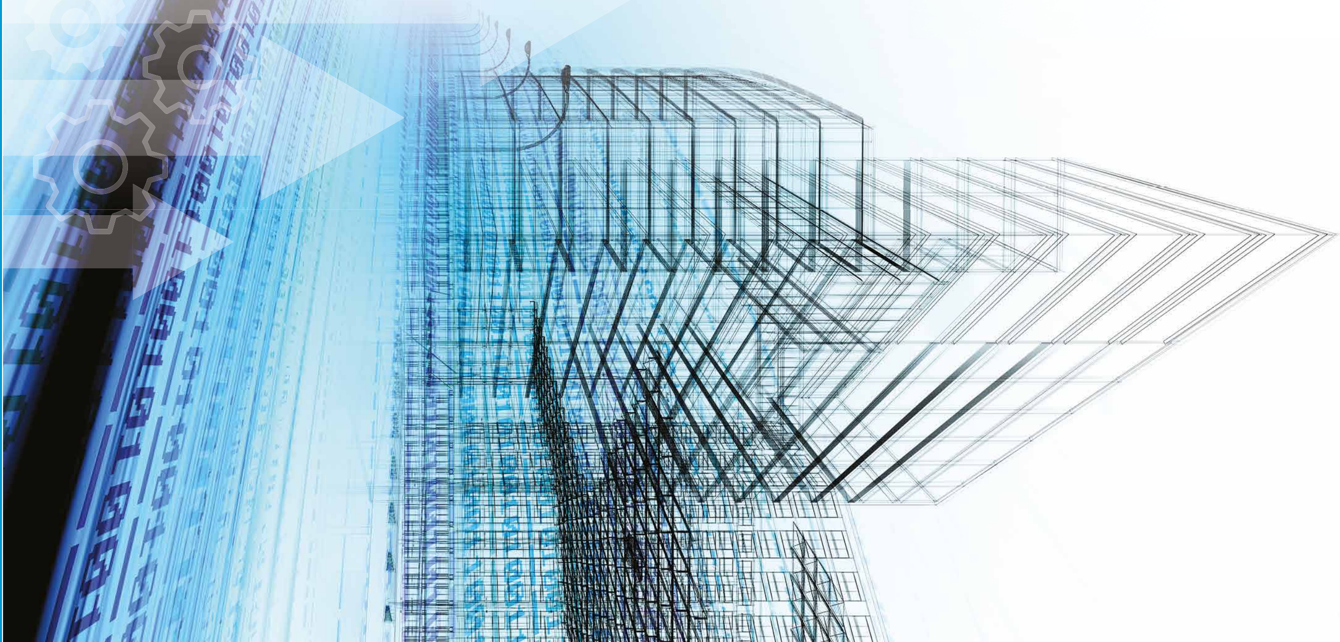
PROPELLING THE POTENTIAL OF ENTERPRISE LINKED DATA IN AUSTRIA

ROADMAP AND REPORT



Propelling the Potential of Enterprise Linked Data in Austria

Roadmap and Report

PROPEL is funded by the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT) and the Austrian Research Promotion Agency (FFG) under the program "ICT of the Future" between November 2015 and December 2016.

Imprint

Published by:

edition mono/monochrom
Zentagasse 31/8, A-1050 Vienna
fax/fon: +431/952 33 84
edition-mono@monochrom.at

ISBN:

978-3-902796-55-4

Editors:

IDC Central Europe GmbH, Semantic Web Company GmbH,
Vienna University of Economics and Business, Vienna University of Technology

Written by:

Dr. Javier Fenández, Dr. Elmar Kiesling, Dr. Sabrina Kirrane, Mag. Julia Neuschmid,
Mag. Nika Mizerski, Univ.-Prof. Dr. Axel Polleres, Dr. Marta Sabou, Thomas Thurner,
Dipl.-Ing. (FH) Peter Wetz

Design & Layout:

Susan Härtig (Semantic Web Company)

Print:

druck.at, Leobersdorf, Austria

Copyright:

Table of Contents

AUTHORS

Dr. Javier Fenández

Javier Fenández has joined Vienna University of Economics and Business (WU) in January 2015 as a postdoctoral research fellow under an FWF (Austrian Science funds) Lise-Meitner grant. His expertise includes compression of Linked Data, and his current research focuses on efficient Linked Data access, archiving and processing. He was leading a W3C member submission on efficient, binary RDF interchange (HDT) and has published over 20 papers in the areas of Linked Data, RDF and its applications.

Dr. Elmar Kiesling

Elmar Kiesling, PhD (Head of Linked Data Lab, Information and Software Engineering Group, Institute of Software Technology and Interactive Systems) joined Vienna University of Technology (TUW) as a postdoctoral researcher in 2012 after completing his PhD at the University of Vienna. He has contributed numerous scientific publications in a wide range of research areas including Linked Data, operations research and management science, decision analysis, agent-based simulation, innovation, and blended learning.

Dr. Sabrina Kirrane

Sabrina Kirrane joined the Vienna University of Economics and Business (WU) as a postdoctoral researcher in September 2015. Prior to taking up the position at WU she was a researcher at the Insight Centre for Data Analytics, Ireland. Her PhD focused on the problem of access control for the Web of Data. Before that she spent several years working in Industry on topics around data integration and security, such as system security requirements their and implementation in an application service provider environment. Dr. Kirrane is a guest editor for the Journal of Web Semantics special issue on Security, Privacy and Policy for the Semantic Web and Linked Data and a co-organiser of the PrivOn workshop series on Society, Privacy and the Semantic Web – Policy and Technology.

Mag. Julia Neuschmid

Julia Neuschmid joined IDC in 2015. As a senior research analyst, she is responsible for the analysis of trends and developments in the IT services market. She has several years of experience in applied research in the field of technology, which she leverages for her activities with IDC. As a researcher, Julia has worked on several Austrian and European research projects, some of which she has coordinated. These were focused on urban technologies and topics such as smart cities, mobility, active assisted living, and open data. She presents the results of the research on a regular basis at national and international conferences. Julia is PMP certified, and has a master's degree in geography from the University of Vienna.

Mag. Nika Mizerski

Nika Mizerski joined the Semantic Web Company in 2015. Nika studied International Business Administration and Slavonic Languages in Austria, Poland and the US. She started her career as a consultant at a Big Four company with a strong technological focus. That led Nika to a well-known regional CRM software vendor, where she accompanied a major rebranding process as Product Marketer. Her marketing talent in a professional context got awarded from the Executive Academy of the Vienna Business Administration University. Since 2012, you can follow Nikas marketing perspective @ marketingpotentials.com.

Univ.-Prof. Dr. Axel Polleres

Univ.-Prof. Dr. Axel Polleres joined Vienna University of Economics and Business (WU) in September 2013. Before he worked at TU Vienna, Univ. Innsbruck, Univ. Rey Juan Carlos (Spain), DERI/NUIG (Galway, Ireland), and Siemens AG. His research topics include ontologies, query and rule languages, Semantic Web, Web services, knowledge management, Linked Open Data and configuration technologies. He has more than 100 scientific publications and has actively contributed to international standardisation efforts in the W3C where he co- chaired the W3C SPARQL working group. He is also currently participating in W3C's RSP community group, the RDF Data Shapes working group as well as representing WU in W3C's advisory committee.

Dr. Marta Sabou

Marta Sabou is a Senior Researcher at the Vienna University of Technology. Previously she was an Assistant Professor at the MODUL University Vienna and a Research Fellow at The Open University. She holds a PhD from Vrije Universiteit Amsterdam. She won the IEEE Intelligent System's Ten to Watch Award (2006) for her PhD work and the Linked Data Cup at i-Semantics 2012. Her major areas of research are Semantic Web, Linked Data, Human Computation. Her work has been published in high-impact conferences (WWW, Int. Semantic Web Conferences), the J. of Web Semantics, and IEEE

Intelligent Systems. She is a key member of the Semantic Web community where she acts as an organizer of major conferences (e.g., PC Chair for ESWC'15, Resources Track Chair ISWC'16) as well as editorial board member for three core journals in this area (SWJ, IJSWIS, JoDS).

Thomas Thurner

Thomas Thurner started his career at Schlumberger Industries' R&D-Department as developing engineer for energy consumption meters. Thomas started as an 'early mover in Austrian free radios' in the early 90ies and was thereafter one of the founding partners of Vienna's first community radio. Beginning of 2002 Thomas and some colleagues launched the SpinOut-Company "Team Teichenberg" active in the fields of audio exchange, streaming, eLearning and podcasting. From 2006 to 2008 he was in charge of a two year project on setting up an innovation hub on behalf of Telekom Austria. Since early 2008 Thomas is coordinating the Semantic Web Company's transfer activities as well working on public relations and communication campaigns. Thomas is also heavily engaged at Semantic Web Company's Open-Data activities and consulting, where he is active in community building and consulting for a growing Linked Open Government Data (LOD) scene in Austria and beyond.

Dipl.-Ing. (FH) Peter Wetz

Peter Wetz (Linked Data Lab, Information and Software Engineering Group, Institute of Software Technology and Interactive Systems) became enthusiastic about Linked Data technologies while working as a junior researcher on several research projects at the Know-Center and Graz University of Technology. In 2013 he joined TU Wien as a member of the Linked Data Lab and as an associate to the Doctoral College Environmental Informatics where he applies Linked Data technologies to Environmental Data streams as part of his Doctoral Thesis.

1. MANAGEMENT SUMMARY

In times of digital transformation and considering the potential of the data-driven economy, it is crucial that data is not only made available, data sources can be trusted, but also data integrity can be guaranteed, necessary privacy and security mechanisms are in place, and data and access comply with policies and legislation. In many cases, complex and interdisciplinary questions cannot be answered by a single dataset and thus it is necessary to combine data from multiple disparate sources. However, because most data today is locked up in isolated silos, data cannot be used to its fullest potential.

The core challenge for most organisations and enterprises in regards to data exchange and integration is to be able to combine data from internal and external data sources in a manner that supports both day to day operations and innovation. Linked Data is a promising data publishing and integration paradigm that builds upon standard web technologies. It supports the publishing of structured data in a semantically explicit and interlinked manner such that it can be easily connected, and consequently becomes more interoperable and useful.

The PROPEL project – Propelling the Potential of Enterprise Linked Data in Austria – surveyed technological challenges, entrepreneurial opportunities, and open research questions on the use of Linked Data in a business context and developed a roadmap and a set of recommendations for policy makers, industry, and the research community. Shifting away from a predominantly academic perspective and an exclusive focus on open data, the project looked at Linked Data as an emerging disruptive technology that enables efficient enterprise data management in the rising data economy.

Current market forces provide many opportunities, but also present several data and information management challenges. Given that Linked Data enables advanced analytics and decision-making, it is particularly suitable for addressing today's data and information management challenges. In our research, we identified a variety of highly promising use cases for Linked Data in an enterprise context. Examples of promising application domains include 'customization and customer relationship management', 'automatic and dynamic content production, adaption and display', 'data search, information retrieval and knowledge discovery', as well as 'data and information exchange and integration'. The analysis also revealed broad potential across a large spectrum of industries whose structural and technological characteristics align well with Linked Data characteristics and principles: energy, retail, finance and insurance, government, health, transport and logistics, telecommunications, media, tourism, engineering, and research and development rank among the most promising industries for the adoption of Linked Data principles.

In addition to approaching the subject from an industry perspective, we also examined the topics and trends emerging from the research community in the field of Linked Data and the Semantic Web. Although our analysis revolved around a vibrant and active community composed of academia and leading companies involved in semantic technologies, we found that industry needs and research discussions are somewhat misaligned. Whereas some foundation technologies such as knowledge representation and data creation/publishing/sharing, data management and system engineering are highly represented in scientific papers, specific topics such as recommendations, or cross-topics such as machine learning or privacy and security are marginally present. Topics such as big/large data and the internet of things are (still) on an upward trajectory in terms of attention. In contrast, topics that are very relevant for industry such as application oriented topics or those that relate to security, privacy and robustness are not attracting much attention. When it comes to standardisation efforts, we identified a clear need for a more in-depth analysis into the effectiveness of existing standards, the degree of coverage they provide with respect the foundations they belong to, and the suitability of alternative standards that do not fall under the core Semantic Web umbrella.

Taking into consideration market forces, sector analysis of Linked Data potential, demand side analysis and the current technological status it is clear that Linked Data has a lot of potential for enterprises and can act as a key driver of technological, organizational, and economic change. However, in order to ensure a solid foundation for Enterprise Linked Data include there is a need for: greater awareness surrounding the potential of Linked Data in enterprises, lowering of entrance barriers via education and training, better alignment between industry demands and research activities, greater support for technology transfer from universities to companies.

The PROPEL roadmap recommends concrete measures in order to propel the adoption of Linked Data in Austrian enterprises. These measures are structured around five fields of activities: 'awareness and education', 'technological innovation, research gaps, standardisation', 'policy and legal', and 'funding'. Key *short-term* recommendations include the clustering of existing activities in order to raise visibility on an international level, the funding of key topics that are under represented by the community, and the setup of joint projects. In the *medium term*, we recommend the strengthening of existing academic and private education efforts via certification and to establish flagship projects that are based on national use cases that can serve as blueprints for transnational initiatives. This requires not only financial support, but also infrastructure support, such as data and services to build solutions on top. In the *long term*, we recommend cooperation with international funding schemes to establish and foster a European level agenda, and the setup of centres of excellence.

PROPEL was funded by the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT) and the Austrian Research Promotion Agency (FFG) under the program "ICT of the Future" between November 2015 and December 2016. The following organizations were involved in the collaborative research: IDC Central Europe GmbH, Vienna University of Economy, Vienna University of Technology, and the Semantic Web Company.

2. SCOPE AND METHODOLOGY

2.1. About PROPEL

The PROPEL project – Propelling the Potential of Enterprise Linked Data in Austria – surveys technological challenges, entrepreneurial opportunities, and open research questions on the use of Enterprise Linked Data (ELD) in a business context. Shifting away from a predominantly academic perspective and an exclusive focus on open data, we conceive Linked Data as an emerging key disruptive technology for efficient enterprise data management in the rising data economy. PROPEL was funded by the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT) and the Austrian Research Promotion Agency (FFG) under the program "ICT of the Future" between November 2015 and December 2016. The following organisations were involved in the collaborative research: IDC Central Europe GmbH, the Semantic Web Company, the Vienna University of Economics and Business and the Vienna University of Technology.

The PROPEL exploratory study draws a picture of the Austrian Linked Data (LD) landscape, including characteristics of data-driven industries, enterprise data management requirements, technological challenges, and identifies recommendations for stakeholders from research and industry regarding the efficient use of Enterprise Linked Data for data and information management.

The target groups of the document are policy makers, businesses / industry, and the research community. The document is structured in eight main parts. *Chapter 2* describes the concept of (Enterprise) Linked Data as well as the objectives and methods used for this study. *Chapter 3* describes key market forces for data and information management in Austria covering business, technological, social and legal aspects. The chapter further discusses the impact that market forces have on data and information management. A comprehensive characterisation of data-driven industries and a discussion on their predisposition in terms of Enterprise Linked Data adoption is provided in *Chapter 4*. The sectorial analysis is expanded upon in *Chapter 5* in the form of a demand side market analysis, which highlights the data challenges that organisations are currently facing, and derives enterprise requirements for data and information management. While, *Chapters 3 to 5* focus on the market/industry potential of ELD, *Chapter 6* examines trends within the research community and identifies gaps between existing research activities and the market needs. *Chapter 7* brings these insights together in one comprehensive roadmap for Enterprise Linked Data in Austria, and provides recommendations for industry, the research community and policy makers on how to integrate and use enterprise data more effectively. The roadmap concludes with *Chapter 8*, which brings together the key results and compiles "the complete picture".

For an interactive discovery of the Enterprise Linked Data roadmap please visit the project website https://www.linked-data.at/.

2.2. (Enterprise) Linked Data

The success of many businesses is increasingly becoming a function of effective and efficient leveraging of information. At the same time, the growth in information volume, velocity, variety and complexity and new information use cases make information management even more complex than it has been in the past.

Linked Data (LD) is a promising data integration and data interchange paradigm. It is a method of publishing structured data so that it can be interlinked and become more useful in both business operations and strategy. It builds upon standard web technologies that enable data from different sources to be connected and queried. (Bizer, Heath, Berners-Lee 2009)

Although LD has mainly been used as a concept for publishing structured data on the web, it also has significant disruptive potential in enterprise settings. LD principles and technologies can be applied by businesses to create value: (i) *internally*, for instance to facilitate data and legacy system integration, and (ii) *beyond organisational boundaries*, enabling efficiencies in terms of cooperative engineering or supply chain integration.

Linked Data provides guidelines for publishing and consuming data on the web to facilitate the interlinking of data just like web pages. Beyond open data, Linked Data technologies and principles enable flexible bottom-up integration within and between enterprises, a development frequently referred to as Enterprise Linked Data (ELD) (Wood et al., 2013; Passant, 2010). Furthermore, Linked Data is closely related to emerging big data technologies[1] and may serve as a complementary technique for organizing large volumes of polystructured data.

1 Big data is a new generation of technologies and architectures designed to economically extract value from very large volumes of a wide variety of data by enabling high velocity capture, discovery, and/or analysis"(Sabharwal, 2016).

2.2.1. Principles and Technology Stack

LD is based on a set of four basic principles: (i) use Uniform Resource Identifiers (URIs) to name things; (ii) use Hypertext Transfer Protocol (HTTP) URIs so that people can look up things; (iii) when someone looks up a URI, provide useful information, using the standards; and (iv) include links to related URIs so that people can discover more things. The growing adoption of these principles in recent years, particularly among the research community, as well as the increasingly common practice of embedding metadata into Hypertext Markup Language (HTML) pages (such as promoted by schema.org), have led to a vast volume of RDF data being made openly available online. (Auer et al. 2013; Heath and Bizer 2011; Bizer et al. 2009)

The Linked Data stack includes some the following technologies and standards:

RDF: The Resource Description Framework is a standard model for data interchange on the web that can also be used in a corporate environment. A variety of data schemes can be mapped with RDF, which enables data processing irrespective of the underlying data model.

URI: The Uniform Resource Identifier is used to identify an entity. Every concept is included only once in a knowledge model, but can be linked multiple times depending on the context.

SPARQL: SPARQL is the standard query language for RDF, and enables application developers and data scientists to query and traverse graphs.

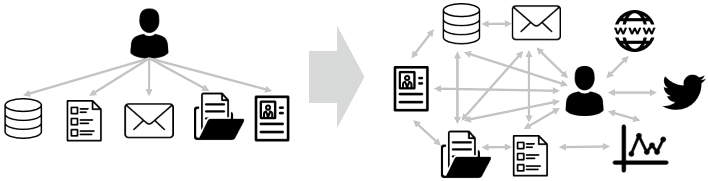
SKOS: The Simple Knowledge Organisation System is a standard for building knowledge models. It includes a set of modelling rules and is based on RDF. As its name suggests, SKOS can be used by subject-matter experts without specialized skills in knowledge modelling.

OWL: The Web Ontology Language is an advanced knowledge modelling standard that is also based on RDF. OWL enables knowledge engineers to model complex domains in the most specific way.

2.2.2. Enterprise Linked Data

Business-critical information is often gathered in information systems, such as Enterprise Resource Planning (ERP), Customer Relationship Management (CRM) and Supply Chain Management (SCM) systems. Integration across these systems as well as integration with the abundance of other information sources is still a major challenge.

Figure: Leaving a conventional enterprise data setup towards Linked Data



Source: PROPEL, 2016

In the last decade, the prevalent data integration approaches were primarily based on XML, Web Services and Service Oriented Architectures (SOA). Even though classic SOA architectures are well-suited for transaction processing, other technologies that can be employed for enterprise data integration are also available. In particular, the use of the Linked Data paradigm for integrating enterprise data is a highly promising approach.

The Linked Data principles has been designed to facilitate the specification of dynamic descriptions for data, and to construct simple links between data so as to build a comprehensive knowledge base within an organisation. Such a knowledge base can be used to aid decision making and to find solutions to complex questions. In a conventional setup, users access different applications that each manage their own data. Users who must interrogate multiple sources of data to resolve a problem are forced to consult each application separately and then undertake the sometimes tedious process of consolidating the results. Enterprises that deploy a Linked Data strategy, can interlink distributed data sources, and even enrich the internal data repository with external data, e.g. from social media, statistics, open data, Wikipedia, etc. in order to develop a comprehensive data and information space.

"Organisations are facing fragmented data and the desire to connect this data, and linked data provides a mechanism for doing that. LD is an interoperability layer – an added layer of intelligence – on top of existing systems, without disrupting the existing system." (Linked Data Expert and Pioneer, UK, PROPEL interview 2016)

2.3. Objectives and Research Plan

The PROPEL project assesses enterprise data integration needs and both the economic and innovative potential of Linked Data for business. We base our research on a comprehensive market and industry analysis that includes the identification of relevant market forces and their impact on data and information management, the identification of criteria that can be used to systematically examine industrial characteristics with a view to enabling the classification of industries based on their potential for LD adoption. Informed by the market forces and industry analysis, we examine the data and information challenges that enterprises are facing and formulate user stories. In parallel, we examine LD research and development (R&D) activities by analysing scientific trends, involvement of industry in the research community and existing standardisation efforts. The outputs from the industry, market, sectoral and technical analysis feed into and enable the creation of an integrated picture of the drivers and barriers for LD adoption in Austrian enterprises. All these activities are enriched with stakeholder integration measures and activities to disseminate the project results to potential LD adopters and the LD R&D community. The overarching aim is to develop a roadmap that consists of objectives, measures, and recommendations for furthering ELD development and adoption.

The objectives and key research questions are:

Objective 1: The characterisation of industries/domains (such as finance, manufacturing, media, health, etc.) according to defined criteria in order to identify the industries/domains with highest potential for the adoption of the (E)LD concept.

Key question:

- *Which industries are the most likely to adopt LD technologies?*

Objective 2: An investigation into the data and information management challenges that industries are currently facing, and their formulation as use cases.

Key questions:

- *What are the key challenges in data management and integration from a demand-side perspective?*
- *What are examples of promising use cases?*

Objective 3: Analysis of the Linked Data community, its current research and development activities as well as open challenges in regards to LD technologies and standards.

Key question:

- *What are technological and standardisation opportunities and challenges?*

Objective 4: The development of an integrated roadmap based on the industry, market, and Linked Data community analysis.

Key question:

- *What recommendations are necessary for enterprises, policy makers and researchers in order to propel the adoption of LD in enterprises?*

Objective 5: The elicitation of requirements from all stakeholder groups and the dissemination of the research results through various channels to the R&D community and potential LD adopters.

Key question:

- *Who are key stakeholders and how can they be reached?*

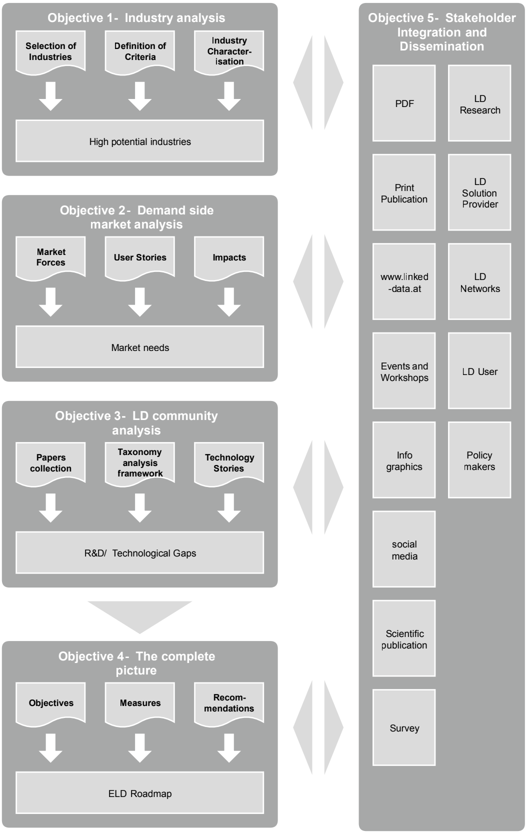


Figure: PROPEL Research Plan

2.4. Tools and measures

The research elicits input from the following key stakeholder groups:

1. *Business/industry:* these are enterprises that are in need of optimised data and information management in order to develop new and enhance existing products and services, and improve efficiency and customer experience. These stakeholders are potential adopters of the Linked Data concept;
2. *Research sector:* the research community consists of universities and other research organisations that have been involved in researching and developing the Linked Data concept over the past years. These stakeholders are interested in identifying research gaps and opportunities, and establishing new collaborations with industry;
3. *Policy makers:* these are stakeholders involved in initiatives to support innovation in data and information management such as (inter)national (public) research funding bodies. This group also includes national and international bodies that work on standardisation in data and information management.

The methods used in the course of this research project are described below. Although the primary aim was to elicit requirements, the interviews and the stakeholder workshop also served to make Austrian enterprises aware of the potential benefits of Enterprise Linked Data and how it can accelerate their business.

2.4.1. Conceptual Framework for Sectoral Analysis

In order to assess the potential of LD in enterprise, we develop working hypotheses that link industry characteristics to the propensity to adopt Linked Data technologies. Based on these working hypotheses, we systematically identify industries whose structural and technological characteristics align with the technical characteristics and principles of Linked Data. Industries identified as promising in this analysis: (i) have structural characteristics that suggest that they may benefit strongly from adopting LD technologies; and (ii) are characterised by high technological readiness for LD. The goal of this analysis is to provide a comprehensive high-level overview from a generic international perspective. The hypothesis-driven, industry-level analysis then serves as a starting point for the development of the roadmap and informs the development of guidelines for the industry expert interviews and the stakeholder workshop, which resulted in the analysis presented in *Chapter 5*.

For the classification of industries, we rely on the economic activity classes defined by the 2008 ÖNACE[2] system, which is derived from the statistical classification of economic activities in the European Community (NACE rev. 2[3]). NACE is a framework for collecting and presenting a large range of statistical data according to economic activity in the fields of economic statistics (e.g. production, employment and national accounts) and in other statistical domains developed within the European statistical system (ESS). Statistics produced on the basis of NACE are comparable at European and, more generally, at the world-wide level. The use of NACE is mandatory within the European statistical system[4].

We base our industry classification on the top level sections defined in the ÖNACE framework and make selective use of the more detailed level 2 classes for more fine-grained analyses of the Telecommunications (J61), Engineering (M71), and Scientific Research and Development (M72) domains. It is worth noting that we do not cover the entire set of economic activities defined within NACE, but focus on 15 (out of 21 fields) that are both (i) sufficiently homogeneous in their internal characteristics to allow for integrated analysis and (ii) potentially highly relevant for the development and adoption of Linked Data in the Austrian context[5]. The following table summarizes the list of industries that we cover in our analysis and their corresponding ÖNACE classifications.

Section	Code ÖNACE	ÖNACE Section
4.2.1 Resource Industries	A, B	Agriculture, forestry and fishing, mining and quarrying
4.2.2 Manufacturing	C	Manufacturing
4.2.3 Energy	D	Electricity, gas, steam and air conditioning supply
4.2.4 Construction	F	Construction
4.2.5 Retail	G	Wholesale and retail trade, repair of motor vehicles and motorcycles
4.2.6 Transport and Logistics	H	Transportation and storage
4.2.7 Tourism	I	Accommodation and food service activities
4.2.8 Media	J	Information and communication

2 ÖNACE 2008
3 http://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF
4 http://ec.europa.eu/eurostat/statistics-explained/index.php/NACE_background
5 Industries defined in NACE, but not considered in our analysis are: Water supply, sewerage, waste management and remediation activities (E); Real estate activities (L), Administrative and support service activities (N); Arts, entertainment and recreation (R), Other services activities (S), Activities of households as employers (T), Activities of extraterritorial organisations and bodies (U) . The latter categories cover a broad spectrum of highly specific activities with varying characteristics and requirements. Therefore, a general high-level assessment of the potential of Linked Data is difficult in these sectors.

4.2.9 Telecommunications	J61	Telecommunications
4.2.10 Finance and Insurance	K	Financial and insurance activities
4.2.11 Engineering	M71	Architectural and engineering activities; technical testing and analysis
4.2.12 Research and Development	M72	Scientific research and development
4.2.13 Government	O	Public administration and defence; compulsory social security
4.2.14 Education	P	Education
4.2.15 Health	Q	Human health and social work activities

The methodology for analysing industries with respect to Linked Data adoption potential relies on desk research, an internal survey, industry expert interviews, and statistical data. The analysis was predominantly based on desk research, which was conducted both for the theory-guided development of working hypotheses and for the analysis of the industries along the explanatory dimensions of the working hypotheses. In addition, to validating the consistency of our valuation, we conducted an internal questionnaire-based survey in which the project team members independently evaluated industry characteristics along the identified structural and technological dimensions. In order to obtain additional information, we also consult domain experts from the respective industries. In addition, we relied on quantitative indicators such as average company size, average revenue, R&D intensity, investments in software, etc. from statistical sources[6].

2.4.2. Qualitative interviews

Qualitative, semi-structured, interviews on both a national and international level were conducted; in total 22 interviews in spring 2016. These qualitative interviews were performed with research and industry experts in the Linked Data area, and employees from companies that are struggling to deal with data and information management. The scope of the interviews was to discover potential use cases – in the form of user stories – for Enterprise Linked Data. Interviews covered questions on the organisation, data integration challenges, existing technologies and Linked Data. The general interview guide is presented in *Annex A*.

6 Statistik Austria Wirtschaftsdaten, Austrian Research and Technology Report 2015 etc.

2.4.3. Workshops

A stakeholder workshop that was conducted in May 2016 served to synthesise the information that was gathered in the previous steps of the project. The half-day workshop focused on Austrian needs and requirements. The workshop had a good mix of 22 participants from the private sector (solution providers and users), and the research sector; technicians and strategic people. While the first part of the workshop provided an overview of the PROPEL project and Linked Data technology through presentations, the second part consisted of round table discussions where participants discussed business and technical aspects of data management within their organisations. During the workshop we identified additional challenges that enterprises are facing in data and information management, potential use cases for improved data and information management and recommendations. The workshop agenda is presented in *Annex B*.

2.4.4. Internal project workshops

The consortium joined at a number of 10 internal project workshops where the research plan and structure were developed, and the intermediate analysis and assessment results were presented, discussed and further developed. In detail:

- The kick-off meeting marked the start of the project and set out the goals and objectives of the project, including the first brainstorming on the roadmap for Linked Data in Austrian enterprises.
- The industry analysis workshop drew a clear picture of relevant businesses/ industries, the key objectives of the industry analysis, and defined the methodology.
- The visionary workshop was used to share existing know-how and experience on ELD among the consortium, and to draw first visions for the roadmap.
- The market and technology analysis workshop outlined the detailed objectives and methodologies used for the demand-side market analysis and research community analysis.
- A series of four roadmapping workshops were conducted in order to integrate the analysis results, to derive recommendations for stakeholders, and to define development paths for the further adoption of the LD concept in Austrian enterprises.
- The follow-up meeting focused on the exploitation and dissemination of the PROPEL project results.

2.4.5. Data and information management survey

A user survey was conducted among Austrian IT and Line of Business managers. Firstly, market surveys by global analyst firms were collected and analysed to get a broad understanding of the market and its forces on data management and related technologies. Secondly, a quantitative online survey was conducted (n=133), in order to determine the status quo of Linked Data awareness in Austrian enterprises as well as the potential and challenges of data and information management.

2.4.6. Linked Data community analysis

The focus of the community analysis was threefold: (i) to analyse the research topics and trends emerging from the five major publishing avenues for Linked Data publishers (ISWC, ESWC, Semantics, SWJ, JWS) over the last 10 years (i.e. from 2006 to 2015 inclusive); (ii) to examine the existing standardisation activities around Linked Data within the W3C; and (iii) to investigate industry involvement in the aforementioned venues. To this end, it was necessary to gather not only metadata relating to the conference and journal proceedings but also metadata relating to the venue itself. In the case of the conferences we extracted data concerning the organising committees, tracks, call for papers, programme committees, workshops, tutorials, sessions and sponsors. While for the journals we extracted the editorial boards along with calls for papers in relation to regular calls and special issues. This data was subsequently used to create a dictionary of research topics that are of interest to the community. This dictionary was then used to count the number of occurrences of each of the topics in the conference/journal papers from the past 10 years, thus highlighting the general trends in these topics.

2.4.7. Technological Foundations

The outputs from the market, industry and technical analysis were examined not only in isolation but also in relation to one another by deriving a set of foundational technologies (based on existing taxonomies and topics) that can be used to cross correlate the information gathered during the market analysis, interviews and the stakeholder workshop:

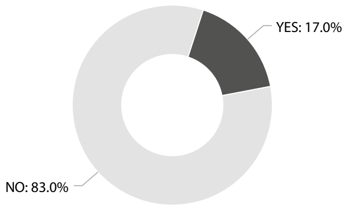
- Analytics
- Computational linguistics and natural language processing
- Concept tagging and annotation
- Data integration (incl. linking, mapping, fusion, etc.)
- Data management (incl. archiving, versioning, provenance, etc.)

- Dynamic data and streaming
- Extraction, data mining, text mining and entity extraction
- Formal logic / formal languages / description logics / reasoning
- Human computer interaction and visualisation
- Knowledge representation and data creation / publishing /sharing /reuse
- Machine learning
- Ontology/ thesaurus/ taxonomy management
- Quality (incl. data curation, inconsistency and anomaly detection, correctness)
- Recommendations
- Robustness, scalability, optimisation and performance
- Searching, browsing and exploration
- Security and privacy
- System engineering (incl. platforms, system integration, programming, etc.)

3. MARKET FORCES

While the concept of Linked Data is well known in the research sector, it has not gained widespread adoption among Austrian enterprises. According to the Data and Information Management Survey (2016) more than 80% of survey respondents are not familiar with the concept of Linked Data, and only 2% has already tested Linked Data.

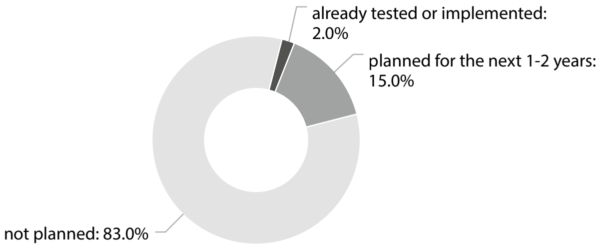
Figure: Are you familiar with the concept of Linked Data?



Source: Data and Information Management Survey, 2016

Out of the organisations that are familiar with the concept of Linked Data only 2% have already tested or implemented LD within their organisation, however an additional 15% are planning to do so in the upcoming 1-2 years.

Figure: What is the status quo of Linked Data within your organisation?



Source: Data and Information Management Survey, 2016

The PROPEL market analysis, stakeholder workshop, and interviews show similar results: enterprises have a need to deliver a better customer experience, in order to support product and service innovation and to optimise business processes. Enterprises are very much concerned about finding ways to manage their data and information more efficiently, but generally speaking they are not aware of Linked Data as a potential solution. The increase in demand for optimised data and information management leads to a growing market for solutions that add value to existing data. Therefore, this chapter describes key forces on the Austrian IT market that nowadays impact data and information management. Market forces cover aspects related to business, technology, processes and methods, users, and standards.

