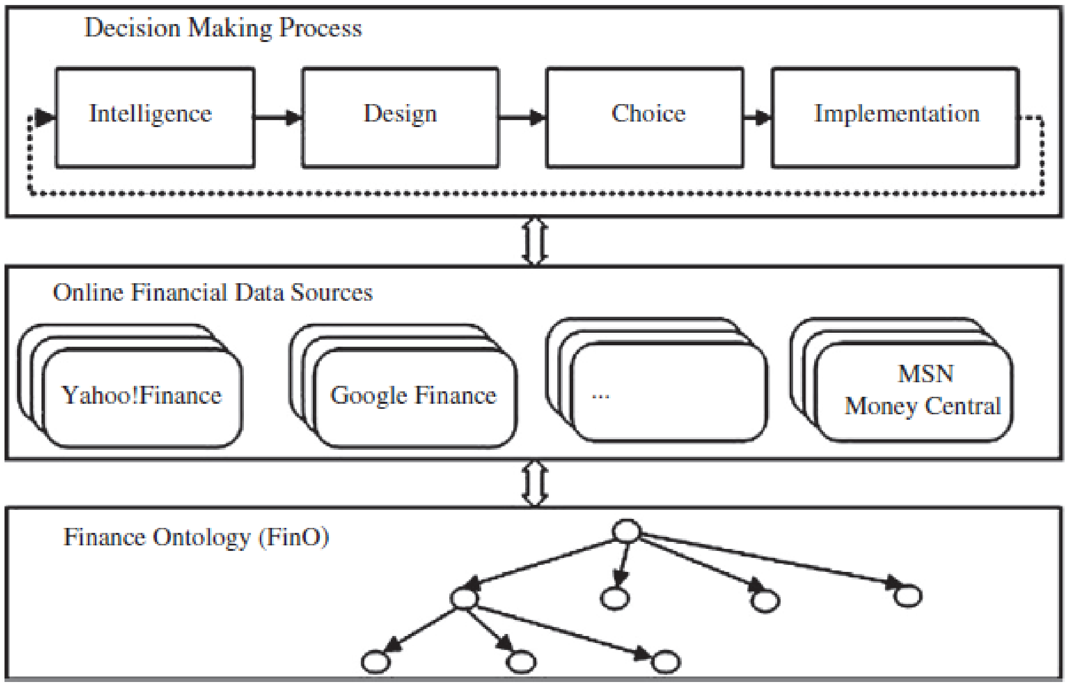


Figure 12: OFFDM: Ontology-based Framework in support of Financial Decision-Making (Du and Zhou, 2012)

The decision-making process starts with the intelligence phase, where data is collected. In the design phase possible solutions and their pros and cons are taken into consideration. In the next step one solution is chosen and implemented in the last fourth step of the decision-making process. In the first step data is gathered from diverse online resources, whereas FinO facilitates the ontology mapping between the different data sources. Furthermore, it supports the financial decision-making by addressing data quality (Du and Zhou, 2012). FinO "interoperates" various "financial data sources" via ontology mapping (Du and Zhou, 2012) and is created using the single-shared ontology mapping approach, which means that there is a common ontology to which every specific ontology is mapped. It provides a common layer to ensure access to several ontologies and to allow for information exchange according to the semantic principles. For instance, at the concept level concepts of FinO are mapped to concepts from the ontologies used by individual online financial resources, such as Google Finance or Compustat. At the instance level a missing value in one source can be identified in another of the resources, using the mapping of their ontologies. FinO is to be developed top-down or according to the hybrid approach, since finance is a well-established domain. Existing ontologies, which could support the development of FinO are mentioned below. Du and Zhou (2012) develop such ontology of income statement, based on those ontologies and other online financial resources, such as Google Finance, MSN Money Central etc. It is shown below.

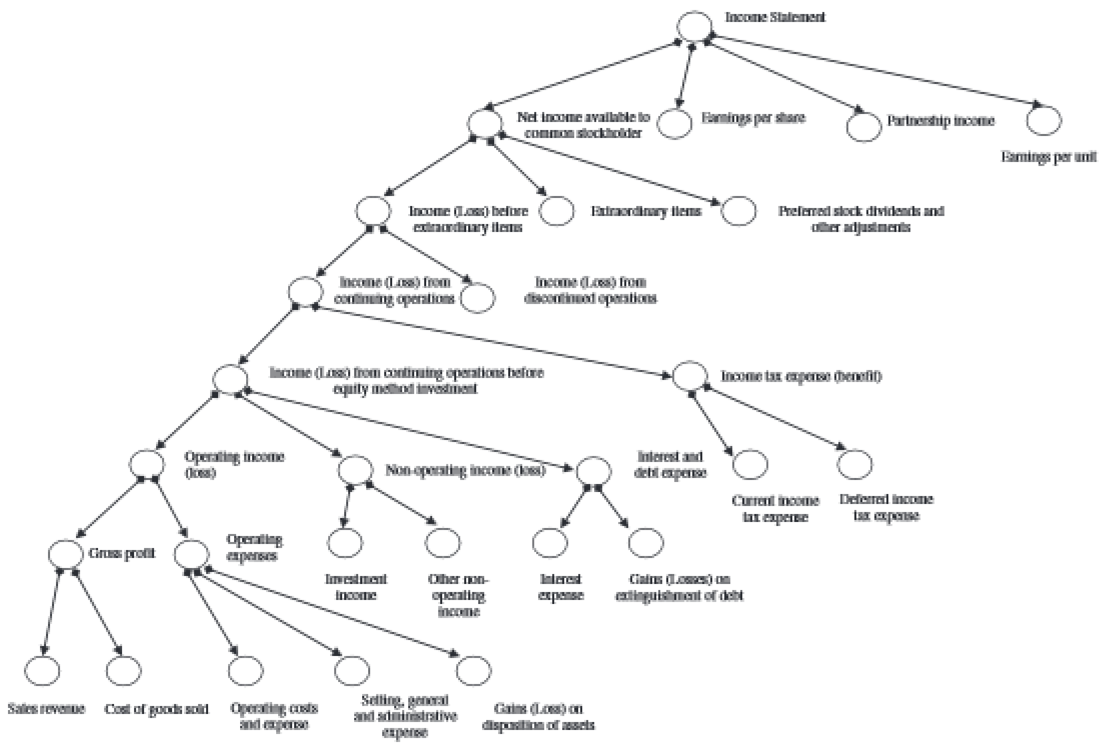


Figure 13: Ontology of income statement (Du and Zhou, 2012).

Data quality is a key factor for decision-making performance. Studies show that 20% of asset managers, investment bankers and hedge fund professionals spend 25% to 50% of their time in data validation efforts, which keeps them from focusing on tasks contributing to the bottom line (Valiante, 2008). Therefore improving automated validation processes can significantly improve decision performance, provided that the data quality will not suffer from it. Moreover financial data need a synergetic semantic alignment of a variety of resources in order to improve its quality, since it has one unique feature – redundancy. Financial data is duplicated across online sources such as Google Finance, Yahoo!Finance, MSN Money Central and Compustat, which contain partially the same financial data but are yet complementary. The data is heterogeneous across the different sources although the financial domain is highly regulated. This illustrates the value-added of the aforementioned framework of Du and Zhou (2012).

More than these ontologies have been developed and applied in the financial domain (Sujanani et al., 2005). develop an ontology to facilitate the communication among agents in a multi-agent financial investment system. Kanellopoulos et al. (2007) promote an ontology to facilitate the predictions of firms with fraudulent financial statements. Kingston et al. (2006) develop an ontology to support investigation in detection of fraudulent financial sites. As a support for developing of new ontologies several ontologies within the financial domain can be used:

- SUMO[14] (Suggested Upper Merged Ontology) – it represents high-level concepts and relationships between concepts in the financial domain.
- LSDIS[15] (Large Scale Distributed Information Systems) Finance Ontology – extension of SUMO.
- Finance Ontology[16] – written in OWL to represent the knowledge in the financial services domain incl. the financial statements.
- XBRL US GAAP Taxonomies v1.0[17] – provides concepts from financial statements and relationships between them (Du and Zhou, 2012) The data instances are official financial statements, encoded in XBRL. They are then submitted to the SEC by publicly traded companies (Zhu and Wu, 2014).

Furthermore an ontology-driven framework for product information can be used likewise in the financial industry for aggregating financial product information from a variety of sources. Due to unawareness of syntactical differences and language dependency, current search engines cannot allow web-wide parametric search (Nederstigt et al., 2014). Therefore it is difficult for customers to identify (financial) product that would best match their needs and often they are forced to make a decision only based on price characteristics. However this is suboptimal to the buyer as well as to the seller (Nederstigt et al., 2014). FLOPPIES would enable the build of a web-wide knowledge base that includes (financial) product attributes and enables product comparisons over a variety of data sources. The sellers don't need to comply to the same data format when publishing data because the framework is ontology-based and can get product information itself without the need for a template (Nederstigt et al., 2014). This framework can thus enormously ease the comparison over financial products, given their complexity and diversity.

Another application of ontologies and LD can be found in bank stress testing. Bank stress testing relates to modeling and analyzing "exceptional but plausible" risk scenarios (Hu et al., 2014). Such risks can be for instance the 2008 banking crisis, the bankruptcy of Lehman Brothers etc. One of the major reasons for the crisis is that the financial stakeholders failed to model the "exceptional but plausible" scenarios in bank stress testing (Hu et al., 2010). Those events are highly important and beyond normal expectations and predictions and are therefore called "Black Swan" events (Taleb, 2011). There are three major challenges in the prediction of "Black Swan" events:

[14] http://seigmakee.cvs.sourceforge.net/viewvc/sigmakee/KBs/FinancialOntology.kif?view=log.
[15] http://lsdis.cs.uga.edu/projects/meteor-s/wsdl-s/ontologies/LSDIS_FInance.owl.
[16] http://www.fadyart.com/ontology.html.
[17] http://www.xbrl.us/Pages/US-GAAP.aspx.

- Current stress testing methods rely on probability-based models and use historical financial data, such as the Value-at-Risk measure (Jorion, 1997). However “Black Swan” events are very rare and usually have no precedents.
- It is difficult for stress testing scenario designers to imagine all kinds of possible scenarios, since “Black swan” events are too rare to imagine and there is a groupthink within a profession that is not easy to overcome.
- The scenarios must be checked for plausibility. However it is not easy to design plausible stress testing scenarios due to the risk complexities of those events and their interactions.

In order to overcome the first challenge, (Hu et al., 2014) develop OESM, Ontology-based Event-driven Scenario Model, which gives a formal representation of stress testing domain knowledge. Furthermore they propose two analysis methods, based on OESM for scenario recommendation and plausibility check to address the last two challenges. The ontology-based scenario model and the two analysis methods are being addressed as BESST. BESST is shown in the figure below.

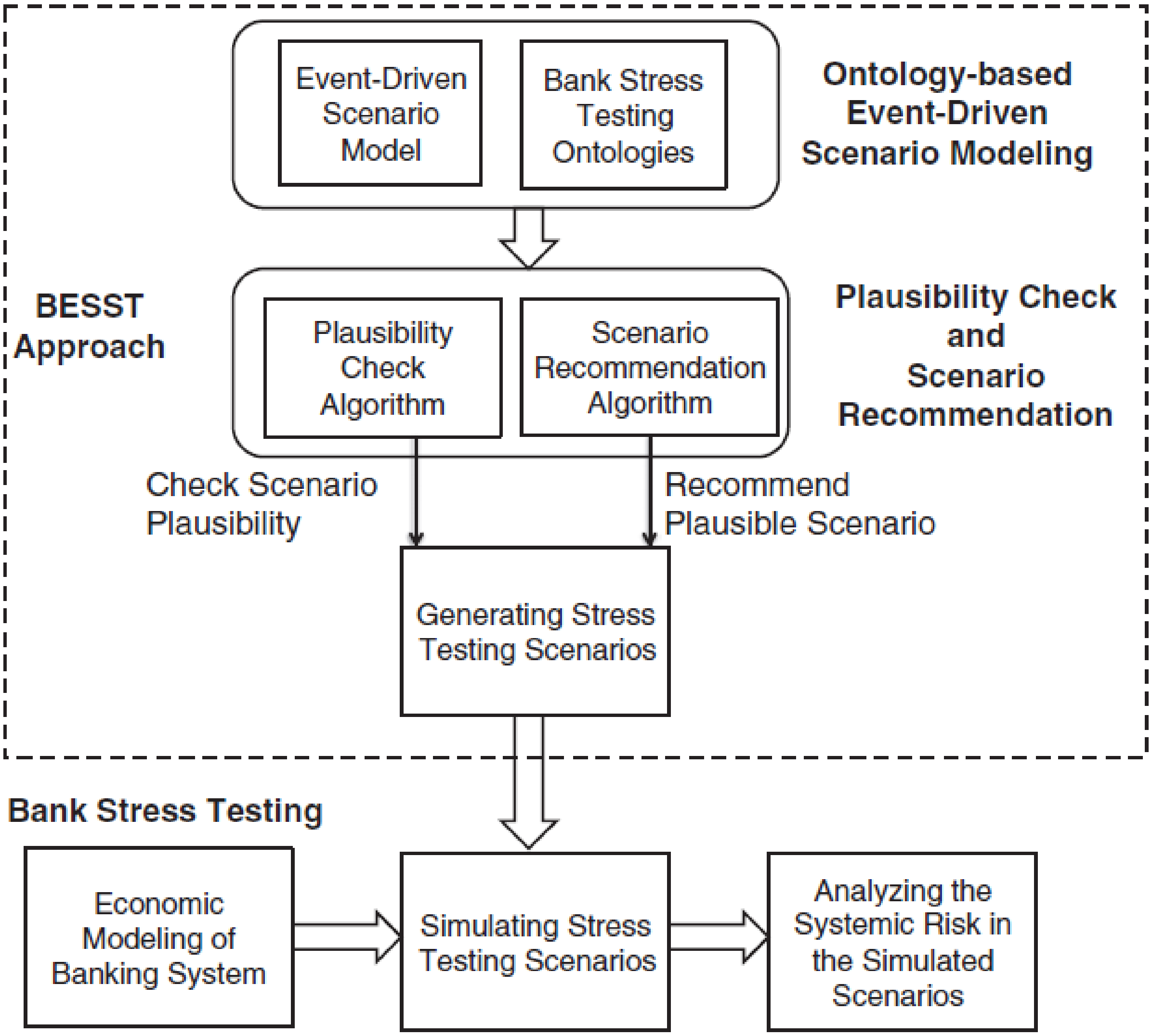


Figure 14: BESST for stress testing (Hu et al., 2014).

As shown in the figure, it is a multistage process that consists of economic modeling of the banking system, simulating stress testing scenarios and analyzing the impacts. Hu et al. (2014) are the first to introduce ontology and conceptual modeling into bank stress testing scenarios. Furthermore BESST is the first non-probability-based method for modeling and analyzing “exceptional but plausible” stress testing scenarios without the use of historical data. Therefore the BESST approach is an important addition to among others financial risk management (Zopounidis et al., 1997). Moreover when the ontology is formalized with logic languages, such as first-order logic, it can support logical deduction mechanisms, used to derive new facts and check the logical consistency on those facts. Therefore ontologies support knowledge representations and logical deductions of risk events, and consistency checks, which are all essential to bank stress testing (Hu et al., 2014).

Another development in the financial industry is proposed by (Chan and Franklin, 2011) for financial sequence prediction. As mentioned before humans are inherently limited in their ability to process information (Simon, 1982). Therefore there is the need for a decision support system to bring to foresee hidden regularities and trends in price movements included in financial news. The proposed system is again based on semantics and can extract knowledge from text, with limited human intervention (Chan and Franklin, 2011). This text-based decision support system (DSS) extracts event sequences from shallow financial text patterns and gives a prediction of the likelihood of the events occurrence using a classifier-based inference engine. It assess incoming, even new and unseen, event sequences in its predictions. Both explicit and implicit information can be accessed depending on the needs of the user. DSS uses a corpus of financial documents, extracts all of the event sequences from the texts and predicts interesting and unseen relationships between them. Traditional financial text mining techniques or knowledge discovery from text (KDT) are based on keywords and their main goal is to fill in values for predefined template slots. DSS on the other hand provides an "automated means of learning shallow event patterns from text" (Chan and Franklin, 2011). For example "shortfall", "risk of default", "resignation" would give an indication of possible fall in the company's stock price, even in a case of reported sound financial figures. Opposed, "merger", "acquisition", "alliance" indicate an upward trend (Chan and Franklin, 2011). Furthermore methods exist to automatically discover semantic relationships between concepts (words or phrases). They are based on semantic mining, i.e. discovering of relationships on the bases of the concepts co-occurrence in the same documents/ in the same vicinity of each other within documents. One of them is the empirically proven Web mining approach by (Roussinov and Zhao, 2003) It has been applied for automated summarization of meeting messages. For this method not only the concept co-occurrence is employed but also the context is taken into account within a new developed approach called context sensitive similarity discovery (Roussinov and Zhao, 2003). Nevertheless the experiments confirm that the context is crucial for the consistency of the similarity relationships as indicated in the previous chapters.

7 Summary of the advantages of LD.

As discussed in the previous sections Web 2.0 mashups work against a fixed set of data sources, whereas the LD applications operate on top of an "unbound, global data space." Thus an answer of a query can be provided based on the information of unlimited (number of) sources, which enables different perspectives to come into play. Furthermore the "decentralized publication model" (Tummarello et al., 2007) allows for anyone with information at his disposal to make it available. With LD all those sources are considered when providing an answer of an user's query. Not only can then applications use all the information provided but using it does not necessitate changes in the application code. This contributes connecting different data depositories currently available on the Web into a single global information space. As touched upon in previous sections, this is arguably the main benefit of LD from users' perspective according to Bizer (2009) and also one of the main building blocks of the Semantic Web.

Not only are data re-use barriers lowered with LD but it also allows for higher efficiency in data re-use. This is due to the new exchange paradigm associated with it. Traditional standards foresee exchanging messages when the relevant event has occurred, e.g. after delivery, an invoice is sent. However with LD no exchange of information takes place but moreover referencing to it. Instead of exchanging data, data is kept by its source and it's simply linked to when querying for information (Folmer and Krukkert, 2015). This reduces redundancy in information exchange by simply linking it. Furthermore in this way only one stakeholder needs to maintain certain data (a specific dataset). All other stakeholders can simply query the already maintained "up-to-date" data as needed.

LD also improves cross-sectorial interoperability. This is a very important issue, since B2B or B2G collaboration typically goes across sectors. However standards are usually developed within a specific sector, which in turn very often leads to interoperability problems. An example for interoperability issues can be shown on the Dutch invoicing. The Dutch governmental organizations set a target of receiving up to 80% of the invoices electronically by the end of 2014. The target for 2010 was 10%. However every sector has its standard for invoice messaging. This creates a barrier for the Dutch government to cope with all those different invoicing standards. A major source of such issues are the different vocabularies and ontologies that are being used and as highlighted in the previous sections this leads to huge interoperability barriers. However "it is not realistic to believe that all those [standards] will be replaced by a single European or worldwide standard." as Sonntagbauer states (Gusev and Mitrevski, 2011). LD provides a more realistic solution to this issue; not per se keeping data at the source but communication with no restriction to format or layout. Such solution implies not the enforcement of a standard but a 'semantic model' to which all electronic invoices must comply. Therefore it vanishes interoperability issues but without demanding a switch to another standard from the different domains. The implementation of this semantic standard is a great advantage and its implementation not only on national but also on European level is planned for the coming years. Several standards should at least be supported and a mapping to the semantic model must be ensured. LD improves the transaction-based enterprise interoperability, since it addresses issues of the current XML-based interoperability such as lack of dynamics and cost of implementation, adoption and cross-industry exchange (Folmer and Krukkert, 2015). This then leads to significant operating cost savings. Savings resulting from solving interoperability issues amount to "at least one billion dollars" within the automobile sector in the US and about 3.9 billion within the electro technical industry (Steinfield et al., 2011).

The greatest potential for LD according to Folmer and Krukkert (2015), however, is seen in "specifying semantics formally". This not only contributes to interoperability improvements but the data semantics can be now used not only for semantic search, but for data integration purposes as well, i.e. linking ontologies and schemas. That way legacy systems remain unaltered, while different standards are overcome by (semi-) automated generation of transformation schema's. However the creation of an ontology is a difficult and time consuming process, since it requires a massive amount of knowledge and time to organize a high number of concepts properly. Therefore as mentioned before there are automated and semi-automated approaches. Meijer et al. (2014) propose ATCT, Automatic Taxonomy Construction from Text, which is a framework for the automatic building of a domain taxonomy from text corpora. The taxonomy is built in four steps. First, from a corpus of documents, terms are extracted. These can be articles on a certain theme, documents from different data sources, or company's internal documents. Then the most relevant of those terms for the specific domain are selected using a filtering approach. In the third step the selected terms are disambiguated and concepts are generated. In the final step relations between the concepts are set. This is done with a subsumption technique, which uses the concept co-occurrence in a text (Meijer et al., 2014). This process is shown in the figure below.

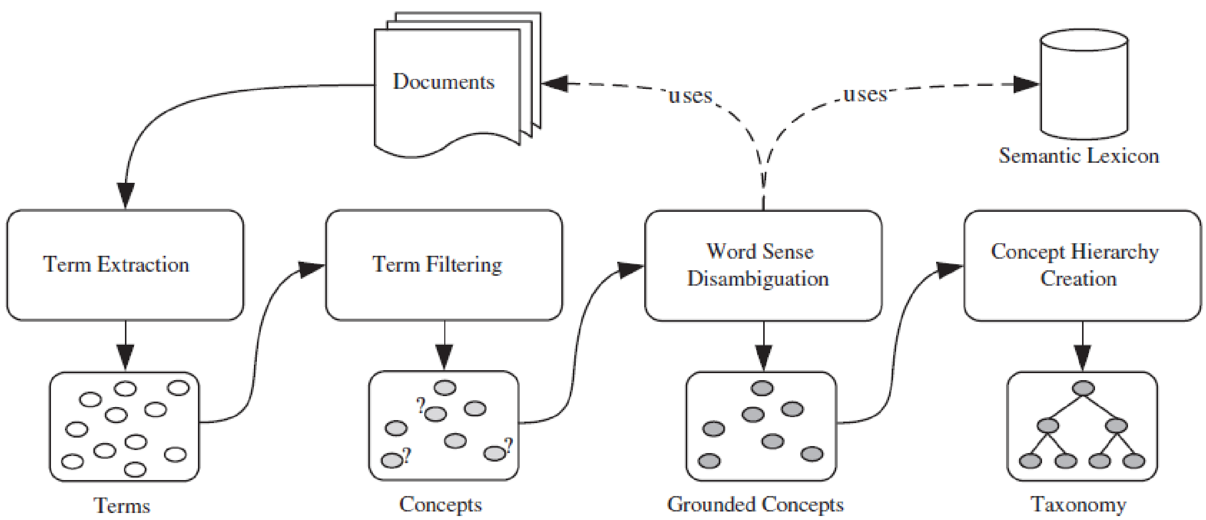


Figure 15: The ATCT framework (Meijer et al., 2014)

In the case of ATCT, the knowledge acquisition is automatic and it provides up-to-date knowledge that can be used in the decision process in real-time businesses in a variety of tasks. The taxonomy can then be further developed into becoming an ontology. Ontologies are useful not only for information search but also for classification and navigation through data (Berners-Lee et al., 2001). Moreover within a specific domain they play an important role in improving information consistency, reusability, systems interoperability and knowledge sharing. Next to that they can improve information organization, management and understanding (Fensel and Brodie, 2003). The aforementioned taxonomy, ATCT can be used for summarizing information from different data sources, for support recommendation systems, for faceted search applications (Vandic et al., 2012) or for filtering, enriching and improving the quality of data used in support systems (Meijer et al., 2014).

As mentioned earlier current search engines are keyword-based and fail to work with syntactical differences. Furthermore they are language-dependent. Using semantics (and LD) helps in overcoming these issues. Next to that LD could also cope with (slightly) different semantics in communication and different versioning. Therefore it enables better reuse of existing data (Folmer and Krukkert, 2015). A recent development is FLOPPIES, which is a Framework for Large-scale Ontology Population of Product Information in E-commerce Stores. It is ontology-driven framework that uses the tabular data on Web store product pages and creates a structured knowledge base of product information. The proposed ontology is OntoProduct and it is mapped to the GoodRelations ontology for e-commerce. Linking to already existing ontologies or ontology mapping is really advantageous and essential, as explained in the previous sections. Since product information is formalized in the ontology, better product comparisons can be carried out and more advanced recommendation applications can be build. Furthermore there is no more need for the Web stores to provide data in a certain format. Search engines will be able to pull this information from the Web stores themselves. This information can be then easily aggregated over a variety of sources of product information (Nederstigt et al., 2014). Thereby the information heterogeneity problem will be solved, which would lead to massive improvements in business information exchange (Ng et al., 2000). Furthermore the product retrieval capabilities of the consumers will increase due to the more intelligent product search engines. The engines will provide more reliable rankings, since they can reason about product attributes' values and how they relate to each other. For instance, if "HSPDA" is faster than "3G" and "3G" is faster than "GPRS", a semantic search engine will be able to derive from that that "HSPDA" is faster than "GPRS". The created knowledge database is understandable not only for humans but also for machines, which enables (semi-)automatic reasoning (Nederstigt et al., 2014). FLOPPIES can be applied to different products, also in gathering financial product information. For instance, if company A has a higher profit margin than company B and company B – higher than company C, then a semantic search engine can reason that company A has a higher profit margin than company C.

Semantics can also enhance the world of web services. Adding semantics to web services' in- and output and defining relationships leads to reduced manual effort required for discovering and using web services. The so called semantic web services use ontologies to provide a more powerful method for the discovery of web services (Pedrinaci, 2010). This is of high importance since Web services have become common in the areas of B2B integration and enterprise application integration (Narock et al., 2014). Furthermore many scientific domains have to deal with increased data volumes. This makes it impractical to copy data and perform local analysis. Instead scientific analysis can take place on the Web through online tools that combine and mine pools of data (Goble and De Roure, 2009). This is where LD comes in. Service provenance can be exposed as LD thus enabling both, human and machine dereferencing of web service execution details. Service application details can be linked to domain specific semantic information published by other parties. This linkage can then be used for answering of more detailed provenance questions (Ding et al., 2011).

However intelligent applications are needed to exploit the implicit semantics and provide more insightful analysis. Nebot and Berlanga (2012) propose the building of a data warehouse. A data warehouse is "a decision-support tool that collects its data from operational databases and various external sources, transforms them into information and makes that information available to decision makers in a consolidated and consistent manner" (Kimball and Ross, 2002) It is useful in traditional business analysis and decision-making processes (Nebot and Berlanga, 2012). The method proposed by Nebot and Berlanga (2012) allows for efficient analysis and exploring of large amounts of semantic (linked) data, so that sophisticated queries can be expressed and evaluated.

The reason this is beneficial lies in the limited human abilities in information processing. The efficient market hypothesis (EMH) in finance illustrates price movements as unbiased or rational reflections of all past or current publicly available information about future earnings expectations. However the models of bounded rationality claim that humans are inherently limited in their ability to process information (Simon, 1982). Therefore there is the need for a decision support system to bring to foresee hidden regularities and trends in the financial domain. If it is based on semantics, it can extract knowledge from text, with limited human intervention (Chan and Franklin, 2011).

Furthermore there are clear advantages in using LD specifically in the financial domain and business processes. An object-oriented approach, with an extended semantic modelling capability turns out to be better suitable for modelling the financial accounting systems. Financial accounting information nowadays is used to perform many tasks, such as reporting profit, calculate tax, evaluate investment options, performance analyses, compute value of assets and design financial strategies. Therefore the development of a computerized financial accounting system, is of critical importance for managers and stakeholders of business firms (Chu, 1992). A data model can be judged by its semantic expressiveness, i.e. the ability to capture entities with their specific characteristics and features, as well as the relationships between them (Hammer & McLeod, 1981). Chu (1992) furthermore shows that relational data models are not an effective tool for modelling the financial accounting systems, although it is currently dominant at the market. This is caused by the lack of intrinsic capability to model generalizing abstractions and complex structures, and the lack of control over accounting procedures. An object-oriented model however seems to be "an extremely good fit to accounting systems" (Chu, 1992), which proves that it is not only applicable within the development of new applications but also in conventional business areas (Chu, 1992). Moreover it enables high reusability. An overview of the aforementioned advantages associated with LD is provided below.