3.1. Austrian IT Market

3.1.1. Moderate economic growth

Generally speaking, the economy of a country is one of the most important factors influencing IT investments. The overall economic development in Austria is stable and moderate. According to Statistics Austria (September 2016), the Austrian economy expanded by 1% in 2015[7]; which represents a stronger growth than in previous years. Challenges include the 5.7% unemployment rate of the working population - the highest in many years - (Austrian Chamber of Commerce, 2016), and the threat to the European Union's stability posed by financial and geopolitical crises, leading to generally low business confidence and uncertainty on the part of IT vendors and end users alike. However, all things considered economic growth is expected to accelerate in the coming years. Austria's gross domestic product (GDP) is expected to grow by 1.6% in 2016 and by 1.5% in 2017 (Austrian Chamber of Commerce, 2016). Such an economic development will lead to an increase in IT investments - including data and information management solutions - that have been held back for several years in both the private and public sectors. A positive economic development would lead to a growth in IT spending, a relaxation of the IT market and would ease the pressure on IT vendors, including providers of data storage, data management, and analytics. (IDC, 2016c)

3.1.2. Cost reduction and process efficiency

Although Austrian organisations are increasingly returning to growth, they still focus primarily on costs. Cost reduction or process efficiency in IT or business processes - including efficiency in data management and exchange - remain the leading drivers of investment (IDC, 2016c). Despite the significant impact of technological and business developments in the IT market, the macroeconomic environment will continue to exert downward pressure on IT spending. Austrian organisations are ambivalent about efficiency improvements and expansion. On the one hand, they are under pressure to reduce costs. On the other hand, IT investments can transform business and drive

7 STATISTIK AUSTRIA Annual Data from 19.09.2016.

growth. According to an IDC end-user survey in 2015, the biggest internal challenges named by Austrian end users are reducing costs (50%). This is followed by integrating IT and line of business (47%), staying up to date with rapid technological development (37%), and digital transformation (29%). The biggest challenges for business and management are reducing the time to market for new products and services (57%), increasing efficiency (47%), and improving the infrastructure for higher productivity and cost control (32%). (IDC, 2015)

3.1.3. Digitisation and digital transformation

One of the greatest drivers for the development of the market is digitalisation. Digitisation of everything and the increase in the number of data producers are driving greater demand in data capture, management, and analysis software. Digital transformation goes even further and affects all components of an organisation: the business model, leadership, the organisation's culture, processes, data, and people. The digital transformation is already part of Austrian organisations' agendas, and 55% of organisations are actively dealing with the topic (IDC, 2016b). Cloud computing, data analytics, and process automation are key drivers of digital transformation, and related software and services will account for most IT-growth in the 2015–2020 time period. Digital transformation is simultaneously an opportunity and a risk for Austrian organisations. Vendors can support organisations with their digital transformation; however, they themselves have many things to learn, and they need to adapt their portfolios and IT services to new customer requirements.

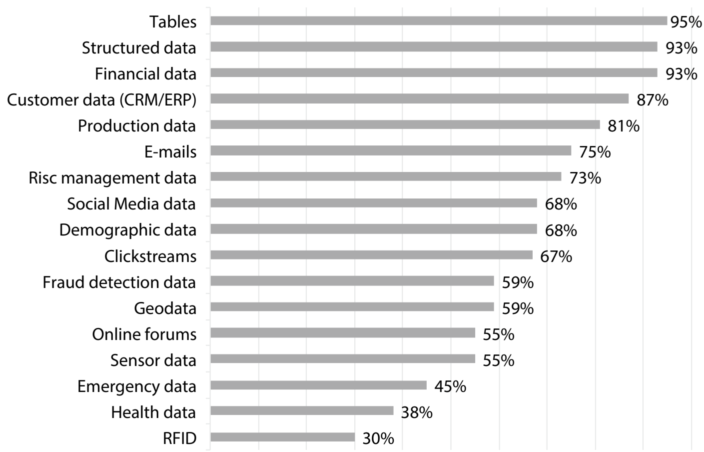
3.1.4. Technological advancement outpaces technological adoption

New technologies, including data management and processing solutions, are ready to use, yet the Austrian IT market remains reserved and conservative (IDC, 2016c). In-fact, in 2015, technological advancement outpaced technological adoption on the Austrian market. Some organisations have tried novel digital technologies in isolated areas, but few have started to make significant changes. The fear of economic recession, security concerns, and insecurity in their ability to use emerging technologies in an efficient manner keep new investments on hiatus. Many potential users have projects in the pipeline that will use technologies such as data analytics or cloud computing; however, they have taken a wait-and-see attitude, in the hope that investment in new technologies will become less risky. Decisions on new IT investments are often slow, and postponed due to doubts regarding the business development. These uncertainties result in an increasing number of smaller and more agile projects.

3.1.5. Data driven networked global economy

In an increasingly networked global economy, there is a growing need to break up data and information silos, and to collaborate and share data across organisational boundaries. Linking data across datasets and sites (resulting in Linked Data) is crucial to making data exploitable as a whole rather than as isolated unrelated datasets. "Data has been in fact regarded as the oil of the new economy. Consequently, data linking and content analytics are the key technologies to refine this oil so that it can actually drive the motor of many applications." (W3C, n.d) Technologies that support such collaboration, such as industry standards for structured information exchange, are plagued by fragmentation, limited interoperability, and restricted flexibility and extensibility. Tighter integration on the data level could yield substantial benefits in terms of operational efficiency, automation, optimisation and, ultimately, competitive advantage.

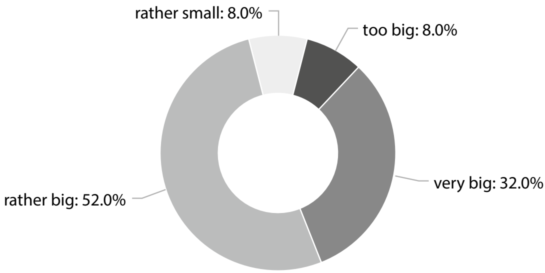
Figure: What data do you currently analyze in your organisation?



Source: Data and Information Management Survey, 2016

According to the Data and Information Management Survey (2016) almost all organisations analyse structured data in the form of tables, financial data, customer data, and production data, with systems such as CRM or ERP. Unstructured data is analysed less. For example, e-mails are analysed by two thirds of survey respondents, followed by social media data. Online forums are analysed by 55% of organisations. There is potential for digging deeper into the analysis of unstructured data, especially in combination with structured data. Efforts for data management are big for Austrian organisations (92%), and only small for a minority (8%).

Figure: How big are your efforts for data management within your organisation?



Source: Data and Information Management Survey, 2016

3.1.6. Citizen demand for online services is growing

In 2015 according to Statistics Austria 82% of Austrian households were equipped with an Internet connection. Technological developments connect citizens with work, family, leisure and government. High Internet adoption among Austrian households implies adaptation in the enterprise architecture, redefinition and expansion of services, linking virtual value chains, cost reduction in public relations, increase in safety requirements, promotion of new jobs, implementing automated man-machine interfaces and disappearance of the temporal and spatial limitations at work. The importance for the enterprise to interact with their customers, i.e. to create positive customer experiences (CX) and managing CX, is increasing. Enterprises use new channels to exchange data and information with their customers, and analyse their customer behaviour to provide personalised offerings and recommendations.

3.1.7. Data security and privacy

Common barriers to the adoption of new technology include security and privacy concerns, as well as the related challenges with respect to collecting, using, and managing customers' personal data. A shifting regulatory environment is another area of concern across industries (refer to the General Data Protection Regulation in the following section). That being said, for the overall IT market security is a key driver as new data management, mobility and cloud models require increased security, and threats are constantly increasing in volume and sophistication. 78% of Austrian organisations believe that IT security will become more important in the next years, while for the remaining 22% IT security will have the same importance as it has now, and no organisations believes that IT security will become less important for their organisation (IDC, 2016a). Line of business and IT executives will need to approach these ongoing challenges with awareness, flexibility, adaptability, and responsibility.

"Conditions that can positively impact Linked Data concepts, data integration and exchange are regulations and policies." (Founder of Austrian Start-up on Semantic Technology, PROPEL Interview 2016)

3.1.8. General Data Protection Regulation (GDPR) is a game changer

The General Data Protection Regulation (GDPR) is a new piece of EU legislation that was agreed in early 2016 and regulates personal data management within organisations. The regulation will take effect in 2018, and European organisations are starting to prepare for GDPR compliance. For security and storage software vendors the regulation means a substantial opportunity with a total market opportunity of 3200 million Euro in Europe (Brown, 2015). The severity of fines, coupled with the substantial changes in scope, will drive enterprises to radically shake up their data protection practices, seeking the assistance of new technologies to assist with compliance. The complexity involved in addressing this shift makes it more important than ever for organisations to understand the regulatory requirements and how to action their business to prepare for them. These regulations require organisations to take a fresh look at how they manage data and information, specifically where they store data - in the cloud or on-premise - and how they leverage it for data processing and analytics. (Christiansen et al., 2015)

3.1.9. Cloud technology stimulates growths

Austrian organisations look to use IT to create growth by increasing efficiency and reducing costs. Cloud technology, in particular, is stimulating growth in the Austrian IT market. Cloud offers shared computer resources (software, infrastructure, platforms) and data to computers and other devices on demand. It combines efficient use of multi tenant (shared) resources, radically simplified "solution" packaging, self-service provisioning, highly elastic and granular scaling, flexible pricing, and broad leverage of Internet standard technologies and APIs. Clients are looking to modernise their infrastructure and make it flexible. This will facilitate the fast deployment of new applications to drive growth, and make it easier for employees to access company systems and data from remote locations. Demand for cloud services in Austria is increasing, making it one of the fastest-growing technology segments in the country. The increase in cloud spending is due in part to its cannibalisation of traditional services delivery. Organisations are interested in the pay-as-you-go model, which aligns better than license models with fluctuating business demands. Data are a critical enterprise resource that nowadays can be stored efficiently and cost effectively in the cloud. Enterprises are starting to move their data to the cloud, especially hybrid cloud solutions, i.e. a mix of private and public cloud solutions. (Gens, 2015)

"(Spatial) data infrastructures become more and more cloud-based." (LD Researcher and project manager, Czech Republic, PROPEL Interview 2016)

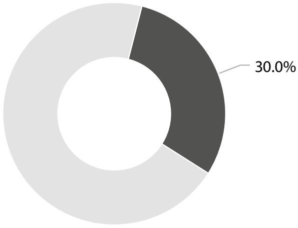
3.1.10. The Internet of Things as an innovation enabler

The Internet of Things (IoT) is a term that describes the ability of devices to connect and communicate via the Internet from any location. It is: "a network of networks of uniquely identifiable endpoints (or "things") that communicate without human interaction using internet protocol (IP) connectivity" (Turner et al., 2015). The IoT market is maturing and going beyond its initial focus of connecting more and more things. Data management is fast becoming the overarching theme, with analytics and the IoT platform emerging as the main requirements of 31% of organisations surveyed that have already launched IoT solutions, and the additional 43% looking to deploy in the next 12 months. Worldwide spending on IoT is expected to grow at a 17% compound annual growth rate (CAGR) between 2015 and 2019. In Central & Eastern Europe the fastest growing IoT category will be smart buildings, where IoT technology that utilises advanced automation and integration is being used to measure, monitor, control, and optimise building operations and maintenance. From a vertical industry perspective, manufacturing and transportation led the way in worldwide IoT spending. Both industries have been connecting their supply chains, products, customers, and even workers for some time now, and really embrace the value of business outcomes. Until 2019, the industries forecast to have the fastest IoT spending growth will be insurance, healthcare, and consumer. (Press, 2016)

3.1.11. Rising data volume, variety, and velocity

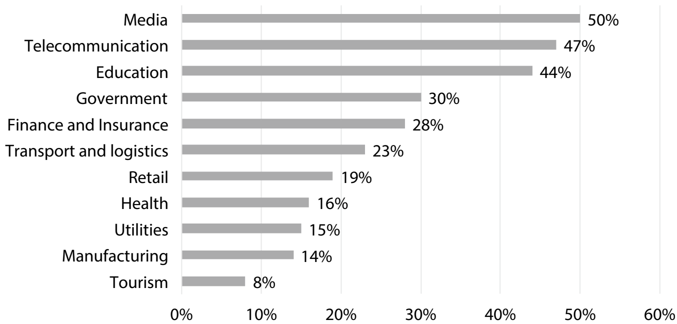
Austrian organisations say that in the last 12 months the volume of data that is processed has increased 30%. Data growths differs significantly from industry to industry. Processed data volume increased the most in the media sector (50% in the last 12 months), followed by the telecommunication sector (47%) and the education sector (44%). (Data and Information Management Survey, 2016)

Figure: How did the data volume in Austrian organisations develop in the past 12 months?



Source: Data and Information Management Survey, 2016

Figure: Growths of processed data volume per industry



Source: Data and Information Management Survey, 2016

Enterprises have a huge volume and variety of internal data as well as data that is created and stored elsewhere. Still a lot of data cannot be used to its fullest potential because it is locked up in data islands/silos. As the amount of data is constantly growing the term "big data", which is characterised by the four Vs, is pervasive in industry (ibid):

- Volume: enormous growth of the amount of data;
- Velocity: the speed of data processing;
- Variety: huge variety in data formats (arbitrary > relational > plain text);
- Value: generation of added value from data (from data to information, knowledge, and better decisions).

With big data comes a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data by enabling high velocity capture, discovery, and/or analysis (Sabharwal, 2016).

"Private companies are highly interested in big data and the analysis of real-time data". (Head of R&D at a technology provider, Italy, PROPEL Interview 2016)

The characteristics of Linked Data technologies naturally lend themselves to solve big data issues, especially heterogeneity or variety of big data. Linked Data focuses on integrating data from different sources, while big data is primarily about the processing and analysis of big data volumes. Therefore, both approaches are highly complementary.

"LD can be big and small, can help to bring additional structure and meaning to big data sources." (Linked Data Expert and Pioneer, UK, PROPEL Interview 2016)

From a market perspective, it is very important to be aware of the big data market when talking about data and information management. The big data technology and services market represents a fast-growing multibillion-dollar worldwide opportunity growing at a compound annual growth rate of 23.1% over the 2014-2019 forecast period. These figures are more or less in line with trends in Austria, even though in countries such as Germany and USA the developments go much faster than in Austria. Worldwide the big data analytics technology and services market continues to exhibit strong momentum as businesses continue their transformation into data-driven intelligence. This in turn is driving strong growth in infrastructure, software, and services revenue attributed to big data. (Nadkarni and Vesset, 2015; Nadkarni 2015)

An online study among Austrian IT and Line of Business managers collected and analysed the status quo of big data in Austrian enterprises. 36% of respondents say that big data and analytics is seriously discussed within the organisation. For 28% big data was not an issue to date. Considering that in 2014/2015 big data was not relevant for 39%; and in 2013 it was not relevant for 62%. (IDC, 2016d), there is a clear trend towards increasing interest in big data solutions.

3.1.12. Content Analytics and Cognitive Systems

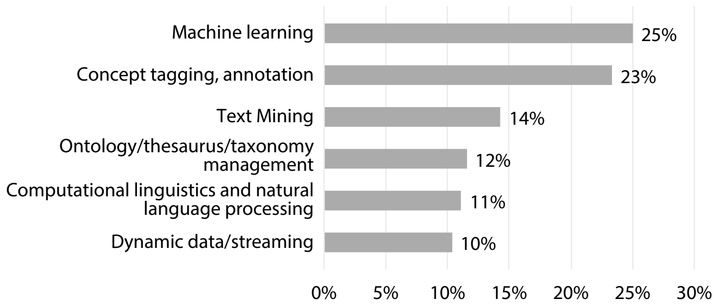
Many organisations are continually looking for ways improve the efficiency and productivity of their knowledge workers, to increase sales, reduce costs, or understand their customers better by using various types of automation. So technologies and approaches that extract meaningful information from data are developing. A lot of data such as text, video, images, social media, websites, mobile and audio data in organisation is unstructured. Content analytics straddles the line between the unstructured and the structured information worlds. It extracts the elements of meaning from unstructured information and presents them in a more structured format so that they can be combined and analysed in concert with structured data. Content analytics is also used to extract meaningful information from (big) datasets and can be used to normalise data across both databases and collections of content in order to find relationships that cross the boundaries of the source collections. This technology categorises and tags documents to emphasize what they are about. It extracts names of people, places, products, and things as well as time, opinions, sentiment, and geographic location, and adds this additional information as metadata to the search index. (Feldmann et al., 2012)

The overall content analytics market is expected to grow significantly until 2018, with a CAGR of 17%. The demand for natural language understanding and processing is growing in applications ranging from enterprise search and social media monitoring to cognitive applications and virtual assistant solutions. (IDC, 2014)

Cognitively enabled applications are being developed today by organisations, enterprises, and technology vendors in a wide variety of industries. These applications learn about us, our likes, dislikes, what we do, and then use that learning to answer questions, predict actions, and make recommendations. The technology components use tagging, natural language processing, content aggregation, knowledge representation, reasoning, and machine learning in order to provide expert assistance in a wide range of areas and contexts. The system actually learns and adapts to new situations and challenges, as it ingests data and receives feedback about the predictions and recommendations that it makes. (Chua et al., 2016)

Organisations in Austria are interested in technologies that can add value to their data. 25% of survey respondents in Austria are testing or using machine learning, concept tagging is used by 23%, and text mining by 14%. All of which are part of the Linked Data foundational technologies.

Figure: Which technologies are you currently using?



Source: Data and Information Management Survey, 2016

3.2. Impact on Enterprises Data and Information Management

Market forces and user demands have certain impacts on enterprises, and on how data and information is or can be managed within organisations. The key impacts are outlined below.

Figure: Market forces and impacts on enterprise's data and information management

MARKET FORCES

- Economic development
- Digital transformation
- Technological innovations
- Data-driven global network
- Regulations
- Digital life of citizens
- Security and data privacy

User Requirements

- Customization
- Search
- Exchange
- Automation

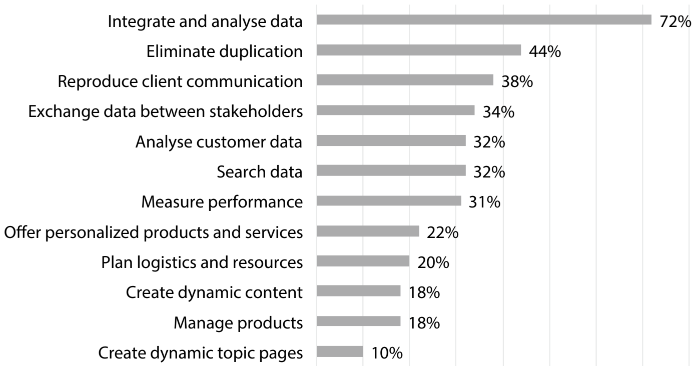
IMPACTS
on Data and Information Management

- Investments in IT and innovative data & information management solutions
- IT investments are increasingly business-driven and require collaboration
- New offerings and providers on the market, incl. providers of LD and foundation technologies
- Linked Data concept gains recognition and is taken into account as a concept that supports digital transformation and efficiency
- Reserved culture and security concerns are inhibitors of adopting new solutions for data and information management
- Open data is growing and provides potential for LD to enable more semantic-focused open data publishing
- New professional skills are needed for innovative data and information management
- Optimized business intelligence and decision-making
- General Data Protection Regulation provides an opportunity for optimizing data and information management
- Uptake of cloud solutions that can also be used for data infrastructures
- The Internet of Things brings new challenges for data management as well as an opportunity for LD
- Big Data hype and the opportunity to bring additional structure to big data with LD concept

Source: PROPEL, 2016

According to the Data and Information Management Survey (2016) organisations see the biggest potential for optimisation in the way they integrate and analyse data (72%), followed by the elimination of duplication (44%), and the ability to reproduce past communication with customers via different channels (e-mail, social media, phone, etc.) (38%).

Figure: Use cases with highest potential for optimisation

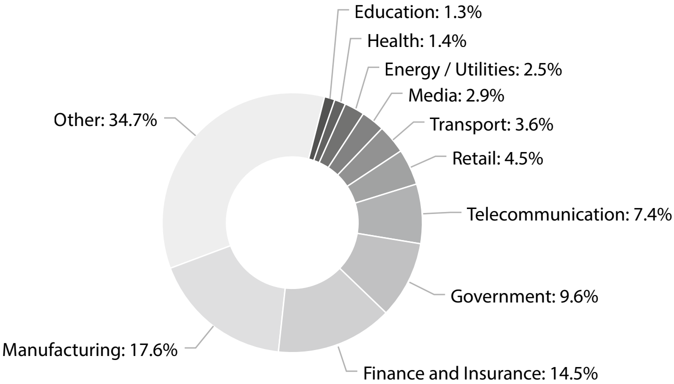


Source: Data and Information Management Survey, 2016

3.2.1. Growing IT Spending in Austria

Due to the digital transformation, the overall positive economic outlook and citizen demand we see stronger interest in business transformation to drive growth, as Austrian organisations realise that continued cost reduction and efficiency improvements are not enough to ensure future success. This recognition, combined with the realisation that digital transformation will change the world, means that growth, efficiency, and cost reduction are on the same agenda as transformation, and are offsetting the negative impact of the economic downturn on IT services spending. Geopolitical changes or significant changes to European Union policies (e.g., United Kingdom leaving the European Union and any domino effects following from this departure) could have a negative impact. On a positive note, IT spending is growing in Austria. According to an IDC end-user survey, 70% of the respondents recorded that IT spending in 2016 will be the same as, or greater than in 2015. IT spending in 2016 concentrates mainly on consulting and professional services around cloud, application development, education, and systems integration. Also the software market has positive outlook. Industries with the highest IT spending are manufacturing (16%), finance and insurance (15%), and government (10%). (IDC 2016c)

Figure: IT-Spending Austria 2015 per industry (ÖNACE classifications)



Source: IDC, 2016c

Top IT projects in 2016 (IDC, 2015) are the introduction or updating of enterprise resource planning, and the examination and optimisation of IT and data security in connection with training and consulting activities. IT security is always an issue. Current forecasts estimate that by 2019 25% of spending on IT security will be driven by the European Union and other jurisdictional data regulations, e.g. the GDPR (Christiansen, 2015). Cloud is on the agenda of many organisations, starting with the evaluation of cloud services and ending with their adoption. Organisations also want to invest in their web business. This includes updating websites/platforms, designing and implementing new web services, and creating new and multichannel customer experiences. Furthermore, organisations want to start making better use of mobile applications, social media, and data analytics. Projects to optimize communication and collaboration internally and with external stakeholders are high priority. The business outlook for the next years is neutral to positive, which restores organisations' business confidence. This creates a good environment for the uptake of optimised data and information management solutions and provides an opportunity for Linked Data.

Market forces:

- Economic growth
- Cost reduction and process efficiency
- Digital transformation
- Data driven networked global economy
- Citizen demand for online services is growing
- Data security and privacy

- General Data Protection Regulation (GDPR)
- Cloud technology stimulates growths
- The Internet of Things as an innovation enabler
- Rising data volume, variety, and velocity, and big data
- Content Analytics and Cognitive Systems

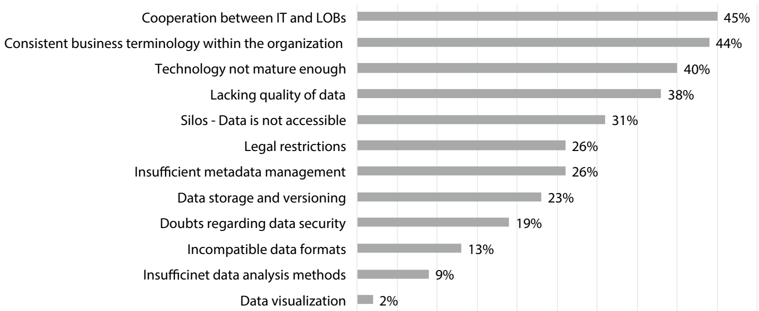
3.2.2. IT investments are increasingly business-driven

Considering the digital transformation, the need to break down data silos and the aim to develop an organisation wide information space, IT departments team up with Lines of Businesses (LOBs) and function as key enablers of the digital transformation of the organisation. The IT budget is often spread across different departments, and vendors more and more try to address their clients in a holistic way (i.e., addressing both IT and LOBs) in order to create clear value propositions. Although IT and LOBs are starting to work together, even greater collaboration is essential for success. The cooperation between IT and LOBs is seen as the biggest challenge in data and information management in almost every second organisation (45%); especially in sectors such as utilities, media and tourism. Related challenges for survey respondents are a consistent business terminology (44%), data and information silos (31%), and insufficient metadata management (26%). Anyway, traditional infrastructure projects are changing, and IT is no longer in the back office, but is becoming a key business component.

Market forces:

- Digital transformation
- Data driven networked global economy
- Citizen demand for online services is growing

Figure: Enterprise's' data and information management challenges



Source: Data and Information Management Survey, 2016

3.2.3. New portfolios and players on the Austrian market

Low demand for innovation from organisations led to high competition between IT providers, and makes them rethinking and updating their portfolios (e.g., providing cloud solutions and new consulting services, digital transformation, and improved customer experience). More and more suppliers are converting or extending their technology stacks to address data management and analysis needs. At the same time leading solution providers seek to maintain their position by developing comprehensive solutions and new go-to-market paths.

The ability to leverage modern data management and analysis solutions to develop an integrated view of customer activities and business operations can provide competitive differentiation to companies across industries. Therefore, there is a predicted and increasing demand of enterprises to embrace emerging solutions related to data management and analysis while keeping the implementation costs low. This demand leads to the creation of a rich ecosystem of new and incumbent suppliers in fields such as business analytics, big data and analytics, semantic technologies, cognitive, cloud, etc.

Austria has a disproportionately large and growing community of pioneering Linked Data developers and solution providers. Despite considerable success in the early phases, however, mainstream adoption among businesses has been slow, despite the fact that many enterprises today have realised that they have a "data problem". This shows that the market is still immature, but rapid development can be expected. Many of them have collected vast amounts of heterogeneous data of various kinds and store them in disparate systems. Enterprises invest large amounts of resources in order to unlock the considerable value that is hidden in these "data silos". In this context, LD technologies have strong potential as enablers that facilitate integration of heterogeneous data sources, provide the means for efficient retrieval, and support advanced data-driven analytics.

"Linking data has great value, especially from a business perspective. It is an essential task that is absolutely necessary in all domains." (Founder of Austrian Start-up on Semantic Technology, PROPEL Interview 2016)

Market forces:

- Economic growth
- Cost reduction and process efficiency
- Digital transformation
- Data driven networked global economy
- Citizen demand for online services is growing

- Data security and privacy
- General Data Protection Regulation (GDPR)
- Cloud technology stimulates growths
- The Internet of Things as an innovation enabler
- Rising data volume, variety, and velocity, and big data
- Content Analytics and Cognitive Systems

3.2.4. Linked Data concept gains recognition

The need to digitally transform, the need for efficiency and the data-driven economy lead to increased adoption of efficient approaches for data management such as the concept of Linked Data. While the Linked Data concept has been around for a while, it is increasingly being used in real world applications. Linked Data has attracted strong interest from the scientific community and is increasingly being adopted by governmental bodies, national and international organisations, and a wide range of online communities.

For example, the BBC is using Linked Data for the automatic tagging of sport stories, storing the semantic annotations to improve news accessibility and findability. Facebook is now using Wikipedia and DBPedia to expand its understanding of the world around us. Graph search queries like "find pictures of my friends at national parks" require that Facebook understands what national parks are, what their names are, and where they are located throughout the world before one can even start sorting through the pictures in the albums. Google's knowledge graph uses the Semantic Web and Linked Data as a starting point and then builds on top of it with more than 500 million objects and 3,5 billion facts about and relationships between these objects. (Schubmehl, 2013)

"Linked Data principles are pushed a lot by the European Commission, and some national governments strongly support Linked Data approaches. For example in the frame of the Czech e-Government initiative land use data is published as Linked Data." (Researcher at the University of West Bohemia, Czech Republic, PROPEL Interview 2016)

Austria in particular has been at the forefront of the movement towards open data in the public sector. Increasingly, such data is being published as Linked Data, which makes it much more useful and creates the potential for a range of innovative applications.

When it comes to the adoption of new technologies Austrian organisations are generally more reserved than other western European countries. This "wait and see attitude" of most organisations is an inhibitor for the adoption of LD in Austria. This attitude is also driven by security and privacy concerns when it comes to new data management solutions.

On the other hand new opportunities arise with new data management regulations. The European General Data Protection Regulation (GDPR) for example, puts enormous pressure on organisations to look deeper into their data management in the upcoming year in order to be compliant with the new regulation before May 2018. Essential questions are: What data is available within my organisation? Where is it stored? Who is responsible? Are there any copies? This intensive exploration of the data ecosystem of an organisation can also be taken as an opportunity for data and information management optimisation and adoption of the LD concept.

Also new technologies around cloud, big data, the internet of things and cognitive systems can provide an opportunity for LD. Big vendors and established technology providers mainly focus on solutions around big data and machine learning that are becoming more and more serious. This provides a chance for the Linked Data concept as it can enrich big data solutions with more structure and semantics. Big data and the concept of Linked Data can be used together under the umbrella of “structured big data”. Also the Internet of Things (smart meter/grid; sensors in production) is a big chance for Linked Data technology as a means to support the management of new and complex data. So we clearly see that concepts around linking and integrating data are of high value and certainly promising.

Market forces:

- Economic growth
- Cost reduction and process efficiency
- Digital transformation
- Technological advancement outpaces technological adoption
- Data driven networked global economy
- Citizen demand for online services is growing
- Data security and privacy
- General Data Protection Regulation (GDPR)
- Cloud technology stimulates growths
- The Internet of Things as an innovation enabler
- Rising data volume, variety, and velocity, and big data
- Content Analytics and Cognitive Systems

3.2.5. Open data initiatives

Open source software ranging from data integration and management software to advanced analytics and reporting tools will continue to be broadly adopted. We expect open source components to represent the core of many major data solutions. (Nadkarni and Vesset, 2015, 7; Nadkarni, 2015)

Open data technologies in general, and the Open Government Data (OGD) movement in particular, are novel web-enabled socio, political and technical movements that bring together important groups of stakeholders: companies/governments that provide their datasets as open data, developers who create applications based on O(G)D datasets and users from the public at large who make use of these applications. The City of Vienna, for example, has championed an active OGD policy from 2011 onwards, publishing a wide range of datasets in domains as varied as population statistics, health care, or sports and culture (data.wien.gv.at). As a result, it enabled businesses and social entrepreneurs alike to create over 200 software applications (i.e., apps) making use of these datasets. A recent study of the apps based on Viennese OGD (Koczanski and Sabou, 2015), (Koczanski and Sabou, 2015a) found that these apps were built as a result of both business and social innovation. Additionally, they allow citizens (and tourists) to be more involved in the city's life and therefore implicitly leading to economic benefits. Besides these economic benefits, there is evidence that open government data applications help foster sustainability efforts in cities as well.

Although the benefits drawn from Open (Government) Data are many, using these datasets to create apps is often hampered by the fact that these datasets are published in non-semantic formats. Linked Data technologies could enable a more semantics-focused publishing of O(G)D sources contributing to their further easy reuse.

Market forces:

- Data driven networked global economy
- Citizen demand for online services is growing

3.2.6. Professional skills

Digital transformation and the adoption of modern technologies lead to high demand and fierce competition for skilled staff among Austrian enterprises and IT providers associated with data and information management solutions and business analytics services. The lack of professional skills and IT experts are key challenges for the Austrian market. The skill gap is across the board, from data scientists to data and information architects to data security experts. Further IT education and training is required, with the support of education and training organisations as well as the IT industry itself. Vendors that have created talent management strategies to successfully handle the skill shortage issues will be in the driving seat to handle client demands in the future.

Market forces:

- Digital transformation
- Data driven networked global economy
- Citizen demand for online services is growing
- Data security and privacy
- General Data Protection Regulation (GDPR)
- Cloud technology stimulates growths
- The Internet of Things as an innovation enabler
- Rising data volume, variety, and velocity, and big data
- Content Analytics and Cognitive Systems

3.2.7. Data-driven decision making and innovation

The market growth of data and information management solutions is strongly driven by digital transformation initiatives, the data-driven economy, and strong citizen demand for new services and products, in which data with integrity, data that is trusted, available, secure, and compliant, is a critical success factor. In many cases, datasets impose limitations towards answering complex and interdisciplinary questions and require a combination of data from multiple and in many cases distributed sources. Organisation's ability to build robust "data pipelines" in and out of their organisations influences how they can compete in the market. In 2016, and over the next several years, enterprises will focus on two types of "data pipelines": Identifying, syndicating, and linking internal data (some of that might also be delivered to the marketplace); linking and enriching internal data with valuable external data sources that can enhance their understanding of, and drive good decisions in the marketplaces (Gens, 2015a).

By adopting modern data management principles, a set of advantages are provided in terms of both day-to-day operations and advanced business analytics:

- Interlinked versions of datasets may be maintained or regularly updated facilitating the provision of access to the latest version of the available datasets;
- With linked data analytics, the produced data are made discoverable for further private or public use in the future;
- Linked datasets eliminate redundancies and reduce complexity in data collection and management;

- Cost and time savings;
- Optimisation of products and services;
- Easy replication of an analysis with up to date datasets;
- Decision making based on sound analytics;
- Competitive advantage.

"There is definitely a need to integrate distributed databases, and LD seems to be among the most suited technologies to do so - the question is how fast we get there." (Linked Data Expert and Pioneer, UK, PROPEL Interview 2016)

With innovative data management approaches and solutions new use cases and applications for advances analysis and decision-making become real, such as consumer behaviour tracking, where existing or new data is used to understand customer behaviour and then target clients with more specific offers; more effective analysis of business operations and communication, e.g., transparency in financial reporting and CSR; and the interaction with customers, where the focus is on the transformation of customer interactions, multichannel development, and the creation of excellent customer experiences.

"Linked Data is more and more used by organisations who want to improve efficiency, reduce costs, and be faster. Drivers are decision-makers and policy makers." (Project Manager at a leading technical solution provider in the field of environment, Italy, PROPEL Interview 2016)

4. SECTORAL ANALYSIS OF LINKED DATA POTENTIAL

The goal of this part of the roadmap is to *systematically assess the potential of Linked Data technologies* on a *per-industry basis*. In assessing the potential in various application domains, we do not ground our analysis in specific technological characteristics of Linked Data, but instead take a high-level perspective focusing on industry characteristics. Such an approach allows us to develop structural and technological profiles that in turn enable us to asses the respective potential for Linked Data adoption in various sectors. By developing these industry profiles and relating them to key principles and characteristics of the Linked Data paradigm, we provide a comprehensive analysis across a wide spectrum of potential enterprise application domains. The methodology of this analysis is outlined in Section 2.4.1. Overall, we develop fifteen industry profiles, each one characterised along a set of nine dimensions that together serve as an indication of the industries' susceptibility towards Linked Data adoption.

4.1. Working Hypotheses

We propose nine key dimensions that impact the potential for LD adoption in industry and explicate our expectations by formulating a set of working hypotheses based on structural industry characteristics and the role of information technology and information in the industry and its environment. In formulating these hypotheses, we relied on theoretical models (e.g., technology adoption and diffusion models), empirical studies on the adoption of comparable technologies (e.g., big data, cloud computing), and arguments based on general reasoning and LD-specific considerations.