Advantages for LD application within (financial) business processes	
Statement from the literature	Description
More perspectives come into play	LD applications operate on top of an "unbound, global data space" and not against a fixed (predefined) set of data sources (Bizer et al., 2009)
"AAAAA" principle	"decentralized publication model" (Hitzler and Janowicz, 2013) and therefore "Anyone can say Anything about Any topic"
Flexibility, reduced re-use barriers	Making use of information from additional data sources does not necessitate changes in the application code (Tummarello et al., 2007)
Consistency, promoting the Semantic Web	Contributes connecting different data depositories currently available on the Web into a single global information space (Bizer et al., 2009)
Reduced redundancy in information exchange	new exchange paradigm, simply linking to information (Folmer and Krukkert, 2015); only one instance has to manage/maintain certain data (a specific data set)
Vanishes cross-sectorial interoperability issues	This leads to significant operating cost savings (Folmer and Krukkert, 2015) & (Steinfield et al., 2011)
Automatic knowledge acquisition and up-to-date knowledge	Ontology development automatic (Meijer et al., 2014); Can be used for decision-making processes in real-time businesses

Allows for (semi-) automatic reasoning	Specifies semantics formally; The created knowledge database is understandable not only for humans but also for machines (Folmer and Krukkert, 2015) & (Nederstigt et al., 2014)
Solves information heterogeneity problem	This leads to massive improvements in business information exchange (Ng et al., 2000)
Reduced manual effort for discovering and using web services	By adding semantics to web services' in- and output and defining relationships (Pedrinaci, 2010)
Human and machine dereferencing of web service execution details	When service provenance is exposed as LOD (Ding et al., 2011)
Building of a data warehouse	allows for efficient analysis and exploring of large amounts of semantic (linked) data, so that sophisticated queries can be expressed and evaluated (Nebot and Berlanga, 2012)
"Antidote" for the human bounded rationality	a decision support system to bring to foresee hidden regularities and trends in the financial domain (Chan and Franklin, 2011) helps the human bounded rationality in information processing abilities;
Advantages for LD application within the financial domain	
Better (financial) product comparisons; more advanced recommendation applications;	(financial) product information is formalized; no need for providers to use templates or specific data format (Nederstigt et al., 2014); provision of aggregated information over a variety of sources
More reliable product rankings	Engines can reason about product attributes' values and how they relate to each other (Nederstigt et al., 2014)
Object-oriented approach with an extended semantic modelling capacity is an "extremely good fit to accounting systems"	Not as relational data models, since they lack intrinsic capability to model generalizing abstractions and complex structures, and also lack control over accounting procedures (Chu, 1992)

Table 9: Summary of the advantages for LD application within (financial) business processes

8 Summary of disadvantages associated with LD.

In the previous sections the LD concept and principles were explained in the context of the Semantic Web technology. In the previous chapter the advantages of LD application within (financial) business processes were highlighted. However in order to be able to assess the potential for LD, the disadvantages associated with its application must be likewise taken into consideration.

For instance, the decentralized publication model of LD was until now considered in its one side as an advantage, as it allows for different views and perspectives of and additional information from various sources. However the question arises whether those sources are trustworthy and/or relevant for the user (the user's query). As mentioned earlier the user has not selected the sources explicitly. Even if the user did select them, the absence of statements about the resources arises issues around the selection made. Therefore concerns arise about how to add and remove sources from an integrated, entity-centric view, how to meaningful map data from different data sources, how to deal with licensing issues, since copyright law does not apply to data and/or is treated differently depending on jurisdictions and how to ensure identification and making available of the most relevant or appropriate data to the user's needs. It is crucial to ensure non-violation of privacy by at the same time benefiting from the Web as a single global database (Bizer et al., 2009). A current issue associated with LD is therefore the absence of statements about the encountered resources (Tummarello et al., 2007), which would allow for proper assessment and selection thereof. "Anybody can say Anything about Any topic (Hitzler and Janowicz, 2013) opens up a lot of opportunities but can at the same time endanger quality and reliability of provided query responses. One of the LD principles tries to tackle this issue. It promotes including yield of metadata about the resource when publishing the data. However this is usually information that can only limited be interpreted by the users in their role of outsiders and thus only solves part of the problem. A more complete solution of this problem proposed by (Tummarello et al., 2007) is Sindice. It is "a lookup index over resources crawled on the Semantic Web" (Tummarello et al., 2007). Thus it provides clients with ranked order of statements, according to the current query, while offering additional information for each source (resource) in order to enable an informed choice. Furthermore if current applications support the LD principles, integration with Sindice is trivial (Tummarello et al., 2007).

Next to the trustworthiness issues, there are further challenges that could hamper/could have hampered future/current LD implementation, such as global entity consolidation policies and voluntary or involuntary denial of services due to queries complexity or excessive data etc. (Tummarello et al., 2007). An example of occurred denial of service is

Geonames on February the 2^{nd} 2007. In fact this is also "the first distributed denial of service on the Semantic Web"(GeoNames Blog, 2007) and therefore worth mentioning.

However a solution is proposed by (Tummarello et al., 2007), which provides an alternative to full site crawling and thus prevents a complete denial of service. A crawling ontology is proposed that helps in indexing and operating over excessive quantity of LD[18]. Furthermore it was showed that LD contributes connecting different data depositories into a single global information space. However in the case of LD and not LOD, the RDF links might just lead to "dead ends" for a specific user query.

Next to that however beneficial it might be in terms of reducing the redundancy in information exchange, the associated with LD new exchange paradigm also brings some challenges to foresee. There are both conceptual and technical difficulties (Folmer and Krukkert, 2015). Since this work has clear focus on conceptual advantages and disadvantages of LD, the conceptual challenges will play a major role. Some technical issues will be mentioned due to the strive for completeness. However they will be mentioned and briefly explained without the excess of technical details. First of the problems arising with the application of LD is on a legislative basis. The exchange of "business documents" whether in physical or electronic form is made explicit in current legislation. For example an invoice is defined by the Dutch tax authority as "a document that contains…". However LD promotes keeping data at the source instead of copying them and/or physically exchanging documents and information. Therefore the application of LD requires not only a shift in the mind set but also change in formal definitions. Another concern is security. Since data is no longer exchanged but linked to, if a SparQL endpoint is open, the data can be accessed by anyone, even a competitor within the B2B businesses, since endpoints don't have any security or access policies in place. As a current solution a web-service in between is proposed. It can solve this, as it works like an api (Folmer and Krukkert, 2015). The api is an application program interface and is a set of routines, protocols and tools for software application development. It defines how the software components interact with each other and can control or restrict access.

A purely operative challenge for LD present the legacy systems. As many of them don't support semantic web technology and LD, they have to be replaced and/or adjusted (Folmer and Krukkert, 2015). A solution for this was somewhat touched upon in previous chapters, namely using LD and the formal specification of semantics for data integration purposes, such as (semi-) automated generation of transformation schema's, which would leave legacy systems unaltered.

The main benefit from LD, as highlighted in the previous chapter, is making semantics explicit and thereby machine interpretable. However there are limitations of the LD

[18] http://sindice.com/srobotsfile

expressiveness. For example it can be made explicit that one concept is the "sameAs" another one and a machine can interpret this. However two concepts can be the same within a certain context but really different in different contexts. This again requires a document to specify the different concepts relevant. Let's take a staffing agency for example. For such agency the workers are seen as resources in the same way as the chips are considered resources for a semiconductor company. Therefore semantically the concept "workers" is the "sameAs" the concept "resources" within the context of the staffing industry. Furthermore the concept "chips" is the "sameAs" the concept "resources" within the semiconductor domain. However the concepts "workers" and "chips" certainly do not refer to the same concept outside this contexts. This requires an additional specification of the concepts and a precise definition of the contexts within which different concepts are the same (Folmer and Krukkert, 2015). Such specification can be made, for instance, by specifying that a worker is a person and a chip is a device. Then a machine can immediately detect that both concepts are resources but are not equal.

In case they are the same but have different URIs they are called URI aliases (Bizer et al., 2009). Furthermore this could become somewhat more complex if other relations between concepts are to be specified, such as "similarTo", parenting relationships etc.

Another difficulty with LD is finding a "semantic web" counterpart of a "mandatory field". And as far as a solution was proposed to all of the aforementioned issues and concerns, currently there is no solution of this one [as of February 2015]. With XML a field can be made mandatory in the XML schema. For instance an invoice cannot be send without an invoice number. However with LD no data is being exchanged. Instead the data is kept by the source and can be linked to. This is a critical issue because an invoice for instance is not valid if there is no unique number to validate it. So, even if the exchange paradigm mentioned before is accepted and an invoice does not have to be a "document" that is being exchanged, the issue remains how to ensure that parties know which information is mandatory and needs to be provided and furthermore ensure its provision (Folmer and Krukkert, 2015).

In terms of the financial domain, some of the aforementioned advantages again have another side. For instance, it was promoted that LD and semantic frameworks can enable better (financial) product comparisons and more advanced recommendation applications due to provision of aggregated information over a variety of sources. However the question arises whether this is really applicable to financial products, as usually their features are extremely sensitive information. Even if the framework is designed in a way to extract only non-sensitive, non-competitive information, the benefit is questionable since it might not be suitable for product comparisons any more. The same applies to product rankings.

Next to that Object-oriented approach with an extended semantic modelling capacity was promoted as an "extremely good fit to accounting systems" (Chu, 1992) and its advantages over the relational data models were illustrated. However relational data models are still currently dominant in the market. So, it is still doubtable whether there is enough incentive for this to change on global basis in the near future. Nevertheless LD only makes sense if it is the dominant solution in a certain context or domain.

Below an overview of the mentioned disadvantages for LD application within (financial) business processes is provided.

Disadvantages for LD application within (financial) business processes	
Statement from the literature	Description
Trustworthiness issues	A disadvantage of the decentralized publication "Anybody can say Anything about Any topic" opens up a lot of opportunities but can at the same time endanger quality and reliability of provided query responses (Tummarello et al., 2007) & (Bizer et al., 2009)
Questionable relevance of encountered resources	The user does not select the sources explicitly (Bizer et al., 2009).
Difficulty in proper assessment of resources	Due to absence of statements about the encountered resources (Tummarello et al., 2007)
Difficulty in proper selection of resources	Even including yield of metadata about the resource when publishing the data does not solve the issue completely, since it is information that can only limited be interpreted by the users in their role of outsiders (Tummarello et al., 2007)
Global entity consolidation policies may hamper LD application	LD has the potential of connecting different data depositories currently available on the Web into a single global information space. However in the case of LD and not LOD, the RDF links might just lead to "dead ends" for a specific user query (Tummarello et al., 2007) & (Bizer et al., 2009).
Voluntary or involuntary denial of services	due to queries complexity and/or excessive data is possible. Too complex queries or queries over too excessive data can lead to denial of service. This might endanger service level agreements (Tummarello et al., 2007) & (Bizer et al., 2009) & (Chu, 1992).
The exchange of "business documents" is made explicit in current legislation.	Therefore the application of LD requires not only a shift in the mind set but also change in formal definitions (Folmer and Krukkert, 2015).

Security concerns	Due to SparQL endpoints. If an endpoint is open, the data can be accessed by anyone, even a competitor within the B2B businesses, since endpoints don't have any security or access policies in place (Folmer and Krukkert, 2015).
Legacy systems concerns	A purely operative challenge Many of them don't support semantic web technology and LD, so they have to be replaced and/or adjusted (Folmer and Krukkert, 2015).
Limitations of the LD expressiveness	Different concepts can be the same within a certain context but really different in different contexts. This requires an additional specification of the concepts and a precise definition of the contexts within which different concepts are the same (Folmer and Krukkert, 2015) & (Hammer & McLeod, 1981).
Complexity increases if there are different types of relations	Other possible relations between concepts are "similarTo", parenting relationships etc. (Chowdhuri et al., 2014) & (Zhu and Wu, 2014) & (Bizer et al., 2009).
Difficulty with finding a "semantic web" counterpart of a "mandatory field".	Currently there is none [as of February 2015]. It is a critical issue because an invoice for instance is not valid if there is no unique number to validate it (Folmer and Krukkert, 2015).
Disadvantages for LD application within the financial domain	
Limited applicability of the product comparisons, recommendation applications	Features of financial products are usually extremely sensitive information. Even if the framework is designed in a way to extract only non-sensitive, non-competitive information, the benefit is questionable since it might not be suitable for product comparisons any more. The same applies to product rankings (Nederstigt et al., 2014) & (Bizer et al., 2009).
Object-oriented approach with an extended semantic modelling capacity is not currently dominant at the market	Relational data models are still currently dominant in the market. So, it is still doubtable whether there is enough incentive for this to change on global basis in the near future. Nevertheless LD only makes sense if it is the dominant solution in a certain context or domain (Tummarello et al., 2007) & (Chu, 1992)

Table 10: Summary of the disadvantages for LD application within (financial) business processes

9 Adoption of Linked Data.

The adoption of Linked Data is hypothesized, based on chapter 2.3. It starts with the model of technology diffusion by Zmud and Apple (1989). Using this model the stage of adoption is framed, positioned in between initiation and adaption. Furthermore, the adoption process is visualized, using the TAM (Fishbein and Ajzen, 1975). Next to that, the extension of this model with network externalities by Song et al. (2009) is included as well. Moreover, the mentioned multi-item measurement scale by Davis (1989) is chosen for correct capture of perceived usefulness and perceived ease of use. Conceptually this is shown in the figure below.

Figure 16: Model of technology diffusion (Zmud and Apple, 1989).

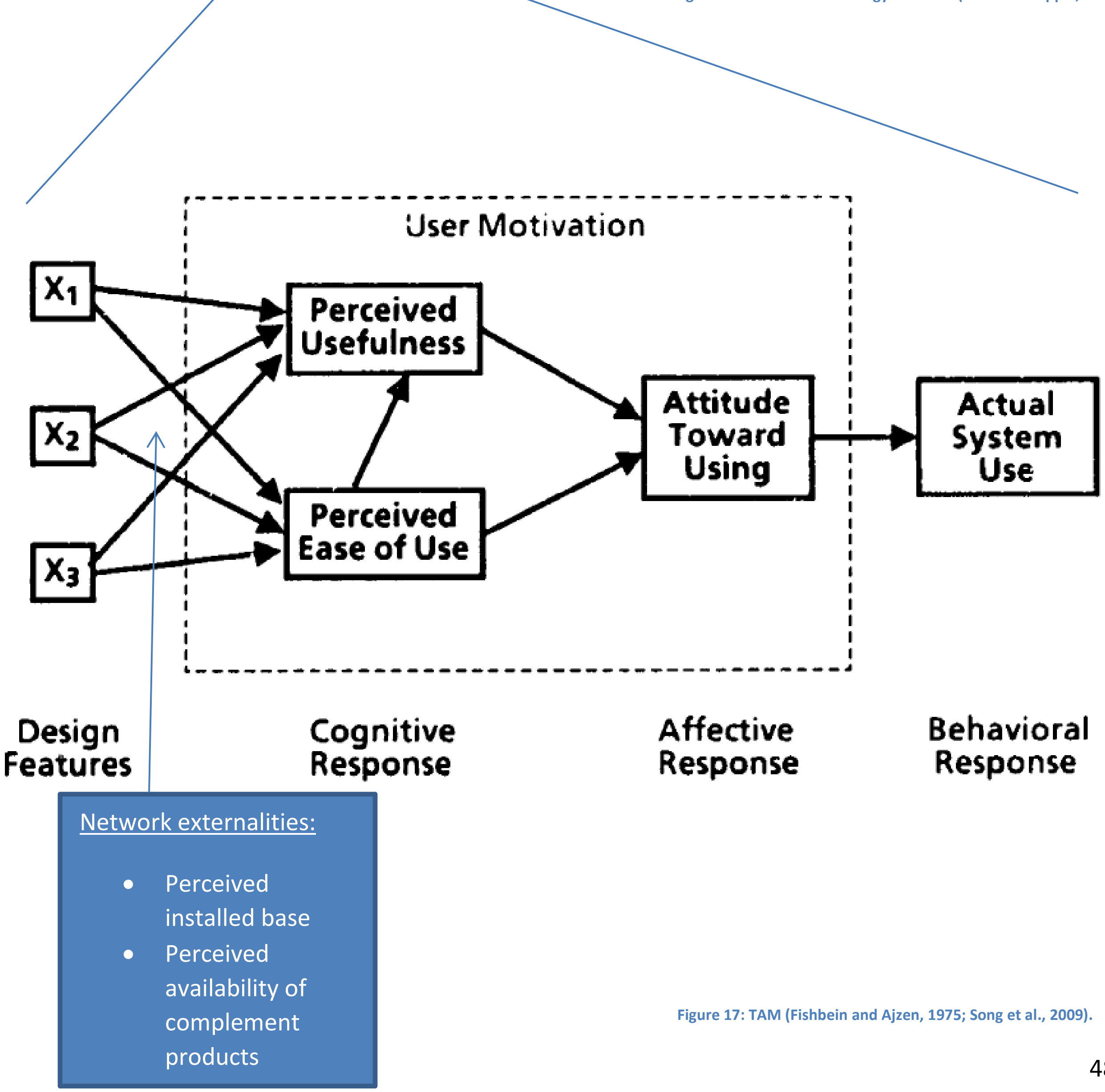


Figure 17: TAM (Fishbein and Ajzen, 1975; Song et al., 2009).

From the overview of advantages and disadvantages, as summarized in chapter 7 and 8, a list of factors that affect the adoption and diffusion of LD is created. The factors are derived from dominant studies and rearranged or summarized into fourteen variables with positive and twelve variables with negative impact on adoption, using grounded theory approach. For the overview of this list, the respective variables and their definitions, please refer to the Appendix.

Twelve of the defined variables with positive effect on adoption refer to different dimensions of perceived usefulness, whereas only "easy access" can be connected to the scale for ease of use in terms of flexibility of interaction with the technological innovation. The first defined criteria "full information" refers to improved job performance and can be therefore accounted for increased perceived usefulness, as defined by Davis (1989). Furthermore, productivity increases due to the reduced redundancy in information exchange, associated with LD. Enhanced effectiveness on the job within the perceived usefulness dimension is characterized through four characteristics of and potentials for LD, i.e. data re-use potential, the possibility for better (financial) product comparisons, advanced recommendation applications and reliable product rankings. Furthermore, LD is an antidote for human bounded rationality in information processing, vanishes cross-sectorial interoperability issues, allows for automatic knowledge acquisition and (semi-)automatic reasoning, which makes it easier to do one's job. Next to that, one finds it useful on the job due to the provision of up-to-date knowledge and the related time and resource savings. "Extremely good fit to accounting systems" (Chu, 1992), which captures the beneficial fit in comparison with classical relational data models can be directly related to attitude toward using. Furthermore, necessary changes in formal definitions (current legislation), security concerns and current market dominance of relational models negatively relate to "attitude toward using".

The rest of variables with negative impact have somewhat more factors influencing adoption through perceived ease of use than the ones with positive impact. Four of the former correspond with ease of use dimensions. "Limited expressiveness" and the absence of a "semantic web equivalent" of a "mandatory field" may lead to no easiness of getting LD technologies "to do what one wants them to do", as defined by Davis (1989). Furthermore, linked "closed" data can reduce the flexibility of interaction with the technology. Moreover, complexity automatically hampers easiness to use, which is the last of Davis' items on the measurement scale for perceived ease of use. Several of the variables represent dimensions of perceived usefulness. The trustworthiness issue can reduce job performance opposite to "full information". Furthermore, a possible denial of services can decrease productivity. Questionable relevance of encountered resources can reduce effectiveness on the job, whereas limited applicability of information aggregation within the financial domain reduces usefulness in the job. The related with LD altering of legacy systems reduces easiness "to do the job".

All of this leads to the hypothesis that the adoption of LD depends most heavily on the perceived usefulness with its dimensions defined, using the multi-item measurement scale by Davis (1989). Furthermore, it is claimed that the adoption of LD can be enhanced by a higher availability of complementary products, such as ontologies and vocabularies, and increase in the perceived installed base. Statements about the installed base of LD are not easily dispensed, since LD does not always correspond with Open Data. However, statements about the size of the installed base of LD, based on LOD can be found in chapter 6.2. The complete overview of the proposed adoption model for LD is shown on the next page.

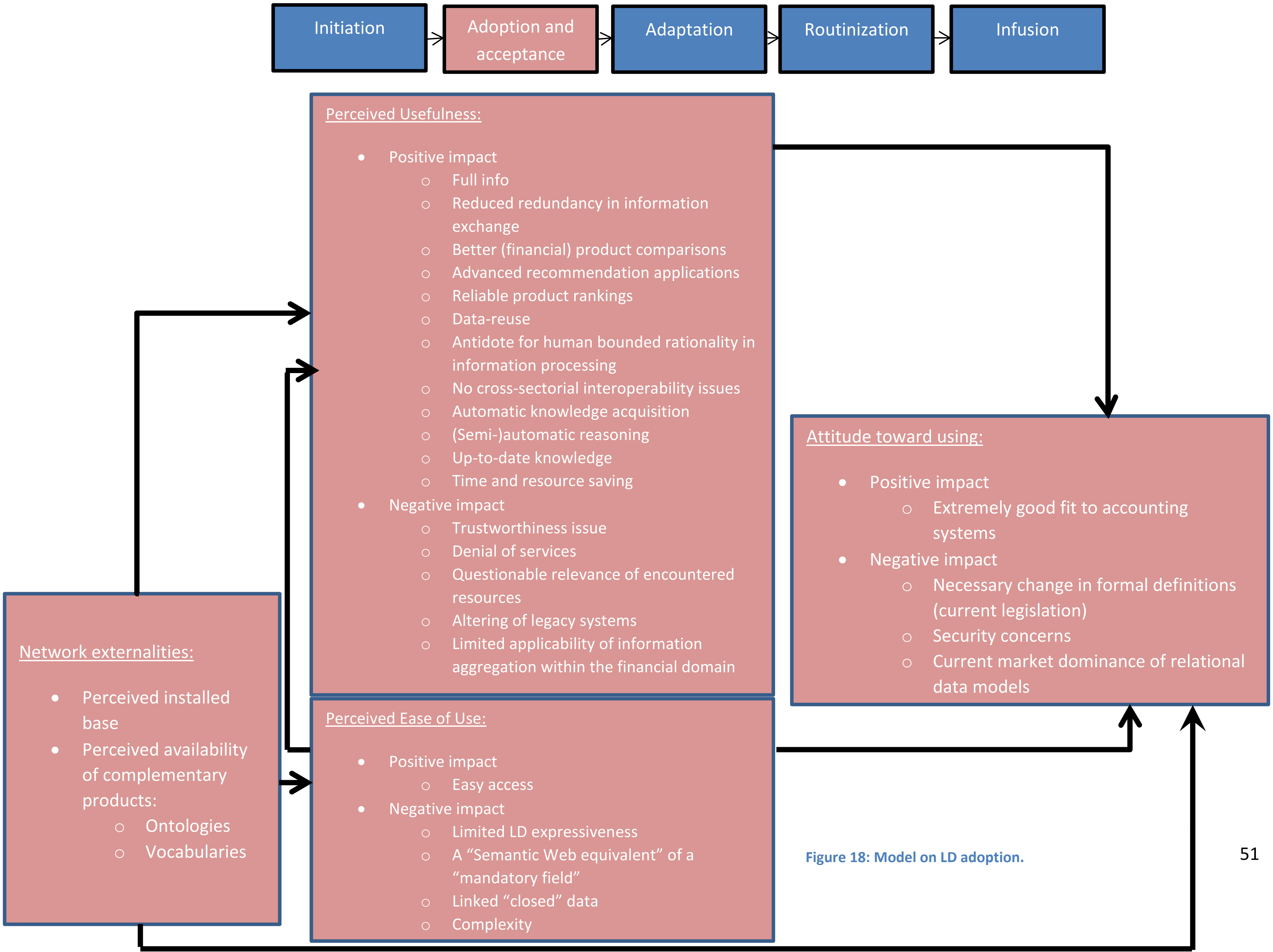


Figure 18: Model on LD adoption.

10 Business reporting and Linked Data – a case study.

In the following LD will be considered in the context of XBRL and it will be of interest whether it can help in reaping the desired benefits of XBRL, not only from a governmental perspective but also from the perspective of an (individual) investor.

10.1 XBRL basics and terminology.

XBRL stands for eXtensible Business Reporting Language and it is an international open large-scale data standard (Zhu and Wu, 2014), (Oude Luttighuis and Krukkert, 2009) for the electronic communication of business and financial data (Chowdhuri et al., 2014). XBRL covers all financial data flows. It is therefore not limited to reporting flows but can include invoice and payment flows as well (Oude Luttighuis and Krukkert, 2009). It is used by many organizations for data production and exchange (Zhu and Wu, 2014) Furthermore, it is (Zhu and Wu, 2014)mandatory for all tier-1 public companies to report their financial statements in XBRL, with a phased-in schedule based on company size, but for all by October 31th, 2014. Financial statements are an important source of information, used by investors for assessing the companies' financial health (Deller et al., 1999). XBRL provides "a technical framework for machine-readable financial reports composed of XBRL taxonomies and XBRL instance documents" (Chowdhuri et al., 2014). XBRL taxonomies represent dictionaries of "tags". "Tags" define particular reporting items in a financial statement. The taxonomy document consists of an XML schema and a collection of associated linkbases (XML Linking) (Zhu and Wu, 2014). The schema describes and delineates the concepts of financial reporting, e.g. Net Income or Assets. More importantly, each concept is uniquely defined. The linkbases can express relationships between concepts, i.e. calculation, definition and presentation or associate concepts with additional information, i.e. label and reference links. The XBRL taxonomy defines the concepts, after which facts are quantified with instance documents. A fact is "a single reportable piece of information", such as Net Income or Revenues (Chowdhuri et al., 2014). The facts are described by multiple elements – a container element as a root element, an item element for a single business fact or measure and a context element, specifying the relevant reporting period, industry segment etc. The attribute "contextRef" links the reporting item with its appropriate context. Furthermore a distinction between the XBRL basis and XBRL taxonomies is necessary. More importantly experts claim that a technical implementation of XBRL without the XBRL taxonomies is a "meaningless expense" (Oude Luttighuis and Krukkert, 2009). The XBRL basis offers a framework for structuring data (taxonomies) and instances, whereas the taxonomies define the data structures (Oude Luttighuis and Krukkert, 2009). In many sectorial message standards unlike XBRL, no distinction between the basis and the actual data definitions is made. Moreover the basis remains implicit to a high extent, with the exception of ebXML and StUF (Oude Luttighuis and Krukkert, 2009).

Many countries that adopted the standard, designed taxonomies matching their respective jurisdictions, such as the US, Brazil, Belgium, Denmark, Chile, The Netherlands, Germany etc[19]. The evolution of XBRL specifications and local efforts are coordinated by a non-profit consortium of over 600 organizations across the globe, called XBRL International.[20] Furthermore, the XBRL US[21] oversees the development of the taxonomies per US GAAP[22] (XBRL.ORG 1 and XBRL.ORG2). The GAAP taxonomy has two versions up until now.[23] The one from 2009 contains 13452 elements and 41651 presentation and calculation links among those, whereas the revised taxonomy in 2011 has 15967 elements and 54823 such links. It is recommended that companies use the latest versions when submitting their financial statements to the SEC but are at the same time allowed to extend the taxonomy with additional data elements and relationships to represent their business more accurate (Zhu and Wu, 2014).

XBRL complies to the criteria of "Open standards, i.e. the process for creating a list of open standards" from 2008, which is open, usable, has potential and considerable impact. Open points to the working group of XBRL International, which offers members and non-members to submit comments. Furthermore, members participate in the decision-making process. The XBRL specifications and documents are available on the website of XBRL International free of charge.[24] Their use is also free of charge. Next to that, there are no restrictions in re-use of the standard. Moreover, XBRL is sufficiently mature, which indicates its usability. XBRL International is established in 1998, has over 550 members, such as IASB, SEC, AICPA and several central banks. It is used by many groups in different counties and it is therefore a stable base that is growing consistently. In terms of potential, a lot of activities are undertaken in the area of open-source XBRL implementations.[25] It has a high impact since its potential benefits of implementation are considerable and exceed its costs, whereas risks remain manageable. This is, however, only the case if a suitable taxonomy is available or can be developed by the involved parties with maximum reuse of existing XBRL taxonomies (Oude Luttighuis and Krukkert, 2009). Nevertheless, in order to assess the potential, a chain-wide business case analysis for each chain is recommended (Oude Luttighuis and Krukkert, 2009). More details on the respective factors of this analysis are provided in the next chapter.

---

[19] https://www.xbrl.org/the-standard/why/who-else-uses-xbrl/
[20] http://www.xbrl.org/AboutXBRL.
[21] XBRL US is responsible for the data quality of XBRL on an international level
[22] GAAP stands for Generally Accepted Accounting Principles, adopted by the US Securities and Exchange Commission (SEC)
[23] As of April, 2015.
[24] http://www.xbrl.org
[25] http://www.xbrlwiki.info/index.php?title=Open_Source_and_XBRL

10.2 XBRL – the status quo. Benefits and challenges.

It is expected that XBRL will improve accessibility, create transparency and efficiency in the dissemination of financial data (Zhu and Wu, 2014). Therefore it should ease the necessary data curation and allow for more accurate and less resource-intensive cross-company comparison analyses (Kämpgen et al., 2014). It allows for automatic processing and is consequently more amenable than traditional representations such as PDF, HTML and text documents (Kämpgen et al., 2014). Furthermore it can increase the information quality in financial analytics. However it still does not solve the problem of data integration from various sources, e.g. of company background information, balance sheets, stock quotes, necessary for a holistic view on companies (O'Riain et al., 2012). Furthermore, Semantic heterogeneity issues impediment the realization of the government's vision for financial reporting, as well as hamper the benefits of the (individual) investors. Therefore the full potential magnitude of XBRL could not be grasped until now.