The developed working hypotheses are based on five structural attributes used to characterize each industry:

1. Internationalisation
2. Knowledge intensity
3. Operational complexity
4. Networking
5. Openness

Furthermore, we use four attributes to characterize each industry from a more technical perspective:

6. Data drivenness
7. ICT intensity
8. ICT innovation

9. Web technology adoption

In order to develop our working hypotheses, we relate these industry characteristics to principles and characteristics of the Linked Data paradigm[8], such as:

- *Decentralisation*: Linked Data creates a globally distributed federated dataspace.
- *Linking, sharing and reuse*: Globally unique identification of entities and the ability to use different schemata in parallel facilitates global data sharing.
- *Self-descriptiveness*: Linked Data relies on semantically explicit shared vocabularies that ease the integration of data from different sources.
- *Flexibility and extensibility*: Terms from different vocabularies can be mixed, multiple schemas can be used in parallel, and there is no technical distinction between data and metadata. Linked Data modelling therefore requires little a-priori structural commitment and can be performed in a bottom-up manner, growing graph-based data models dynamically along the way.
- *Openness*: A wealth of Linked Data is publicly available and can be linked to and used to enrich internal Linked Data.

Against the backdrop of these high-level technological characteristics of Linked Data, we derive nine working hypotheses that link structural and technological industry characteristics to the propensity to adopt Linked Data.

4.1.1. Internationalisation

[H1] "Internationalisation is positively associated with a propensity to adopt LD technologies."

4.1.1.1. Rationale:

Internationalisation involves the geographical dispersion of people, goods, processes, infrastructure etc. The requirements of organisations in more internationalised industries are therefore more likely to align with the characteristics of the Linked Data paradigm because:

- LD enables globally distributed data infrastructures and processing facilities.
- LD creates globally distributed data spaces.
- LD facilitates international intra- and inter-organisational cooperation.

8 Derived from Bizer et al. 2009, Heath et al. 2011.

- LD supports organisations' ability to cope with heterogeneity and may support them in adapting flexibly to local environments (e.g., languages, cultures, regulations, consumer needs etc.).

4.1.2. Knowledge intensity

[H2] "Knowledge intensity is positively associated with a propensity to adopt LD technologies."

4.1.2.1. Rationale:

Knowledge-intensive industries primarily rely on intellectual capabilities rather than on physical inputs or natural resources (cf. Powell and Snellman 2004). In these industries, managing, developing, and supporting these knowledge-based capabilities is vital to the business. Linked Data can be a key enabler in this context. Specifically, organisations in knowledge-intensive industries:

- are more likely to have a need for sophisticated knowledge management tools.
- are more likely to draw benefits from knowledge-based systems that involve knowledge representation and reasoning.
- often have to deal with complex, heterogeneous information.
- have a higher need for flexible data models.
- are more likely to benefit from distributed knowledge bases[9] on the web scale.
- are more likely to draw on data outside their organisation.
- are more likely to benefit from making their own data inter-operable by adopting established vocabularies and thesauri.

4.1.3. Operational complexity

[H3] "Operational complexity is positively associated with a propensity to adopt LD technologies."

4.1.3.1. Rationale:

Complex operations are associated with notions such as dispersion, heterogeneity, fragmentation, concurrency, and dynamic interaction. In environments that exhibit these characteristics, Linked Data can excel as a tool for managing the inherent complexity whereas in more simple and well-structured environments, LD technologies

9 "A distributed knowledge base is a systemically coherent set of knowledges, maintained across an economically and /or socially integrated set of agents and institutions." (Smith, 2000)

provide more limited immediate benefits. Hence, we expect that adoption is more likely in operationally complex environments.

4.1.4. Networking

[H4] "Networking within industries is positively associated with a propensity to adopt LD technologies."

4.1.4.1. Rationale:

More networked industries that require strong coordination of activities between independent organisations (e.g., joint research and development, cooperative knowledge sharing in collaborative and coopetitive settings, vertical integration of tightly interlinked supply chains etc.) are characterised by intense intra- and inter-organisational coordination and information exchange. Therefore, networked industries will be more likely to adopt LD because they:

- require data management approaches for non-hierarchical networks of independent institutions in which no single entity prescribes a uniform data model. (Linked Data does not necessitate the use a single shared infrastructure or commitment to a single standardized data model.)
- can benefit from a shared information space.
- require cost-efficient means for the integration of internal and external data.
- have more pressing needs for a distributed data sharing infrastructure.

4.1.5. Openness

[H5] "Openness is positively associated with a propensity to adopt LD technologies."

4.1.5.1. Rationale:

Linked Data is closely associated with the concept of openness. In fact, the term Linked Open Data (LOD) is frequently used for Linked Data which is released under an open license, which does not impede its reuse for free (Berners-Lee 2010). Although Linked Data does not necessarily have to be open and LD technologies are increasingly used to process closed data internally (Cobden et al. 2011), part of its value stems from the positive economic network effects that openness provides. We therefore expect that industries that already are more open in terms of making use of and publishing open data, engaging in open innovation etc., have a higher propensity to adopt Linked Data.

4.1.6. Data-drivenness

[H6] "Data-intensiveness is positively associated with a propensity to adopt LD technologies."

4.1.6.1. Rationale:

Organisations in "data-intensive" industries produce, transform, and add value to data and/or make heavy use of it in the execution of their core business and support processes. In many such industries, data itself is the core product or the core product is heavily based on data and analytics. These "data-driven" industries are more likely to adopt LD because:

- they have a need for sophisticated data management, integration, and analysis
- can benefit from extracting insights from combined internal and external data.
- they need the flexibility to model and make use of all kinds of data.
- the operations and analyses performed on the data are highly dynamic.

4.1.7. ICT-intensity

[H7] "ICT-intensity is positively associated with a propensity to adopt LD technologies."

4.1.7.1. Rationale:

Highly ICT-intensive industries will heavily rely on information and communication technology systems to add value to their products, support their operations, and shape consumer experiences. Organisations in ICT-intensive industries that already rely heavily on established systems therefore:

- are more likely to benefit from "pay as you go" technologies that integrate and leverage their existing infrastructure rather than migrating from it.
- are more likely to have or acquire the competences required for embarking on LD projects.
- are more likely to benefit from efficiency gains in their data infrastructure.

4.1.7.2. Caveat:

ICT-intensive industries that have adopted ICT technologies a long time ago rely heavily on their complex "legacy" system infrastructure in the execution of their business processes. This introduces strong inertia and hampers migration to new technologies. However, we expect the net effect of ICT intensity on LD adoption to be positive, i.e., that ICT-intensive industries are more likely to adopt Linked Data than those in which ICT plays a very limited role.

4.1.8. ICT innovation

[H8] "ICT innovation is positively related with a propensity to adopt LD technologies."

4.1.8.1. Rationale:

Organisations operating in dynamic environments that require innovative ICT solutions are both more likely to have the capabilities and resources to adopt LD technologies and to benefit from them. In particular, highly innovative industries with regards to ICT tend to:

- be culturally more open to change.
- have a higher need for creative change, which requires flexibility and modifiability of IT systems at low cost.
- be more likely to have the technological maturity required for LD adoption.
- be more likely to accept the risks of LD adoption.
- be more likely to benefit from the flexibility provided by LD (i.e., from avoiding relational databases, which are more resistant to change and less responsive to heterogeneous needs).
- be more likely to make use not just of predictably structured data, but to leverage and integrate heterogeneous, poly-structured data (a job that LD technologies are designed for).
- be more likely to have a need to experiment and discover new ways of creating value.
- operate in dynamic environments (market needs, competitive pressures, regulatory requirements etc.) that necessitate the agility and flexibility that Linked Data provides.

4.1.9. Web technology adoption

[H9] "Prior adoption of web technologies is positively associated with a propensity to adopt LD technologies."

4.1.9.1. Rationale:

The Linked Data vision evolved from the traditionally document-centric web context. Based on web technologies and principles, it aims to create a web of data. We expect industries that already are "web-centric" and make intense use of the web and associated technologies (Semantic Web technologies in particular) to be more likely to adopt Linked Data technologies because they

- are more likely to be able to leverage existing LD initiatives
- can more easily adapt their existing technical infrastructure
- are more likely to obtain immediate value in existing application contexts
- are more likely to benefit from publishing their data on the web

This hypothesis extends H3 in that web technologies are a particular subset of ICT. Due to their significance in the context of Linked Data, we consider them as a separate dimension.

4.2. Analysis

This section presents the results of our sectoral analysis. Based on the working hypotheses developed in the previous section, we evaluate and discuss each industry along the identified dimensions and conclude each analysis with an overall assessment. At the end of the section, we synthesise and summarise our findings across industries.

4.2.1. Resource industries

4.2.1.1. Structure

The resource industries that are a key part of the primary economic sector, include agriculture, forestry, fishing, mining, and quarrying. Internationalisation in this economic sector varies. On the one hand, there is a large number of locally operating entities whereas large global players such as Monsanto or Syngenta are on the other side of the spectrum. Internationalisation mainly manifests itself through the trade of resources.

Economic development is often portrayed as a stepwise progression, first from an agriculturally-based to an industrial economy and then finally to a knowledge economy. This model neglects that resource industries such as agriculture, forestry, and mining are also themselves -- to varying extents -- knowledge-intensive. Operations in the resource industries rely heavily on chemical, biological, geological, meteorological knowledge etc. and developments. Therefore, although resource industries are not quite on par with the most knowledge-driven industries, we consider the overall level of knowledge-intensity at least medium.

Operational complexity varies in this sector, and is quite high in some instances. In the agricultural sector, for instance, developing an understanding of biochemical processes that are relevant for high yields requires intense R&D. On the other hand, the foundational operations of this industry (planting seeds, catching fish, cutting trees) are not particularly complex.

Interconnectedness and inter-organisational information flows are high, which can partly be attributed to the intense regulation characterising this industry. Producers are obliged to follow strict reporting guidelines in order to ensure, for instance, food safety. Typical stakeholders of the involved processes are farmers/fishers (production, harvesting), traders (collection, storage), food companies (processing, manufacturing), retailers, consumers/customers, associations, and public authorities. Notably, there has been work on standardising data exchange formats in order to ease structured information flows in the agricultural domain, such as AgroXML (Kunisch et al. 2007). Conversely, this shows that there is a need to get a common sense of how to transmit data between involved stakeholders along production and distribution channels while reducing redundancies (Schmid et al. 2013, Holster et al. 2012).

Currently, willingness to publish open data is low in the resource industry. Much of the data generated by companies in this sector would be useful for others, such as sensor measurements, are rarely published. However, many datasets have been published under terms such as "environmental open data" or "agricultural open data", but these generally only include conceptual meta-information which is difficult to utilize in core business processes. A potential motivator would be to provide transparency by opening up data which needs to be provided to public authorities (for obligatory reporting) anyway. We suspect that this sort of reporting is still conducted based on outdated technology (heterogeneous data formats, large overhead in the process, etc.). Hence, there are potentially large opportunities to improve the current state via standardisation and "open data best practice" efforts. A study conducted by the Open Data Institute (ODI) and Global Open Data for Agriculture and Nutrition (GODAN) also concludes that *"there are still obstacles to realising open data's full potential in agriculture and nutrition"*, but also shows uses cases where open data can help (Carolan et al. 2015).

4.2.1.2. Technology

Traditionally, the resource industry has not made intense use of data to support decisions and improve its products. Recently, however, there has been an increasing uptake of concepts that involve data processing to improve operations in the resource industry (e.g., precision agriculture, big data in agriculture[10]). Data analytics represent an opportunity to optimize operations and improve decision making. There is also a large number of startups that target this field, mostly focusing on the agricultural domain

10 Note that currently many efforts focus on agriculture, but they can similarly be applied to other resource industries, such as forestry or fishery.

(e.g., Agsquared, Fieldview, Vital Fields, Gamaysa, and Oada). As of now, the average player in this industry does not have particularly data-driven operations, while large global players, on the other hand, appear to increasingly make use of data to improve operations.

The resource sector is presently also not particularly ICT-intensive and the share of value added by ICT it is generally relatively low. Schmid et al. (2013) provide an overview of ICT and technology adaptation in agriculture in most EU countries, which confirms this result.

Traditionally, the resource industry has not been particularly active in terms of using ICT to drive innovation. Recently, however, we have seen a number of initiatives that aimed at improving operations in this sector (Sonka et al. 2015, Jespersen et al. 2014). An example for innovation driven by ICT in this sector is precision farming, which aims to increase yields by means of data analysis. Mobile and wireless technologies are key enablers in this context, because typically remote areas (e.g. acres, sea, forests) are used for production. A study commissioned by the world bank highlights many similar examples (George et al. 2011).

The resource industry's focus on web technologies is also low. Most processes typically do not involve web technologies, but there are some exceptions. An example for the use of web technologies is AGROVOC, which is a controlled vocabulary including concepts - among others - of agriculture, forestry, and fishery (Caracciolo et al. 2013). However, the vocabulary is mostly used by researchers and librarians and we couldn't find indicators of adoption by industry.

4.2.1.3. Assessment

Overall, the resource industries have historically not been particularly ICT-intensive, innovative, data-driven and open. However, the prevalence of cross-organisational information flows and the high potential for data-driven and open innovation, which are gaining traction within this industry lately, suggest that LD technologies could make important contributions in this domain.

4.2.2. Manufacturing

4.2.2.1. Structure

The manufacturing industry, which forms large parts of the secondary sector of the economy, has become highly globalized. This development was driven by growth in foreign competition, rapid increase in foreign demand, attractive markets abroad, and changes in global political and economic conditions which have created new opportunities (NRC 1990).

Manufacturing is also *highly knowledge-intensive* and continues to become even more so as production becomes more and more automatized in developed and developing economies. Further, R&D expenditure, which is frequently used as an indicator of knowledge-intensity (Willoughby & Galvin 2005, Coviello 1994), is the highest among all industry sectors in Austria (Statistik Austria – Bundesanstalt Statistik Österreich 2016, p. 13)[11].

Operational complexity of manufacturing processes is typically high (ElMaraghy 2011). Manufacturing complexity includes several dimensions such as size, variety, information, uncertainty, control, cost and value of the environment. Finding an appropriate trade-off between the increasing importance of mass customisation and mass production (Kamrani et al. 2010) is an important goal for manufacturing enterprises. The demand for shorter production life cycles and frequent changes in products, processes, and volumes further increases production complexity (Mattsson 2011).

Manufacturing is highly process-oriented and integrating supply chains with various stakeholders requires strong coordination and intense information flows. Even though such information flows tend to be somewhat more streamlined than in other industries, studies suggest that the development of stable external communication channels with suppliers is increasingly becoming crucial for manufacturing enterprises (Forza & Salvador 2001).

Openness, e.g., in terms of innovation processes and particularly in terms of the willingness to publish open data is presently very limited in manufacturing. Primarily, this lack of openness can be attributed to concerns over the protection of trade secrets, thereby losing competitive advantages. However, there is some "Open Manufacturing Data" available, for instance, provided by data.gov, the Open Data portal launched by the U.S. government[12]. These data sets are, however, limited to aggregated statistical data and are therefore not likely to be very valuable for individual manufacturing enterprises. .

4.2.2.2. Technology

The manufacturing industry operates with large amounts of data, but it is not yet fully "*data-driven*". Data as a source for improving operations (optimizing workflows, pro-

11 It was not possible to obtain R&D expenditure data by industry sector on the European level due to the fact that some countries do not publish it and hence it is not possible to compute aggregates.

12 https://www.data.gov/manufacturing/

duction, etc.), for instance, is theoretically acknowledged, but not been extensively used (Shao 2014). However, developments such as Industry 4.0 strongly aim at utilising data to create intelligent solutions throughout the production lifecycle and to optimize production plants and supply. Manufacturing is therefore likely to become significantly more "data-driven" in the future.

Intensity of ICT use within the manufacturing industry varies greatly. In Austria, for instance, 31.4% of manufacturing businesses have dedicated ICT specialists on staff. The increased adoption of ICT in manufacturing will be influenced by the current worldwide move towards flexible industrial production, as reflected by relevant initiatives around the globe. Introduced in Germany, *Industry 4.0*[13] is a vision for a more advanced production system control architecture and engineering methodology (Bauernhansl et al, 2014). Similar initiatives for modernising industrial production have been set up such as the "*Industrial Internet Consortium*" in the USA or the "*Factory of the Future*" initiative in France and the UK (Ridgway et al, 2013). Due to these developments, ICT usage can be expected to continue to grow on a massive scale and improve manufacturing and production systems.

Manufacturing is widely perceived as a conservative industry with relatively *moderate IT-based innovation*. Due to intense global competition, however, manufacturers are increasingly under pressure to restructure and innovate. Studies have found evidence that an increase in a company's IT input is associated with an increase in innovation output, however, suggest that IT-based innovation facilitates value creation and establishing a firm's long-term success (Kleis et al. 2012).

Manufacturing today largely relies on specialized software deployed in isolated production environments. The *role of web technologies* within the core processes of manufacturing is therefore presently fairly limited. Still, there is a growing tendency towards integrating web technologies into manufacturing systems, e.g., for mass customisation or to support collaborative activities in various stages of product development -- e.g., design, process planning, production, or distribution (Shaw 2001, Tian et al. 2002). Opportunities to facilitate new business models via web- and information-based systems are not exploited to their full extent in manufacturing yet.

4.2.2.3. Assessment

Overall, Linked Data could solve a range of important problems in the manufacturing industry and the structural characteristics of this domain suggest that it has a high potential for Linked Data adoption. As manufacturing becomes not only increasingly

13 http://plattformindustrie40.at/

automated, but also more and more integrated into cyber physical systems, Linked Data technologies could create a uniform data platform and facilitate interoperability. In such Industry 4.0 scenarios, LD could enable cyber-physical systems to cooperate flexibly, both internally and across organisations. Indeed, Linked Data concepts are at the basis of ensuring the integration of data between participants to the Germany wide Industrial Data Space initiative[14].

4.2.3. Energy

4.2.3.1. Structure

Energy, a primary sector industry that has traditionally been predominantly nationalised, has become more and more internationalised in most parts of the world. The exchange of electric and fossil energy is increasingly brokered transnationally and in the European Union in particular, the energy market was subject to a strong drive towards liberalisation which in turn increased internationalisation. Most recently, the "Third Energy Package" (European Commission 2014) liberalised energy markets further in order to foster competition.

The energy industry is knowledge-intense, but less so than many other industries because conventional energy production is based on stable, mature, and long-running technologies. On the other hand, energy efficiency and renewable energies have become increasingly relevant as the general public and policy makers advocate a more sustainable use of environmental resources. These emerging areas within the energy industry are based on new and innovative technologies and involve intense R&D activities. They are hence considerably more knowledge-intense than the traditional energy sector.

The complexity of operations in the energy sector is moderate. On the one hand, the delivered product is clearly defined and unique. The sales processes are well-established and involve limited innovation. There is also limited cross-sectoral activity generating operational complexity. On the other hand, the complexity within interconnected power systems has dramatically increased due to deregulation of the markets (Willis & Philipson 2005).

Networking within the energy domain is high, both in terms of energy networks (grids, pipelines etc.) and on the inter-organisational level. Internationalisation and complex value chains, as well as the involvement of a multitude of stakeholders (e.g., suppliers,

14 Industrial Data Space Initiative: https://www.fraunhofer.de/en/research/lighthouse-projects-fraunhofer-initiatives/industrial-data-space.html

customers, grid operators, NGOs, regulators, R&D institutions etc.) also increase the complexity of information flows that require adequate technical solutions.

The level of openness in the energy sector is typically is low. Most data is stored internally and not made publicly available if not explicitly required by regulators. However, there are a number of initiatives, such as the Open Energy Information portal[15] or Reegle (Bauer et al. 2011) that provide open energy data on an aggregated level.

4.2.3.2. Technology

The energy sector already and is becoming increasingly data-driven. This applies to all stages of the energy value chain, from, e.g., hydrocarbon exploration, which generates large amounts of data and the industry needs new technologies and approaches to integrate and interpret this data (Farris, 2012), to energy trading. The data-intensity of the latter can be attributed to increased globalisation and competition within the market, which requires effective means for price forecasting in order to improve decision-making at the corporate level (Weron 2014). Efficient methods for the analysis of big data are essential to continuously improve energy price prediction models or other energy-related analytics[16]. Similarly, cloud-based techniques, such as the "Dynamic Demand Response" platform, enable demand forecasting and make use of (big) data analytics in the energy market (Simmhan et al. 2013).

ICT use in the energy sector is generally high. Recent developments to improve energy efficiency are strongly dependent on ICT, which facilitates the transition to an energy-efficient, low carbon economy. ICT-based solutions already contribute towards achieving the energy and climate objectives involved in the transformation to a low-carbon society. Examples include technologies such as smart grids, home automation, or smart metering, each of which are primarily based on ICT (ICLEI 2011). In line with this development, IT-based innovation in the energy sector is increasing. Such innovations are also strongly supported by the European Union in the course of the work programme for "Secure, Clean and Efficient Energy" which provides a budget of €194 million for 2016 and 2017[17].

In terms of web-focus, we need to differentiate between two areas: traditional energy production and energy efficiency and renewable energy. The web focus in the former is typically relatively low, as energy producers do not primarily rely on any web-based technologies to run their business. Increasingly, however, the importance of web tech-

15 http://en.openei.org/wiki/Main_Page
16 http://solarindustrymag.com/online/issues/SI1501/FEAT_03_Analytics-And-Big-Data-Are-Changing-The-Energy-Market-Map.html
17 http://ec.europa.eu/programmes/horizon2020/en/h2020-section/secure-clean-and-efficient-energy

nologies is growing in this traditional area, e.g., in the context of smart meter introduction, which allow energy providers to make data available to the consumer in near real-time. In the energy efficiency and renewable energy domain, the web already plays a more prominent role, as exemplified by data exchange portals such as Reegle[18].

4.2.3.3. Assessment

Overall, given the high level of internationalisation and interconnectedness, the increasing levels of knowledge-intensity and data-drivenness, the pressing need for efficient data integration and analysis, and a drive towards ICT-based innovation in many areas of the industry, we consider the Energy sector a promising LD application domain.

4.2.4. Construction

4.2.4.1. Structure

The construction industry, which is part of the secondary economic sector, is characterised by a high level of internationalisation. Many construction firms already operate in international markets to either trade their design services (Reina and Tulacz 2015a) or construction products and services (Reina and Tulacz 2015b, Gajendran et al. 2013). Typically, the main goal of internationalisation in the construction industry is to either outsource core operations, or to offer the company's products or services in the international market in order to expand beyond the limitations of the domestic markets (Jung et al. 2010).

Companies operating in the construction industry see their core business as being primarily driven by knowledge, design, assembly, or a combination of those (Gajendran et al. 2013).

Construction operational complexity is typically high due to the involvement of many different stakeholders in the process of design, planning, and actual construction. According to Wood & Gidado (2008) project complexity, which we can use as a measure of operational complexity, is also high due to several factors such as a large number of interacting parts, interaction, interdependencies, and interrelationships between parts of a construction project. The authors also confute the argument that the construction process is an ordered, linear, and predictable flow of activities and suggest that this perception may explain the poor success rates of construction projects. Instead, the construction process should be seen as a complex, dynamic phenomenon

18 http://www.reegle.info/

in a nonlinear setting. Similarly, other studies identify typical characteristics of complex systems when analysing operations in construction (Bertelsen 2003).

Construction is a highly networked industry with strong intra-organisational information flows. A multitude of stakeholders are involved in both the design and planning stage as well as in the actual building stage. Zeng et al. (2007) highlight the importance of information flows in construction industry projects. Their study also reveals that information asymmetries lead to shortcomings in quality management.

The willingness to publish open data is currently very limited in the construction industry. The private construction sector has not been very active in participating in open data initiatives due to commercial sensitivities, the need to anonymize data, and resources needed to set up licensing and agreements[19]. Usually, opening up data is seen to provide benefits to the society, however, for private construction companies providing open data the economic benefit is not immediately apparent. Discussion about how this benefit could be generated and how innovation from open data can arise in the construction industry is still in its infancy. Experts see it as free research and development or as a possibility to bring in people from different sectors to analyse the data[20].

4.2.4.2. Technology

Construction is not particularly data-driven, but has a high potential to become more so in the future due to advances in Building Information Modelling (BIM). A general issue is that a large part of the construction knowledge is tacit. Therefore, in order to make use of this knowledge, it needs to be codified. Only then, data analysis methods can be applied. In this context, BIM and the rise of big data analytics are promising developments. For instance, measuring building performance with sensors and performing analyses over the obtained data can help to improve the efficient lifecycle management of buildings (e.g., energy use, behavioural patterns). A study on big data conducted by the McKinsey Global Institute further points out that advanced simulations can reduce the number of production- and construction design changes as well as the overall cost of construction. Moreover, sensor-based real time data allows usage patterns to be derived and enables manufacturers to improve demand forecasts and product development (Manyika et al. 2011).

Construction is generally a team effort, hence, knowledge and knowledge management techniques are essential. A large part of the domain knowledge in the construc-

19 http://brebuzz.net/2015/12/09/how-can-open-data-help-to-develop-the-construction-industry/
20 http://brebuzz.net/2015/11/25/utilising-data-open-big-data-in-construction/

tion industry is well-suited for explicit expression. Examples of such knowledge include construction plans, design briefs, sketches, engineering drawings, performance specifications, conditions of contract or bills of quantities (Egbu & Robinson 2008). Whereas 80% of the knowledge used during concept design is tacit and can therefore not be readily handled by ICT, this figure is reversed during the detailed design stage where 80% of the knowledge is explicit and can be codified (Al-Ghassani 2003). Use of ICT is already prevalent in the planning and design process of construction projects. Due to the fact that many different parties are involved in these processes, ICT also facilitates communication- and knowledge-sharing processes, for instance, via document transfer. However, for the actual process of building planned facilities, the impact of ICT has been more limited. Even though companies are aware of the benefits, the adoption of ICT still faces significant barriers (Gajendran et al. 2013).

In terms of innovation, IT applications are typically used to support the execution of processes, for instance, via document management systems. However, they are rarely used to foster innovation. A concept that has attracted significant attention recently is BIM, which has significant potential for IT-based innovation in the construction sector. BIM involves innovative software that helps to create building models that help to improve the planning and construction phases by reducing uncertainty, or simulating and analysing potential impacts (Smith 2007).

The adoption of web-based technologies ("web focus") of the construction industry is generally low. Use of web technologies does not go beyond online project-management and collaboration software (Nitithamyong & Skibniewski 2004) and cloud services for efficient document transfer over the web.

4.2.4.3. Assessment

Overall, whereas the limited adoption of web technologies, moderate ICT-driven innovation and low willingness to publish open data do not inspire confidence in a rapid take-up of LD technologies in the construction industry, other factors are more favourable. In particular, the high level of internationalisation, knowledge-drivenness, intense use of ICT, interconnectedness and complexity suggest that there is a significant potential for LD adoption in the construction industry. LD could, for instance, play a key role in improving collaboration among the stakeholders involved in construction projects. It may also contribute towards making this industry more data-driven and efficient.

4.2.5. Retail

4.2.5.1. Structure

Retail, which forms part of the tertiary sector of the economy, has become highly *internationalised* since World War II. Prominent examples of globally established retailers include Wal-Mart, Metro, Inditex, H&M, or Kingfisher (Elsner 2013). According to Deloitte (2015), an indicator of this high internationalisation is that, 66% of the top 250 worldwide retailers operate in more than one foreign country. Further, 23.4% of revenues stem from foreign operations. The report also highlights that European retailers have the most international scope: In 2014, the top 250 European retailers had a presence in 16.8 countries on average.

Retail can be considered a moderately *knowledge-intense* industry. According to Uppenberg (2009), R&D intensity, which we use as a proxy, is very low in the EU compared to the US and to other commercial services, such as transport, storage, and communication or financial intermediation.

Operational complexity is primarily introduced via increased globalisation and management of multiple locations of retailers. Other than that, typical processes such as the handling of claims or collaborative product data maintenance may be seen as fairly complex, but are typically implemented efficiently.

Networking and cross-organisational information flows are prevalent in the retail sector. Retail businesses need to communicate with suppliers, vendors, distributors, producers, and the customer. These information flows are essential to the retail business and their complexity should not be neglected. IT facilitates tightly integrated, agile supply chains that allow retailers to become extremely responsive to market developments by "link[ing] customer demand to manufacturing, and linking manufacturing to distribution" (McAfee et al. 2004), a strategy famously implemented, for instance, by Zara (Ghemawat et al. 2003).

The willingness of the retail industry to publish and *open* up their data is typically low. Despite considerable academic interest and market research in the field of retail data mining (with some datasets publicly available[21]), enterprises are typically not willing to publish their data, which is often considered a major source of competitive edge. Still, there are some efforts such as the Consumer Data Research Center[22], which publishes Open Data related to the retail sector on a limited scale.

---

21 http://recsyswiki.com/wiki/Grocery_shopping_datasets
22 https://www.cdrc.ac.uk/

4.2.5.2. Technology

The retail industry is highly *data-driven*. On the one hand, data plays a crucial role for retailers because product and supply chain data is big and therefore data management and all the challenges involved are important. On the other hand, customer data and its integration with other sources bear great potential for improving the efficiency of processes. In addition, due to the recently increased interest in big data analytics, the industry identified large potential in analyzing social media and online browsing patterns or introducing new location sensing technologies to improve store layout and product mix (Ericsson 2015, Tene & Polonetsky 2013).

ICT-intensity in retail is generally high. As an OECD working paper points out, ICT is a dominant success factor in the retail industry and has been a key enabler of efficiency improvements since the 1990s (Eröcal 2005). Developments such as larger store formats and new logistics systems in retailing also necessitate increasing reliance on ICT-based systems to understand consumer behavior and track goods.

More generally, *ICT-based innovations* play a key role in transforming the retail industry. Online shopping and mobile technologies have transformed the ways retailers approach (potential) customers. A recent Ericsson report (Ericsson 2015) also highlights the opportunities created by IT and ICT innovations. The authors mention that ICT innovations already played a large role in retail by creating industrialized global supply chains. The advent of social media increased engagement of the retail industry with customers and extended interactions beyond just being "transactions".

The *web-focus* of the industry and the level of adoption of web technologies is also typically high. This is underlined by the ubiquity of electronic and mobile commerce, often with real time stock information (e.g., Amazon). Traditional brick and mortar retailers also increasingly engage with customers via online and mobile platforms and aim to migrate towards multi-channel retailing. A recent report of a European Commission expert group also highlights the increasing significance of online retailing (Sundström & Reynolds 2014). Overall, the web-focus of the retail industry is therefore strong and increasing.

4.2.5.3. Assessment

Overall, due to the high level of ICT intensity and IT-based innovation, the strong role of existing web technologies, the high level of internationalisation and the strong need for IT-based integration, the retail industry is a promising domain from a Linked Data perspective, even though knowledge intensity and openness are fairly limited.

4.2.6. Transport and Logistics

The transport and logistics sector encompasses a diverse set of activities including rail, road, pipeline, sea, water, air and space transport. These activities can be further subcategorized into specific forms of passenger and freight transport. Furthermore, the logistics sector also consists of a broad range of support activities for transportation, including warehousing and storage, cargo handling, postal and courier activities etc.

4.2.6.1. Structure

Today, transportation and logistics is comprised of activities on all geographic scales, ranging from taxi and public urban passenger transport to large-scale trans-border freight systems. The sector has become increasingly "internationalised" since at least classical antiquity. The geographic expansion of the Phoenician, Minoan, and Greek thalassocracies, for instance, were driven by naval power and trade routes (Abulafia, 2014). The Roman empire developed an extensive road infrastructure that was vital to its expansion to the interiors (Forbes, 1993). Albeit strongly motivated by the supply and distribution of military resources, the construction of this extensive transportation infrastructure also established long-distance trade relations and was central to cultural interactions that shape European history to this day. On an even larger scale, the silk road opened long-distance political and economic relations between the civilisations of China, the Indian subcontinent, Persia, Europe, the Horn of Africa and Arabia (Bentley, 1993). This historic perspective highlights that the internationalisation of transportation and logistics is not a recent phenomenon. From a more contemporary perspective, innovations such as the intermodal transport container supported the post-war boom in international trade and were a major element in globalisation (Levinson, 2008). Today, worldwide logistics and supply chain management are instrumental in economic globalisation, coordinating activities and synchronizing supply with demand. As far as passenger transport is concerned, global travel demand has also increased dramatically and contributed significantly to increasing internationalisation of the sector in recent years (cf. Section 4.2.7). Overall, there has been a continuous drive towards increasing globalisation of the transportation and logistics sector in recent decades.

Transportation and logistics covers a wide range of activities with varying degrees of *knowledge-intensity*. In complex supply chains, members' combined knowledge may well be the most significant source of value creation (Hult et al. 2003), although transportation and logistics itself does not necessarily become highly knowledge intensive in this setting. Management of complex transportation networks can involve highly knowledge intensive activities, but in terms of provision of basic freight and passenger transportation services, the sector is not particularly knowledge intensive. Overall, knowledge-intensity of the *transportation and logistics* sector falls in an intermediate range.

Transportation and logistics scores high in the *operational complexity* dimension, which is evident in the Oxford Dictionary on-line definition of Logistics as "*the detailed organisation and implementation of a complex operations*[23]". Supply chain management aims at effective and efficient flow of goods and services, which essentially is contingent upon the coordination of complex operations.

In terms of *networking intensity*, transport and logistics also falls within the high end of the spectrum. Supply chain management is concerned with the flow of physical goods as well as information between supply chain members' organisations[24], which requires both physical (e.g., road, rail, pipeline etc.) and information networks. Supply chain members' combined information and experience may well be the most significant source of value creation in today's competitive environments. Efficient supply chain networking is therefore essential to leverage the combined knowledge pool that idiosyncratic to the specific chain and, thus, difficult to imitate (Hult et al. 2003).

Openness in the sector varies. Examples of increasing openness can be found in the public transport information domain, where many - often government-held - companies have opened up network and schedule information and started to provide real-time open access to traffic information. Another example relevant to the transportation sector is the proliferation of open geographic information pioneered by projects such as OpenStreetMap[25], a collaborative project to create a free editable map of the world. On the other hand, most transportation and logistics companies have exhibited limited tendency to open up and information is for the most part closely guarded and exchanged via proprietary interfaces.

4.2.6.2. Technology

Transportation and Logistics is an inherently *data-intensive* domain, as the movement of physical goods and people is mirrored by an increasingly dense trail of records. The adoption of technologies such as bar codes, Radio Frequency ID (RFID) systems[26], automated warehouse and distribution systems etc. have led to a proliferation of increas-

23 https://en.oxforddictionaries.com/definition/logistics, retrieved Nov. 20, 2016.
24 https://emmansonme.wordpress.com/2012/03/12/the-role-of-information-technology-in-logistics-and-supply-chain-management/, retrieved Nov. 20, 2016.
25 https://openstreetmap.org, retrieved Nov. 20, 2016.
26 According to SAS, RFID systems generate up to 1,000 times the data of conventional bar code systems. http://www.sas.com/en_us/insights/big-data/what-is-big-data.html, retrieved Nov. 20, 2016.

ingly granular data. Overall, transportation and logistics companies today generate and have to cope with enormous amounts of data[27].

Transportation and Logistics are characterised by high levels of ICT involvement in key processes that are used to manage complex information flows instrumental in the efficient and effective provision and coordination of activities. Whereas in the 1980s, information flows within organisations and between supply chain members were largely paper based, information and communication technology adoption played a crucial role in addressing business needs for reduced cycle times reduction, just-in-time production, supply chain integration etc. Today, *ICT innovation* is perceived in the logistics industry as a major enabler and source of competitive advantage, which, for instance, is evident in significant investments by logistics companies in big data technologies[28]. In the passenger transportation sector, ICT also has significant disruptive potential, as illustrated by the rise of information-centric transportation network companies such as Uber.

Web technologies have improved customer access to real-time information and changed processes in both the passenger and freight transportation sector. Web-based parcel tracking, online booking and ticketing, and access to real-time traffic information are examples of web technology adoption in this sector. However, web technologies have not fundamentally transformed processes in the industry in fact adoption has been slow in some areas. Overall, we estimate that web technology adoption in the transportation and logistics sector has not been above an average level compared to other industries.

4.2.6.3. Assessment

The structural characteristics of the Transportation and Logistics sector, i.e., high levels of internationalisation, operational complexity and intensely networked processes aligns favourably with characteristics of Linked Data technologies. From a technology perspective, Transportation and Logistics also ranks highly in terms of Linked Data

27 UPS, for instance, received 69.4 million daily tracking requests in 2015. https://pressroom.ups.com/assets/pdf/pressroom/fact%20sheet/UPS_General_Fact_Sheet.pdf, retrieved Nov. 20, 2016.

28 DHL - Big data in logistics www.dhl.com/content/dam/downloads/g0/about_us/.../CSI_Studie_BIG_DATA.pdf, retrieved Nov. 20, 2016

Accenture - Global Operations: Big Data Analytics in the Supply Chain https://www.accenture.com/us-en/insight-global-operations-megatrends-big-data-analytics, retrieved Nov. 20, 2016

Going big on Big data: How the Logistics industry is enhancing its supply chain visibility with real-time analytics

http://www.igt.in/blog/going-big-on-big-data-how-the-logistics-industry-is-enhancing-its-supply-chain-visibility-with-real-time-analytics/, retrieved Nov. 20, 2016 adoption potential with its data-intensive processes that involve intense use of ICT and the perception of IT as a technology enabler for innovation and competitive advantage. Linked Data could provide the means for a global distributed data infrastructure that could drive innovation in the sector, with the caveat that the sector has not been on the forefront with respect to web technology adoption and openness.

4.2.7. Tourism

4.2.7.1. Structure

The NACE classification framework does not recognize tourism as a discrete 'industry' because the term is typically interpreted based on the status of the consumer, i.e., the person traveling rather than the supply sides that industries are grouped by in NACE. The closest ÖNACE Class is I *Accommodation and food service activities*, which captures only part of tourism-related products and services, as well as activities not directly linked to tourism. For our analysis, we rely on the definition put forth by the World Tourism Organisation, which defines Tourism as a demand-side phenomenon that refers to the activities of visitors and their role in the acquisition of goods and services. Viewed from the supply side, Tourism is defined as the set of productive activities that cater mainly to visitors. A visitor is a traveller taking a trip to a main destination outside his/her usual environment for less than a year and for any main purpose (business, leisure or other personal purpose) other than to be employed by a resident entity in the country or place visited[29].

Internationalisation of the tourism sector can be viewed from multiple perspectives. On the one hand, the tourism industry is fundamentally associated with locations, as visitors are motivated to travel to particular places in order to experience the - often unique - cultural and/or natural qualities of a particular destination. In that sense, tourism is a highly localized and businesses operating in the sector (e.g., by providing transportation, lodging, information, and entertainment service) are tightly linked to particular geographic locations. On the other hand, global travel demand is growing dramatically and international tourist flows have reached record levels in recent years[30]. As leisure travel becomes more desirable and economically viable for a growing audience, the amount of travellers from emerging economies is growing significantly. At the same time, new destinations market themselves to an international market and

29 Extracted from Organisation for Economic Development (OECD), Commission of the European Communities, United Nations and World Tourism Organisation, (2008) *Tourism Satellite Account: Recommended Methodological Framework*, OECD, United Nations Publications and World Tourism Organisations, p. 1, par. 1.1.

30 International tourist arrivals have increased by 4.4% and reached a record 1.2 billion in 2015. http://media.unwto.org/press-release/2016-01-18/international-tourist-arrivals-4-reach-record-12-billion-2015

popular destinations increasingly attract visitors on a global scale. This rise of mass tourism is (for better or worse) frequently associated with the "productisation" of destinations (e.g., through the provision of standardised travel packages), a development that is paralleled by the internationalisation and concentration of tourism service provision through internationally operating booking platforms, travel agents, information service providers, airline networks, hotel chains etc. Large parts of the tourism industry therefore have a strong international dimension and global providers satisfying tourists' information and service needs are increasingly in a strategically dominant position to capture value and monopolize and monetize tourists' attention. In the light of these broad developments - and despite significant variation - we consider the tourism sector strongly internationalised overall today.

Knowledge-intensity varies significantly within the tourism industry, which involves both activities that do not strongly rely on knowledge (e.g., basic lodging) and highly knowledge-intensive activities such as highly individualised trip planning services, event management etc. We estimate that overall knowledge-intensity in the sector is at an average level.

Due to the broad definition of *Tourism* as an industry, *operational complexity* also varies greatly within it. Integrated high-end accommodation, catering, and travel services may involve the coordination of many activities and stakeholders to provide a seamless service and are hence operationally quite complex. On the other end of the spectrum, basic lodging services can be provided without particularly complex operational processes. Overall, we consider operational complexity within the tourism sector at an average level.

High levels of *networking intensity* are evident in complex value flows that involve a multitude of stakeholders (e.g., agents, booking platforms, payment providers, insurance providers, advertisers, information services providers, accommodation service providers etc.). The provision of integrated services also often requires intense networking and coordination which leads to intense cross-organisational information flows. Finally, information flows between stakeholders in the tourism industry and government institutions also plays a major role. Overall, we consider tourism a highly networking-intensive industry.

Driven by the adoption of concepts such as open innovation, crowdsourcing, and co-creation, there is a considerable drive towards more *openness* in the tourism industry (Egger et al., 2016). This openness can consist in making corporate boundaries more porous to enable service innovation, or in accommodating external input and co-creating personalised customer experiences in services. On a related note, the rise of the “sharing economy” and the rise of new challengers such as airbnb also had a profound impact in challenging traditional providers that provide highly standardized tourism services. By comparison, the proliferation of open data in the tourism industry has been less significant. Many tourism-related statistical data sets are openly available. Geospatial data, which is among the most widely used categories of open data, is highly relevant in touristic applications and frequently reused in the development of touristic applications (e.g., points of interests, public transport information etc.). Overall, we consider the tourism sector to be very open relative to other industries.

4.2.7.2. Technology

Data-intensity in the tourism industry is high. Historically, airline reservation systems were key drivers in the development of early database systems (IBM IMS, SABRE) in the late 1960s (Singh 2009, p. 37). Today, every transportation reservation, hotel stay, car rental etc. leaves a data trail. Mobile-phone based tracking yields highly detailed motion profiles that reveal where and how in a city users are moving, which places of interest they visit, at what times, in which sequence etc. This provides a basis for new data-intensive tourism applications and services and has the potential to drive change in the industry. Overall, even though many parts of the tourism sector remain moderately data-intensive, control over and access to data has become a major competitive factor in the tourism industry, which leads us to rate the overall data-intensity of the tourism sector as high.

Whereas certain areas (e.g., booking systems) have leveraged ICT already very early in its development, *ICT intensity* has historically not been particularly high in major parts of the tourism sector. Small and medium-sized hotel businesses, for instance, were relatively late to adopt ICT for purposes other than mundane office tasks. Today, however, ICT plays a major role in virtually any tourism-related operation, being it large-scale international service companies or small actors in developing countries that heavily rely on their smartphones to conduct business. Whereas the tourism industry has always been relatively rigid in its attitude to innovations (Egger et al. 2016, v), *ICT-based innovation* today acts as a major driver for change. The rise of the Internet has already severely disrupted the tourism industry. Instead of availing oneself of the services of a local travel agents to plan a holiday, the majority of travellers today research options online, draw upon the experience of other travellers, and book the individual services using online booking platforms directly. ICT continues to act as an enabler for innovation in the tourism industry, e.g., through mobile apps and services. Overall, ICT-based innovation is a major competitive force in the sector.

Web technology adoption is also high in the tourism industry, which is not surprising given that the web is the primary source of travel-related information and the primary distribution channel for tourism services today. This is true for both small and medium-sized tourism companies and large online booking services, which have become highly concentrated today, with a few global companies handling a dominant share of the transactions. These services are typically relatively fast to (partly adopt) new web technologies. For instance, the majority of these services include schema.org mark-up on their web pages, which allows them to incorporate to supply structured data to search engines to improve their appearance in search result listings.

4.2.7.3. Assessment

Overall, the tourism sector's structural characteristics are favourable from a LD perspective. High levels of internationalisation, networking and openness point to a need for technologies that can address the resulting requirements. Knowledge-intensity and operational complexity are not as high as in other industries, but growing. The technological environment, which is characterised by high technological intensity and dynamic, appears even more favourable for Linked Data adoption. Overall, the industry-specific preconditions are therefore highly favourable.

4.2.8. Media

4.2.8.1. Structure

The media industry, which forms part of the tertiary economic sector, is concerned with the creation and publication of information and the provision of entertainment services via a range of channels including radio, print, TV and online media.

Internationalisation within the media sector varies as consolidation and deconsolidation are concurrent phenomena in this industry. Mergers of media companies are more frequent on a regional level. A number of media companies provide their services on a global level, however, often these companies focus on covering particular regions they operate in. Overall, the distribution and trade of media is becoming more and more internationalised (Havens et al. 2009).

Knowledge-intensity in media is very high. Knowledge is at the core of the value added by media companies. The provided services are based on knowledge-intensive processes and primarily focus on the production and distribution of knowledge via various channels (Kok 2004).

The operational complexity of media agencies strongly varies based on their size and nature. The typical processes are well-established and do not change substantially over time. Usually, technological innovations will ease the flow of operations which further decreases their complexity.

Intra- and inter-organisational information flows are important in the media industry. In some cases, media producers obtain information from agencies that operate on a global level. Many of them, on the other hand, work independently and produce their own original content. In areas such as film production and the gaming industry, the creation of media products may involve the orchestration of a large number of contributors as well as intense collaboration and information flows.

Publication of open data is a rare phenomenon in the media industry. Media companies do not see any immediate economic benefit in publishing their data openly and hence refrain from doing so. However, there are some initiatives, such as BBC Things, which aim at providing *"data on the places, people and organisations that appear in BBC programmes and online content"*[31]. For the BBC, this initiative primarily helps to provide better search results for search engines, which can follow the links to the (Linked) Open Datasets, and for their customers, who can accurately find articles. Furthermore, developers can also use the data to build new websites and applications.

4.2.8.2. Technology

The industry is increasingly becoming data-driven and new developments, such as data journalism, illustrate that content producers are aware of the potential of data. In fact, data analysis is often used to build the foundation of journalistic stories and underpins their credibility. Furthermore, data can further add value to the products by enabling content usage to be measured very precisely. Data-driven analytics also help to gain better insights into consumer behaviour or to become more responsive to emerging issues.

ICT is ubiquitous in the media industry and vital in the creation and distribution of media products and services. Furthermore, ICT affects the inner processes of media companies as well as the design of products and value structures (Hagenhoff 2006).

The media sector is currently facing technological disruption as traditional and established companies are challenged by new competitors that leverage new technologies and their agility due to their lean organisational structure. IT-based innovation plays a crucial role in this disruption. Such innovations include, but are not limited to, indi-

31 http://www.bbc.co.uk/blogs/internet/entries/afdf2190-4e60-3dfc-b15f-fc17f88c85a1

vidualized information goods, interactive information goods, print on demand techniques, mobile information goods, and digital rights management (Hagenhoff 2006).

By nature, the web-focus of the media industry is particularly high. More and more agencies use online channels for the distribution of their content. For instance, print magazines provide articles via the web and TV broadcasters develop sophisticated video-on-demand platforms for their audience. Each of these developments is web-based and has continually gained market share vis-à-vis traditional channels. Many established media companies were not prepared for this paradigm shift, which requires the development of innovative concepts to deliver and monetize their content.

4.2.8.3. Assessment

Overall, media is one of the most promising industries from a Linked Data perspective, with a wide range of applications that can potentially benefit from a paradigm shift towards LD. The high level of industry dynamics in this sector, where mergers and acquisitions as well as splits and spin offs are commonplace, for instance, poses difficult challenges with respect to the maintenance and integration of data resources and infrastructures. This creates ideal opportunities for LD technologies as a means to overcome these challenges. Thomson Reuters, for instance, which formed through Thomson Corporation's merger with Reuters Group in 2008, took a LD approach to join up its specialized data assets without having to migrate them to a single database or data warehouse (Open Data Institute 2016).

Other factors that contribute to the potential for LD technologies in the media sector include the high level of knowledge-intensity, the ubiquity of heavily IT-supported processes, and the disruption through ICT-driven innovation that many areas of the media sector are currently facing. Finally, media is becoming increasingly networked and data driven and web technologies are already in widespread use throughout the industry.

4.2.9. Telecommunications

4.2.9.1. Structure

Telecommunications involves both the supply of goods (e.g., telecommunication equipment) and the provision of services. It can therefore be classified as part of both the secondary and the tertiary economic sector. The industry used to be dominated by national and regional suppliers and operators, but has become highly internationalised in most parts of the world today. It has been swept up in rapid deregulation and innovation over the past decade and former government monopolies are now facing competitors both on a national and international level (Clegg & Kamall 1998).

Information flows are key in the knowledge economy. As telecommunication businesses provide the infrastructure for the exchange of data, the sector is seen as a pillar of the knowledge-based economy; it is also itself highly knowledge-intensive (Brinkley 2006).

Operational complexity within telecommunications is also high. There are numerous interdependencies, vulnerabilities, and incompatibilities between components and services that create complexity that needs to be managed. In fact, the interdependencies of products and services, sales channels, marketing campaigns, and networking and business processes, along with the IT systems needed to support it all, have become unsustainable for most telecom operators (Arias et al. 2014; Rehse et al. 2010).

Telecommunications is a naturally networked industry with a multitude of stakeholders contributing to typical business processes that require information exchange on the organisational level.

Open Data has made limited inroads into the telecommunications sector so far. Some initial steps to open up telecommunications data, such as call data records, were made. One example is the Telecom Italia Big Data Challenge which provides its data sets as Open Data in cooperation with the Open Data Institute[32].

4.2.9.2. Technology

The telecom industry, which revolves around the transmission of data, is also itself highly data-driven. Data analytics pave the way for personalised services, network optimisation, and location-based service provision, to name just a few. The real-time aspect is particularly important for communication service providers. A study by IBM shows that only 13 percent of their respondents from the telecom industry, far fewer than respondents from other industries, have not started any big data activities yet (Fox et al. 2013).

ICT plays a key role in the telecommunications industry and since the dawn of the internet, telecommunications and information technology have been converging. As a consequence, IT-based innovation has been prevalent, both in core telecommuni-

32 http://theodi.fbk.eu/openbigdata/

cation services and in terms of non-traditional business models, such as Internet of Things applications, payment, and evolving communication technologies. [33]

The telecommunications industry, which has been disrupted by the internet, has not been at the forefront in adopting web technologies and practices. Web-based consumer facilities are today mainly used to reduce operation costs rather than to provide innovative integrated web-based services or to transform telecommunications companies' business models. Overall, web adoption in the sector is relatively moderate.

4.2.9.3. Assessment

Overall, given its internationalisation, knowledge-intensity, networked nature, and operational complexity, the telecom industry could potentially be an attractive target for Linked Data technologies. It is, however, entrenched with established standards and technologies optimized to fulfil its specific technical requirements.

4.2.10. Finance and Insurance

4.2.10.1. Structure

The Finance and Insurance industries supply services to consumers and businesses and form a major part of the tertiary economic sector. They are characterised by a high level of internationalisation (Schoenmaker 2014), with a large number of multinational enterprises operating, e.g., in the European Union. Still, a large share of the firms operating in this sector are national companies (national banks, insurance brokers, etc.). Main drivers for the internationalisation are cross-border information flows, common international standards, and linking infrastructures across borders (Hannoun 2006). However, it is important to note that as of 2013, internationalisation had not reached the pre-crisis level of 2009 yet (Lund 2013).

Financial services are highly knowledge-intensive, driven by a tremendous number of rules, best practices, and regulations (Brinkley 2006). In addition, organisations in this industry rely heavily on professional expertise relating to a specific functional subdomains.

Operational complexity is high due to numerous rules and regulations that govern lending, finance, and insurance. Indeed, studies prove that the main drivers of complexity in the industry are regulation, channel proliferation, systems fragmentation, product proliferation, and geographic expansion (Wyman 2015). These factors both

33 http://www2.deloitte.com/us/en/pages/technology-media-and-telecommunications/articles/telecommunications-industry-outlook.html

increase operational costs and undermine decision making as well as decrease the influence managers have over their firms.

The financial services industry is highly interconnected with strong inter- and intra-organisational information flows are business-critical, as well as highly regulated and governed. Financial organisations also typically rely on a large branch network, with offices dispersed over large geographic areas. Therefore, information flows are essential to the operation of the business.

Uptake of the open data concept has been slow in the financial and insurance industry so far. Most of the data is very sensitive due to concerns related to privacy and competitiveness and it is generally not desirable to provide this data to the public. There are, however, notable exceptions. Startups such as the Open Bank Project have been supported by the Open Data Institute and aim to create an open banking platform. Another example is Quandl, a data platform providing financial and economic data for analysts, which offers also open data from central banks, exchanges, brokerages, etc[34]. In the United Kingdom, the HM Treasury and Cabinet Office has commissioned a report on data sharing and open data in the banking industry that concluded that greater access to data has the potential to help improve competition in UK banking and proposed a set of measures to achieve these benefits[35].

4.2.10.2. Technology

Financial services are highly data-intensive, as they revolve primarily around transactions. Data adds value by enabling sophisticated analytics that allow for superior pricing strategies, risk analysis, and mitigation strategies. Data is also heavily used to obtain deeper insights into transactions and better understand customer behaviour. In the financial trading sector, analytical methods are crucial for bottom-line profitability. Data security is equally vital, which underlines the data-driven nature of this industry.

Financial service enterprises are both primary sources of information and users of innovative technology (Windrum & Tomlinson 1999). Hence, ICT-intensity is very high in the financial and insurance sector. Given the extensive rules governing the business, the audit requirements etc., it would be very difficult to operate in these markets without ICT involvement (Dapp 2014).

IT-based innovation in this industry is high, despite the fact that the IT infrastructure in the financial sector is frequently built upon non state of the art technology. This

34 https://www.quandl.com/open-data
35 https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/382273/141202_API_Report_FINAL.PDF

conservative stance towards IT is increasingly being challenged by smaller "FinTech" companies that are heavily based on ICT to make financial services more efficient and strive to introduce new ways to deliver value to the customer. Online banking has become prevalent and more and more financial services companies are investing in mobile apps, as well as self-service kiosks. Additionally, IT plays a crucial role in financial trading. Smaller organisations in the sector are increasingly taking advantages of the lower cost of cloud based solutions (Ericsson 2014). The high level of ICT investment is also reflected in the global ICT spending statistics reported in a study by the Information and Communications Technology Council (ICTC 2012).

The focus on web technologies is increasingly prevalent in the financial services domain. Customers increasingly want to take charge of their own finances, in their own time. Very often, it is difficult to speak to a financial advisor in a branch for various reasons (shorter opening hours, staff downsizing, etc.), hence, a web presence is vital. Mobile payments and online banking are based on web technologies and also accepted by many customers. Also, in the insurance sector, online insurance plays an important role driven by a company's infrastructural flexibility and degree of business integration (Cata & Lee 2006).

4.2.10.3. Assessment

Overall, the finance and insurance industries could benefit significantly from LD adoption to address growing data problems. Relatively high levels of internationalisation, strong inter- and intra-organisational information flows, the growing focus on web technologies, and the high level of operational complexity are aspects that suggest considerable potential. The conservative stance towards IT-based innovation and the inertia introduced by the large-scale legacy infrastructure, however, could be a major stumbling block for adoption. Short- and long-term benefits and the increasing dynamicity in the industry may help to overcome this major barrier.

4.2.11. Engineering

4.2.11.1. Structure

Engineering forms part of the secondary economic sector. It is a very diverse sector which makes a general assessments difficult. We will therefore study it by example of production systems automation and make illustrative references to this field.

The level of *internationalisation* in this industry is very high, because most enterprises in the sector are multi-nationals acting on a global level. Furthermore, barriers to global trade became lower which in turn opened up business opportunities on a global scale. According to case studies, project offices and/or strategic alliances with local partners are typically chosen forms of market entry. This helps enterprises to get access to market knowledge, which is crucial especially in early stages of entering foreign markets (Brandl 2010).

Engineering generally involves highly *knowledge-intensive* processes that rely strongly on software and the associated knowledge encoded in this software.

Operational complexity in engineering is high. Important factors which lead to high complexity in production systems are demand fluctuation as well as high product quality, low cost, short lead times, and high customisation requirements (Efthymiou et al. 2012). Mostly, this applies to large manufacturers whereas SMEs tend to operate in niches with more limited complexity.

Cross-organisational information flows are intense, especially due to the similarly high level of internationalisation and division of labour. In particular, there is a need to streamline the work of a large and diverse set of stakeholders which span diverse engineering disciplines (mechanical, electrical, software), make use of a diverse set of (engineering) tools and employ terminologies with limited overlap (Schmidt et al., 2014). This requires dealing with heterogeneous and semantically overlapping engineering models (Feldmann et al, 2015). Therefore, a key challenge relies on intelligently solving data integration among the various stakeholders involved in the engineering and operation of systems both across engineering domain boundaries (horizontal integration) and between different abstraction levels (business, engineering, operation) of the system (vertical integration). The use of Semantic Web and Linked Data technologies for engineering settings has been investigated to length in (Biffl 2016).

The level of openness in terms of data publication is low and can be separated into two levels: First, within an enterprise's cooperation consortia, data is often shared. Second, the core intellectual property is closed data, that is, not shared with anyone. This may even be closed to employees to prevent leakage of know-how. Sharing data to the public - as is the original sense of Open Data - is not common in the engineering industry because this may provide useful insights also for competitors on the market.

4.2.11.2. Technology

The engineering domain is highly data- and technology-driven. Technologies need to fit the processes and consequently also the data, such as parameters, settings, configurations, need to fit. Moreover, as the paradigm of Industry 4.0 is emerging, the implementation of feedback-loops by incorporating additional data sources such as sensor-measurements plays a crucial role. Generally, the idea is to use data for analysis and for streamlining of machine-human interactions (Tayal 2016). Data adds value through providing, transforming, processing, analysing, and aggregating meaning and plays an essential role throughout the complete process.

ICT-intensity is also generally high, given that ICT is essential for the automation of production and manufacturing systems. Studies show that ICT is essential for creating supplementary services relating to machinery. These services allow for new business models, increase global competitiveness, and create jobs for skilled workers. Examples are "available warranties", "agreement on continuous improvement", "pay-on-production", and "warranted life cycle costs". In addition, ICT adds value by enabling innovative measurement and control technology as well as machine-to-machine communication (Fraunhofer 2012).

IT innovation in engineering is also high. However, despite intense efforts, research results from academia are often not implemented and put into operation. One example is the research on "software agents", which was emphasized by many organisations over years without being adopted in practice, partly due to security and traceability concerns. Still, progress in form of faster CPUs, increased memory and storage capacities contributes to a high level of IT innovation this economic sector.

The web focus of the industry is low because traditionally, production systems were always isolated from the web for security reasons. However, more recently the drive towards "Industry 4.0" has aimed to generate value and increase efficiencies by means of internet- and web-based technologies in the engineering domain. A key in this context is to integrate value-added processes which themselves are based on web-technologies. Examples include online product catalogs which are relevant for (i) sales and marketing departments on the producer side, and (ii) solution designers, engineers etc. on the user side in order to answer questions regarding, for instance, driver compatibility.

4.2.11.3. Assessment

Engineering is highly internationalized, knowledge-intensive, operationally complex and characterised by a high level of networking. These favourable structural factors suggest a high propensity to adopt LD. However, the very limited openness of the sector suggests that adoption of LD as an internal technology for ELD/LED is more likely to succeed than the adoption of LOD. Similarly, technology factors such as data- and ICT-intensity and innovation also point at a high potential for ELD/LED adoption.

4.2.12. Research and Development

4.2.12.1. Structure

Research and development (R&D) is challenging to demarcate as a distinct area of economic activity. The most widely adopted definition has been put forth by the OECD in the Frascati Manual on the measurement of Scientific, Technological and Innovation activities. The latest version of this manual (OECD, 2015) categorizes R&D into basic research, applied research, and experimental development and defines them as follows:

- *Basic research* is experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundation of phenomena and observable facts, without any particular application or use in view.
- *Applied research* is original investigation undertaken in order to acquire new knowledge. It is, however, directed primarily towards a specific practical aim or objective.
- *Experimental development* is systematic work, drawing on knowledge gained from research and practical experience and generating additional knowledge, which is directed to producing new products or processes or to improving existing products or processes.

The NACE classification includes scientific research and development as a sub-category of professional, scientific and technical activities (M) and distinguishes research and experimental development of natural sciences and engineering (M72.1) and research and experimental development on social sciences and humanities (M72.2). In our analysis, we follow these definitions and include both public and private sector research and development activities in our considerations. This includes R&D activities conducted at universities, research institutes, industrial research centres etc. We do not include corporate R&D as a functional unit within organisations in other industries here, but cover them in the respective industry sections.

Due to the ubiquitous and universal nature of knowledge, scientific research has in some sense inherently been a global endeavour throughout history. In practical terms, however, limited means for information exchange have historically led to significant duplication of effort and a multitude of parallel discoveries. The foundation of scientific journals in many fields in the 18th century and increasing mobility triggered a process of internationalisation. Today, international cooperation has become the norm in most scientific fields, either through informal cooperation, projects, researcher mobility etc., or in institutionalized and jointly governed long-term efforts (e.g., CERN, ESA, IIASA etc.). Overall, however, the global research geography is still characterised by regional clusters, as truly global cooperation is contingent on the availability of appropriate funding. Funding priorities and policy also drive the formation of thematic research clusters and industrial policy plays an important role in that. A recent shift towards knowledge and innovation generation in regions outside Europe, North America and Japan can also be observed. Countries that are often considered in their relatively early stages of economic development, are increasingly driving global knowledge and innovation development (Serber and Wise, 2010).

Research is, by its very definition, a knowledge-intensive endeavour which aims at the creation of knowledge as its main outcome. This applies to a similar extent to development, which aims to put the created knowledge into practice.

Research, essentially a creative process, is inherently unstructured and complex. The extent of operational complexity varies among disciplines, with a spectrum ranging from low (e.g., philosophical discourse) to high (e.g, natural science research that involves complex machinery and requires strong coordination between a multitude of stakeholders and activities in order to conduct a single experiment).

As a globally distributed endeavour, research and development as a whole is highly networked, which is reflected in research cooperations and institutionalised events such as conferences and workshops. On the level of individual research topics and initiatives, intense inter-organisational information exchange is commonplace.

"Openness" is increasingly considered an essential core value and a moral imperative by many researchers. The scientific method itself is founded on the idea that others must be able to replicate - and potentially falsify - results. Sharing of research results is hence one of the most treasured principles in contemporary academia and global dissemination of results in scientific journals has both become standard practice and a key indicator of scientific achievement. However, these dissemination channels are presently mostly controlled by a small number of major publishers driven by commercial interests. Dissatisfaction with this status quo has led to a growing open access movement in recent years. More broadly, the idea of open science has gained momentum, which involves not only the dissemination of final results, but also the documentation and sharing of the input data, processes, and tools used to derive them. Industrial research and development activities, on the other hand, are often highly secretive when commercial interests and competitive considerations are involved.

4.2.12.2. Technology

Many areas of research and development are highly and increasingly data-intensive. This is reflected in the notion that data exploration has emerged as a "fourth paradigm" of science, implying that computational thinking will transform scientific discovery which had previously been driven by empirical, theoretical, and computational approaches (Hey et al. 2009).

ICT-intensity varies among research domains, but even fields that have traditionally been characterised by limited adoption rely heavily on ICT today (e.g., the digital humanities). Whereas formerly, ICT was often used as a tool to conduct research more efficiently (e.g., through literature databases, document management systems, statistical software packages etc.), computational methods have today fundamentally transformed many fields of research. Techniques such as numerical simulations, large-scale data analysis and data mining, optimisation etc. have created entirely new fields of research in various domains (e.g., computational biology, chemistry, genomics, astronomy, economics, social science etc.). Overall, R&D is characterised by high levels of ICT intensity and ICT-driven innovation.

It is not surprising that the web emerged from research, addressing a pressing need for global information sharing. Today, despite growing commercial interests, academia remains heavily involved in the development of web technologies and standards. On the other hand, the fundamental principles of the web are not generally ingrained in today's research culture and dissemination practices.

4.2.12.3. Assessment

Overall, we consider research and development one of the most promising domains. In addition to an existing wide array of proven LD applications, general aspects such as the high levels of knowledge, ICT, and data intensity, the high degree of internationalisation and cooperation, and the tendency towards openness and web-savviness suggest that conditions for LD adoption are excellent.

4.2.13. Government

4.2.13.1. Structure

Government has traditionally been seen as part of the tertiary sector, including services such as social security, policy, education, transport services, health, etc. (Mohanty & Behera 1996). Note, however, that it is difficult to categorize all the activities carried out by a government. As such, the International Standard Industrial Classification of All Economic Activities (ISIC) by the United Nations (United Nations 2008) does not provide a specific code for this special activity as a whole, but considers the classification of each of the underlying activities (e.g. public hospitals within the general "Hospital activities" category, i.e. health activities). Nevertheless, the ISIC division (no.84) includes some activities (e.g. foreign affairs, defence activities, public order) that, although they disregard the institutional status, constitute core processes in an administration.

The level of internationalisation of the governmental sector varies widely on a global scale and is difficult to measure. On the one hand, obviously, the activities of local and national governments are more or less concentrated on a national scope. On the other hand, foreign policy and supra-national negotiation has historically led to strong alignment of many policy areas (e.g. economy, security and education). Finally, the role of governments has become highly internationalised within politico-economic unions, such as the EU, which operates through supranational institutions, and other kinds of bilateral agreements. Overall, the governmental sector involves both local and national activities and areas of strong international cooperation.

Knowledge-intensity varies heavily. On the one hand, the governmental sector can be seen as an area where extensive knowledge (know-how, regulations, strict information processes, etc.) is required to operate in a legal and efficient manner. On the other hand, public administration sectors do not fit the traditional classification of knowledge-based industries. For example, Lee and Has (1996) proposed a categorisation based on three R&D measures: i) the ratio of R&D-to-sales, ii) the ratio of R&D personnel to total employment, and iii) the ratio of professional R&D personnel to total employment. These measures cannot be applied to government directly (educational services, health care, and public administration sectors, etc.), which would require a specialized classification framework of knowledge-intensity (Kanagarajah 2005).

The complexity of government operations varies considerably, ranging from low (e.g., register for a public library) to high (e.g., collection of taxes). The most "critical" operations in a government (economy, health, defence, etc.) tend to involve diverse stakeholders, potentially across agencies, units, companies and countries, which results in considerable operational complexity.

Public administrations are also a prototypical example of complex multi-level organisations composed of individual units that need to interact both internally and externally (with citizens, enterprises, international organisations, etc.). These interactions are executed in – often strictly and formally defined – processes and result in various heterogeneous information flows.

Finally, openness and transparency are of fundamental importance in democratic societies and open government principles have been adopted in many countries around the world, often implemented by "freedom of information" legislation. Nonetheless, its practical implementation varies significantly. As of September 2013, 95 countries had laws establishing some procedure to publish and request open data (Right2info.org, 2013). In Europe, the Directive 2013/37/EU (amending Directive 2003/98/EC known as the 'PSI Directive') regulates the re-use of public sector information, encouraging the Member States to make as much information available for reuse as possible. As of July 2015, all EU member states had to implement this directive. The way and extent to which members implemented this directive varies widely, from adopting specific measures or adapting their legislative frameworks. Open Data portals (such as publicdata.eu, a Pan European data portal, and data.gv.at in Austria) provide means to access open government datasets and attracted significant interest in recent years. To sum up, despite resistance in many areas, the willingness to publish open data in the governmental sector can be seen as high relative to most private industries.

4.2.13.2. Technology

Although generalisations are difficult, large parts of the governmental sector tend to be highly data-intensive. Government decisions, policies, and rules are (ideally) driven by (statistical) information. Although this information can be intangible (such as social and political trends), critical decisions such as economic policies, public health measures, etc. should be based on data analytics. Benefits expected from data-driven government include sounder decisions and policies, optimized fraud and error detection, improved services, efficiency, and public perception of the government (Lutes 2015).

ICT plays an important role for public sector innovation, in particular in the context of e-government developments (Layne and Lee 2001). E-government can be defined as the concept of using ICT within governments to improve the efficiency (cost reduction), quality, functionality, and user satisfaction of the services provided by the government to its citizens (Ndou 2004). In addition, e-government is tightly associated with open government, which enables governments to be more transparent, open, accountable, and agile (McDermott 2010). Nowadays, ICT-enabled public sector innovation is in many government's digital agenda, especially within Europe in the "Europe 2020" strategy (European Commission 2010), and Austria, where the Platform Digital Austria[36] plays a catalytic role.

Although many areas of government have traditionally been inert to change and conservative in their IT strategy, government has, on the other hand, been a major driving force of IT innovation, as, for instance, in defence, health, energy, space and support via research programmes (Eurostat 2013). In an urban context, the ambition to lever-

36 http://www.digital.austria.gv.at/

age IT innovation to improve quality of life and resource efficiency is reflected in the "smart city" concept that is increasingly being adopted by government institutions. Overall, IT innovation in the government sector is not very equally distributed, but there is a tendency towards intensified IT-based innovation in the future.

As far as the relationship between government and the web is concerned, web-based applications and services have been used to provide information and services to citizens, businesses and other governmental entities for many years (Layne and Lee 2001) in the context of e-government and open government initiatives. A review of methods to evaluate current E-government services is provided by Osman et al. (2014). (Semantic) Web technologies are also increasingly applied to ease access to governmental and public-sector data (Heath and Bizer 2011). In particular, the Government Linked Data Working Group[37] of the World Wide Web Consortium (W3C) fostered this adoption by providing best practices for publishing Linked Data, i.e., guidelines to facilitate development and delivery of open government data as Linked Open Data (World Wide Web Consortium 2014).

4.2.13.3. Assessment

Overall, it is not surprising that the governmental sector has been interested in Linked (Open) Data from an early stage. Given the structural premises of the domain, the technical need for interoperability, flexibility and data integration, and the drive towards transparency and openness, the governmental sector could play a pioneering role in LD adoption.

4.2.14. Education

4.2.14.1. Structure

The education "industry" is part of the tertiary economic sector. Its level of internationalisation is relatively moderate, particularly in primary and secondary education, which are run by public schools and therefore fall within the responsibility of federal governments. Tertiary education (e.g., undergraduate and postgraduate education) is more internationalised due to increased student, programs, and institution mobility (Naidoo 2006). Main motivations for internationalising higher education are increased profit, access provision, and curriculum enhancement, among others (Altbach and Knight 2007).