It is important to note that XBRL is a derivative of XML and therefore inherits its key benefits, such as platform independence, extensibility and flexibility. The extensibility feature allows for new elements to be added easily by companies without altering the accounting standards. This is necessary since federal regulators cannot generate a complete list of financial reporting items due to specifics in respective businesses or jurisdictions. As mentioned earlier with XBRL any jurisdiction can develop an own reporting taxonomy as an input in the data standard for exchanging business data for companies (Zhu and Wu, 2014). Furthermore, that way it allows companies not only to comply to the mandatory disclosure requirements but to represent their unique financial situation more accurate. With XBRL information is displayed on the Web like HTML and context is added to the content. This allows for searches within a specific context and results in faster search processes and a higher probability of relevance of the returned results (Chowdhuri et al., 2014).

However in the inheritance of XML lie also the main challenges for XBRL in terms of semantic interoperability. Due to its inheritance relation with XML, XBRL is very technical and cannot be used directly by investors for their decision-making, if relevant technical expertise is lacking. Furthermore, the extensibility of taxonomies leads to challenges in XBRL data interoperability. Especially the possibility for inter-firm comparisons is limited due to the heterogeneity of XBRL data. One of the many sources of heterogeneity are element definitions conflicts. For instance cash generated by operating activities is used by Apple with the element label: "NetCashProvidedByUsedInOperatingActivities" and by IBM with "NetCashProvidedByUsedInOperatingActivitiesContinuingOperations". Furthermore, contextual conflicts may impede the grasping of the XBRL potential. This is the case when different conventions are used by companies to identify a specific context element, e.g. the reporting period, use of decimal point, precision etc.

Within concepts of the value chain, the biggest challenge to leverage the value of XBRL is the processing hurdle (Rayport and Sviokla, 1995). Information processing has three levels:

- Information extraction
- Information interpretation
- Comparative evaluation

XBRL information extraction refers to the fact that investors must be able to identify and extract relevant financial elements from XBRL documents and will therefore need a software to view them, since the data structure of XBRL is technically challenging. Furthermore, investors need to contextualize XBRL data in order to be able to interpret it and make informed decisions. Again software is needed to process XBRL data to inform varying application contexts. Finally, it should be possible to routinely compare financial performance indicators across companies. However, it is challenging to decipher the relationships between data elements in the annual reports. Again a software solution can be chosen for this but it has to overcome the semantic heterogeneity, amongst others.

The heterogeneity between different XBRL fillings is addressed by Chowdhuri et al. (2014). An Ontology-based Framework for XBRL-mapping and Decision-making (OFXD) is proposed. It is capable to overcome semantic heterogeneity by mapping the financial classes and calculate ratios based on those mappings. These ratios can be then used within the decision-making by (individual) investors. OFXD collects instances (.xml) and their corresponding taxonomies (.xsd), retrieves values from the instance documents using the standard US GAAP taxonomy and the custom taxonomies and integrates them into the XBRL ontology – XBRLOnt. This ontology is used to support financial decision-making. The underlying conceptual model looks as follows:

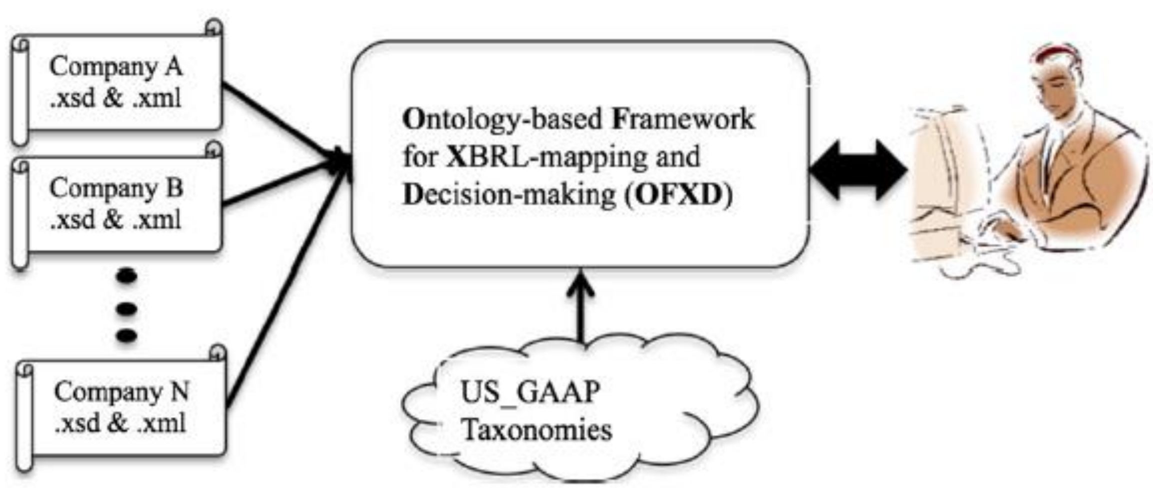


Figure 19: Conceptual model of OXFD (Chowdhuri et al., 2014).

A central point in this framework is that the creation of an ontology is automatic and based on values extracted from the XBRL-format financial statements. This is done by measuring the semantic similarity with the Jaccard Similarity Coefficient between the data names in EDGARD's HTML-based financial statements and their equivalents in the XBRL documents. Based on this, semantically similar class names in different XBRL fillings are mapped as equivalent classes. Furthermore, the time dimension is managed by "contextRef", which is an attribute that states whether the data is reported on a cumulative basis (income statement and statement of cash flow members) or it captures a specific point in time (balance sheet items) (Chowdhuri et al., 2014). This improves the XBRL interoperability for financial decision-making.

Due to its complexity, the number of parties that have XBRL skills is limited. For instance, experts estimate fifteen to twenty organizations in the Netherlands possessing such but more on international level (Oude Luttighuis and Krukkert, 2009).

Some of the aforementioned issues can be addressed by Linked Data, especially the semantic interoperability. Previous studies consider the advantages of using RDF over XML, where the knowledge is defined as an RDF ontology (Decker et al., 2000). Furthermore, approaches have been proposed to convert XBRL taxonomies and instance documents into their RDF or OWL equivalents, e.g. (Bao et al., 2010), (García and Gil, 2009) and (Raggett, 2009). Garcia and Gil (2009) propose an approach to move XBRL data to the Semantic Web by exploiting XML semantics. Bao et al. (2010) use OWL to translate the data model of XBRL into a Semantic Web representation, where the concepts are transformed into classes and the arcroles into properties. This involved the creation of three ontologies – an XBRL ontology, taxonomy ontology and an instance ontology. Since an RDF ontology includes classes, individuals and properties and XML – element, instance and relationships, a respective mapping is needed, i.e. elements correspond to classes, instances – to individuals and their relationships are defined by a properties (Chowdhuri et al., 2014).

Furthermore, it is interesting to find a way to integrate XBRL and non-XBRL data, such as governance scores, stakeholder assessments, market shares etc. into the decision-making process (Chowdhuri et al., 2014). This again can be addressed by LD as it allows the integration/aggregation of different sources, independent of their format. This will be covered in the following chapter.

10.3 Resolving the XBRL challenges with LD.

As mentioned in the previous section, XBRL does not solve the data integration problem per se. However data integration is crucial for the creation of a holistic view on companies (Kämpgen et al., 2014). There are various reasons for this. First of all formal semantics are limited within XBRL (Wenger et al., 2011) and relationships within concepts are only textually described (Kämpgen et al., 2014). Furthermore, it is difficult to compare financial information from different XBRL documents because the accounting and regulatory organizations do not align their taxonomies of financial concepts. Next to that, no unique companies identifiers exist across different reporting sources, which makes collecting information about a company challenging (Kämpgen et al., 2014). Some sources refer to a company using its name, or address, or its CIK[26]. Furthermore, other finance-related Open Data, e.g. stock quotes, background information, is published based on different data models.

LD could solve this issue. Applications built using the LD principles, as described in previous chapters, can be used amongst others for background information analysis, for multi-company KPI analysis and cross-data-sources KPI analysis (Kämpgen et al., 2014). Background information analysis refers to identification, acquisition and analysis of distributed company information, such as the company's address, industry and founding date. Multi-company KPI analysis allows for a comparisons of relevant KPIs over time for several companies, e.g. the stock market price for companies within a certain industry. Cross-data-sources KPI analysis, on the other hand, allows comparing these KPIs, which values origin from heterogeneous datasets and would have been non-comparable without extensive manual effort otherwise, e.g. earnings per share from the balance sheets, prices per from electronic stock quotes and total assets as published using US-GAAP.

Kämpgen et al. (2014) propose a development, i.e. Financial Information Observation System (FIOS), which uses LD for the access and representation of financial data. Since there is no standard representation of XBRL data as LD, financial data is modelled using the existing LD vocabularies, such as SKOS, FOAF and the DBpedia ontology and entities (concepts) from different sources are linked. Data is consolidated, i.e. made explicit and merged, so that access to the information is provided independently from the specific distribution across sources. Furthermore, analytical aggregate queries can be performed on this data and quality checks can be made using SPARQL. Three different interfaces are used to perform analysis on the integrated financial LD. The advantage here is that FIOS is not too general for financial analysis, unlike semantic search engines but also not too complicated for domain experts, unlike the analysis tools such as the SPARQL package for R by van Hage and Kauppinen. LD principles allow for access to data in a "standard and modular way" (Kämpgen et al., 2014). New data can be easily added due to the flexibility of

[26] Central index key used by SEC to uniquely identify companies.

the RDF schema; the crawler needs only to be allowed to reach further entities. Access to data is simplified, resulting in entity consolidation due to the formal semantics, e.g. explicit equivalent statements (Kämpgen et al., 2014).

Other development that aim at resolving the XBRL challenges with LD is the Rhizomik Semantic XBRL demo, which generates RDF representations of XBRL data (Garcia and Gil, 2010). The Business Intelligence Cross-lingual XBRL (BIXL) demonstrator is able to retrieve facts from unstructured text in fillings (O'Riain et al., 2012). Furthermore, the pipeline from Midas is capable of extraction and linkage of information about entities, e.g. company, key people etc. from semi-structured XML documents (Burdick et al., 2011).

All in all, data integration across sources is made possible with LD, which consequently enables background information analysis, multi-company KPI analysis, as well as cross-data-sources KPI analysis. Furthermore the same data can be shown using different interfaces and complying to the requirement: "Overview first, zoom, details on demand" (Kämpgen et al., 2014). Finally, own reports can be created and desired analysis performed. The main benefits are a flexible schema, standard access, expressive queries and formal semantics.

10.4 The XBRL case within the developed conceptual model.

Advantages	Score	Disadvantages	Score
Full information	++	Trustworthiness issues	--
Easy access	+	Questionable relevance of encountered resources	--
Data re-use	++	Linked (closed) data	N/A
Reduced redundancy in information exchange	+	Denial of services	--
No cross-sectorial interoperability issues	++	Necessary change in formal definitions	--
Automatic knowledge acquisition	++	Security concerns	--
Up-to-date knowledge for decision-making in real-time businesses	++	Altering of legacy systems	--
(Semi-)automatic reasoning	++	Limited LD expressiveness	-
Time and resource saving Reduced manual effort	++	Complexity	-
Antidote for human bounded rationality in information processing	++	"semantic web" equivalent of a "mandatory field"	--
		Current market dominance of relational data models	-
Specifically within the financial domain			
Better (financial) product comparisons	++	Limited applicability of information aggregation within the financial domain	-
Advanced recommendation applications	++		
Reliable product rankings	N/A		
"extremely good fit to accounting systems"	++		

Hereby "++" indicates a very strong advantage of LD application within XBRL, "+" indicates a strong advantage, "0" indicates no relevance for the financial domain, "-" indicates strong disadvantage for LD application, "--" indicates a very strong disadvantage and "N/A" means that this line item does not apply to the XBRL case study. Therefore the implementation of the developed conceptual model on LD adoption looks as follows.

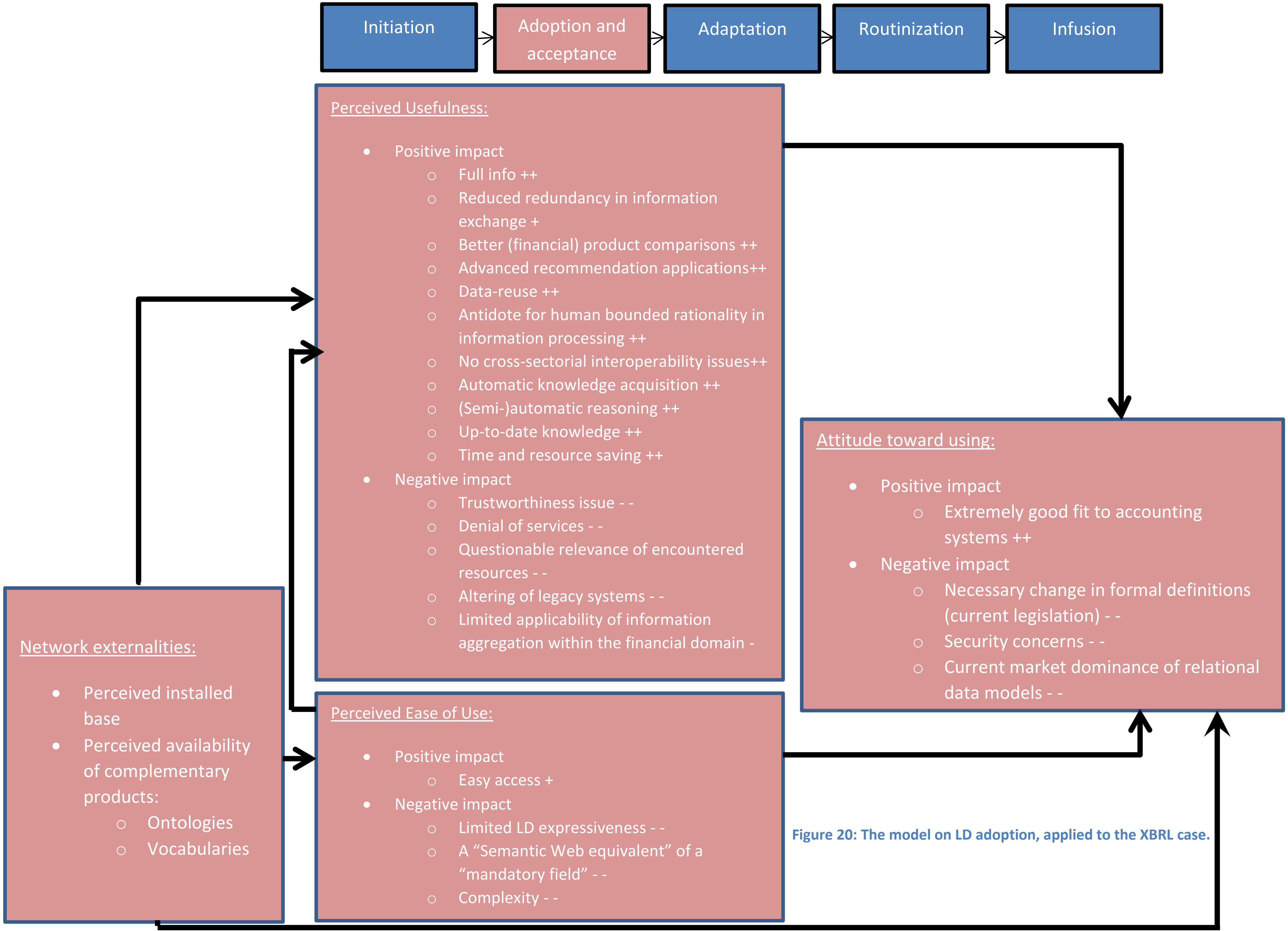


Figure 20: The model on LD adoption, applied to the XBRL case.

11 Insights from the interviews with financial experts.

According to experts it all started out with XBRL in 1998. Furthermore, XBRL started out as XFS – eXtensible Financial Statement reporting language but since for computers (machines) there is no difference between financial and business information, later on it has been converted to eXtensible Business Reporting Language. Initially built for reporting and data exchange outside companies, it is expected to find use for internal reporting as well within couple of years.

The interviews with financial experts led to very interesting insight, as they were partly confirmatory to already made statements but also partly additive to the mentioned content. For instance, experts describe XBRL taxonomies as more than just data definitions. From their perspective XBRL taxonomies are “semantic networks of everything that is related to the business reports”, i.e. the data definitions, the relationships between them. Furthermore, they include data validation mechanisms and data quality allowance; could include also business rules. Next to that, nowadays they also contain layout renderings. So, XBRL taxonomies externalize the business logic, formalize the data definitions and the layout in a global standard.

When asked about advantages of XBRL, experts compare it with a mobile phone, since it doesn’t matter which brand or vendor one chooses for, there are standards for communicating, exchanging messages, making phone calls etc. So, XBRL stands for standardization of data, standardization of the reporting process all over the world. It is not restricted to financial information but can support all kinds of business information. Moreover it makes it more efficient, cheaper, less error-prone. Expert opinions reconfirm the advantages of XBRL in improvement of accessibility of data. Transparency as well but only where owners of the data are willing to make this data transparent. For instance, accountancy in every country is mandatory, so financial statements are to be filed within business registries and send to the chamber of commerce. Not many companies are willing to provide more data than is required due to fear of competitors etc. So, usually transparency only goes as far as the mandate goes. An example of higher accessibility and transparency enabled by XBRL is the small company ISeeMore. It collects data in XBRL format and allows for fast benchmarking that can be performed on different devices, even on an iPhone. So, one of the main advantages of XBRL pointed out by experts is that it is a global open standard and if all of the information is made available in this same standard, with the same technology/ technology standard, benchmarks across companies, domain and countries are easier to perform. Another point is the content – or what is being reported. It is dependent on local accountancy rules (local GAAP), e.g. the Dutch annual reports have different definitions for profits and revenues and costs than the Germans, French etc. For the big public-quoted companies/ stock-valued companies there is a slightly bigger advantage, due to the presence of the International Accounting rules (IFRS), which makes it easier to make cross-country comparisons.

In chapter 10 it was pointed out that efficiency in financial decision-making can be enhanced, if XBRL and non-XBRL data can be combined. Experts reveal that this is already possible and realized. Therefore it is possible to combine the insights from XBRL data, such as financial statement data elements with data, such as governance scores, stakeholder assessments, market shares etc.

Another main advantage of XBRL is that the content and the context of the data is linked to the data itself. So, if a party sends data to another one, it can ensure the same correct interpretation of the data. On the other hand, when human beings communicate, every single one of them has different interpretation of the (exchanged) data. With XBRL there is no more difference in the interpretation, because the content of the data is described in the taxonomy. Furthermore, the system to system communication and exchange of data is very robust and integer.

The statement that XBRL is a derivative of XML, which was taken from Zhu and Wu (2014), experts would phrase differently. They would see XBRL not as a derivative per se but as an advanced implementation, using XML. Furthermore, research claims that XBRL is due to its roots in XML, too technical and a challenge for investors to make use of it in their decision processes. Experts disagree because they claim that one doesn't need to read the instance documents, since there are enough standard software products in the market to collect XBRL data, validate it, process it, store it in a database, compare it and make beautiful report formats. Such applications are software products from Semansys with integrated controls. So, if a financial statement goes through such, it automatically sees if all the calculations are correct, if there are links between the income statement, balance sheet and attachments etc., even if subsidiaries are involved (e.g. Shell). This is done instead of recalculation by humans in India. Another applications are offered by ISeeMore, Creaen and Batavia.

However there are issues with XBRL that are recognized by specialists. One of the main problems is the semantic heterogeneity, as mentioned in Chapter 10. This is where experts also see potential for the positioning of LD and the Semantic Web principles. Using those in XBRL is currently in the beginning stage. So, ontologies, prevalence engines, Semantic Web type of approaches are not yet considered and implemented in the XBRL space. Experts claim, however, that there are some experiments already and some individuals that understand the potential benefits in such developments. On the other hand, they claim that whole 99% percent of the XBRL people don't understand this yet. Currently a solution for the semantic heterogeneity is proposed in the context of XBRL. XBRL Nederland is in the middle of developing a Trusted Taxonomy Centre, where all taxonomies are to be stored. In the Netherlands companies are not allowed to use a company (private) extension in public, i.e. use it for data exchange/reporting purposes outside the company borders. For reporting purposes a company in the Netherlands is allowed to use only the taxonomies, published by the Dutch government. If there is a Trusted Taxonomy Centre, private extensions can be allowed in public, as long as they are published there.

Another disadvantage of XBRL is that one could put so many elements in a tag that this could lead to a huge amount of data to store and therefore a huge computer power is necessary to handle that data. The reason for that is that one extracts the relevant information from the taxonomy but is also able to put own/additional information as needed, for instance the currency, the year or the region, linked to a specific transaction. People do it and since the exchange is easy to do, people put more data in it – usually if the people have the possibility to store more information, they think they also need more information. Furthermore, experts share that XBRL started by tagging the face of the financial statement, e.g. profit or losses and the balance sheet but in currently are in the middle of implementing XBRL GL (XBRL Global Ledger). XBRL GL enables every individual transaction to be directly tagged to the system. Furthermore, the formula link base is in the development phase. Thereby, continuous auditing and continuous monitoring are enabled. Each transaction can be audited and monitored in the moment it is put in the system. IBM for instance is doing continuous monitoring. But if continuous monitoring requires continuous internal auditing. It will be done in the same moment, in which the data is being processed. For instance, if an invoice comes electronically, it will be directly processed through the system and can be monitored and audited simultaneously. The complete three-way-match can be controlled by XBRL on invoice level or on individual line item level.

LD could further support those potentials and solve some of the challenges. One of those potentials is connected with the accounting rules and the taxonomies. If for instance, there is a different definition for profits, via mechanisms like the Semantic Web and LD, meta-definitions can be made comparable. Comparable meta-definitions would enable, for instance, a comparison of a bio-tech company in the Netherlands with a bio-tech company in Germany or anywhere. By definition, the quality of the question (taxonomies, meta-data, semantic definitions) prescribes the quality of the answer (outsource, means comparability of data).

Furthermore, the enforcement of a global standard and the prohibition of private extensions can endanger the foreseen benefits of XBRL in regards to flexibility and extensibility. Semantic Web and LD type of approaches can again help here. For instance, if a company does not agree with a definition of profit in the accounting rules and wants to add its own definition and applies Semantic Web/LD type of implementations, then it will be forced to check whether the same definition it is looking for already is applied by other company or group of companies and then it can re-use the existing additional element "profit". Again, this shows enhanced efficiency in data re-use. Next to that, cross-border, cross-sectors and domains harmonization and linking of data can be additional benefits of LD and Semantic Web approaches within XBRL. There is an initiative in XBRL Europe – "Europe XBRL Business Registries". It is to make comparable the data entity structures of entity information from different countries. This can be solved with standardization (the XBRL way) but it is less dynamic than if it would be solved with LD and Semantic Web technologies.

Another benefit that was pointed out is the way information/data is treated. Experts even foresee that the financial statements will disappear in this form in the near future. So, instead having one way stream of information, coming from the organizations to the outside world, the outside world is going to ask for specific information. For instance, doing a tax filing for a small company on paper would require filling in of 1600 data elements. Doing the same with XBRL would require 400 data elements. So, experts expect that there will be a message in XBRL coming from the tax authorities that directly links to the computer, asking for those specific data elements in the taxonomy. If the company agrees, the system goes and gets all this information, stored in the company. Furthermore, there are automatic controls running and validation mechanisms. The aim of this is to develop a process within two years, where after sending a tax filling with XBRL, an answer from the tax authorities comes within 2 weeks. This period currently amounts three months in the US and in Denmark three weeks. So, XBRL does all the controls, normally done by hand, thereby speeding up the process and enhancing efficiency.

Another potential for LD lies in the XBRL taxonomies. Currently XBRL taxonomies are "reasonable limited straight-forward collection of metadata", so there is no intelligence in them. If those taxonomies move further to ontologies, in a technology point of view even more advanced implementations than the ones today can be put in place. Then one could reason about the data or even have automatic reasoning. However, experts also expressed some concerns. They shared that from their experience, financial people all over the world are "terribly traditional"; they sometimes still work with pen and paper, don't even send emails via smartphone because that's modern innovative technology. Dutch banks in the Netherlands, for instance, still receive papers and pdfs for credits and loan reporting and have employees rekeying the data. Next to that, they expressed their concerns about security. In their understanding if a company has LD outside the company, maybe the competitive hedge is also outside the company. If everyone can link to the data and even competitors can make use of the data, it might become a problem.

Furthermore, XBRL often is seen as a burden by companies. Reporting is not tolerated in general and with XBRL this means an extra activity that needs to be paid for. Another point is that companies currently are the preparers of XBRL data but do not consume XBRL data. Therefore they currently cannot really benefit from it. So, one of the challenges is to create benefits for the companies by enabling consumption of XBRL data.

Another issue that was pointed out is that the people would need to change when they work with XBRL. Making use of possibilities like continuous monitoring, continuous auditing etc. requires the use of new tools, such as expert systems, pattern recognition, process mining, data mining etc. However, the concern is that people are reluctant to change. This challenge cannot be solved with LD, since LD is a new concept, which requires change as well. This process of change is seen in following phases:

1) Development of a common standard

2) Development of a common set of reporting definitions, taxonomies

3) Development of standard software to do the work

4) A lot of data in XBRL (chamber of commerce etc.) and enabling of companies and individuals to make use of them. For instance, the M&A process would be supporting by enabling comparison of data with just one click. With few simple clicks the top five shortlisted companies that comply to the specific requirements (minimum net profit of… etc.) can be selected.

The status is different in different countries. For instance, in the UK all of the financial statements information is publicly available. One could therefore download all the filing of all the entities in the UK. Furthermore, using the industry codes, one could select only companies in a specific sector or division. In the Netherlands, Netherland SBR enforces the chamber of commerce to make the data from the financial statements publicly available on their website. It is expected within the next few months. Furthermore, it is mandatory for all tier-1 public companies to report their financial statements in XBRL, with a phased-in schedule based on company size, but for all by October 31th, 2014, according to Zhu and Wu (2014). However, the experts indicate that there are different mandates. There was one for VAT reporting, where companies are obliged to do it with XBRL; another one was from last year on corporate tax. This year one goes on financial reporting to the chamber of commerce. On a global scale it is in the parliaments to be turned into legislation. It is mandated that the small companies in the Netherlands have to file their financial statements with XBRL from the fiscal year 2016 onwards, the midsize companies in 2017 and the large companies in 2019. This is due to the fact that the Transparency Directive in Europe states that in 2020 all listed companies in Europe should file their financial statements in XBRL.

Experts see LD and XBRL as very familiar to each other and possibly the next step in the system to system data exchange. “I think it’s not XBRL and LD, I think it’s both together”. They acknowledge that the combination LD and XBRL would result in referring to data instead of copying it, powered by LD and no need to copy the whole dataset but just extract the data elements that are needed and getting the content next to the data, powered to XBRL.

In regards to LD adoption, the interviewed financial experts agreed upon the involved factors, as well as their impact, respectively negative or positive, as shown in the analysis so far. Furthermore, they confirmed linked (closed) data as not relevant for the XBRL case. Interesting fact was that they could imagine an application of product rankings, also within the XBRL environment. Next to that, experts seem to be less concerned about security issues in regards to LD. For the XBRL case a strength of “--” was opposed. The experts, on the other hand, did not see it as a problem. All in all, it is to consider that all of the given weights are case specific and since there are many application possibilities within the XBRL environment, an assessment per application is recommended. This way the specific purpose, potentials and risks can be taken into consideration.

12 Future Research and Conclusion Remarks.

This work examines the factors that affect the adoption of LD by using a ground theory approach. The adoption and acceptance phase is framed, based on the model of technology diffusion by Zmud and Apple (1989). The adoption itself is examined, using the TAM (Fishbein and Ajzen, 1975). Furthermore, the developed LD adoption model includes the extension of the TAM with network externalities, as discussed by Song et al. (2009). Moreover, the main elements of the TAM, as applied to LD are to be quantified, using the multi-item measurement scale by Davis (1989).

A total of fourteen factors with positive and twelve factors with negative impact are derived from the dominant research, based on a systematic literature review and rearranged or summarized into the TAM factor categories. Those provide a better understanding of the factors that diffuse LD as a technological innovation.

Twelve of the defined variables with positive effect on adoption refer to perceived usefulness, whereas only “easy access” can be connected to the Davis scale for ease of use (1989) in terms of flexibility of interaction with the technological innovation. The factors that correspond with different dimensions of perceived usefulness are “full information”, reduced redundancy in information exchange, data re-use potential, the possibility for better (financial) product comparisons, advanced recommendation applications and reliable product rankings. Furthermore, LD is an antidote for human bounded rationality in information processing, vanishes cross-sectorial interoperability issues, allows for automatic knowledge acquisition and (semi-)automatic reasoning, which makes it easier to do one’s job and therefore corresponds again with increased perceived usefulness. Up-to-date knowledge and the related time and resource savings further enhance perceived usefulness. The last fourteenth factor with positive impact on LD adoption, i.e. that LD is an “extremely good fit to accounting systems” (Chu, 1992) in comparison with current relational data models, can be directly related to attitude toward using. Out of the factors with negative impact, necessary changes in formal definitions (current legislation), security concerns and current market dominance of relational models negatively relate to “attitude toward using” respectively.