37 https://www.w3.org/2011/gld/wiki/Main_Page

Education is highly knowledge-intensive by its very definition. The delivered product, i.e., knowledge itself is an intangible asset. Media that embody and transport this asset have traditionally been physical objects (e.g., books), but in recent years there has been a strong shift towards digital media. Ongoing R&D in the area of education also contributes to the industry's knowledge-intensity. For instance, e-learning and how to perform it properly (didactics) is an intensely researched topic. In addition, there has been a significant amount of investment into software and web-based technologies (Fromm and Kern 2000).

Operational complexity in education is fairly low, because typical processes in education are well defined and do not undergo major changes (Kvavik et al. 2005, p. 25). Complexity may be added due to the increase in globalisation (e.g., management of international students with different degrees), but this will not significantly drive the overall complexity in this industry.

Networking is relatively limited due to the fact that only few parties are involved in the key processes. Typically, the processes are already determined and stable and therefore do not change frequently. Again, tertiary education is characterised by a significantly higher dynamics and a higher level of interaction and networking.

The level of openness in terms of data publication is medium for the educational sector. On average, educational institutions do not appear to be eager to publish open data, although some universities have made explicit efforts to publish university related data as Linked Data, as evident from the Linked Universities[38] project.

But in line with the identified shift towards more digitalized delivery of education, openness may also become increasingly important. Terms like "open education", "open data in education", and "open knowledge" became popular recently and show that sharing and openness are widely considered as important values within educational culture. It is important to note that there are two perspectives on Open Data in education: i) simply reusing already available "open data" sources for education purposes; and ii) utilising Open Data which is generated by education institutions (e.g., student data, course data, institution data)[39]. The latter is more relevant for our evaluation of the current state of open data in education. Furthermore, we observe that learning analytics, which is highly related to utilising open educational data, is still in its infancy. Hence, we conclude that there is still a large untapped potential to open up educational data.

38 Linked Universities project: http://linkeduniversities.org/
39 http://de.slideshare.net/MariekeGuy/edtalk2

4.2.14.2. Technology

Education involves a lot of content production and is knowledge-intensive, but educational processes (i.e., teaching and learning) has not been particularly data-intensive compared to key processes in other domains. Even though a lot of data such as student data (how many students are enrolled in which courses, what are their grades, etc.) likely exists, it is typically not being exploited systematically. Another dimension is data generated by assessments such as the PISA study (or more recently and also relevant, the results from the new 'Zentralmatura' which was introduced in Austria). Such data is potentially useful for analysis, but currently is not a major factor in the quality of delivering education. Rather, the data is used to influence policy and decision making at the administrative level. However, it is to be expected that the education market will become increasingly globalized due to innovations in education. Data-drivenness could then potentially significantly increase. For instance, the performance of students under varying conditions can be evaluated, which in turn can be used to improve didactic designs and ways of how teaching is conducted[40]. At the same time, it will be crucial to account for ethical and privacy considerations when tracking learners (Pardo and Siemens 2014, Chatti et al. 2014).

Furthermore, education has become increasingly ICT-intensive in recent years. Again, especially in tertiary education and in an increasingly connected world (mobile technology, massive open online courses, e-learning, etc.), ICT is a key enabler for educational innovations. At a minimum, more and more schools have so called laptop-classes where each student is equipped with his/her own laptop and teaching of most classes involves these laptops. Beyond that, there is a large stream of research towards improving e-learning and web-based learning methods, including technology-enhanced learning (TEL), e-learning, learning analytics, mobile learning, gamification, and MOOCs. This results in new ways to deliver education to students. On top of that, there is also a large number of startups that aim to disrupt the education sector [41,42,43]. Moreover, the administration of educational institutions is also typically ICT-intensive and involves specialized software to efficiently manage curricula, time tables, and staff. On average, and if we add primary and secondary education to the equation, ICT usage and innovation is intense in many parts of education, but arguably not as dominant as in some other industries.

40 http://radar.oreilly.com/2011/07/education-data-analytics-learning.html
41 http://www.inc.com/ilan-mochari/16-startups-that-will-disrupt-the-education-market.html
42 http://edtechreview.in/news/2080-education-startups-companies-funding-2014-2015
43 http://techcrunch.com/2016/01/09/how-startups-are-solving-a-decades-old-problem-in-education/

The web-focus of the education sector is considerable and evident, for instance, in the increasing interest in online courses and remote learning. Still, we need to make a distinction between "web focus of companies operating in the education industry" and "web focus of the services which are provided by companies operating in the education industry". Most statements made above refer to the latter. For the former, a clear statement is difficult to make since "companies operating in the education sector" are a cross-section of public sector institutions (education providers), IT companies (software), publishers (providing books/materials), etc. Hence, it is difficult to make general statements about their web focus.

4.2.14.3. Assessment

Based on structural and technological industry characteristics, we do not expect education to be particularly prone to Linked Data adoption. The educational sector is knowledge-intense and partly relatively open, but also geographically fragmented, not particularly data and ICT intensive, and not particularly innovative in ICT use. Despite this cautious assessment from a structural and technological perspective, there may be good use cases in this sector but we expect that fostering adoption will be challenging.

4.2.15. Health

4.2.15.1. Structure

The health industry is highly diverse and spans both the secondary (e.g., manufacturing of pharmaceutical products) and tertiary (e.g., health care services) economic sectors. Health systems have traditionally developed on the national level, whereas the pharmaceutical industry has been, and is continuing to become, more internationalised. Internationalisation of healthcare is also growing due to the liberalisation of institutions, which has a significant impact for the delivery of health services. For instance, developments in the British healthcare system have triggered a shift away from direct state provision towards regulating private providers. This process facilitates internationalisation (Holden 2003).

The healthcare domain is highly knowledge-intensive (Southon 2003). Medical and pharmaceutical practice and public health policy making require in-depth medical knowledge about diseases, mechanisms and pathogenesis, therapies and interactions, and interpretation of lab tests[44].

44 Segen's Medical Dictionary. S.v. "medical knowledge." Retrieved November 5 2016 from http://medical-dictionary.thefreedictionary.com/medical+knowledge

Operational complexity is typically very high in health care. Intricacies include guidelines that need to be followed and regulations that need to complied to. Studies report that 90 percent of healthcare executives expect high or very high levels of complexity until 2015, but more than 40 percent are unprepared to deal with those complexities (Cortada et al. 2012). Systemic issues, citizens' access to information over the web, and increasing regulatory presence significantly contribute to the high level of perceived complexity (ibid).

The health domain is characterised by a high degree of interconnectedness and inter-organisational information flows. Many stakeholders and organisations, such as hospitals, pharmacies, researchers, drug manufacturers, and patients, are involved in typical health processes. Furthermore, regulations and guidelines involve a multitude of stakeholders and large amounts of information that are transmitted across organisational boundaries. Overall, the health industry is characterised by an abundance of data, although access limitations due to privacy and security requirements for personal information apply. Health analytics aims to improve the quality of health care delivery, detect diseases at early stages, and manage specific health populations. Examples of data that is available for such analytics include clinical data, activity and cost data, pharmaceutical R&D data, and patient behavior data (Groves et al. 2013). The volume of health data is expected to grow dramatically in the years ahead (Cottle et al. 2013).

The willingness to publish open data and to follow open data principles varies greatly. In the pharmaceutical industry, openness is limited, as the business is essentially based on intellectual capital and licensing. In areas that are considered less relevant from a competitiveness perspective and where large cost savings can be achieved through sharing, initiatives to open up data and knowledge exist, such as exemplified by the OpenPHACTS[45] project (Groth et al, 2014). In the health care provision domain, a number of open data initiatives exist, as for instance the "Open Data Strategy 2013–2016" of Queensland Health (Queensland Government 2013) or the Health section of data.gov.uk[46] (i.e., the open data initiative of the UK government). The provided data sets primarily focus on aggregated statistical data. Openness in the medical domain is more prevalent in academia (which, however, does not fall under the health industry definition according to the NACE framework).

4.2.15.2. Technology

The health care sector has traditionally fundamentally relied on data in research, diagnosis, and practice. A large share of the data, however, was kept in physical patient

45 Open PHACTS project: https://www.openphacts.org/
46 https://data.gov.uk/data/search?theme-primary=Health

records. In recent years, e-health has spurred the proliferation of electronic medical records as health care continued to become increasingly data-intensive. Data-centric approaches allow physicians to profile their patients, to recognize the best treatments for disease, and to help researchers discover potential cures by mining big data.[47] From genome sequencing to building personalised health profiles from the data in patients activity trackers, data-driven medicine promises to not only improve the speed and accuracy of diagnosis for genetic diseases, but also unlock the possibility of personalised medical treatments[48]. Against the backdrop of these broad developments, we consider the healthcare sector highly data intensive.

Techniques to effectively manage explicit knowledge, such as practice guidelines, decision support systems, tools for empowering patient choice, and reference databases are crucial for successful knowledge management in this industry (Wyatt 2001). Accordingly, intensity of ICT is high, as information (management) systems are typically used to organize and manage knowledge and to facilitate communication (Waling 2006).

In the pharmaceutical industry, advances in ICT have dramatically affected drug discovery and design. In the healthcare industry ICT, particularly e-health, is expected to reduce costs, improve services, and cause behavioural change (Houghton 2002). The term e-health stands for the application of ICT to improve access and quality of clinical and business processes by healthcare organisations, personnel, and patients. It gained significant popularity in recent years. In line with these developments, IT-based innovation in the health industry has been growing. Studies also highlight the potential of e-commerce and internet-based technologies in drug marketing and distribution (Houghton 2002). Additionally, a 2013 study by McKinsey states that in the US, more than 200 new businesses have developed innovative healthcare applications since 2010. Approximately 40% of these aim at direct health interventions or predictive capabilities (Kayyali et al. 2014).

Adoption of web technologies within the healthcare industry, on the other hand, varies greatly. A number of web-based solutions to provide improved services for medical staff and patients have been developed, but the overall role of the web as a platform to tackle inefficient paper-based information sharing practices and for managing patient information has been limited. This may partly be explained by the sensitive nature of medical records and the associated privacy concerns. In the pharmaceutical industry, the impact of the web (e.g., as a marketing platform) has also not been particularly

47 http://med.stanford.edu/iddm/overview.html
48 http://www.forbes.com/sites/jenniferhicks/2016/07/27/artificial-intelligence-and-data-driven-medicine/#74b69e185af8

profound. However, Semantic Web technologies have seen widespread adoption in the pharmaceutical industry across organisations and has already led to significant changes in drug discovery research, as exemplified by the Open PHACTS[49] project. This platform was built in collaboration with a large consortium of major academic and commercial organisations involved in drug discovery and brings together pharmacological data resources from a multitude of stakeholders in an integrated, interoperable Linked Data infrastructure.

4.2.15.3. Assessment

Favourable factors for Linked Data adoption in the healthcare sector include its highly knowledge-intensive and operationally complex nature, the strong need for networking and information exchange, the high levels of ICT intensity and innovation, and the increasingly data-driven nature of medical services. Despite moderate levels of internationalisation, openness and web technology adoption, we consider healthcare a highly promising domain for Linked Data adoption.

4.3. Conclusion

	Internationalization	Knowledge-intensity	Operational complexity	Networking intensity	Openness	Data-intensity	ICT-intensity	ICT innovation	Web adoption
Manufacturing	low	low	low	low	high	medium	medium	high	high
Energy	medium	medium	medium	low	high	low	low	low	medium
Retail	low	medium	medium	low	high	low	low	low	low
Finance and insurance	low	low	low	low	high	low	low	low	medium
Government	medium	medium	low	low	low	low	low	medium	low
Health	medium	low	low	low	medium	low	low	low	medium
Transport and Logistics	low	medium	low	low	medium	low	low	low	medium
Telecommunications	medium	low	low	low	medium	low	low	low	medium
Media	low	low	medium	low	medium	medium	low	low	low
Education	medium	low	high	medium	medium	medium	medium	medium	medium
Tourism	low	medium	medium	low	low	low	low	low	low
Engineering	low	low	low	low	high	low	low	low	medium
R&D	low	low	low	low	medium	low	low	low	low
Construction	low	low	low	low	high	medium	medium	medium	high
Resource Industries	medium	medium	medium	low	medium	medium	medium	medium	high

Figure 4.3: Sectoral characteristics overview (legend: ☐ = low; ▒ = medium; ■ = high).

49 OpenPHACTS project: https://www.openphacts.org/

In this chapter, we conducted an in-depth sectoral analysis of industries' structural and technological characteristics in order to assess their respective propensity towards Linked Data adoption. To this end, we first formulated nine working hypotheses on the industry characteristics that affect the potential for Linked Data adoption. We then evaluated these hypotheses on 16 industries based on the ÖNACE 2008 industry classification schema. Industries identified as promising in this analysis are (i) characterised by a high level of technological readiness for LD and (ii) based on their structural characteristics, may benefit strongly from adopting LD technologies.

Overall, we find that Energy, Retail, Finance and Insurance, Government, Health, Transport and Logistics, Telecommunications, Media, Tourism, Engineering, and Research and Development rank among the most promising industries from a Linked Data technology adoption perspective. All these industries are highly or increasingly data-driven and characterised by strong intra- and inter-organisational information flows. However, limited focus on web technologies could inhibit Linked Data initiatives in most of these industries. Furthermore, with the possible exception of research and development as well as government, most industries have been slow to embrace "openness". The strong emphasis on openness in the academic discussion of Linked Data may therefore contribute towards reservations against the adoption of Linked Data technologies in an enterprise context. Whereas openness can entail strong economic network effects and can be mutually beneficial in many of these industries, it is important to note that the adoption of Linked Data technologies does not necessitate fully or even partly opening up internal data. A stronger emphasis on "Linked Closed Data" may therefore be beneficial for adoption in most of these industries.

At the bottom end of the classification we find the Education, Manufacturing, and Resource industries sectors. All these industries have a relatively limited focus on web technologies, low amount of cross-organisational information-flows, and are not particularly data-driven. On the other hand, these industries are also fairly knowledge-intense and characterised by a high level of IT innovation, which suggest a high potential for Linked Data adoption. Education does not fall into the lower range in any of the dimension except operational complexity, but is only moderately internationalised and characterised by more limited information flows, data-drivenness, and operational complexity.

To sum up, the conducted sectoral analysis and identified eleven industries as particularly promising from a Linked Data adoption perspective. We provide an in-depth discussion based on relevant literature on each industry evaluation.

5. DEMAND-SIDE MARKET ANALYSIS

In this chapter we provide a high level overview of the demand-side analysis conducted as part of the PROPEL project, introduce the user stories, that emerged from the stakeholder interviews and workshop, and discuss the relationship between the user stories and the technology foundations identified in *Section 2.4.7* (i.e. the matching of specific data problems with existing technological solutions). Besides the user stories and the technological foundations mapping the knowledge gained from the interviews, workshop and subsequent analysis was used to inform all other chapters, especially to the description of market forces, sectoral analysis and the roadmap.

5.1. Stakeholder Interviews and Workshop

In order to obtain a deeper understanding of the market drivers and ascertain concrete enterprise data and information management requirements several interviews were conducted and the findings were further elaborated via a stakeholder workshop. The overarching goal of both the interviews and the stakeholder workshop was to investigate the data challenges that organisations are currently facing, and to formulate these challenges in the form of user stories / use cases for Linked Data.

5.1.1. Interviews

In total, 22 face-to-face or phone interviews were conducted with organisations in Austria as well as the United Kingdom, the Netherlands, Italy, Switzerland, the Czech Republic and Sweden. The organisations were diverse not only in organisational type, but also in sectorial focus and LD experience:

- The organisation types consisted of 12 large enterprises, 6 small to medium sized enterprises (SMEs), 2 non-governmental organisations (NGOs), 1 university and 1 international organisation.
- The organisations span several sectors, namely engineering, professional services, research, finance and insurances, transport and logistics, media, government, tourism, and health.
- The level of LD experience that the organisations have differs significantly. One third of the organisations are very experienced with LD (e.g. through active implementation of LD initiatives or because they are a provider of LD solutions or similar solutions), while almost as many organisations have no experience with LD at all. A bit more than one third have some experience (i.e. these organisations have already started to think about adopting LD for certain use cases within their

organisation, or thought about adding elements of LD to their solution portfolio in case they are solution providers).

- 9 organisations are IT-providers - some of them having the LD concept as part of their portfolio. Of the remaining 13 organisations, 4 have already adopted or are just adopting LD for certain use cases within their organisation, and the remaining 9 are potential users.

The interviews were structured into four distinct topics that aimed to ascertain information about:

- **The organisation** - such as their primary business, operational and organisation structure and collaboration both in terms of business processes and software systems.
- **Data integration / exchange / linking challenges and opportunities** – such as the specific organisational or technical challenges with respect to data exchange/ integration within the organisation and with third-party organisations and the potential added value that external data could bring to the business if it was integrated/linked in your company.
- **Linked Data familiarity, drivers and barriers** - Their knowledge of LD, their view of the value of LD from a business and/or technical perspective, the main difficulties encountered in the adoption of LD and potential drivers for the technology.
- **Technology adoption** – If they have a technology roadmap within the organisation and if so the short term and medium term objectives. Information on the rollout of technical solutions within the organisation such as key stakeholders and the primary drivers and barriers.

The interviews provided insights into the challenges that enterprises are facing with data and information management today, which fed into the formulation of more than 60 user stories, describing demand for efficient data and information management solutions from a user perspective.

5.1.2. Workshop

The PROPEL stakeholder workshop, which took place on the 10th of May 2016, consisted of a good mix of participants who come from the private sector (solution providers and users), and the research sector (technicians and strategic people). Linked Data was not a new topic for most of the participants; with the majority having previously heard about it or having worked in the field. The first part of the workshop was dedicated to introducing the PROPEL project, providing an overview of LD and its potential in an enterprise setting, and highlighting some of the open technical challenges. The second part of the workshop was more discussion oriented with facilitators eliciting information on the business drivers and barriers, and the technological challenges and opportunities.

- **Business barriers and drivers** - The general consensus of the group was that the technology itself is not an inhibitor, as it is ready to use. The main barriers related to non functional requirements, such as data quality and trust of the sources, privacy, licensing of the content, clarity in the language and semantics, missing skills and experts who can handle the new technology, general awareness of the technology, and how to convince stakeholders to adopt technologies and/or publish data as Linked Data. While, the primary drivers for the adoption of new technologies in general and LD in particular are the pressure to optimise and to increase efficiency in the organisation. Also legal regulations, standards, and policies are key drivers in data management and exchange.
- **Technological challenges and opportunities** - The technical discussion primarily focused on the issues that could prevent the use of Linked Data in enterprises. The group identified initial challenges in the quality of the data (when integrating data from different sources), security (combining open and closed data) and usability (analyse and visualisation of emerging big semantic data). However, the group recognised the fact that these are common challenges when dealing with big data (variety, volume, velocity) and not a bottleneck when it comes to the use of Linked Data technologies in general. When it comes to Linked Data in particular, two main obstacles were then identified. On the one hand, although some of the potential "customers" are open-minded and could apply Linked Data solutions, there is a lack of education in this paradigm, hence it is hard to find strong arguments (aims, goals, ground-breaking solutions) to carry out a shift from existing technologies and processes. On the other hand, the group acknowledged that Linked Data could help businesses to performed better (more informed) analysis, improve business intelligent and make better decisions (e.g. to improve efficiency).

During the workshop a number of use cases were identified that clearly show the need for optimised data management solutions. For example, in the field of education, product management, engineering to discover failures and mistakes, and advertisement. An important issue is to focus both on the content and on the people. The starting point must be a clear business need including people with specific requirements. Then the technology can provide good solutions. However, workshop participants also agreed that most of the Austrian enterprises are too small, or they haven't reached the technological maturity level, to face the problems addressed by Linked Data technology. In this respect, participants pointed to potential early adopters in the public sector (stating that they are the biggest "company" in Austria), energy sector, public transport (they use and provide open data), and other sectors interested in know-your-customer solutions, such as banking, financial sector and insurance.

5.2. User Stories

Currently, organisations are facing a variety of challenges in data and information management. In order to deal with the current market forces, organisations need to dramatically modernise their IT systems, transform outdated data management infrastructure and replace it with a state of the art information environment able to support knowledge workers in their data to day activities. Based on the interviews and the workshop more than 60 user stories based on current data challenges within enterprises across several different industries were identified. User stories are used to specify the market needs from the perspective of the user. In this section we present a summary of the most common user stories based on our research. The complete list of users stories can be found in *Annex C*.

5.2.1. Customisation and customer relationship management

Optimised customer relationship management is of high demand across multiple industries. Organisations want to increase their revenue by matching available resources with user behaviour and interests, and by addressing customers with personalised offers and recommendations based on customer's interests and behaviour.

Customer profiles A multitude of web data can be sourced and matched with user profiles. These user profiles can be sold to product and service vendors.	*„As a provider of payment solutions I want to automatically analyse customer data such as who buys what, when, where, how often, for how much, etc. so that I can develop and sell customer profiles to companies."*
Dynamic topic pages Adapt websites content based on the user's interest.	*„As a media house I want to display personalised content as precise as possible so that my readers stay as longer on my website."*

Product configuration Customers are provided with highly personalised offers based on a set of information and key data.	*„As an insurance company I want to offer my customers highly customized products so that I can improve customer satisfaction and gain competitive advantages in the marketplace."*
Recommendations Provide recommendations for products based on customer profiles.	*„As an online marketplace provider I want to analyse user data (who buys what, when, where, for how much, etc.) so that I can recommend each user individual products and services that - based on their profile – interests them most."*
Communication Better manage communication with customers across different channels.	*„As a sales representative I want to find out about customer-related prior communication across various channels (E-mail, CRM) so that I can provide a better customer experience; make an informed offer, cross/ up-sell etc."*

5.2.2. Automatic and dynamic content production, adaptation and display

Organisations need to increase content production efficiency and measure their performance. This includes the reduction of manual work on content to a minimum, to avoid duplication, and to support the reuse of existing content based on clear templates and automated processes.

Change management Include changes automatically in different documents and minimise manual work.	*„As a real estate broker I display my rental objects (pictures, description, address) at different real estate platforms, but I would like to maintain changes centrally so that I do not have to update the content in different places."*

Automatic content variations Media content is displayed at different channels in different formats. Transforming content based on preconfigured templates into different variations is cost-saving and raises the publishing frequency.	*„As a media house I want to recycle my content frequently and automatically generate different variations of the content so that content production costs decrease and content efficiency increases.“*
Alignment of Key Performance Indicators (KPIs) Organisations track and compare performance through an extensive set of KPI's (especially the media sector). Established industry standards are often missing and KPI's vary very much from each other and prevent significant industry benchmarks, which is why alignment of KPIs becomes necessary.	*„As a media controller I want to relate changing KPIs to each other that refer to similar actions so that I can compare media consumption in the long-term.“*

5.2.3. Data search, information retrieval, and knowledge discovery

Organisations need to make better use of their internal resources by improving search, information retrieval and knowledge discovery. This includes providing an overview of existing products and offers, and the ability to easily find relevant content and expertise.

Content discovery and reuse Knowledge intense organisations such as consultancy firms or pharmaceutical companies produce a lot of papers and presentations. They often include information that could be reused, if employees would find it.	*„As a market analyst I want to link existing reports on specific technologies that have been published by my organisation based on their attributes/metadata so that I can find existing reports that have been published by my company.“*

Product management As products become more customised it becomes more challenging to maintain an overview of the wide range of existing products within a company. Efficient product management links similar products to each other to gain an overview of the broad product landscape of an organisation.	*„As an insurance company I want to link attributes of product data (client name, insurance price, insurance type, etc.) so that I can keep an overview of the wide range of available products."*
Skill finding Different sources of CVs and professional networks can be merged in a manner that makes it possible to find individuals with specific skills and expertise within a large organisation.	*„As a data scientist I want to identify expertise within a large organisation and be able to pinpoint the relevant experts so that I can identify top trends within the organisation and expertise for the organisation as a whole."*

5.2.4. Data and information exchange and integration

Key elements for the efficiency and success of an organisation are the ability to break down silos, to foster data and information exchange and integration, to be able to eliminate duplication, and to perform holistic analysis and sound fact-based decision-making based on internal and external data.

Enrichment of a data pool with external data Integrating internal data pools with external data (e.g. open data, weather data, Wikipedia, public statistics, newspaper articles, etc.).	*„As a real estate platform provider I want to offer additional geographic, statistic and qualitative data about an apartment's neighbourhood so that the people who use my platform to search for apartments can get a holistic picture of the apartments surrounding and can make a decision if the location fulfils their requirements."*

Internal knowledge portals Integrating data that is available within a knowledge intensive enterprise and making it accessible via one portal.	*„As a business user I want to be able to access all relevant information (ERP, CRM, e-mails, cloud services etc.) within my organisation from a single point to do my job more efficiently and effectively.“*
Data and information exchange between different stakeholders Improving data and information exchange not just within the organisation but with stakeholders outside of the organisation.	*„As a national postal service worker I want to exchange information with my international counterparts so that we can work together on innovating our industry.“*
Data consistency and elimination of duplication Harmonizing internal and external data so that data from different sources can be compared and used together for holistic analysis.	*„As a logistics operator I want to align internal and external delivery IDs so that distribution processes are more transparent in the case of inquiries.“*
Data integration and quality analysis Enriching existing data with external data so that contradictions can be uncovered, quality can be improved and prior insights can be leveraged.	*"As a healthcare worker I wan to integrate disparate medical databases that are hard to integrate, widespread, contain the same data that contradicts each other. So that I can gain insights from other clinical trials."*

5.3. Linking user stories with technology foundations

In this section we discuss how the user stories are connected to the technology foundations that are relevant for the Linked Data concept. In other words, we established a link between the user's needs and technological solutions. In order to exemplify the process, we choose four promising uses cases concerning business process management, human resources, media and publishing and healthcare and pharma.

5.3.1. Business Processes

Data and information exchange among different stakeholders, either departments within an organisation or exchange between different organisations, improves efficiency, effectiveness and flexibility. *Figure 5.1* depicts the mapping between one of the business process use cases derived from our analysis and the technical foundations that could be used to realise it. Although data management, data integration, and system engineering are core to this use case, considering the need for accuracy and the sensitivity of production and logistical data, foundational technologies such as data quality and provenance, security and privacy, are also of high priority. Others such as extraction techniques, knowledge representation, robustness and searching, browsing etc. are highly desirable. While, the ability to deal with streaming data, ontology/thesaurus/taxonomy management and HCI and visualisations are classified as nice to have.

*"**I would like to** be able to exchange information and coordinate production and logistics with suppliers and customers ..."*

*"... **so that I can** improve efficiency, effectiveness and flexibility of my inventory management and operations"*

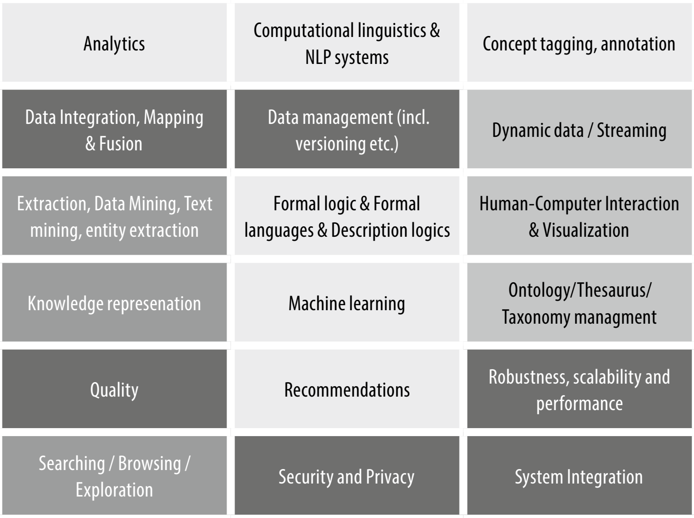


Figure 5.1: Business Processes – (legend: ☐ = not necessary; ☐ = low priority; ☐ = medium priority; ■ = high priority).

5.3.2. Human Resources

Knowledge-intensive organisations benefit from semantic knowledge discovery portals that for example help them to identify expert skills within a large organisation. *Figure 5.2* depicts the mapping between one of the human resources use cases and the technical foundations that could be used to realise it. In such a scenario data management, data integration, and system engineering are all high priority, however so too are extraction techniques, ontology/thesaurus/taxonomy management, computational linguistics and NLP, quality, searching and browsing, and security and privacy. While topics such as robustness and HCI and visualisations are seen as highly desirable and nice to have respectively.

*"**I would like to** identify expertise within a large organisation and be able to pinpoint the relevant experts..."*

*"... **so that I can** identify top trends within the organisation and expertise for the organisation as a whole"*

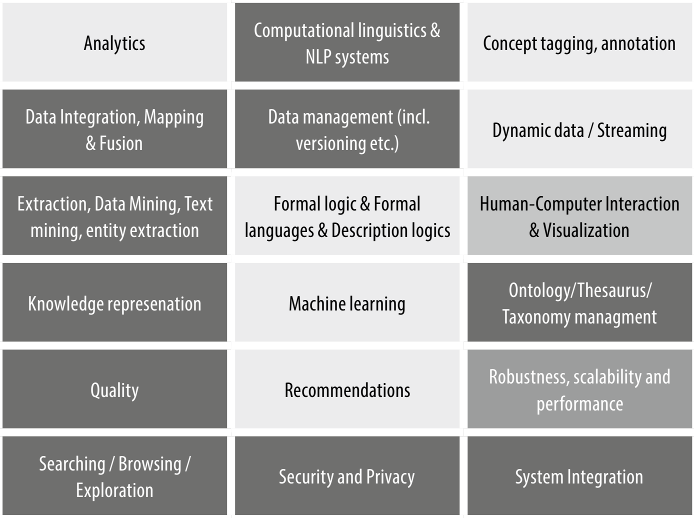


Figure 5.2: Human Resources – Example 3 – (legend: ☐ = not necessary; ☐ = low priority; ☐ = medium priority; ☐ = high priority).

5.3.3. Media and Publishing

One of the strong points of Linked Data technologies is their ability to represent and recognise information entities that belong together or refer to the same concept. Enriched through various kinds of automation workflows, this data can be used to deliver a personalised experience to the user. *Figure 5.3* presents one of the media and publishing use cases and the technical foundations that could be used to realise it. Considering the need to deal with both structured and unstructured content, that may be of a sensitive nature at scale, all foundational technologies fall into either the high or medium priority categories.

*"**I would like to** display personalized content as precise as possible ..."*

*"... **so that** my readers stay as long as possible on my website."*

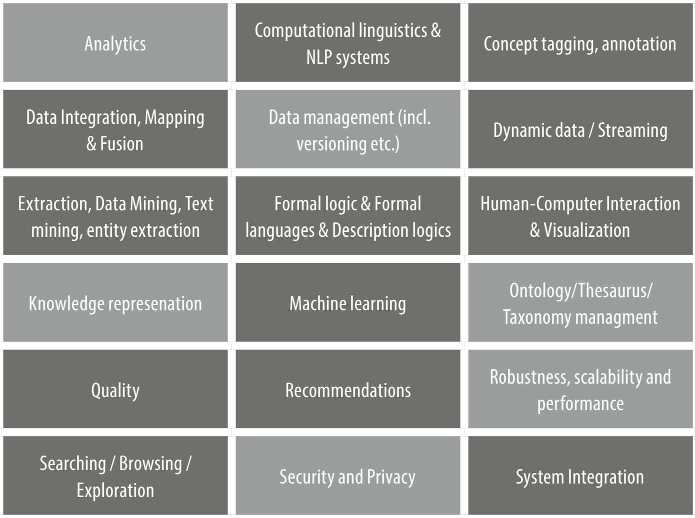


Figure 5.3: Media and Publishing – (legend: ☐ = not necessary; ☐ = low priority; ☐ = medium priority; ■ = high priority).

5.3.4. Healthcare and Pharma

Considering that the healthcare and pharmaceutical sector is predisposed to representing knowledge in the form of dictionaries and taxonomies and the sector is facing major challenges with respect to data integration, it is no wonder that the industry has already started to adopt Linked Data technologies. *Figure 5.4* depicts the mapping between one of the healthcare and pharma use cases that was derived from the interviews and the Linked Data technical foundations. Considering the sheer volume of data that needs to be analysed, coupled with the need for strong security and privacy mechanisms and the dependences of the industry on quality it is not surprising that almost all of the foundations are high priority. Data management and dynamic data and streaming were ranked as medium priority as much of the analysis tends to be analytical in nature and batch oriented.

*"**I would like to** integrate disparate systems that are:*
- *Hard to integrate*
- *Widespread*
- *Contain the same data that contradicts each other"*

*"... **so that I can** gain insights from other clinical trials"*

Analytics	Computational linguistics & NLP systems	Concept tagging, annotation
Data Integration, Mapping & Fusion	Data management (incl. versioning etc.)	Dynamic data / Streaming
Extraction, Data Mining, Text mining, entity extraction	Formal logic & Formal languages & Description logics	Human-Computer Interaction & Visualization
Knowledge represenation	Machine learning	Ontology/Thesaurus/ Taxonomy managment
Quality	Recommendations	Robustness, scalability and performance
Searching / Browsing / Exploration	Security and Privacy	System Integration

Figure 5.4: Healthcare and Pharma – (legend: ☐ = not necessary; ☐ = low priority; ☐ = medium priority; ☐ = high priority).

6. TECHNOLOGICAL STATUS

For many years Semantic Web technologies have been an area of intense research in the academic community. Over the years there has been a number of papers that explore research topics and trends within the Semantic Web community (Berners-Lee et al., 2001, Feigenbaum et al., 2009, Bernstein et al., 2016). In this section, we analyse the academic literature that emerged from the five most popular international publishing venues for Linked Data researchers: the International Semantic Web Conference (ISWC), the Extended Semantic Web Conference (ESWC), the SEMANTiCS conferences, the Semantic Web Journal (SWJ) and the Journal of Web Semantics (JWS). Our analysis, which is limited in scope to the last 10 years (i.e. from 2006 to 2015), examines research trends based on topics:

- identified in the tracks, call for papers, workshops, tutorials and sessions of the top 5 venues; and
- extracted from three seminal papers “The Semantic Web” (Berners-Lee et al., 2001), “The Semantic Web in action” (Feigenbaum et al., 2009), and “A new look at the Semantic Web” (Bernstein et al., 2016).

6.1. Topic and trend analysis

Prior to performing the analysis, we manually extracted the metadata from the ISWC, ESWC and the SEMANTiCS conference and the SWJ and the JWS journal websites from 2006 to 2015 inclusive. Where the data was not available we consulted the Internet Archive Wayback Machine[50]. Once all of the data was gathered we created two separate datasets one containing *venue metadata* (i.e. call for paper line items and the track, session and invited talk titles) and another containing information pertaining to *conference sponsors* (i.e. organisation, website). At the same time, we gathered the full text of the papers from conference proceedings and journals, and we retrieved the paper abstracts and the associated metadata from SCOPUS[51].

6.1.1. Analysis based on venue metadata

Although there are several Natural Language Processing (NLP) tools that can be used to automatically extract topics from text, considering that accuracy is a key factor of this study and the fact that NLP tools often suffer from poor precision and recall, we

50 Internet Archive Wayback Machine, https://archive.org/web/
51 Scopus, https://www.scopus.com/

elected to manually generate and curate a list of topics, based on the venue metadata, that are clearly relevant for the community.