The rest of variables with negative impact influence either perceived usefulness or perceived ease of use. Four of the them have been found to correspond with dimensions of ease of use. Those are “limited expressiveness” and the absence of a “semantic web equivalent” of a “mandatory field”, as well as linked “closed” data and complexity. One of the factors that directly influence the perceived usefulness is the trustworthiness issue, which can reduce job performance opposite to “full information”. Next to that, a possible denial of services, questionable relevance of encountered resources and limited applicability of information aggregation within the financial domain due to (amongst others) sensitive information negatively influence perceived usefulness. Finally, the related with LD altering of legacy systems reduces easiness "to do the job" and therefore corresponds with one of the dimensions of perceived usefulness. All of this leads to the hypothesis that the adoption of LD depends most heavily on the perceived usefulness with its dimensions defined, using the multi-item measurement scale by Davis (1989). Furthermore, it is claimed that the adoption of LD can be enhanced by a higher availability of complementary products, such as ontologies and vocabularies, and increase in the perceived installed base.

The application of these findings within business reporting, illustrated with the XBRL case, shows that all of the factors defined, have the expected positive or negative impact on the process of LD adoption. Linked (closed) data that could lead to RDF links pointing to "dead ends" for a specific user query due to global entity consolidation policies, has been found to not play a direct role in the XBRL context and therefore omitted from the representation. The reason for this is that current XBRL referencing is based on publicly available financial and business information. Furthermore, reliable product rankings are currently beyond the scope of XBRL and therefore accordingly not relevant in the proposed model for LD adoption within this case study.

The aforementioned insights were verified with semi-structured interviews with financial experts. Those delivered interesting points from a practitioners' perspective. Some of their statements ensured additional insights and facts. For instance, experts do acknowledge the potential of LD to solve some of the main issues with XBRL, such as semantic heterogeneity. However, they claim that ontologies, prevalence engines, Semantic Web and LD approaches are not yet implemented in the XBRL space. The reason for this is (amongst others) that 99% of the people familiar with XBRL do not understand this potential yet. The remaining 1%, however, is aware that LD can solve the problem of semantic heterogeneity much more dynamically than XBRL itself and the development of a Trusted Taxonomy Centre. Furthermore, LD could ease the cross-country, cross-domain and cross-company comparisons of XBRL data, which could enable for instance a comparison of a bio-tech company in the Netherlands with a bio-tech company in Germany or elsewhere, without taking away the flexibility by enforcing a global standard and/or prohibiting the private extensions of XBRL. Finally, experts recognize LD and XBRL as very familiar to each other and possibly the next step in the system to system data exchange. They acknowledge that the combination LD and XBRL, as discussed within this work, would result in referring to data instead of copying it, powered by LD and no need to copy the whole dataset but just extract the data elements that are needed and getting the content next to the data, powered by XBRL.

Those findings provide important implications, for both researchers and practitioners. First, this work has an effect of integrating findings from previous studies. Second, LD is defined within the context of the financial services domain. Next to that, a complete overview of (dis-)advantages of this new concept is put together, which is unique to the current research and practice. Next to that, a model for the adoption of LD is developed.

It allows for developers to derive incentives to influence factors that have impact on LD adoption. Next to that, it provides clarity amongst potential adopters about LD potential and loopholes. TNO can use, i.e. already uses, the developed conceptual model on LD adoption for consulting clients on the growing importance of LD, as well as advantages and disadvantages of potential LD application on their specific case study.

Further research is to conduct case studies in different domains, as well as countries to identify possible factors that could also play a role in the adoption of LD on domain or regional bases. Furthermore, LD adoption can be reviewed in the context of the Fishbein model or another technology acceptance model, which could lead to additional factors that play a role in it and possible extension of the developed model within this study. Next to that, a survey on a large basis with LD simulations within the financial domain, can assess the willingness of financial experts to make use of LD application, as well as the caused acknowledgment of their potential within this context.

13 References

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, 84, 888-918.

Antoniou, G., Harmelen, F., 2008. A Semantic Web Printer, 2nd ed. *MIT Press*, Cambridge, MA.

Arndt, J. (1967). Role of Product-Related Conversations in the Diffusion of a New Product. *Journal of Marketing Research* 4(3): 291-295 (August).

Attewell, P., 1992. Technology diffusion and organizational learning: the case of business computing. *Organization Science* 3 (1), 1-19.

Baldwin, A. A., Brown, C. E., Trinkle, B. S., XBRL: an impacts framework and research challenge, *Journal of Emerging Technologies in Accounting* 3 (1) (2006) 97-116.

Bansal, A., Kauffman, R. J., Weitz, R. R., Comparing the modeling performance of regression and neural networks as data quality varies: a business value approach, *Journal of Management Information Systems* 10 (1) (1993) 11-32

Bao, J., Rong, G., Li, X., Ding, L., Representing financial reports on the semantic web, *Semantic Web Rules*, Springer, Berlin Heidelberg, 2010, pp. 144-152.

Belleau, F., Nolin, M.-A., Tourigny, N., Rigault, P., Morissette, J., Bio2rdf : Towards a mashup to build bioinformatics knowledge systems, *Journal of Biomedical Informatics* 41 (2008) 706-716.

Berners-Lee, T., Hendler J., Lassila, O., The Semantic Web, *Scientific American* 284 (2001) 34-43.

Burdick, D., Hernàndez, M.A., Ho, H., Koutrika, G., Krishnamurthy, R., Popa, L., Stanoi, I., Vaithyanathan, S., Das, S.R., Extracting, Linking and Integrating Data from Public Sources: A Financial Case Study*. IEEE Data Eng. Bull*. (2011).

Chau, P.Y.K., Tam, K.Y., 2000. Organizational adoption of open systems: a "technology-push, need-pull" perspective. *Information and Management* 37 (5), 229-239.

Cheng, G., Ge, W., Qi, Y., in: J. Huai, R. Chen, H.-W. Hon, Y. Liu, W.-Y. Ma, A. Tomkins, X. Zhang (Eds.), Falcons: Searching and Browsing Entities on the Semantic Web, *ACM*, 2008, 1101-1102, WWW.

Cohen, W.M., Levinthal, D.A., 1990. Absorptive capacity: a new perspective on learning and innovation. *Administrative Science Quarterly* 35 (1), 128-152.

Davis, F.D., Bagozzi, R.P. and Warshaw, P.R., User Acceptance of Computer Technology: A comparison of Two Theoretical Models, *Management Science* 35(8), August 1989: 982-1003.

Debreceny, R., Farewell, S., Piechocki, M., Felden, C., Gräning, A., Does it add up? Early evidence on the data quality of XBRL fillings to the SEC, *Journal of Accounting and Public Policy* 29 (3) (2010) 296-306.

Decker, S., Melnik, S., Van Harmelen, F., Fensel, D., Klein, M., Broekstra, J., Horrocks, I., The semantic web: the roles of XML and RDF, Internet Computing, *IEEE* 4 (5) (2000) 63-73.

Deller, D., Stubenrath, M., Weber, C., A survey on the use of the Internet for investor relations in the USA, the UK and Germany, *The European Accounting Review* 8 (2) (1999) 351-364.

Ding, L., Michaelis, J., McCusker, J., McGuinness, D., Linked provenance data: a semantic web-based approach to interoperable workflow traces, *Future Generation Computer Systems* 27 (6) (2011) 797-805.

Dingm, Y., Foo, S., Ontology research and development. Part 1 – a review of ontology generation, *Journal of Information Science* 28 (2) (2002) 123-136

Dingm, Y., Foo, S., Ontology research and development. Part 2 – a review of ontology mapping and evolving, *Journal of Information Science* 28 (5) (2002) 375-388

Du, J., Zhou. L., Improving financial data quality using ontologies, Decision Support Systems 54 (1) (2012) 76-86.

Ervin, W., Visualizing uncertainty: the graphical representation of risk in investor communications, *Information Design Journal Document Design* 12 (3) (2004) 208-212.

Fensel, D., Brodie, M., Ontologies: a Silver Bullet for Knowledge Management and Electronic Commerce, 2nd ed. Springer, New York, NY, 2003

Fishbein, M., & Ajzen, I. (1975). Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research. Reading, *MA: Addison-Wesley*.

Garcia, R., Gil, R., Triplificating and linking XBRL financial data. Information Storage and Retrieval (2010).

García, R., Gil, R., Publishing XBRL as Linked Open Data, *World Wide Web Workshop: Linked Data on the Web* (LDOW2009), 2009.

Goble, C., De Roure, D., The impact of workflow tools on data-centric research, in: Tony Hey, Stewart Tansley, Kristin Tolle (Eds.), the Fourth Paradigm: Data-Intensive Scientific Discovery, 2009.

Goodhue, D.L., Understanding User Evaluations of Information Systems, *Management Science* 41(12), 1995: 1827-1844.

Gruber, T., A transitional approach to portable ontology specifications, *Knowledge Acquisition* 5 (2) (1993) 39-41

Hammer, M., & McLeod, D. (1981). Database description with SDM: A semantic database model. *ACM Transactions on Database Systems*, September, 351 – 386.

Holak, S. L. and Lehmann, D. R. (1990). Purchase Intensions and the Dimensions of Innovation. *Journal of Product Innovation Management* 7(1): 59-73.

Hu, D., Zhao, J. L., Hua, Z., Ranking systemic risks in banking networks, *ICIS* 2010 Proceedings, 2010.

IDC, The digital universe decade – are you ready?, from http://www.emc.com/collateral/demos/microsites/idc-digital-universe/iview.htm 2010.

Jorion, P., Value At Risk: The New Benchmark for Controlling Market Risk, *Irwin Professional Publishing*, Chicago, 1997.

Kanellopoulos, D., Kotsiantis S., Tampakas, V., Towards an ontology-based system for intelligent prediction of firms with fraudulent financial statements, *IEEE Conference of Emerging Technologies and Factory Automation*, Patras, 2007

Kimball, R., Ross, M., The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, *2nd edition Wiley*, April 2002.

Kingston, J., Schafer, B., Vandenberghe, W., Towards a financial fraud ontology: a legal modelling approach, *Artificial Intelligence and Law* 12 (4) (2006) 419-446.

Leonard-Barton, D. and Deschamps, I., Managerial influence in the implementation of new technology, *Management Science* 34(10), 1988: 1252-1265.

Nambisan, S., Wang, Y., 2000. Web technology adoption and knowledge barriers. *Journal of Organizational Computing and Electronic Commerce* 10 (2), 129-147.

Ng, W.K., Yan, G., Lim, E-P., Heterogeneous Product Description in Electronic Commerce, *ACM SIGecom Exchanges*, 12000, 7-13.

Oude Luttighuis, P.H.W.M., Krukkert, D., Expert Recommendation XBRL v 2.1., *Forum Standaardisatie*, 2009.

O'Riain, S., Curry, E., Harth, A., XBRL and open data for global financial ecosystems: A linked data approach. *Int. J. of Accounting Information Systems* 13 (2012).

O'Riain, S., Coughlan, B., Buitelaar, P., Declerk, T., Krieger, U., Marie-Thomas, S., Cross-Lingual Querying and Comparison of Linked Financial and Business Data. In: Cimiano, P., Fernàndez, M., Lopez, V., Schlobach, S., Völker, J. (eds) *ESWC* 2013. *LNCS*, vol. 7955, pp. 242-247. Springer, Heidelberg (2013).

Pinsker, R., Li, S., Costs and benefits of XBRL adoption: early evidence, *Communications of the ACM* 51 (3) (2008) 47-50.

Predrinaci, C., Domingue, J., Krummenacher, R., On the integration of services with the web of data, Technical Report kmi-09-06, *Knowledge Media Institute*, The Open University, 2010, (Available at http://kmi.open.ac.uk/publications/pdf/kmi-09-06.pdf, last accessed April 24, 2011).

Raggett, D., How can we exploit XBRL and Semantic Web technologies to realize the opportunities? *Proceedings of 19th XBRL International Conference*, Paris, 2009.

Rayport, J. F., Sviokla, J. J., Exploiting the virtual value chain, *Harvard Business Review* 73 (6) (1995) 75-85.

Rogers, E. M. (2003). *Diffusion of Innovations*, (5th ed.). New York: Free Press.

SEV.GOV, Final Rule: Adoption of Updated EDGAR Filer Manual Final Rule, Retrieved on August 9, 2013 from http://www.sec.gov/rules/final/2009/33-9022.pdf

Shih, H., 2006. Technology-push and communication-pull forces driving message-based coordination performance. *Journal of Strategic Information Systems* 15, 105-123.

Simon, H., Models of Bounded Rationality, Vols. 1 and 2, *MIT Press*, Cambridge, MA, 1982.

Sujanani, A., Ray, P., Paramesh, N., Bhar, R., The development of ontology driven multi-agent systems : a case study in the financial services domain, *Proceedings of the IEEE International Workshop on Business Services Networks*, IEEE Press, Hong Kong, 2005, 1-1pp.

Taleb, N. N., The Black Swan: The Impact of The Highly Improbable, *Random House Trade Paperbacks*, 2011.

Vandic, D., van Dam, J. W., Frasincar, F., Faceted product search powered by the semantic web, *Decision Support Systems* 53 (2012) 425-437.

Wenger, M., Thomas, M.A., Babb Jr., J.S., An Ontological Approach to XBRL Financial Statement Reporting. In: *AMCIS 2011 Proceedings* (2011).

Zopounidis, C., Doumpos, M., Matsatsinis, N. F., On the use of knowledge-based decision support systems in financial management: a survey, *Decision Support Systems* 20 (3) (1997) 259-277.

Valiante, D., Survey finds financial professionals dissatisfied with market data quality, *Wall Street & Technology*, 2008, Retrieved on Mar.1 2011 from http://www.wallstreetandtech.com/articles/208801964.

Wand, Y., Wang, R. Y., Anchoring data quality dimensions in ontological foundations, *Communications of the ACM* 39 (11) (1996) 86-95

Wang, R. W., Strong, D. M., Beyond accuracy: what data quality means to data consumers, *Journal of Management Information Systems* 12 (4) (1996) 5-25.

Zhang, D., Zhou, L., Discovering golden nuggets: data mining in financial applications, *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews* 34 (4) (2004) 513-522.

http://www.fujitsu.com/global/about/resources/news/press-releases/2013/0403-02.html

14 Appendix

14.1 Preliminary outline

Reporting format and language

In consultation with the thesis supervisors at the start of the research project, the writing format of the Master Thesis had to be discussed. Regarding the writing format a distinction between two formats can be made, based on the type of research and preferences of the student, the university supervisor and the external organization:

1. Research article
2. Research report

This master thesis will be delivered in the form of a research report since there is an external organization involved (in this case TNO) and the research is based amongst others on case studies. Therefore it will contain more contextual information about the case studies, in order to enable the organization to grasp the content.

The lay-out of the research report will be based on the standard professional and academic reporting conventions, containing:

I. Management summary
II. Preface
III. Table of content
1. Introduction
2. Literature review
3. Research design
4. Results
5. Conclusions and recommendations
6. Discussion
IV. References
V. Appendices

English was chosen as the reporting language, because the current language of the master program Business Administration (BA) is English and the included case studies will be conducted in English as well. Furthermore most research articles are in English and this makes the writing in English more advantageous.