We started by merging the text from the titles of the invited talks, the line items from the call for papers, the various track names and the session titles, per year per venue. We subsequently went through each line item and noted all of the topics (words and phrases that are relevant for the community) that appeared. It is worth noting that more often than not there were multiple topics per line item. Following on from this we merged all of the topics into a single topic list and removed the duplicates, resulting in a dictionary containing 3421 unique topics. We subsequently used the Semantic Web Company's Poolparty platform[52] to: (i) manually create a coarse grained taxonomy from the dictionary terms; (ii) to construct corpora of academic papers per conference per year; and (iii) to search the corpora for the terms appearing in the taxonomy. In the subsections that follow we examine the conference and journal paper analysis in terms of the high level metadata categories, discuss the coverage of the foundational technologies and finally examine industrial involvement.

6.1.1.1. High Level Metadata Categories

Although many of the topics could be associated with the foundational technologies presented in *Section 2.4.7*, we noticed that there were a significant number of topics that did not fit into the technological foundations that we had identified. Generally speaking, topics could be grouped into the following seven high level metadata categories:

- Applied (e.g. applications, tools, terms relating to usage)
- Business (e.g. profit, business value, impact)
- Domain (e.g. healthcare, media, education)
- Evaluation/metrics/studies (e.g. measure, ranking, negative results)
- Foundation (e.g. knowledge representation, data management, security and privacy)
- Method (e.g. finding, planning, transformation)
- Standards (e.g. SPARQL, standardisation, W3C)

Figure 6.1 provides a summary of the % coverage for each of the aforementioned categories based on the sum of the number of occurrences in all documents for each of

52 PoolParty, https://www.poolparty.biz/.

the topics associated with a given category[53], across the 5 venues for the 10 year time frame under examination.

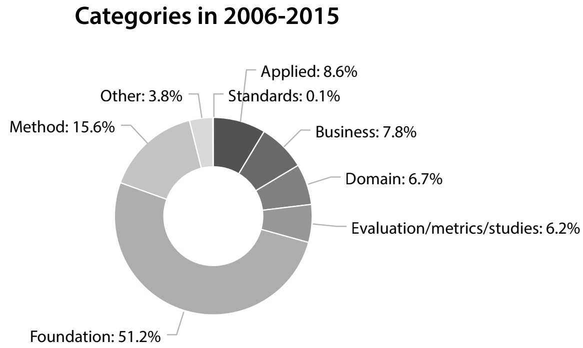


Figure 6.1 % coverage per category across the 5 venues for the 10 year timeframe

Considering the technical nature of the venues under examination it is not surprising that the *technological foundations* represent over half (52%) of the topics that are of interest to the community. The second largest category relates to *methods*, which again is not surprising for academic papers. It was interesting to see that there is very little deviation between the business (7%), domain (6%), applied (8%), and evaluation (6%) terms. Although standards are not well represented in terms of coverage, this can be attributed to the fact that there was very little reference to standards in the conference and journal metadata that was used to search the corpora.

6.1.1.2. Foundational Technologies

IIn order to get an indication of the coverage of the technical foundations identified in *Section 2.4.7*, across the five venue under analysis, we aggregated the number of occurrences for each of the topics associated with a given foundation. *Figure 6.2* provides details on the % coverage for each of the eighteen foundations, across the 5 venues for the 10-year timeframe under examination.

53 Note that one topic can belong to several categories.

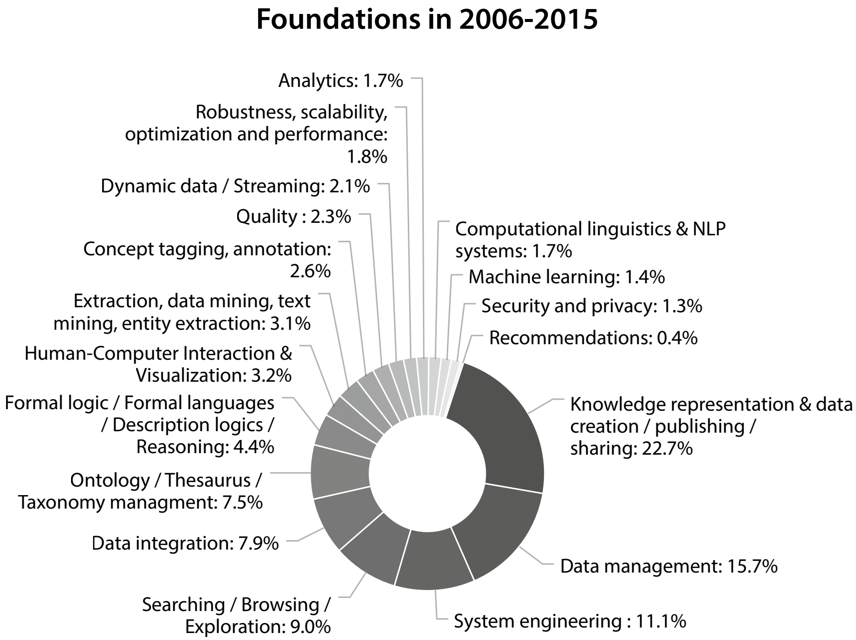


Figure 6.2 % coverage per foundational technology across the 5 venues for the 10-year timeframe

As expected, *knowledge representation and data creation/publishing/sharing* is the top foundation, with almost 23% of the total occurrences in all documents. Note that this foundation includes several topics that are fundamental to the Semantic Web community, i.e. the ability to represent semantic data and to publish and share such data. Next in order of importance, the management of such knowledge (*data management*) and the construction of feasible systems (*system engineering*), constitute almost 16% and 11% of the occurrences, respectively. Important functional areas such as *searching/browsing/exploration*, *data integration* and *ontology/thesaurus/taxonomy* management also figure strongly in comparison to the other foundations (all of them with more than 7.5% occurrences). In contrast, very specific topics of the area, such as *logic* and *reasoning* and *concept tagging* and *annotation* represent a modest 4.4% and 2.6% respectively, and cross-topics, such as *human computer interaction*, *machine learning*, *computational linguistics and NLP*, *security and privacy*, *recommendations* and *analytics* are only marginally represented. It is also worth noting that topics that relate to *quality, dynamic data and scalability* are also under-represented (at around 2%).

In order to gain some insights into the research trends within the Semantic Web Community over the last decade, Figure 6.3 depicts the growth/decline of each of the foundations over the 10-year timeframe.

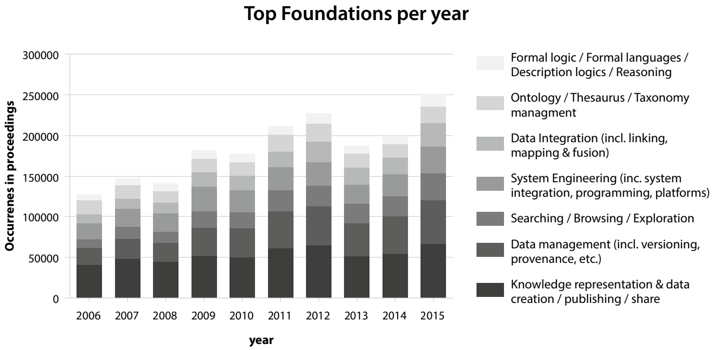


Figure 6.3 Growth/Decline of the foundation technology across the 5 venues for the 10 year timeframe

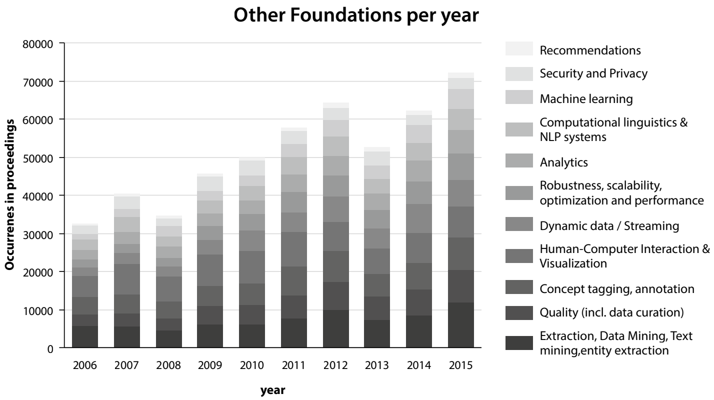
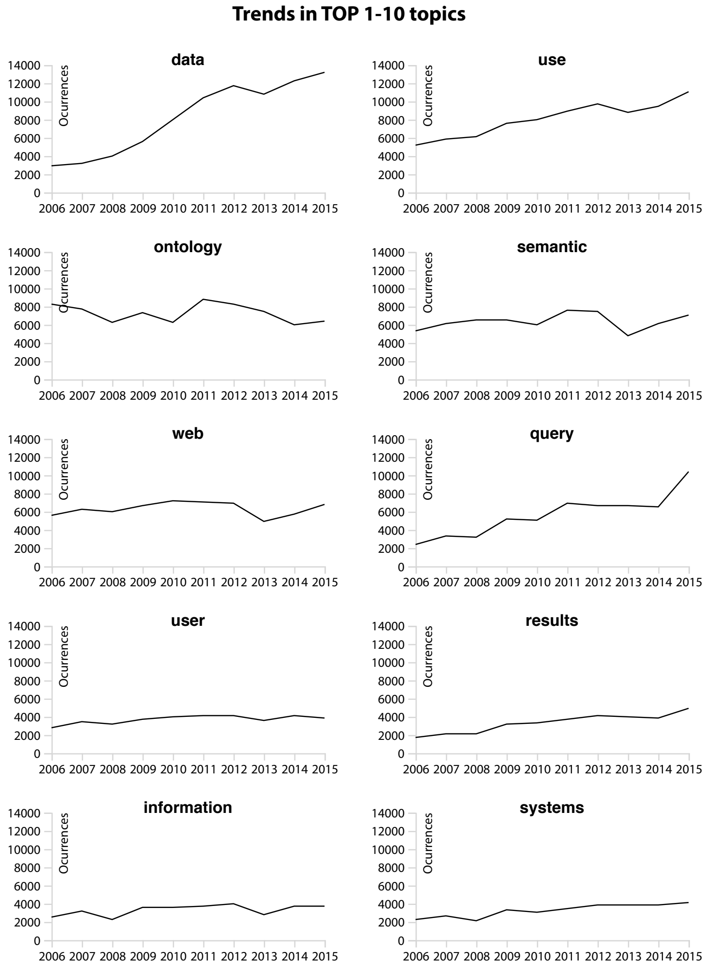


Figure 6.3 Growth/Decline of the foundation technology across the 5 venues for the 10 year timeframe

Although the general trend for all topics shows year on year increases, it is worth mentioning that *robustness, scalability, optimisation and performance, dynamic data/streaming, searching/browsing/exploration* and *machine learning* have increased by more than 200% since 2005. While in contrast, *security and privacy* and *ontology/thesaurus/taxonomy management* have had marginal growth of only 30% for the same period.

Figure 6.4 depicts the growth/decline of the top 20 topics in the papers. As expected, most of the terms reflect general topics (such as semantic, user, results, information, knowledge, etc.) and show a stable behaviour, with peaks and valleys. In contrast, current hot topics such as *data*, *use* and *link* show a noticeable increase in the last ten years, whereas *services* had a major decline from 2009-2013, however in the last two years it seems to be starting to increase in interest.



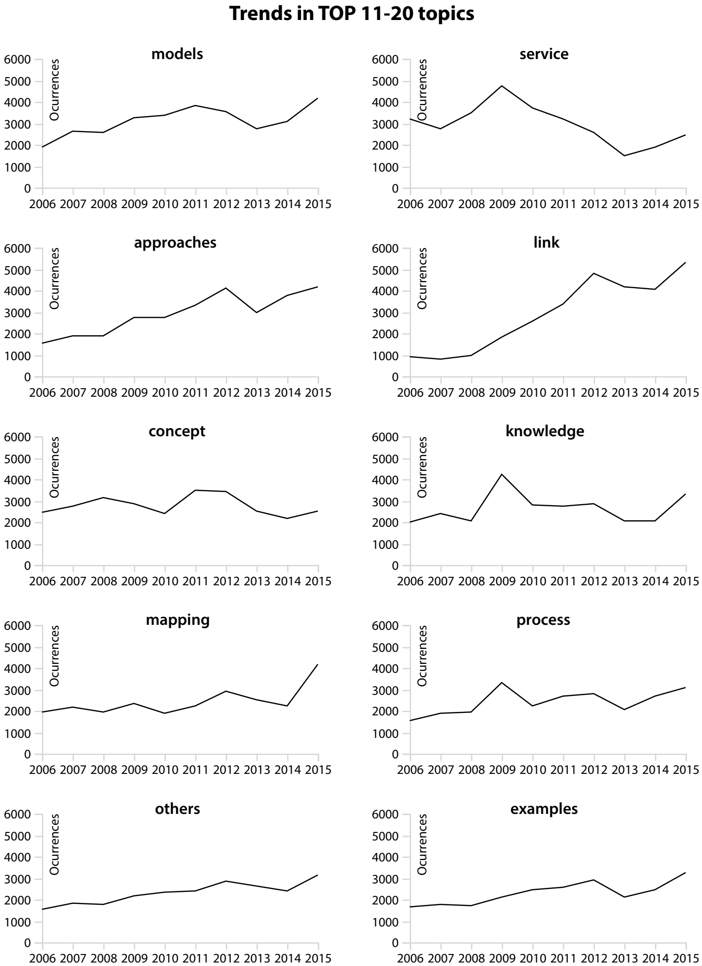


Figure 6.4 Growth/decline of the top 20 topics across the 5 venues for the 10 year timeframe

In turn, *Figure 6.5* focuses on the growth/decline of *the top 10 multiword topics* (those topics composed of more than one word). Interestingly, results show a sharp increase of *linked data* at the expense of *semantic web*. Note also that *natural language* is the top-9 multiword topics, even though this is a cross topic which may be more represented in a different community. Finally, the decrease in the occurrence of web services can also be seen here.

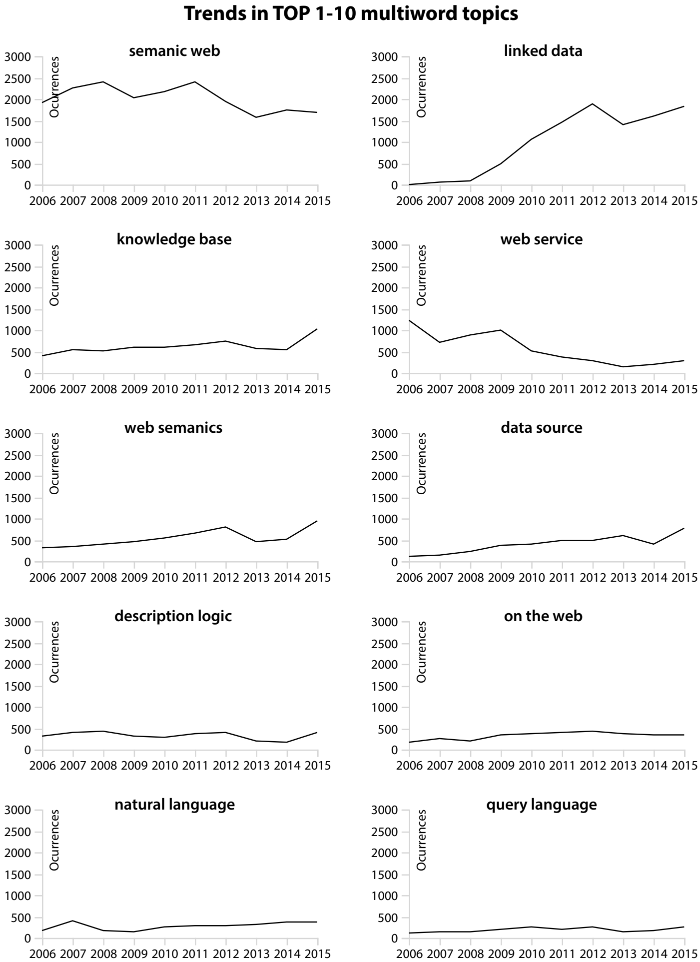


Figure 6.5 Growth/Decline of the top 10 multiword- topics across the 5 venues for the 10 year timeframe

6.1.1.3. Academic vs. Industry topics

To determine the topics that are more relevant from an industry perspective, we performed additional analyses over the metadata extracted from SCOPUS. More specifically, we examined the author affiliations of the papers, classified them as either academic or industry, and tagged each paper that has at least one author with an industry based affiliation as industry and all other papers as academic. *Figure 6.6* provides a snapshot of the topics with the greatest statistical difference in occurrences between purely academic papers and those with an industry partner.

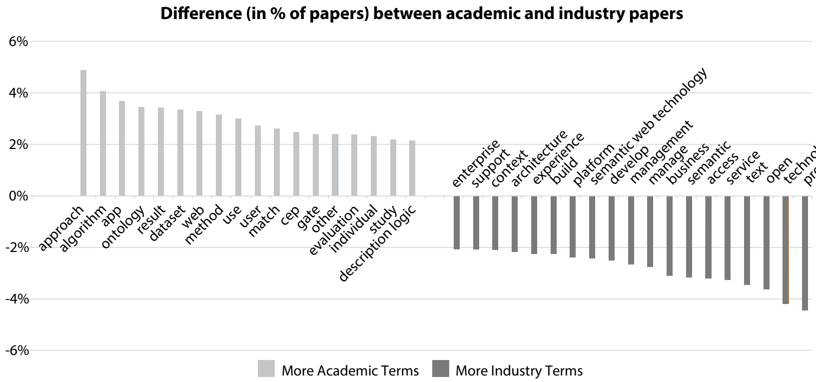


Figure 6.6 Differences in Academic versus Industry topics

As expected, some topics, such as *evaluation*, *algorithm*, *result* or *study* have a traditional academic focus and it is not surprising that they occur more in papers with academic affiliations. In turn, it is not surprising that topics such as *develop*, *technology*, *Semantic Web technologies*, *service*, *build*, *project*, *platform*, *architecture*, *experience*, *business* or *enterprise* are more often associated with industry affiliations. All the same there are some interesting insights with terms such as *open* and *access*, which are unexpectedly more represented in industry. While, *use*, *APP*, *user* and *ontology* remain slightly more important in academia.

6.1.2. Analysis based on seminal papers

In addition to the exploratory analysis, we also performed a targeted analysis based on three seminal papers from the Semantic Web community. In this instance key terms were manually extracted from the text and as before a coarse grained taxonomy was constructed based on the seven high level metadata categories presented in the previous section. In the following subsections we look for evidence of the theses put forth in the three seminal papers "The Semantic Web" (Berners-Lee et al., 2001), "The Semantic Web in action" (Feigenbaum et al., 2009), and "A new look at the Semantic Web" (Bernstein et al., 2016) that formed the basis of our analysis.

6.1.2.1. The Semantic Web (Berners-Lee et al., 2001)

In "The Semantic Web" (Berners-Lee et al., 2001), the authors use a fictitious scenario to describe their vision of the Semantic Web. They subsequently provide an overview of the Semantic Web infrastructure necessary to realise the vision and for the Semantic Web to reach its full potential, focusing on four broad areas of research, namely: *expressing meaning, knowledge representation, ontologies* and *agents*.

The authors identify the need for "*a language that expresses both data and **rules** for **reasoning** about the data*". They continue that there is a need to add "***logic** to the Web—the means to use **rules** to make **inferences**, choose courses of action and answer questions*". Additionally, given that databases may use different identifiers to refer to the same thing there is a need for "*collections of information called **ontologies***" that can be used to express meaning.

Figure 6.7 depicts the evolution of knowledge representation terms such as *rules, reasoning, inference, logic*, and *ontologies* over the last ten years. Although the occurrence of *rules, reasoning, inference* and *logic* are much less than that of *ontologies*, overall the presence of the topic remains relatively stable.

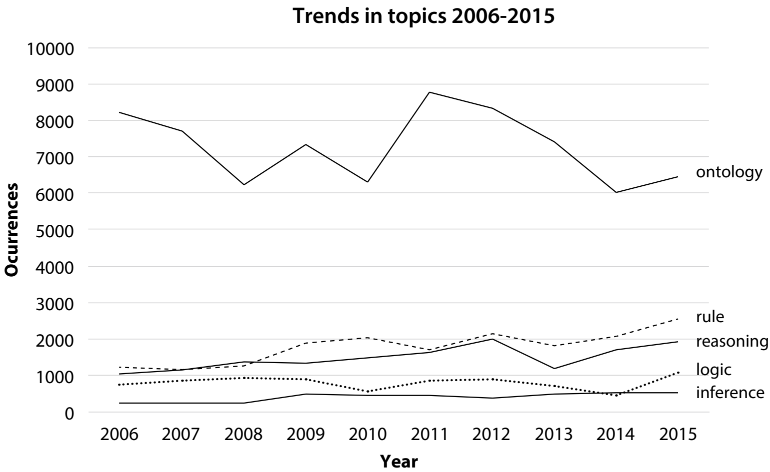


Figure 6.7 Growth/Decline of knowledge representation terms

The paper conjures that the "*real power of the Semantic Web*" will be realised in the form of **software agents** and the "*effectiveness of such **software agents** will increase exponentially as more **machine-readable** Web content and **automated services** (including other agents) become available*".

Figure 6.8 compares the growth and decline of topics such as agents, *automated reasoning* and *intelligent user interfaces* in comparison to topics such as *ontology*, *web-service* and *machine-readable*. Although the occurrence of agents is much less than that of *ontologies* after a dip in 2013 the data suggests that the *agent* topic is on the rise. In turn, the presence of *automated reasoning* remains stable, whereas *intelligent user interfaces* have not attracted attention in the community.

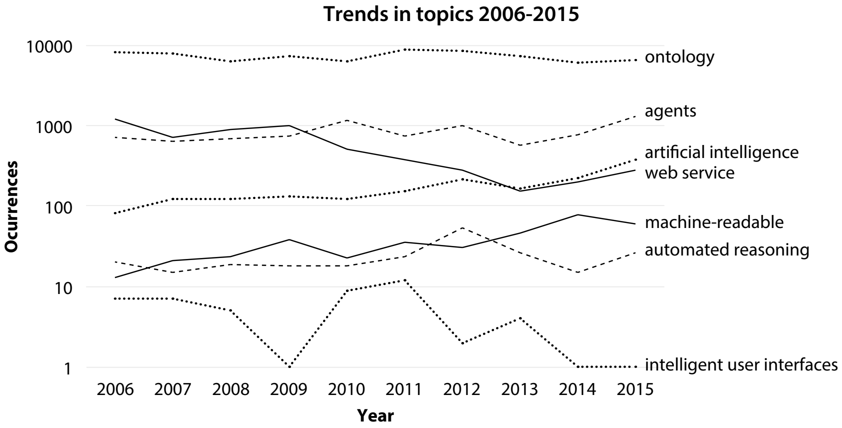


Figure 6.8 Growth/Decline of agent terms

6.1.2.2. The Semantic Web in action (Feigenbaum et al., 2009)

The second seminal paper entitled "*The Semantic Web in action*" (Feigenbaum et al., 2009), refers to the original paper by Berners-Lee et al., 2001, and highlights that the enabling technologies have come of age. The authors primarily focus on the adoption of the technologies in an enterprise setting, stating *that a number of large "**companies** have major projects underway that will greatly improve the **efficiencies** of **in-house operations** and of scientific research" and that "companies are improving the **back-end** operations of **consumer** services*" with the aid of semantic technologies.

Figure 6.9 presents the growth/decline of topics such as *companies*, *consumer, back-end, in-house* and *efficiency*. The number of occurrences of terms such as *companies* and *operations* appear to vary considerably year on year, although *efficiency* does not appear as frequently as one might expect there is a clear upwards trend, however terms such as *back-end* and *in-house* are rarely present in the analysed papers.

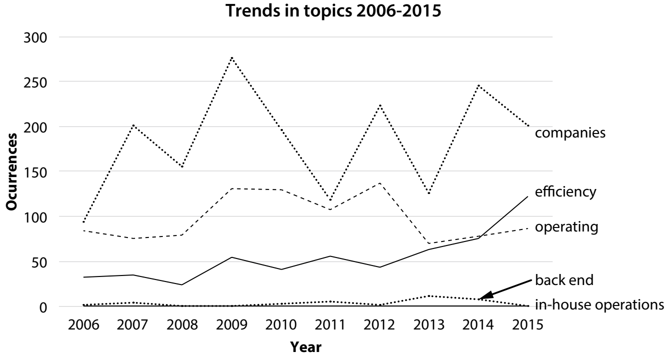


Figure 6.9 Growth/Decline of enterprise terms

The authors state that "*some of the most advanced progress is taking place in the **life sciences** and **health care** fields*" and present two case studies one on **drug discovery** and another on **healthcare**.

Rather than simply focusing on the domains listed by Feigenbaum et al., 2009, *Figure 6.10* presents the top 15 domains based on the number of occurrences of topics that can be directly associated with the domain. Although *Media* takes the number one spot, it is worth noting that the counts are based purely on lookups and context was not taken into account, as such anything that relates to publishing would be categorised as Media. Albeit *Bioinformatics* and *Health* appear to have much less occurrences this might simply be because researchers focusing on these topics tend to publish in other venues (e.g. the journal of biomedical semantics).

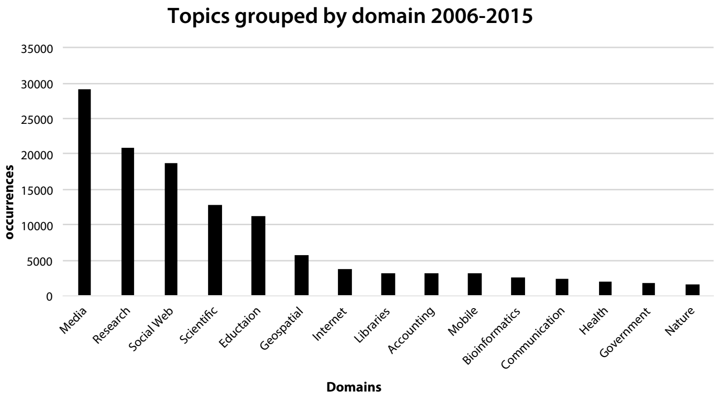


Figure 6.10 Topics grouped per domain

The authors also make reference to the application of Semantic Web technologies stating that "*like an iceberg, the tip of this large body of work is emerging in direct consumer* ***applications***".

Figure 6.11 presents the occurrences of topics that relate to users (i.e. *end users, usability*) and applications (i.e. *application, apps, systems, toolkits*). Although both the *user* and *application* topics remain relatively static, generally speaking the application topics occur much more frequently.

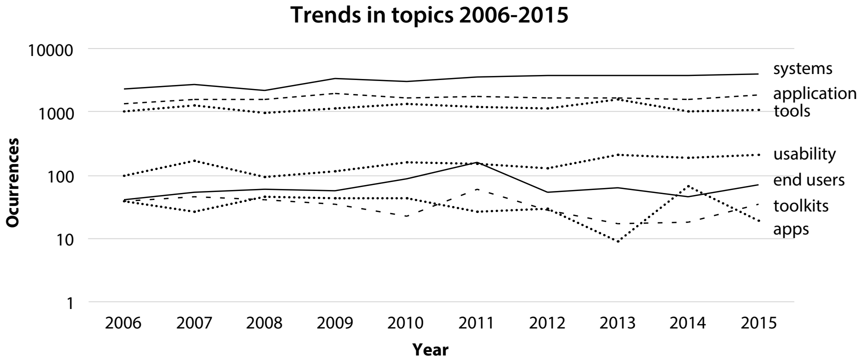


Figure 6.11 Growth/Decline of application versus end users

6.1.2.3. A new look at the Semantic Web (Bernstein et al., 2016)

The most recent paper "A new look at the Semantic Web" (Bernstein et al., 2016), examines past, present and future research that focuses on Semantic Web technologies. The authors claim that *„early research has transitioned into these* ***larger, more applied*** *systems, today's Semantic Web research is changing: It builds on the earlier foundations but it has generated a more diverse set of pursuits".* They highlight the fact that *„the representations that they used became* ***less formal*** *and* ***precise*** *than many early Semantic Web researchers had envisioned"* and *„the semantics, in a sense, becomes more "**shallow**," it could be more widely applicable".*

Figure 6.12 depicts the growth/decline of the aforementioned topics based on our analysis of the corpora. Although there is evidence to suggest that topics such as l*arge/big data* and *applied* are gaining traction within the community, we could not find evidence in the text to support the thesis that the representations are *less formal, less precise* and *lightweight*.

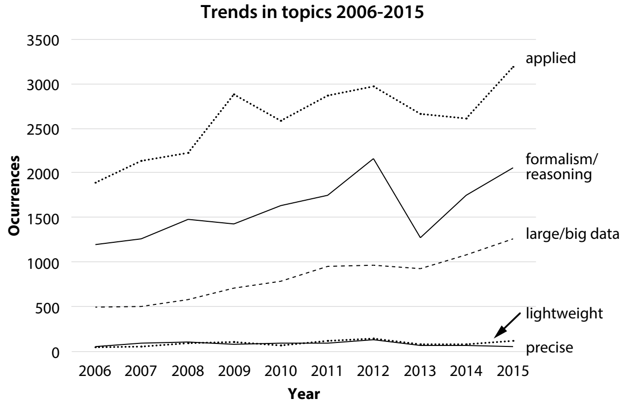


Figure 6.12. Growth/Decline of transition topics

Bernstein et al. further discuss what the future may hold and speculate as to future research growth areas. The authors hypothesise that *"the objective of the next decade of Semantic Web research is to make this vast **heterogeneous multilingual** data provide the fuel for truly **intelligent applications**"* that *"relies less on **logic-based** approaches and more on **evidence-based** ones."*

Figure 6.13 provides some historical data on said topics. Aside from *intelligent applications*, despite some dips in 2012-2013 the general tendency of the other topics is upwards.

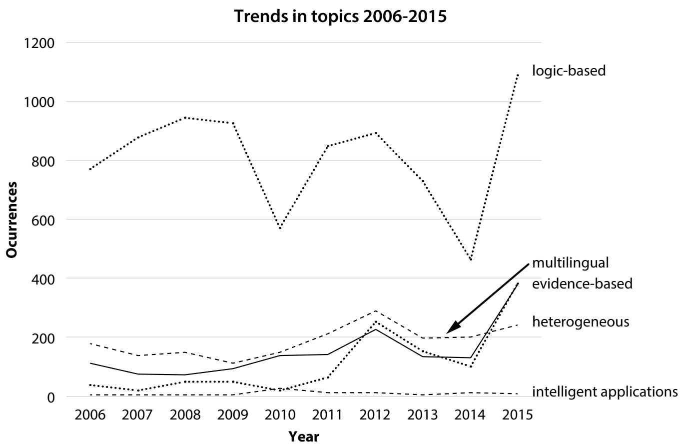


Figure 6.13 Growth/Decline of topics for the next decade

In addition to examining the status quo Bernstein et al. postulate as to the research challenges that the community will need to tackle over the next 10 years. The authors put forward a number of research questions concerning: *representation and lightweight semantics*; data heterogeneity, quality, and provenance; latent semantics; and high volume and velocity data.

6.1.2.4. Representation and lightweight semantics

The authors present the following questions concerning representation issues and lightweight semantics:

- *"How do we leverage these **diverse** representations?"*
- *"How do we coordinate the **diverse** components of structured knowledge that are defined by various parties and that must interact in order to achieve increasingly intelligent behaviour?"*
- *"How do we define **lightweight**, **needs-based**, "**pay-as-you-go**" approaches for describing knowledge?"*
- *"What are the languages and architectures that will provide this knowledge to the increasingly **mobile** and **application-based** Web?"*

Figure 6.14 depicts the trend of the most relevant terms that appear in these open questions. Results show that topics such as *lightweight* and *diverse* seem to be increasing. While *mobile* topics increased steeply up until 2012, the topic appears to be less popular in recent years. Finally topics such as *application scenarios, needs-based* and *pay-as-you-go* do not figure much or at all in the analysed corpora.

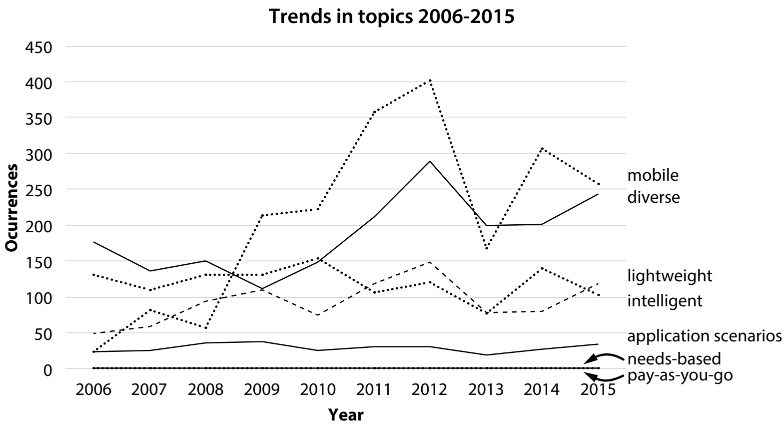


Figure 6.14 Growth/Decline of representation and lightweight semantics

6.1.2.5. Data heterogeneity, quality, and provenance

The authors also raise a number of questions with respect to data heterogeneity, quality and provenance:

- "How do we integrate **heterogeneous** data and particularly how can we understand which data can be integrated to what degree?"
- "How can we represent and assess **quality** and **provenance** of the data?"
- "How do we evaluate whether the **quality** of a particular source is sufficient for a given task?"

As trust is a term that is often used in connection with provenance and quality, in *Figure 6.15 we examine all three terms.* What we notice is a general upward trend with respect to *quality*, a downward trend in relation to *trust* and peaks and valleys concerning *provenance*.

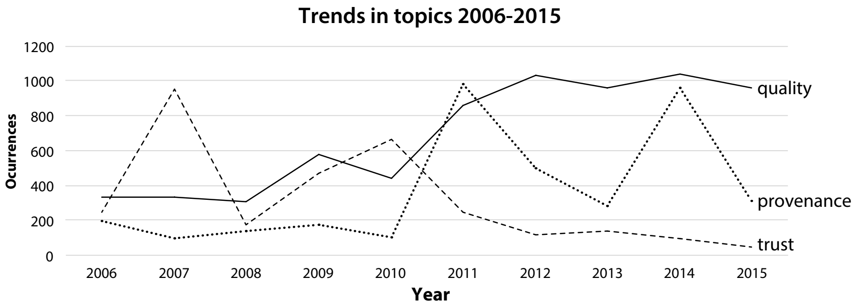


Figure 6.15. Growth/Decline of data heterogeneity, quality, and provenance

6.1.2.6. Latent semantic

The authors also identify the need for tools and techniques that can be used to represent implicit knowledge. When it comes to latent semantic several questions remain unanswered:

- "How much of the semantics can we **learn** automatically and what is the quality of the resulting knowledge?"
- "As **ontologies are learned or enhanced** automatically, what is the very meaning of "**formal ontologies**"?"
- "How do we develop some notion of approximate correctness?"
- "Do similar or different **reasoning** mechanisms apply to the ontologies that are extracted in this way?"

- "How do **crowdsourcing** approaches allow us to capture semantics that may be less precise but more reflective of the **collective** wisdom?"

Figure 6.16 shows a major increase in topics such as *crowdsourcing, collective, ontology learning*, while ontology enrichment, ontology reasoning and ontology quality remain under represented.

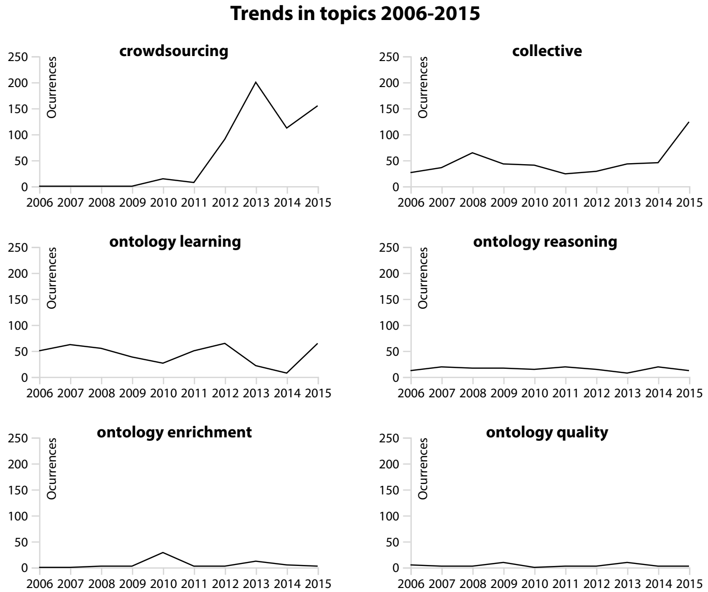


Figure 6.16 Growth/Decline of latent semantics

6.1.2.7. High volume and velocity data

Finally, the authors highlight the need for research to tackle challenges arising from hot topics such as *big data* and the *internet of things*. When it comes to the open research questions raised by the authors, the relation to **data in motion** and **sensors** was explicit, while the need to deal with **big data** was in fact implicit:

- "How do we triage the data in **motion** to determine what to keep and what we may choose, or need, to allow to be lost?"
- "How can our applications integrate constantly changing **sensor** data with fixed data of long duration and high quality semantic provenance?"

Figure 6.17 depicts the number of occurrences of topics that relate to *sensors*, *streaming* and *big data*. In general, both topics are increasing. However, as we can clearly see from the graph, topics such as *velocity* and *large scale* are underrepresented in comparison to topics such as *big/large data, sensors, streams, events, temporal, volume* and *variety*, albeit with some peaks and valleys in more recent years.

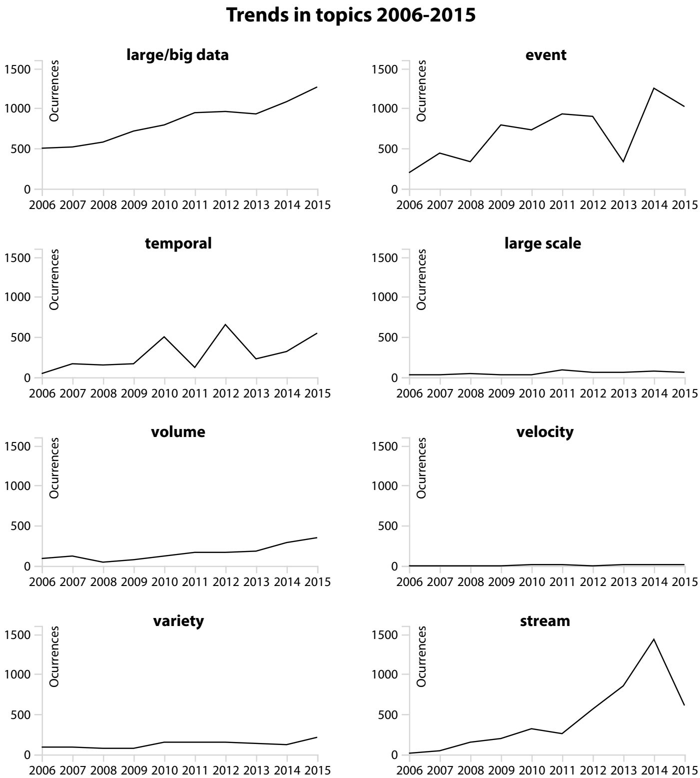


Figure 6.17 Growth/Decline of high volume and velocity data

6.2. Industry relevance / take up

It is clear to see from the conference and journal metadata that Industry is actively involved in the development of the Semantic Web. Grassroots companies like Openlink Virtuoso, the Semantic Web company or Ontotext, to name a few, are current leaders in the Semantic Web arena. However, it is not entirely clear to what extent companies sustainably adopt Linked Data technologies, in which domains, and who are the key supporters of the community. To shed some light on this topic, we performed two complementary studies that focus on (i) the sponsors of the Semantic Web conferences and their involvement; and (ii) the companies outlined in the three aforementioned seminal papers on the Semantic Web.

6.2.1. Sponsors in Semantic Web conferences

Yet again we analysed data from the three main semantic-related conferences, ISWC, ESWC and SEMANTiCS, over the last ten years (2006-2015), however in this instance we focused on the companies as opposed to the topics. We first collected the name of the sponsors from the conference websites. Then, we manually explored the sponsor's website in order to obtain further information about the sponsors, such as their type (e.g. whether they are companies or research projects), their main business area (e.g. media, finance, etc.) and whether the main focus of the sponsor's business is the Semantic Web or not. In the following, we present the main results of the study.

Table 6.1 summarizes the number of sponsors per conference. The second column of the table lists the total number of sponsors in the 10-year period, whereas the last column provides the different sponsors and its percentage over the total. As expected, ISWC attracts slightly more sponsors than ESWC (157 versus 142) and significantly more than SEMANTICS (up to three times more). Interestingly, the number of different sponsors shows that ISWC sponsors tend to repeat (only 54% on the total are different) whereas in ESWC (with almost 60% different sponsors) and SEMANTICS (with almost 70% different sponsors) the sponsors are more sporadic. On average, the number of repeated sponsorships is around 50%. *Figure 6.18* refines this analysis and provides a histogram showing the frequency of the sponsors across all of the conferences. Results show that most sponsors (101 out of 169) only appear in one conference. Some sponsors repeat in two or three conferences, but they rarely appear four or more times. Note that only three sponsors appear in ten or more conferences in the last ten years. These correspond to big companies such as Elsevier, Yahoo and Ontotext.

Conference	Total sponsors	Different sponsors
ESWC	142	84 (59%)
ISWC	157	84 (54%)
SEMANTICS	53	36 (68%)
Total	**352**	**169 (48%)**

Table 6.1 Number of different sponsors per conference in the last 10 years.

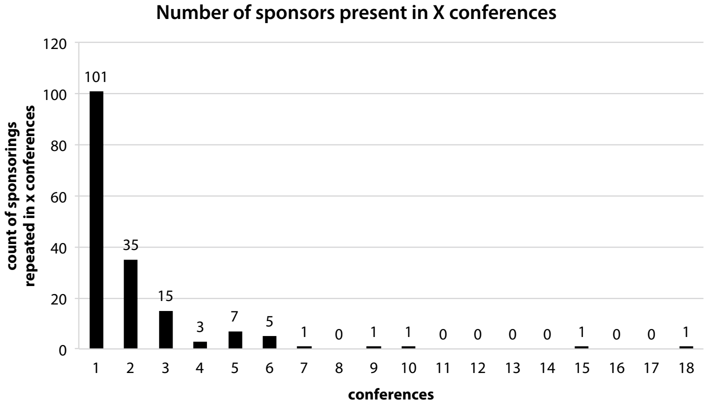


Figure 6.18 Histogram of the number of conferences per sponsor

The distribution of sponsors per conference and year is provided in *Figure 6.19*. In the case of both ESWC and ISWC there is a general tendency of a series of ascending peaks and valleys, with some remarkable numbers in ESWC 2013 (Montpellier, France) and ISWC 2014 (Riva del Garda, Italy). In contrast, SEMANTICS showed a descending tendency since the beginning and up to 2014, where they experienced a huge growth, which was partially lost in 2015.

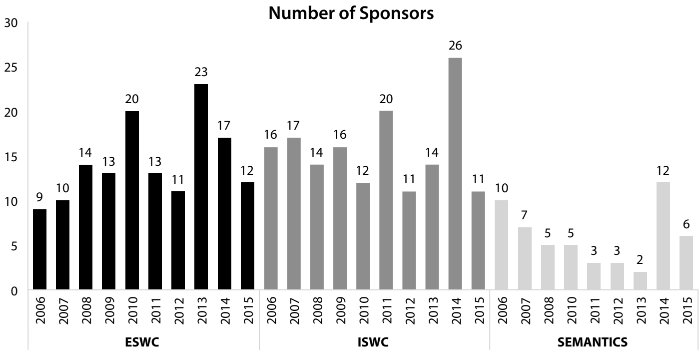


Figure 6.19 Number of sponsors per conference and year

In order to gain more information about the organisations identified, we inspected the websites of the sponsors and manually categorised them into: (a) active companies (i.e. companies that are currently working and active as of July 2016, taking into account name changes, amalgamations and mergers companies); (b) associations or non-profit; (c) EU projects (both finished and currently active); (d) non active companies (i.e. those companies who have creased activities or it was not possible to find a follow-up successor); and (d) other research projects. *Figure 6.20* shows the resultant characterisation of the sponsors. Surprisingly, only 46% of the sponsors are active companies, while almost one third of the sponsors being research projects (28% of the total sponsors were EU projects). It is worth mentioning that 3% of the sponsors are non-active companies, which indicates that the majority of the companies have survived in the last ten years.

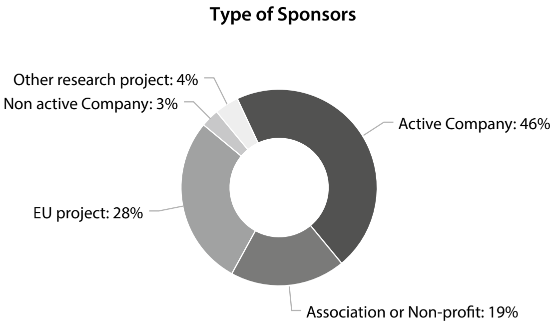


Figure 6.20 Type of sponsors

A fine-grained analysis across each of our five venues is provided in *Figure 6.21*. Interestingly, EU projects constitute the main source of funding for the ESWC conference, which is compatible with the initial European focus of the conference. In turn, SEMANTICS holds the largest number of non-active companies (17% of its sponsorships).

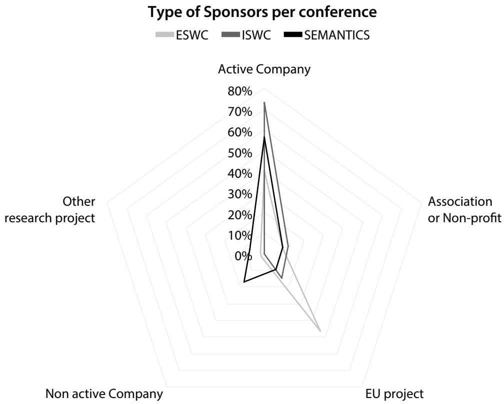


Figure 6.21 Type of sponsors per conference

Note that, in spite of the community's strong emphasis on semantics, sponsors do not necessarily belong to the Semantic Web community, i.e., their business area can touch Semantic Web topics partially, or even not at all, e.g. when the sponsorship is part of a marketing strategy or the sponsor is a local business near the conference venue. Thus, we inspected the websites of all of the sponsors and we categorize their relationship to the Semantic Web, based on information gleaned from their websites, as shown in *Figure 6.22*. In our analysis, we consider that a sponsor is related to the Semantic Web if the sponsor explicitly mentions a Semantic Web-related technology (such as semantic integration, RDF, reasoning, ontologies, etc.) in the description of their activities or products. In turn, if the reference to the technology is implicit or a description can be used to infer that the business could be related to the Semantic Web, we categorise the sponsor as "partially" related to the Semantic Web. Otherwise, we consider that the sponsor is not related to the Semantic Web. Thus, our results, shown in *Figure 6.22*, show that only 23% of the sponsors are not related to the Semantic Web, i.e. more than 75% of the sponsors are involved in Semantic Web technologies, and 50% have a strong relationship to the Semantic Web. Note that our approach is limited to information publicly available and as such we could potentially obtain false negatives when Semantic Web technologies are used but not explicitly mentioned.

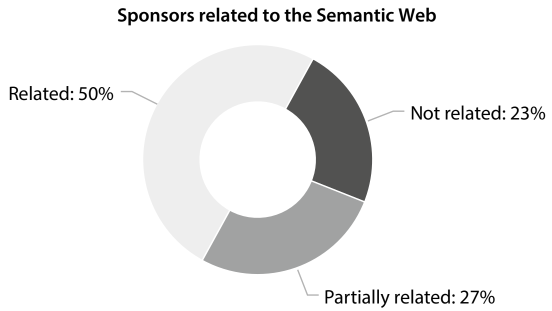


Figure 6.22 Ratio of sponsors related to the Semantic Web

In addition, we analysed the business area of the sponsors and their clients. First we considered the domains arising from our industry analysis: media, finance, medical professional and the pharmaceutical industry, resource and talent management, technology providers and system integrators, product management, supply chains and logistics, energy and insurance. However, in the course of our analysis we also uncovered the following additional domains: security, public sector, and consulting/marketing.

Based on the results of our analysis, we manually annotated the business area of the sponsors, resulting from an analysis of their websites and main products/services. *Fig-*

ure 6.23 provides the distribution of sponsors per domain. It is not surprising that most of the sponsors (54%) are technology providers and system integrators. Media (19%) and the public sector[54] (12%) also figure strongly, while the rest of the areas show a presence ratio between 1-4%.

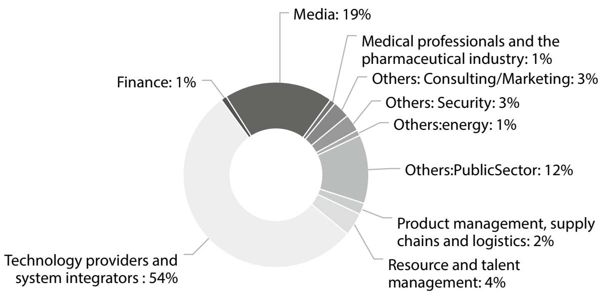


Figure 6.23. Area of sponsors

Following on from this we studied the customers of those sponsors that are related to the Semantic Web. To do so, we inspect the customers/clients that were publicly listed on the sponsor's websites, often in the form of a portfolio or use case. Note that these customers are already using or there is evidence available to indicate that they are potentially interested in Semantic Web technologies, hence by examining their business area it is possible to derive a list of domains that are interested in these technologies.

With this objective in mind, we repeat the analysis of the business domains of these clients, whose results are provided in *Figure 6.24*. Here we can see that media (19%), public sector (21%), the medical professional, the pharmaceutical industry (21%), product management, supply chains and logistics (12%) and finance (7%) constitute the main business areas of the Semantic Web vendors (i.e. technology providers or media-related) who sponsored the three main conferences in the area over the last ten years. The presence of the rest of the areas (e.g. consulting/marketing, energy or insurance), although relatively limited, is also an indicator of the interest and a potential application area for Semantic Web-related technologies.

54 Note that sponsorships from universities and public entities are categorized as "public sector".

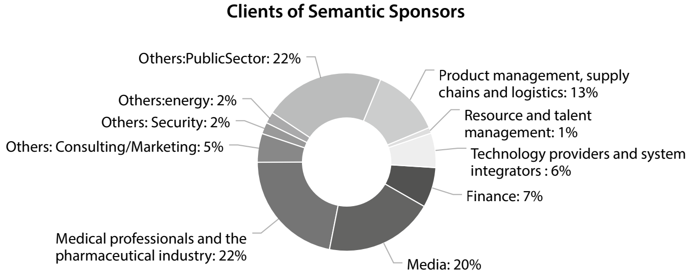


Figure 6.24. Area of the clients of the semantic sponsors

Finally, we studied the companies that appeared both as sponsors and were also mentioned in the conference and journal papers from the ten-year period under investigation. *Figure 6.25* shows the top 10 sponsors and the number of times their name appears in the corpora (both publishers and project sponsors are excluded from the list). As expected, IT leading companies, such as Google, Oracle, Yahoo and SAP, are the top players in the area. Pure semantic-related companies, such as Franz and the Semantic Web company (Poolparty) are in the ten top list but they are mentioned less frequently in the corpora.

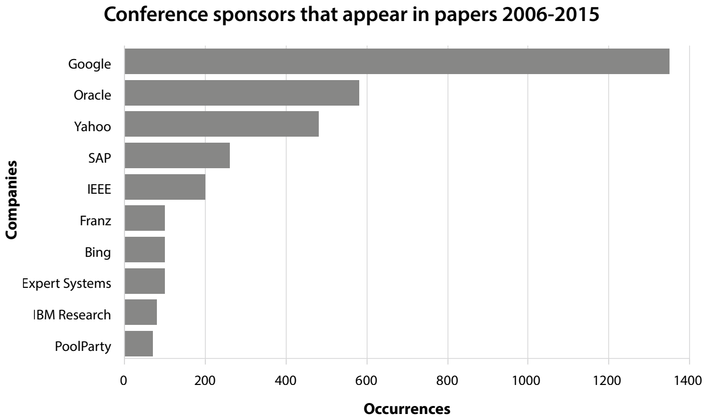


Figure 6.25. Top 10 conference sponsorship companies that appear in the conference/journal papers (2006-2015)

6.2.2. Companies/adopters envisioned in seminal papers on the Semantic Web

As per the topic analysis here again, we studied the three seminal papers on the Semantic Web. Feigenbaum et al., 2009, whose paper focused on corporate applications and consumer use of semantic technologies categorise company involvement as follows:

- **Early adopters** - exploring the Semantic Web and developing prototypes applying Semantic Web tool kits;
- **Software adopters** - software companies which incorporates Semantic Web technologies to their products;
- **Mature companies** - that have deeply and successfully applied these technologies beyond research.

Table 4.2 shows the most frequently cited companies in the seminal papers and their roles. While, *Figures 4.26-4.28* show the number of occurrences of these early adopters, software adopters and mature companies respectively, in the conference/journal papers from the past 10 years. It is worth noting that, the companies that are not mentioned in any papers are not displayed in the graphs.

Early adopters	Software adopters	Mature companies
MITRE Chevron British Telecom Boeing Ordnance Survey Eli Lily Pfizer Agfa Food and Drug Administration National Institutes of Health	Oracle Adobe Altova OpenLink TopQuadrant Software AG Aduna Software Protége SAPHIRE	BBC New York Times Google Yahoo Microsoft Facebook Oracle World Health Organization United Nations

Table 4.2 Company roles in seminal papers

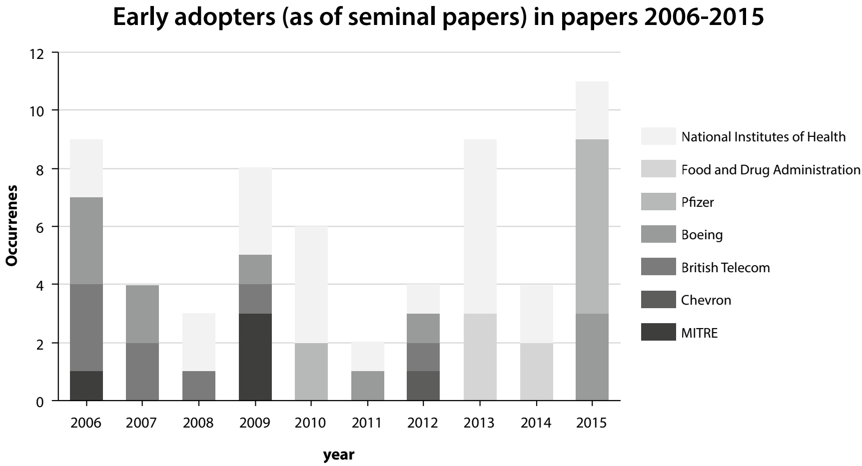


Figure 6.26 Number of occurrences of early adopters in the conference/journal papers (2006-2015)

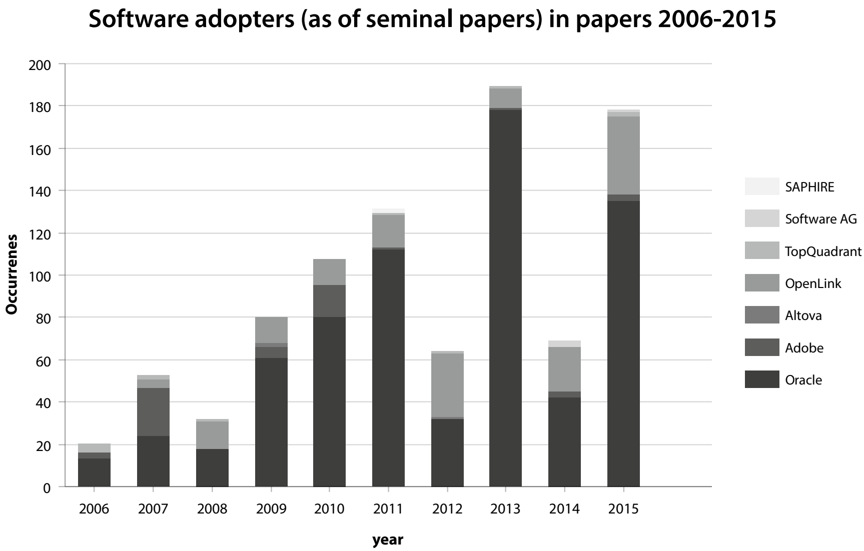


Figure 6.27 Number of occurrences of software adopters in the conference/journal papers (2006-2015)

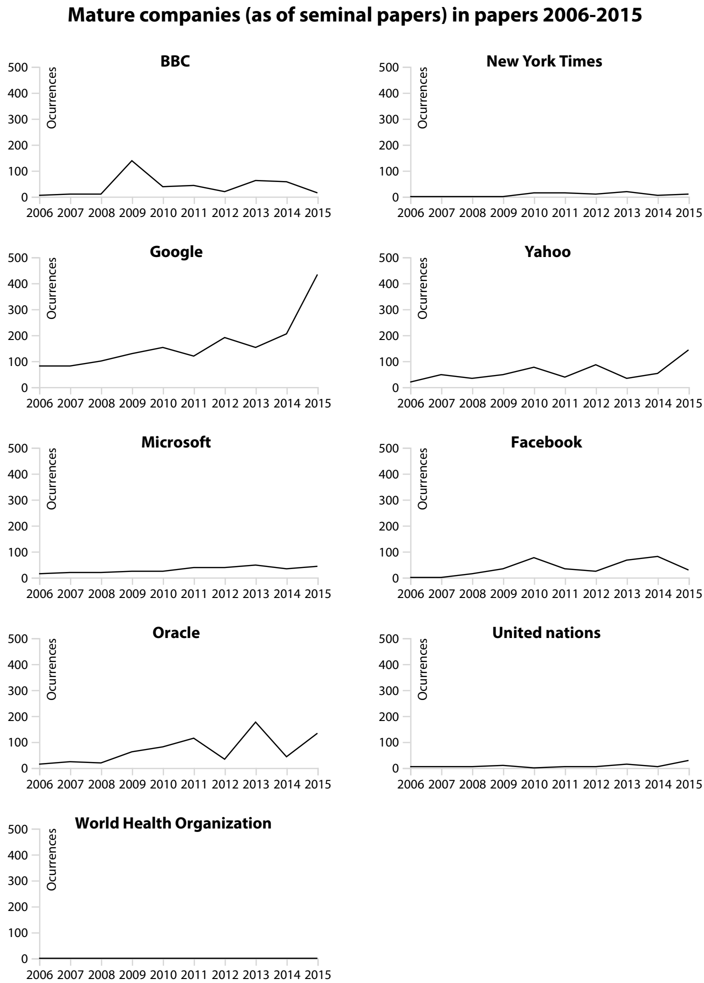


Figure 6.28 Number of occurrences in the conference/journal papers (2006-2015)

As shown in *Figure 6.26* few of the so-called early adopters are really present in the research papers in the last 10 years and, when present, their representation is marginal. One could state that, in view of the results: (a) there was too much trust on these early adopters, who may indeed be entering into the community but at a slower pace, such as Boeing or Pfizer; or (b) the adoption of the technology by such companies is not reflected in research papers.

In contrast, the identification of potential software adopters was more accurate. As shown in *Figure 6.27* companies such as OpenLink, Adobe and Oracle are represented in the research papers. Nevertheless, such presence could be due to their respective tools (such as their RDF stores, e.g. Virtuoso) and cannot be seen as an indication of active and constant contribution in research.

When it comes to the mature companies, results in *Figure 6.28* shows a remarkable presence of such companies in research papers over the last ten years. Although the estimation could be overrated for some companies such as United Nations and the World Health Organisation, most of the so-called mature companies constantly appear across the years. Note also the increasingly large presence of Google, who has made a strong commitment in the development and exploitation of knowledge graphs.

6.3. Standardisation

Although Linked Data technologies have a solid foundation in open standards put forth by the W3C, to date standardisation activities have primarily focused on generic technical concerns. As such the Linked Data technology stack is still incomplete in some key areas, such as privacy, security, robustness and usability, that are prerequisites for widespread enterprise adoption. In the course of our analysis we examined the W3C Semantic Web recommendations and member submissions that have been proposed to date. The overarching objective of this section is to provide an overview of the existing Linked Data standardisation landscape. As the W3C categorisation (Linked Data, vocabularies, query languages, inference and vertical applications) is too coarse grained for meaningful analysis, we compiled a list of all Semantic Web standardisation efforts based on the summary W3C page[55].

6.3.1. Semantic Web Standards

First, we examined the W3C recommendations and member submissions in the context of the well known Semantic Web technology stack (cf. Figure 6.29). Existing standardisation efforts have primarily focused on data exchange, querying and semantics, and unifying logic layers. The vertical and top layers, which relate to cryptography, trust and user interfaces and applications have not resulted in standards thus far.

55 Semantic Web Standards, https://www.w3.org/standards/semanticweb/

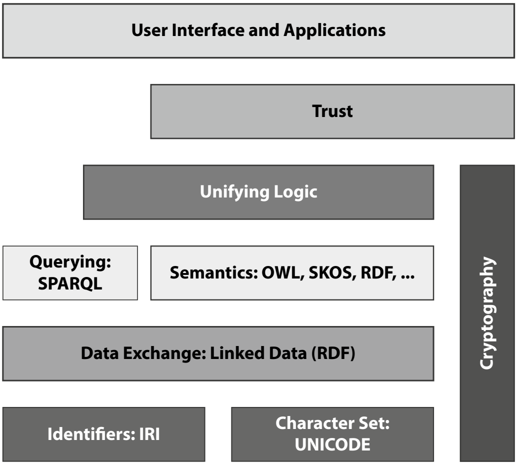


Figure 6.29 Semantic Web Technology Stack

Next, we examined the actual function specified in the abstract of the recommendation, the results of our categorisation are depicted in *Figure 6.30*. Generally speaking, existing Semantic Web standardisations and member submissions can be categorised according to one of the following twelve functions:

- Constraint language
- Schema/ontology/concepts/taxonomy
- Web Services description
- Machine readable data
- Metadata
- Multidimensional data
- Query language
- Rule language
- Transformation language
- Update language
- Transformation language
- Vocabulary

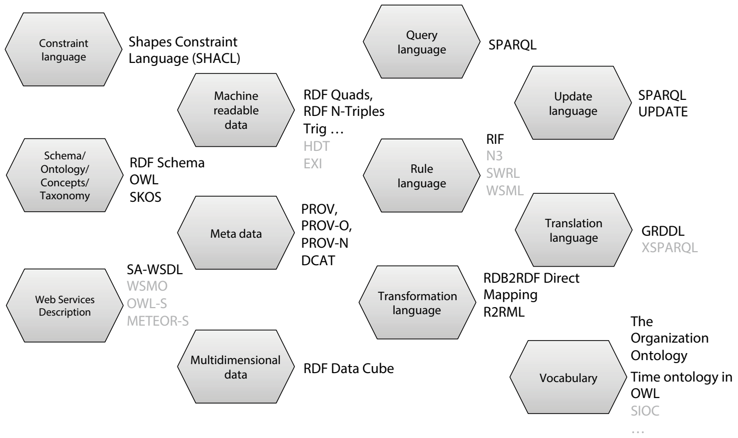


Figure 6.30 W3C Standards and Member Submissions

6.3.2. Gap Analysis

In order to get a better understanding of the coverage over the W3C standardisation efforts in terms of the foundational technologies presented in *Section 2.4.7*, in *Figure 6.31* we overlay the existing W3C standardisation efforts on top of the foundational technologies.

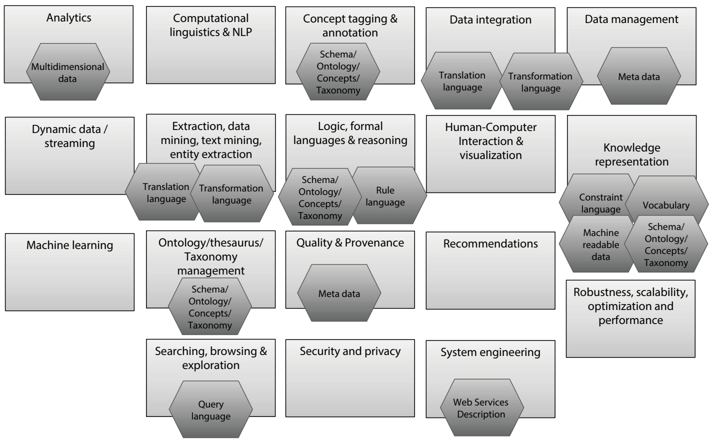


Figure 6.31 W3C Standardisation Gap Analysis

Based on this classification we noticed that there are a number of foundations that are not covered by W3C standardisation efforts (i.e. *Computational Linguistics and NLP*, *Dynamic Data / Streaming*, Human Computer Interaction and Visualisation, Machine Learning, Recommendations, Security and Privacy, Robustness, Scalability, Optimisation and Performance). One explanation is that there are many standards that do not appear under the Semantic Web umbrella that could potentially be adopted for Linked Data e.g. those that relate to security and privacy and visualisation. Additionally, it is worth noting that the W3C focuses on web standards and many of the aforementioned foundations are not web oriented. Having said this a number of community and business groups have been setup to focus on these under represented areas, such as the:

- Ontology-Lexicon community group[56]
- RDF stream processing community group[57]
- Do-not-track community group[58]
- Machine learning schema community group[59]

6.4. Summary

In this section we examined the topics and trends emerging from the research community, the involvement of industry in the prominent publishing venues and existing standardisation activities in terms of the Semantic Web technology stack and also the foundational technologies.

Based on our analysis, we highlight that the community is moving in different directions and at different speeds. Whereas some foundations, such as *knowledge representation & data creation/publishing/sharing, data management and system engineering* are highly represented in the papers, specific topics such as *logics* and *reasoning,* or cross-topics such as *machine learning* or *security* are marginally present.

As expected by the seminal papers, hot topics such as *big/large* data and the internet of things are clearly increasing, as well as those topics regarding *optimisation and performance, implicit knowledge (e.g. crowdsourcing) and quality*. In contrast, *trust*, *mobile topics*, and pay-as-you-go applications are not attracting much attention. There is little evidence to confirm the presence of the early adopters of Semantic Web technologies

56 Ontology-Lexicon community group, https://www.w3.org/community/ontolex/
57 RDF Stream Processing Community Group, https://www.w3.org/community/rsp/
58 Do-not-track community group, https://www.w3.org/community/dntrack/
59 Machine learning schema community group, https://www.w3.org/community/ml-schema/

put forth in the seminal papers, although there is some evidence of the involvement of the identified software adopters and mature companies.

Our analysis also showed that the presence of ontologies, rules, reasoning, inference and logic remains relatively stable, and it is specially tacked by academic affiliations.

In turn, out analysis of the the main Semantic Web conferences sponsors showed a vibrant and active community, where most of the companies are involved in Semantic Web technologies (mainly technology providers and system integrators) that have survived the last ten years. A closer look into the clients of those companies revealed that media, the public sector, the medical professional, the pharmaceutical industry, product management, supply chains and logistics, and finance constitute the main business areas of the Semantic Web vendors.

Finally, we analysed W3C standards and we showed that some specific areas (e.g. *Computational Linguistics & NLP*, *Dynamic Data / Streaming*, Human Computer Interaction & Visualisation, Machine Learning, Recommendations, Security & Privacy, Robustness, Scalability, Optimisation & Performance) are not directly covered by W3C standards. In general, when it comes to standards (be they W3C standards or not), we note that there is a clear need for a more in depth analysis that examines the suitability of non Semantic Web standards in terms of the major functions that are required to propel the adoption of ELD by enterprises.

7. FINDINGS, RECOMMENDATIONS AND ROADMAP

Our in-depth sectoral analysis revealed broad potential for Linked Data adoption across a large spectrum of industries, whose structural and technological characteristics align well with Linked Data characteristics and principles. Based on a set of nine working hypotheses, we analysed 16 industries and identified 11 of them as particularly promising based on their (i) high level of technological readiness for LD and (ii) their structural characteristics, which indicate that they may benefit strongly from adopting LD technologies. Overall, we find that Energy, Retail, Finance and Insurance, Government, Health, Transport and Logistics, Telecommunications, Media, Tourism, Engineering, and Research and Development rank among the most promising. On the other end of the spectrum, we find that the Education, Manufacturing, and Resource Industries have a more limited potential. We provide an in-depth discussion for each industry evaluation in Sections 4.2.

7.1. Market presetting

This section provides a summary from the market/industry analysis that is covered in detail in chapters 3 to 5. A huge volume and variety of data exists within an organisation as well as data that is created and stored elsewhere. But still a lot of data cannot be used to its fullest potential because it is locked up in data islands/silos. The core challenge for most organisations and enterprises in regards to data exchange/integration is to understand internal and external data sources in a fast (even real-time) manner, to save time and costs, and support data-driven decision making. The PROPEL interviews, stakeholder workshop, user survey, and current market reports show that enterprises' need to deliver a better customer experience, to support product and service innovation, and to optimize business processes. Even though Austrian organisations agree on the need for optimised data and information management, including integrating, linking and sharing of data, there is little awareness regarding the term "Enterprise Linked Data". In the past the Linked Data concept has been primarily discussed in the research community - rather than in enterprises - and has primarily been connected to open data - which is often not the case for enterprise data. Nowadays, Linked Data technologies and principles start to enable flexible bottom-up integration also within and between enterprises. In our research we could identify a great variety of excellent use cases for the use of the LD concept and complementing technologies such as big data or cognitive systems within enterprises to support organisations with today's data and information management challenges. Current market forces provide opportunities as well barriers for new data and information management solutions and LD adoption among enterprises, and indicate that the demand for ELD-based data solutions is expected to grow in the future.