APA-style conventions

Compliance to the APA-style conventions according to:

- http://www.csu.edu.au/division/studserv/my-studies/learning/pdfs/apa.pdf
- http://www.uhv.edu/ac/style/pdf/APAQuickGuide6.pdf

including correct citation of consulted articles and other sources is assured, as used within the University of Twente (Universiteit Twente, 2014).

14.2 Preliminary planning

In order to secure completion and a timely appropriate covering of the relevant issues an overall plan was proposed and agreed upon, including five milestones between the start of the master thesis assignment and the student`s graduation. Those milestones include the research proposal (research aim, main research question, research goals and sub-questions), a systematic literature review, semi-structured interviews with LD professionals within and outside the organizational borders of TNO, the development of a conceptual model for the potential of LD, its application on an ongoing case within the financial industry, matching and validating of the model against financial experts' opinions and finally discussion and conclusion. An overview of this overall planning is provided in *figure 9: High level (aggregated) time planning, incl. five milestones.*

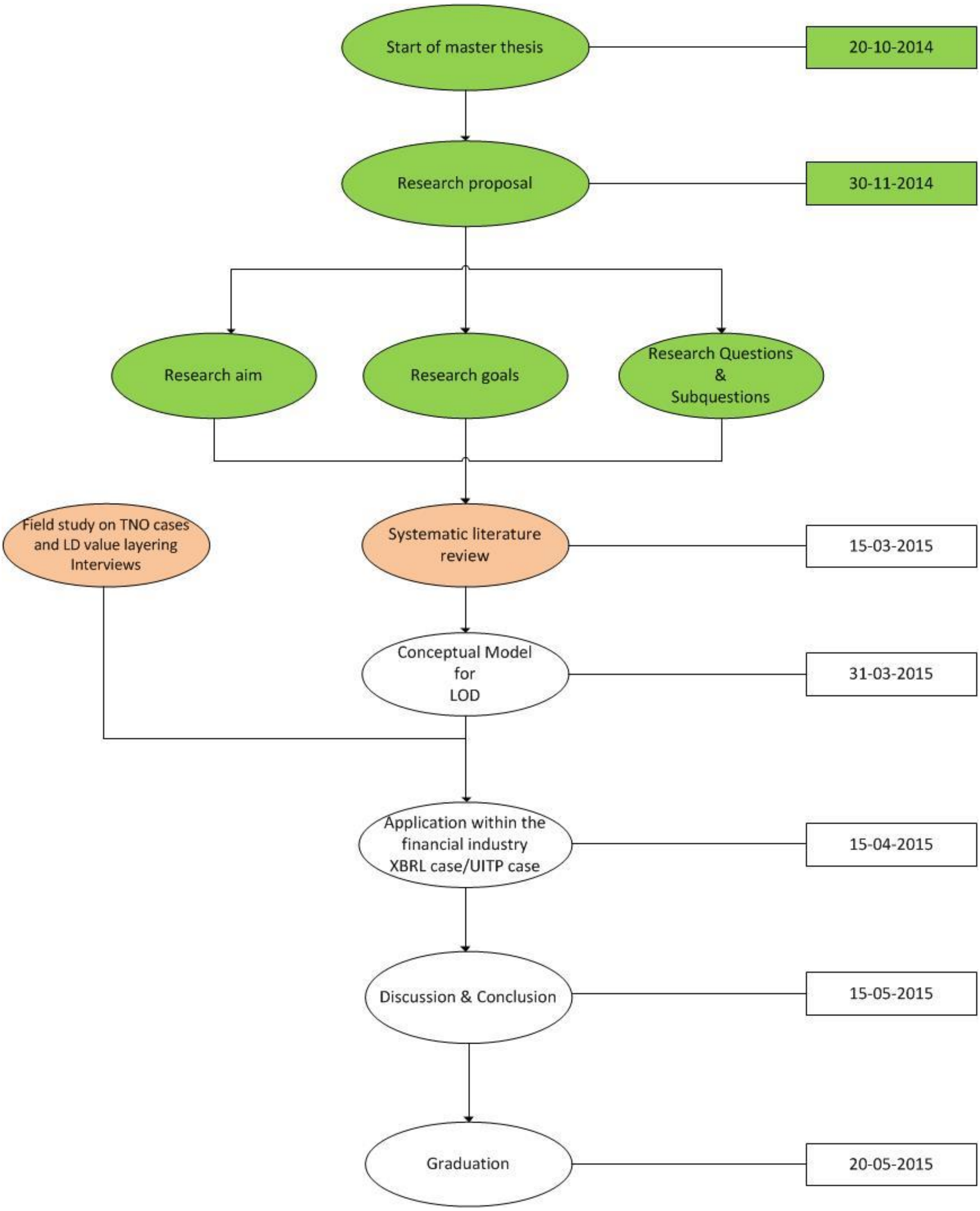


Figure 21: High level (aggregated) time planning, incl. five milestones

Furthermore due to practicability issues a water-flow planning wouldn`t suffice. Therefore a further disaggregation in planning was conducted to ensure parallelism in task execution and a fit within the scheduled time frame. Within this planning each sub-question is assigned a time frame thus ensuring a clear judgment over the work progress at any desirable point of time. At the same time parallelism ensures proper time frame per subject area. An overview of the disaggregated time planning is provided below.

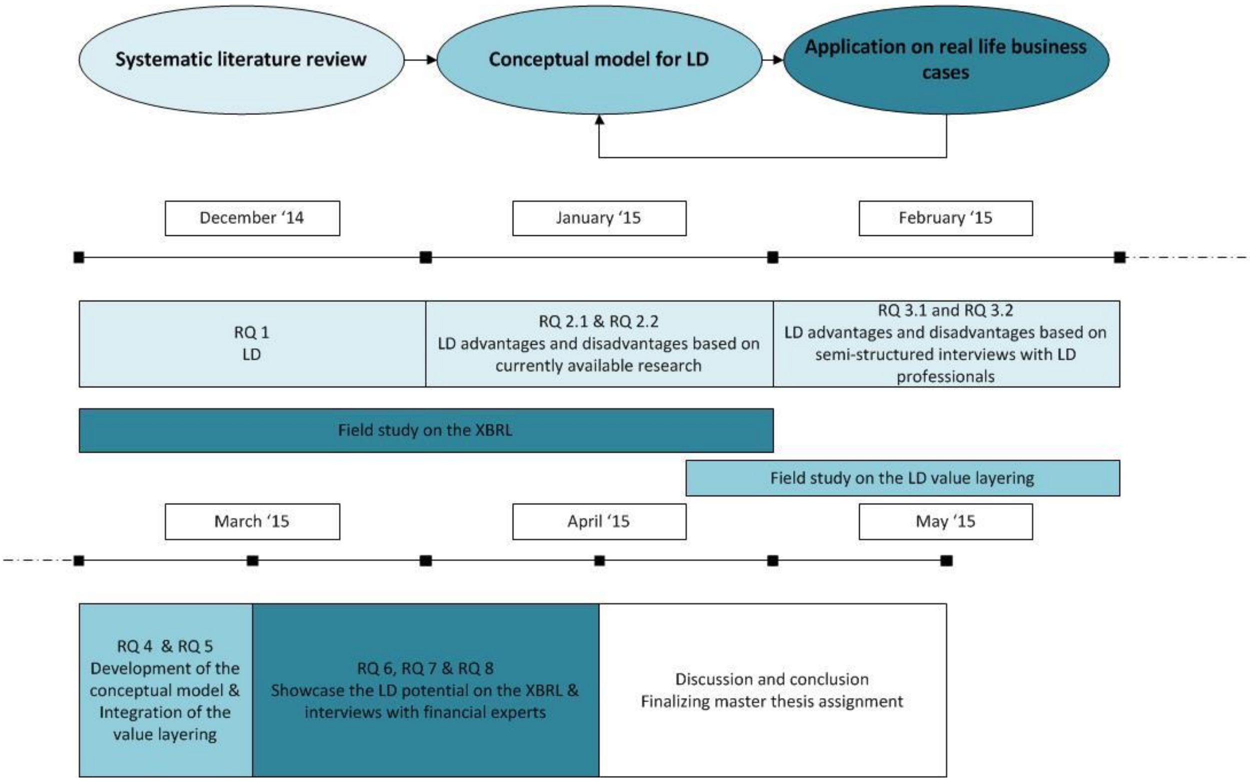


Figure 22: Disaggregated time planning

14.3 List of abbreviations.

LD	Linked Data
RDF	Resource Description Framework
LOD	Linked Open Data
FinO	Financial ontology, as introduced by Du and Zhou (2012)
KPI	Key Performance Indicator
US-GAAP	US Generally Accepted Accounting Principles
FIOS	Financial Information Observation
XBRL	eXtensible Business Reporting Language
CIK	Central Index Key
US SEC	The United States Securities and Exchange Commission
AAA	Anybody can say Anything about Any topic
TAM	Technology Acceptance Model
CS/IS	Computer Science/Information Systems
ASA	Active Scholar Assessment
OWL	Web Ontology Language
URI	Unique Resource Identifier
SUMO	Suggested Upper Merged Ontology
LSDIS	Large Scale Distributed Information Systems
OESM	Ontology-based Event-driven Scenario Model
DSS	Decision Support System
KDT	Knowledge Discovery from Text
B2G	Business-To-Government
XML	eXtensible Markup Language
ATCT	Automatic Taxonomy Construction from Text
FLOPPIES	Framework for Large-scale Ontology Population of Product Information in E-commerce Stores
EMH	Efficient Market Hypothesis
B2B	Business-To-Business
OFXD	Ontology-based Framework for XBRL-mapping and Decision-making
XBRLOnt	XBRL ontology as introduced by Chowdhuri et al. (2014)
BIXL	Business Intelligence cross-lingual XBRL

14.4 Factors for LD adoption.

14.4.1 Factors with positive impact on LD adoption.

Advantages	Explanation	Variable
Full information	More perspectives come into play with LD; The “AAA” principle; LD applications operate on top of an “unbound global data space”; not against fixed set of data sources	Perceived usefulness
Easy access	Flexibility, reduced re-use barriers with LD, no need for changes in the application code.	Perceived EoU
Data re-use	LD contributes connecting different data depositories currently available on the Web into a single global information space and solves sem. heterogeneity	Perceived usefulness
Reduced redundancy in information exchange	New exchange paradigm of LD, i.e. simply linking to information instead of exchanging data;	Perceived usefulness
No cross-sectorial interoperability issues	LD vanishes cross-sectorial interoperability issues. This leads to significant operating cost savings	Perceived usefulness
Automatic knowledge acquisition	Ontology development can be made automatic	Perceived usefulness
Up-to-date knowledge	Up-to-date knowledge for decision-making in real-time businesses	Perceived usefulness
(Semi-)automatic reasoning	LD specifies semantics formally. The created knowledge database is understandable not only for humans but also for machines	Perceived usefulness
Time and resource saving	Reduced manual effort required for discovering and using web services with LD	Perceived usefulness
Antidote for human bounded rationality	LD ontologies can extract knowledge from text, with limited human intervention. For the financial domain: a decision support system for hidden regularities and trends	Perceived usefulness
Better (financial) product comparisons	financial product information is formalized; no need for providers to use templates or specific data format	Perceived usefulness
Advanced recommendation applications	Provision of aggregated information over a variety of sources with LD	Perceived usefulness
Reliable product rankings	Provision of aggregated information over a variety of sources with LD	Perceived usefulness
“extremely good fit to accounting systems”	Object-oriented approach with an extended semantic modelling capacity better match then relational data models	Attitude toward using

Table 11: Factors with positive impact on LD adoption

14.4.2 Factors with negative impact on LD adoption.

Disadvantages	Explanation	Variable
Trustworthiness issues	"Anybody can say Anything about Any topic" can endanger quality and reliability of provided query responses (decentralized publication)	Perceived usefulness
Questionable relevance of encountered resources	The user does not select the sources explicitly. Furthermore proper assessment and selection is difficult due to the absence of statements about the encountered resources	Perceived usefulness
Linked (closed) data	Global entity consolidation policies may hamper LD application; in the case of LD and not LOD, the RDF links might just lead to "dead ends" for a specific user query	Perceived Ease of Use
Denial of services	Voluntary or involuntary denial of services due to queries complexity and/or excessive data is possible. This might endanger service level agreements	Perceived usefulness
Necessary change in formal definitions	For instance exchange of "business documents" whether in physical or electronic form is made explicit in current legislation.	Attitude toward using
Security concerns	Due to a.o. SparQL endpoints. If an endpoint is open, the data can be accessed by anyone, since endpoints don't have any security or access policies in place by default	Attitude toward using
Altering of legacy systems	Many of them don't support semantic web technology and LD, so they have to be replaced and/or adjusted	Perceived usefulness
Limited LD expressiveness	Requires an additional specification of the concepts and a precise definition of the contexts, within which different concepts are the same	Perceived Ease of Use
Complexity	Complexity increases if there are different types of relations between concepts such as "similarTo", parenting relationships etc.	Perceived Ease of Use
"semantic web" equivalent of "mandatory field"	Difficulty with finding a "semantic web" counterpart of a "mandatory field". Currently there is none [as of February 2015].	Perceived Ease of Use
Limited applicability for financial information aggregation	Due to extremely sensitive (financial) information. Applies to product comparisons, rankings and recommendation applications	Perceived usefulness
Currently not dominant on the market	Relational data models and not object-oriented are still currently dominant in the market.	Attitude toward using

Table 12: Factors with negative impact on LD adoption

14.5 Multi-item measurement scale for the adoption determinants

Perceived Ease of Use

Learning to operate CHART-MASTER would be easy for me.

likely	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	

I would find it easy to get CHART-MASTER to do what I want it to do.

likely	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	

My interaction with CHART-MASTER would be clear and understandable.

likely	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	

I would find CHART-MASTER to be flexible to interact with.

likely	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	

It would be easy for me to become skillful at using CHART-MASTER.

likely	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	

I would find CHART-MASTER easy to use.

likely	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	

Figure 23: Multi-item measurement scale for perceived Ease of Use (Davis, 1989)

Perceived Usefulness

Using CHART-MASTER in my job would enable me to accomplish tasks more quickly.

likely	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	

Using CHART-MASTER would improve my job performance.

likely	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	

Using CHART-MASTER in my job would increase my productivity.

likely	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	

Using CHART-MASTER would enhance my effectiveness on the job.

likely	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	

Using CHART-MASTER would make it easier to do my job.

likely	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	

I would find CHART-MASTER useful in my job.

likely	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	|_____|	unlikely
	extremely	quite	slightly	neither	slightly	quite	extremely	

Figure 24: : Multi-item measurement scale for perceived Usefulness (Davis, 1989)