Table: Overview of market forces and impacts on data and information management

Market Forces in Austria	Impact on data and information management and LD
Economic growth Austria's economy is stable and moderate with a positive forecast	The economy of a country is one of the most important factors influencing IT investments. A positive economic development in Austria leads to a growth in IT spending and we expect investments in IT and solutions for data and information management (incl. LD and related technology foundations such as machine learning, etc.). • Investments in IT and data & information management solutions • New portfolios and players on the Austrian market • Opportunity for Linked Data concepts to gain recognition
Cost reduction and process efficiency Cost reduction and process efficiency in IT and business remains the leading driver of investment	Although Austrian organisations are increasingly returning to growth, they still focus primarily on costs. Data and information management solutions and LD can have positive effects in terms of transforming businesses, increasing efficiency and driving growth. • Investments in IT and data & information management solutions • IT investments are increasingly business-driven and require collaboration • Linked Data concepts gain recognition
Digital transformation Digital transformation is part of Austrian organisations' agendas and affects all components of an organisation	Data and information management is a key asset for digital transformation, and concepts around Linked Data can support the transformation process. • Investments in IT and data & information management solutions • IT investments are increasingly business-driven and require collaboration • New portfolios and players on the Austrian market • Linked Data concepts gain recognition • Open Data initiatives are growing • New professional skills for innovative data and information management • Optimized business intelligence and decision-making
Technological advancement outpaces technological adoption The Austrian IT market is reserved and conservative compared to western European	Uncertainties regarding the economic development result in an increasing number of smaller and more agile projects rather than in huge investments and radical changes of systems. This as well as missing innovation culture in some organisations might be inhibitors for the uptake of new technologies incl. Linked Data concepts. • IT Spending is held back • Uncertainties and missing innovation culture inhibits Linked Data adoption • Higher competition among solution providers in the Austrian market

Data driven networked global economy Growing need to break up data and information silos, and to collaborate and share data across organisational boundaries	Tighter integration on the data level of enterprises could yield substantial benefits in terms of operational efficiency, automation, optimisation and, ultimately, competitive advantage. Linking data across datasets and sites (resulting in Linked Data) can make data exploitable as a whole rather than as isolated, unrelated datasets. • Investments in IT and data & information management solutions • IT investments are increasingly business-driven and require collaboration • New portfolios and players on the Austrian market • Linked Data concepts gain recognition • Open Data initiatives are growing • New professional skills for innovative data and information management • Optimised business intelligence and decision-making
Citizen demand for online services is growing High Internet adoption and technological developments that connect citizens with work, family, leisure and government	User demands for new digital products and services lead to redefinition and expansion of services, adaptation in the enterprise architecture, linked virtual value chains, cost reduction in public relations, increased security requirement, promotion of new jobs, automated man-machine interfaces. • Investments in IT and data & information management solutions • IT investments are increasingly business-driven and require collaboration • New portfolios and players on the Austrian market • Linked Data concepts gain recognition • Open Data initiatives are growing • New professional skills for innovative data and information management
Data security and privacy Security is a key driver for the IT market, especially for data management, mobility and cloud models	Common barriers to adoption of new technology include data security and privacy concerns. At the same time security concerns provide an opportunity for solution providers to generate revenue out of their security solutions and services. • Investments in IT and data & information management solutions • New portfolios and players on the Austrian market • Linked Data concepts gain recognition • New professional skills for innovative data and information management
General Data Protection Regulation (GDPR) GDPR regulates data management within organisations and will take effect in 2018	EU organisations are starting to prepare for GDPR compliance. They must take a fresh look at how they manage data and information and think about its optimisation and potential adoption of new data management concepts. For security and storage software vendors this means a substantial opportunity. • Investments in IT and data & information management solutions • New portfolios and players on the Austrian market • Linked Data concepts gain recognition • New professional skills for innovative data and information management • GDPR as an impulse to optimize data and information management

Cloud technology stimulates growths Demand for cloud services in Austria is increasing, making it one of the fastest-growing technology segments	Data and information management solutions, e.g. data infrastructures, become more cloud based and profit from scalability and flexibility. • Investments in IT and data & information management solutions • New portfolios and players on the Austrian market • Linked Data concepts gain recognition • Open Data initiatives are growing • New professional skills for innovative data and information management • Uptake of cloud solutions that can also be used for data infrastructures
The Internet of Things (IoT) as an innovation enabler The IoT market is maturing and going beyond its initial focus on connecting more and more things	Data management is fast becoming the overarching theme when talking about IoT. LD can play a key role in the management of new and distributed data in the Internet of Things. • Investments in IT and data & information management solutions • IT investments are increasingly business-driven and require collaboration • New portfolios and players on the Austrian market • Linked Data concepts gain recognition • New professional skills for innovative data and information management • Optimized business intelligence and decision-making • IoT brings new challenges as well as an opportunity for LD
Rising data volume, variety, and velocity, and big data The amount of data and the need to generate added value from it are growing	Linked Data focuses on integrating data from different sources, while big data is primarily about processing and analysis of big data volumes. Both approaches can complement each other. • Investments in IT and data & information management solutions • New portfolios and players on the Austrian market • Linked Data concepts gain recognition • Open Data initiatives are growing • New professional skills for innovative data and information management • Optimized business intelligence and decision-making • Big Data hype and the opportunity to bring additional structure with LD
Content analytics and cognitive systems Technologies and approaches to extract meaningful information from unstructured and structured data are developing	Cognitive systems complement LD with elements of artificial intelligence and machine learning. Innovative types of data analysis and automation allows enterprises to manage their data more effectively, develop and optimize products and services, to reduce costs, to understand their customers better and to increase revenues. • Investments in IT and data & information management solutions • New portfolios and players on the Austrian market • Linked Data concepts gain recognition • New professional skills for innovative data and information management • Optimized business intelligence and decision-making

In our research we identified several **recommendations for enterprises** that have recently or intend to create links between existing data, whether internal or external, within their organization:

- Create an **information access and analysis strategy** to utilise all important data sources. It is surprising how few organisations have mapped out an information architecture showing the linkages between individual pieces of data and an organisation's overall purpose or mission. This mapping involves understanding how data/information supports a particular decision, which in turn supports a given business process/function, and the organisation's overall purpose. With this architecture in place, you can develop an information access plan and strategy and also identify whether or not you have access to all the critical data needed to support more effective decision making.
- Break down walls between data and offer an **information hub** in which the data itself is urbanized rather than the applications. Each application draws on data from throughout the enterprise by operating through this information hub. ELD is **not a technical task** only. Introducing ELD requires a change in **culture** concerning how data is managed.
- Reduce complexity by offering a **unified methodology** for data exchange between applications. Whether the data is internal or drawn from the Web, Linked Enterprise Data opens up a range of possibilities for consolidating the full range of information available. Deliver new applications based on **standard** modern Web technologies, in this respect Linked Data allows to be reactive and even proactive to market demands.
- Use novel technologies such as **cognitive systems** platforms to create the knowledge bases for cognitive system action. Developing these platforms involves working with unstructured and semi-structured information in order to build up curated information bases and knowledge graphs.
- Make sure the solution that you plan to undertake will be able to help achieve the **desired business outcome** and/or issue that you plan to overcome. Include past project experiences in your design thinking approach, and, if available, include predefined use cases that have been developed for peers within your industry to help develop the optimal use cases(s) for the desired outcome. An enterprise linked-data initiative can start with one use case, and develops over time following a pay-as-you-go data integration model to achieve fast return of investment. Getting more **value out of data** saves resources, and increases revenue, e.g. based on better customer insights.
- **Engage subject matter experts** in the design and implementation processes for the use cases that have been developed. The key here is to establish realistic and achievable expectations. And to do that involve both business and IT executives in varying degrees over time:

- **Lines of business:** LOBs are driving the development of improved data management solutions. It takes people deeply involved with the business to understand how applying new approaches can enable things that were previously impossible or impractical. In addition, LOBs understand where the most potential for return will occur when data is managed more efficiently.
- **IT:** Notwithstanding the aforementioned point, IT-managers should be involved with vendor selection and implementation. It's very hard to sustain a long-term, technology enabled business process change without involving the IT organisation, and this is true even if third-party service providers are used. Because to be effective, data management solutions can't operate in an isolated environment, and that means involvement of the IT department. IT plays a crucial role in helping identify relevant data sources, implementing data management solutions and approaches, working with IT vendors to ensure data and system quality.
- **Specialists:** While these may be a subset, or intersection, of stakeholders already described, certain specialists are key to implementation success. For example, data scientists (for statistical analysis, mathematical modelling, and algorithm development) and subject matter experts, or business analysts (for knowledge representation and understanding how data management and analysis solutions will support or enable better decision making).

7.2. Fields of activities

This section of the roadmap takes input from all other chapters and provides guidance with respect to the actions that are necessary to propel the potential of Linked Data in enterprise. The roadmap is broken into five distinct sections: *awareness and education*, *technological innovation*, *standardisation*, *policy and legal* and *funding*. In each section we provide a description of the field, derive a set of recommendations, examine the recommendations from an Austrian perspective highlighting specific strength, weaknesses, opportunities and threats and present a roadmap that can be used to advance the field.

7.2.1. Awareness and Education

Awareness surrounding the potential of Linked Data in enterprises and the lowering of entrance barriers via education and training were mentioned repeatedly by interviewees and stakeholder workshop participants (Chapter 5). Both awareness and education are crucial to increasing the adoption of Linked Data in enterprises.

7.2.1.1. Description of the field

Linked Data and Semantic Web are well-established autochthonous technologies. However, despite growing adoption in enterprises the technologies still have the image of being experimental, academic and not industry-ready.

This can partially be attributed to the fact that although LD has the potential to become a commodity in software engineering, LD is often not implemented as a standalone solution, but is rather bundled with other technologies in the form of complex solutions.

Based on our demand side analysis there seems to be a general unawareness that together the LD principles and the foundational technologies (that are found in a growing number of real world products and implementations) can be used to develop a full-blown environment for data integration.

Another challenge that needs to be addressed is the fact that Linked Data is like the poor cousin of Big Data and Smart Data[60] both of whom are routinely mentioned in day-to-day discussions on disruptive technologies. Even though Linked Data is an underpinning foundational technology is does not get the same attention. Consequently, it is easier for LD technologists to promote LD in the context of hype technologies instead of letting the unique strengths of the technology speak for themselves.

Furthermore, this lack of individual identity means that LD is not one of the top-listed innovative technologies, which makes it harder to attract change-makers, innovators and pioneers.

7.2.1.2. General recommendations

Awareness

The Linked Data story is often a technology-driven story about technological building blocks, standardisation and integration. It is clear that LD is very strong in terms of technological selling points, however there is no core brand that is clearly identified with it (e.g. akin to the relationship between Hadoop and Big Data). Awareness raising is needed to fill this gap and bring the unique characteristics of LD to the foreground, for example:

60 Smart Data is intelligently prepared originally unstructured data by using knowledge of analytics, devices, and applications, as well as the market environment to create added value.http://www.siemens.com/innovation/en/home/pictures-of-the-future/digitalisation-and-software/from-big-data-to-smart-data-video.html

- Open standardisation
- Environment neutrality
- Flexibility and adaptability

However, brand awareness alone is not enough, in additional there is a need for better marketing of industry success stories, that highlight not only the technical benefits but also the business value and societal impact of LD.

In summary, awareness building needs to focus on: i) carving an identity for Linked Data; and ii) marketing the industrial and societal impact of LD.

Education

From a mainstream adoption perspective, education is key to lowering the barrier to entry for enterprises.

As an effect of the diversification of the field, there are more and more such courses emerging. Sometimes relevant training can be found hidden under an umbrella term such as Big Data, where the programme does not mention Linked Data or Semantic Technologies explicitly, however on closer inspection it can be found hidden under related terms such as "Knowledge Graph[61]" or alike (e.g. The ScaDS Summer School on Big Data[62]). In fact there seems to be a general trend in both formal[63] and informal[64] educational settings whereby the LD topic is bundled under the broader umbrella of Big Data or Data Science.

When it comes to supplementing classical academic education online and offline training courses and webinars play a vital role. There are a number of education offerings that focus specifically on LD in the narrower sense that appeal to academics and industry professional alike:

- ESWC Summer school[65]
- Reasoning Web 2016, Summer School Aberdeen[66]
- MOOC: Inria Introduction to a Web of Linked Data[67]

61 https://en.wikipedia.org/wiki/Knowledge_Graph
62 by Competence Center for Scalable Data Services and Solutions https://www.scads.de/de/Summerschool-2016
63 https://www.wu.ac.at/en/infobiz/teaching/sbwl-data-science/ http://www.southampton.ac.uk/data-science/index.page?http://dsg.kmi.open.ac.uk
64 http://grammars.grlmc.com/BigDat2017
65 http://summerschool2016.eswc-conferences.org/
66 http://reasoningweb.org
67 https://www.fun-mooc.fr/courses/inria/41013/session01/about

- Knowledge Engineering with Semantic Web Technologies[68]

Additionally, the Semantic Web Certification Program offered by the Semantic Web Company is especially relevant for industry professionals:

- PoolParty Academy[69]

7.2.1.3. Austrian perspective

Strengths Few but very active players in the field. Academic and commercial organisations have a long tradition to partner in awareness raising: Vienna Semantic Web Meetup, SEMANTiCS Conference Series	**Weaknesses** Austrian academics and industry professionals have to find partners and customers abroad as there is no "home market" for awareness raising
Opportunities	**Threats**

7.2.1.4. Outlook

Key Goals

- **Ensure a solid foundation for ELD** via awareness raising and educational efforts.
- Foster a **topic wise strengthened group of people** that promote ELD with projects, products and flagship implementations.
- **Sharpen the profile of LD and ELD** in order to have a more consistent and true picture of of the social and economic value.

Roadmap

Short-term (next 2 years):

- **Cluster existing players**
  This study has shown that active players from academia, industry and the public sector are well connected and capable of joining forces in order to intensify existing awareness and education collaborations.

68 https://open.hpi.de/courses/semanticweb2015
69 https://www.poolparty.biz/academy/

- » The impact of awareness and education actions can be raised via the setup of **a cluster of existing players** and **the development of an agenda** that lays out the cornerstones of existing and future activities.
- » As the dialogue regarding the impact of ELD is led on an international stage, the **visibility of Austrian players and technologies in an international context** is important. Here again a clustered appearance will help to gain the attention of an international attendance.
- » The **development of joint ELD projects** that span research, innovation and implementation within the cluster will attract innovators, pioneers and change makers coming from research, business and industry.

Medium-term (next 4 year):

- Relate academic and private education
  - » Develop a **joint certification scheme** which intertwines established and upcoming education offers.
  - » Work on the **exchange and development of curricula, training material and lecturers** between different providers.

7.2.2. Technological Innovation, Gaps and Research Question

The objective of this section is to further elaborate on the findings of the technical analysis presented in *Chapter 6*, considering the market drivers, industry analysis and demand side analysis presented in Chapters 3-5, with a view to identifying concrete goals for the adoption of Linked Data in enterprises.

7.2.2.1. Description of the field

Table 7.1 summarises the key findings of the technical analysis according to the 18 foundational technologies identified in Section 2.4. The *Data Perspective* is derived from our analysis of academic papers from the top five Semantic Web publishing venues (ESWC, ISWC, Semantics, JWS, SWJ), whereby we categorising each of the foundational technologies as high, medium or low coverage, based on the total number of occurrences of all topics belonging to a given foundation. While, the information presented in the *Academic* and *Industrial Perspectives* is also derived from our analysis of three seminar papers that discuss trends in the Semantic Web community (Berners-Lee et al., 2001, Feigenbaum et al., 2009, Bernstein et al., 2016) and the outputs from the PROPEL stakeholder workshop and industry interviews.

Data Perspective: High

Foundational Technology	Academic Perspective	Industry Perspective
Knowledge representation & data creation /publishing /sharing / reuse	All topics are strongly represented in the top 5 Semantic Web publishing venues.	The sharing of public data is often cited as a success story.
Data management	It is the second most represented foundation in the top 5 Semantic Web publishing venues Large/big data topics (as well as volume/variety) are gaining more attraction. Both provenance and temporal topics demonstrate a high degree of fluctuation.	Data maintenance is a challenge for most organisations. Harmonising data across the different systems is extremely burdensome. Big data is cited as both a barrier and a driver. According to our market analysis data management is fast becoming the overarching theme of the IoT.
System Engineering	Although system engineering and technology topics are highly represented in the top 5 Semantic Web publishing venues, specific terms such as back-end and in-house are rarely represented. There is evidence that system engineering topics occur more in papers with industry affiliations.	Compliance with standardisation is important as it facilitates integration and protects against vendor lock in. Many companies are looking to consolidate / merge a number of different systems.
Searching / Browsing / Exploration	This traditionally important foundation, shows a 200% increase in the top 5 Semantic Web publishing venues since 2005.	Several of the use cases arising from the interviews and stakeholder workshop relate to search and information retrieval.
Data Integration	This traditionally important foundation has remained constant over the years.	Data integration is cited as a major driver. According to our market analysis industry standards for structured information exchange, are plagued by fragmentation, limited interoperability, and restricted flexibility and extensibility.
Ontology/Thesaurus/Taxonomy management	This traditionally important foundation, shows a 30% increase since 2005, which indicates that this topic is not growing at the same rate as others Ontologies topics remain slightly more represented in papers with academic affiliations.	Domain expertise is very important when it comes to ontology engineering.

Data Perspective: Medium

Foundational Technology	Academic Perspective	Industry Perspective
Formal logic / Formal languages / Description logics / Reasoning	Formal knowledge representation and reasoning remains stable, however in general it appears much less than ontologies. Agent based research (the primary use case present in Berners-Lee et al., 2001) has remained steady for the last 10 years, and is expected to continue as a prominent topic over the next decade.	Formal knowledge representation did not emerge from the interviews as a key topic.
Human-Computer Interaction & Visualization	User and application topics are under represented and growth remains static. Intelligent user interfaces has not attracted much attention to date. Topics in relation to users are slightly more represented in papers with academic affiliations, whereas topics that refer to applications appear equally in both academia and industry.	Front-end to back-end interface management is a challenge.
Extraction, Data Mining, Text mining, entity extraction	Identified by Bernstein et al. identify the need for tools to extract latent semantics automatically.	Although not mentioned explicitly several of the use cases arising from the interviews need to cater for the extraction of knowledge from structured and unstructured data.
Concept tagging, annotation	The topics has a tendency towards peaks and valleys, but it shows an overall 94% increase in the last 10 years. This foundation consists of very specific topics with few related terms.	Concept tagging, annotation did not emerge from the interviews as a key topic.
Quality (incl. data curation)	There is evidence of an upward trend in terms of quality and a downward trend in trust.	Quality and accuracy are very important to practitioners.
Dynamic data / Streaming	200% increase since 2005. The ability to cater for streaming data is highlighted by Bernstein et al. as one of the big research trends in the next decade that will impact the Semantic Web community.	Although not mentioned explicitly several of the use cases arising from the interviews need to cater for dynamic data/ streaming.

Data Perspective: Low

Foundational Technology	Academic Perspective	Industry Perspective
Robustness, scalability, optimization and performance	200% increase since 2005. Efficiency research is on the up.	Existing implementations are still small in scale as there are many unanswered questions when it comes to efficiency. According to our market analysis the big data technology and services market represents a fast-growing multibillion-dollar worldwide opportunity, which represents an opportunity in terms of variety and a challenge in terms of volume and velocity.
Analytics	Although this topic is only marginally represented in the top 5 Semantic Web publishing venues, it shows a 150% increase since 2005.	Many of the interviewees mentioned that they have a data warehouse. When it comes to analytics interviewees mentioned it can be difficult to push non mainstream technologies According to our market analysis data linking is a crucial precursor to content analysis.
Computational linguistics & NLP systems	Marginally represented in the top 5 Semantic Web publishing venues, nonetheless, "Natural language" is a recurrent topic and it is the TOP-9 multiwords in the papers.	Text classifiers are getting better and although the main focus has been on structured data unstructured data is very important also. According to our market analysis, when it comes to data integration and content analysis unstructured data is currently under utilised as such there is a need for deeper into the analysis of unstructured data, especially in combination with structured data.
Machine learning	Marginally represented in the top 5 Semantic Web publishing venues but it shows a 210% increase since 2005.	Existing applications are very narrow in scope partially due to big data challenges and also rights management.
Security & Privacy	According to our analysis this topic is not well represented in the top 5 Semantic Web publishing venues, with only 30% increase since 2005.	Legislation, regulation, ethics, labour laws, legal issues, security, rights management emerged from the interviews as barriers. Privacy was also mentioned as a driver. According to our market analysis, as a result of the General Data Protection Regulation organisations will need to take a fresh look at how they manage data and information.

Recommendations	This topic is not well represented in the top 5 Semantic Web publishing venues, but it shows a 160% increase in the last few years. This foundation consists of very specific topics with few related terms.	Although not mentioned explicitly recommendations are required for a number of the use cases derived from the interviews.

Table 7.1 Summary of the technical analysis

7.2.2.2. General recommendations

Based on our analysis we make the following recommendations:

- Considering the fact that **legislation, regulation, ethics, labour laws, legal issues, security, rights management** (all of which fall under the *Security & Privacy foundation*) emerged from the interviews as barriers to the adoption of Linked Data in an enterprise setting, coupled with the fact that overall there is a lack of research on *Security & Privacy* topics within the Semantic Web community, we recommend:
  - » An in depth investigation into Linked Data security and privacy **requirements** and how they **differ** from other security and privacy requirements arising from **other disciplines** such as database, information systems etc.
  - » A survey of the state of the art of Linked Data security and privacy research and a study to determine the **level of coverage** the **security and privacy community** can provide with respect Linked Data security and privacy requirements.
  - » The setup of **multidisciplinary teams** to work together on security and privacy topics such that computer scientist can benefit from humanities and social science research and via versa.
  - » Putting a strong emphasis on **public hacking challenges** and **penetration testing** in both national and international projects.
- The **quality and accuracy** of tools and techniques were cited by interviewees as important for the adoption of Linked Data in an enterprise setting. Although the *Quality* foundation appears to have medium coverage in the Semantic Web community and there is evidence of an upward trend, considering its importance to the uptake of Linked Data by enterprises, we recommend:
  - » A study to examine what are the current challenges and opportunities when it comes to both **determining and resolving data quality issues**.

  - Promoting the **release of academic prototypes** under open source licenses and encouraging the **reproducibility** of results.
  - Putting a strong emphasis on **public hacking challenges** and **penetration testing** in both national and international projects.
- Based on our analysis, **big data** has emerged as both a driver and a barrier to the adoption of Linked Data in enterprises. The *Data Management* foundation is the second most represented foundation in the Semantic Web community and there are indications that this topic is gaining more attention from the community. Based on our initial analysis we believe further research is warranted in order to understand current data management challenges from both an academic and applied perspective, to this end we make the following recommendations:
  - An investigation into current data management challenges with a view to streamlining existing data management practices and determining the most **effective means of managing and archiving large amounts of data** and associated metadata such as **provenance and temporal data.**
- There is an obvious overlap between **big data** and the *Robustness, Scalability, Optimisation and Performance* foundation which is still underrepresented in the Semantic Web community despite the fact that it has seen a 200% increase since 2005. As such, we recommend:
  - Putting more **focus on big data management and efficiency** in terms of special tracks at conferences such as the *Empirical Studies and Experiments*, *Evaluations and Experiments,* and the *Replication, Benchmark, Data and Software tracks* that have emerged in recent years.
  - Promoting the use of **existing benchmarks** (and potentially generating **new benchmarks)** that allow researchers and practitioners to compare the expressivity, performance and scalability of their tools to alternatives.
- Although many of the interviewees mentioned that they have need for or currently use **data warehouses**, there was general agreement that it can be difficult to get support for non mainstream technologies in a corporate setting. Although the *analytics* foundation shows a 150% increase since 2005 on the whole it is marginally represented in the Semantic Web community. As such, we recommend:
  - That the community **clearly differentiates** what Linked Data has to offer in terms of analytics in general and the combination of Semantic Web technologies with *Computational linguistics & NLP systems* and *Machine Learning* in particular.

  - » In the course of this study it became clear that although there are are number of different tools that used NLP techniques to **extract topics** from documents and to automatically link them with other topics, further research is needed in order to **improve the precision and recall** richer topics beyond named entities. In addition, there is a need to better understand the diffusion of topics from academia to industry and visa versa.

- Another foundation that emerged as both underrepresented and static in terms growth is *Human-Computer Interaction & Visualisation*. One could argue that this is because the foundation transcends many domains, however this foundation also includes topics such as users and applications, which is slightly disconcerting considering the importance of said topics from an enterprise perspective. Consequently, we recommend:
  - » The encouragement of **multidisciplinary teams**, and both **humanities and social science** and **human computer interaction researchers** to actively engage with the Semantic Web community and for the community to encourage participation via special tracks at conferences, special issues of journals etc.
  - » An in depth investigation into specific **visualisation requirements** and how they **differ from visualisation requirements** arising from other disciplines such as database, information systems etc.
- The **consolidation/merger of a number of different systems** was identified as a strength of Linked Data technologies. In the same context one of the interviewees highlighted that compliance with standardisation was a key requirement as it facilitates integration and protects against vendor lock in. Although there appears to be good coverage of the *System Engineering* foundation especially in papers that have at least one industry affiliation among the authors, commonly used terms such as back-end and in-house are rarely represented. As such, we make the following recommendations:
  - » A study to better understand the different **systems** that are referred to in academic papers, the different **application areas** and their maturity in terms of r**eal world deployment**.
  - » An investigation into potential **industry showcases** in terms of startups, projects, applications and systems based on Linked Data technologies, including concrete details of the technologies used and the lessons learned.
- Although there is an increasing interest in using machine learning techniques to analyse big data, this topic in marginally represented in the community, but it shows a 210% increase in the last ten years. On the other hand, an interviewee

that is currently using machine learning techniques indicated that existing applications are very narrow in scope partially due to big data challenges and also rights management. Thus, we recommend:

 » Fostering the collaboration with machine learning researchers, **promoting machine learning topics in the CFP,** with special emphasis on the combination of machine learning and Semantic Web technologies, privacy preserving machine learning and the robustness of machine learning algorithms.
 » Promoting the **release of prototypes applying machine learning techniques to big data** under open source licenses and encouraging the **reproducibility** of results.

- Although recommender systems and recommender techniques are required for several of the use cases mentioned in the interviews, it is the least represented foundation in the community, albeit showing a 160% increase in the last years. As such, we recommend:

 » An in depth investigation into Linked Data and recommender system models and techniques and how they **differ** from **other disciplines** such information retrieval.
 » An investigation into **current tools and practical applications** that offer recommendations on top of Linked Data techniques.

- Although not mentioned explicitly several of the use cases arising from the interviews need to cater for dynamic data/ streaming, additionally the seminal papers identified this area as one of the big research trends in the next decade. However, despite a 200% increase since 2005, it is still underrepresented in the Semantic Web community, with no W3C standardisation efforts besides the RDF stream processing community group. Thus, we recommend:

 » An investigation into current research challenges for preserving and managing dynamic linked datasets, determining the most effective way to **store, synchronise, represent changes and query evolving graphs**.
 » Promoting the **adoption of versioning and preservation techniques** in the community that allow researchers and practitioners to synchronize their data and perform time-based queries on evolving linked datasets.

Promoting the **continuation of standardisation efforts**, joining forces with novel and active communities such as the Internet of Things.

7.2.2.3. Austrian perspective

Table 7.2 Provides a snapshot of the strengths, weaknesses, opportunities and threats from an Austrian perspective.

Strengths • Austria is a strong innovator, with innovation performance above or close to that of the EU average[70]. • There are several national and european projects (ADEQUATe: Analytics & Data Enrichment to improve the QUAliTy of Open Data, DALICC:Data Licenses Clearance Centre, SPECIAL:Scalable Policy-awarE linked data arChitecture for prIvacy, trAnsparency and compLiance, LinDA:Enabling Linked Data and Analytics and SDI4APPS: cloud-based framework for data integration) that are already tackling some of the open research questions.	**Weaknesses** • Austrian organisations are rather conservative and usually not among the pioneers when it comes to adoption of new technology. As such, technology development is outpacing technology adoption in Austria. • Compared to Western Europe (UK, Germany), and US Austria is usually behind when it comes to innovation and technology adoption.
Opportunities • The General Data Protection Regulation will come into effect in May 2018, if Austria were to invest in related topics we could our Enterprises could gain competitive advantage. • Austria already has a number of relevant competency centres (i.e. The Semantic Technology Institute (STI) Innsbruck, The TU Linked Data Lab, The WU Privacy & Sustainable Computing Lab), that could be further developed in order to form centres of excellence.	**Threats** • Shortage in numbers of suitably qualified science and engineering graduates entering postgraduate studies and the market. • Cuts to core fundamental and applied research budgets.

Table 7.2 Summary of the technical SWOT

7.2.2.4. Outlook

Key Goals

- Develop **technical competencies** in the foundational technologies that have been identified as key **drivers** and/or **barriers** for Linked Data adoption in enterprises, such as:
  - » *Security and Privacy*
  - » *Quality*
  - » *Data Management*
  - » *Robustness, Scalability, Optimisation and Performance*
  - » *Analytics*
  - » *Human-Computer Interaction & Visualisation*
  - » *System Engineering*
  - » *Dynamic data / Streaming*
  - » *Computational linguistics & NLP systems*

- Foster the development of **multidisciplinary teams** and **public private partnerships** such that research on the aforementioned technologies is informed by societal and business requirements, which in turn increases the impact of the developed tools and technologies.
- Provide a platform for Austrians to **showcase technological results** from nationally and internationally funded projects, such that it encourages discussion on both current and future challenges and opportunities, supports inter project collaboration, and fosters innovation and excellence.
- Support the setup and development of **centres of excellence** that can demonstrate how **technical competencies** come together in order to advance the adoption of technology in an enterprise setting, starting with **Linked Data** and moving beyond to **Big Data Analytics** and the **Internet of Things**.

Roadmap

Short-term (next 2 years): Capacity Building

- Incentivise organisations to develop both offline and online **training** in the form of tutorials and courses, tailored to cater for the **skills and competencies** that are required for the adoption of Linked Data in enterprises.
- Launch preparatory studies in the foundational technologies that have been identified as key **drivers** and/or **barriers** for Linked Data adoption in enterprises, which are necessary in order to gage the maturity of existing technologies, identify technology gaps, and to understand how technology from other domains could potentially be used to bridge the gaps.

Medium-term (next 4 year): Collaboration

- Initiate projects that are necessary to ensure technical competencies in the foundational technologies that have been identified as key **drivers** and/or **barriers** for Linked Data adoption in enterprises.
- The organisation of national events that promote greater awareness of Linked Data technologies, provide a platform for the dissemination of research outputs connected to the Linked Data, support discussion on barriers and drivers, and enable **public-private and interdisciplinary** matchmaking.

Long-term (next 7 years): Development of centres of excellence

- Provide the supports necessary to assist in the setup and development of centres of excellence, including but not limited to financial, administrative, communication, innovation.ancial, administrative, communication, innovation.

7.2.3. Standardisation

Linked Data has a solid foundation in open standards proposed by the W3C, however given the importance of standardisation to industry players and the gaps identified in Chapter 6 of the roadmap it is clear that much more needs to be done.

7.2.3.1. Description of the field

Table 7.3 summarises the key findings of the standardisation analysis according to the eighteen foundational technologies. The *Data Perspective* is derived from our analysis of academic papers from the top 5 Semantic Web publishing venues (ESWC, ISWC, Semantics, JWS, SWJ), whereby we categorising each of the foundational technologies as high, medium or low coverage, based on the total number of occurrences of all topics belonging to a given foundation. While, the data presented in both the *W3C SW Standards* and *Other Standards & Groups* are based on the analysis of the existing standardisation efforts.

Data Perspective: High

Foundational Technology	W3C SW Standards	Other Standards & Groups
Knowledge representation & data creation/ publishing/ share/ reuse	• The Resource Description Framework (RDF) a language for representing information about resources • The Shapes Constraint Language (SHACL) for describing and constraining RDF graphs • The Organisation Ontology for the publishing of organisational information	
Data management	• The PROV Data Model for provenance interchange on the web • The Data Catalog Vocabulary for interoperability between data catalogs published on the Web	
System Engineering	• Semantic Annotations for Web Services Description Language and XML Schema that allows description of additional semantics of WSDL components	
Searching / Browsing / Exploration	• SPARQL a set of specifications that provide languages and protocols to query and manipulate RDF graph content	
Data Integration	• RDB2RDF a mapping from relational representation to an RDF representation • R2RML: RDB to RDF Mapping Language • GRDDL a mechanism for Gleaning Resource Descriptions from Dialects of Languages	
Ontology/Thesaurus/Taxonomy management	RDF Schema (RDFS) Language vocabulary and formal semantics OWL Web Ontology Language vocabulary and formal semantics	

Data Perspective: Medium

Foundational Technology	**W3C SW Standards**	**Other Standards & Groups**
Formal logic / Formal languages/ Description logics / Reasoning	• Rule Interchange Format (RIF) Production Rule Dialect • RDF Schema (RDFS) Language vocabulary and formal semantics • OWL Web Ontology Language vocabulary and formal semantics	
Human-Computer Interaction & Visualisation		
Extraction, Data Mining, Text mining,entity extraction	• RDB2RDF a mapping from relational representation to an RDF representation • R2RML: RDB to RDF Mapping Language • GRDDL a mechanism for Gleaning Resource Descriptions from Dialects of Languages	
Concept tagging, annotation	• Simple Knowledge Organisation System (SKOS), a common data model for sharing and linking knowledge organisation systems • RDFa (Resource Description Framework in Attributes)	
Quality & Provenance	• The PROV Data Model for provenance interchange on the web • The Data Catalog Vocabulary for interoperability between data catalogs published on the Web	
Dynamic data/ Streaming		RDF stream processing community group

Data Perspective: Low

Foundational Technology	**W3C SW Standards**	**Other Standards & Groups**
Robustness, scalability, optimisation and performance		
Analytics	• RDF Data Cube Vocabulary links statistics to related dataset and concepts	
Computational linguistics & NLP systems		Ontology-Lexicon community group
Machine learning		Machine learning schema community group
Security & Privacy		Do-not-track community group
Recommendations		

Table 7.3 Summary of the standardisation analysis

7.2.3.2. General recommendations

Based on our initial analysis it is clear that foundational technologies such as *Computational Linguistics & NLP, Dynamic Data / Streaming, Human Computer Interaction & Visualisation, Machine Learning, Recommendations, Security & Privacy, Robustness, Scalability, Optimisation & Performance*, are not part of the core Semantic Web standards. Having said that there are other standards that do not fall under the Semantic Web umbrella that could potentially be relevant for Linked Data e.g. those that relate to security and privacy and visualisation. Additionally, when it comes to standardisation generally speaking it is difficult to measure the degree of coverage of a particular topic offered by existing standards. Based on our initial analysis we recommend the following:

- A study into the **effectiveness of existing standards** and the **degree of coverage** they provide with respect the foundation they belong to and also in terms of relevant use cases and requirements.
- An investigation into the **suitability of alternative standards** in order to determine if there are any existing standards (that are recommended by either the W3C or other standards bodies) that could potentially be applied in the context of the Linked Data technical foundations.
- Promoting the setup and development of a certification scheme for tools and products that adhere to the standard.

7.2.3.3. Austrian perspective

Table 7.2 Provides a snapshot of the strengths, weaknesses, opportunities and threats from an Austrian perspective.

Strengths • There are a number of Austrian organisations already actively involved in standardisation.	**Weaknesses**
Opportunities • Setup and develop a standardisation certification scheme.	**Threats** • Standardisation bodies decide to focus their efforts on other technologies.

Table 7.4 Summary of the standardisation SWOT

7.2.3.4. Outlook

Key Goals

- Identify gaps and priorities in terms of technical standards that are necessary to propel the potential of Linked Data in enterprises.
- Encourage both industry and academia to get more involved in standardisation.
- Support the setup and development of a standardisation certification scheme.

Roadmap

Short-term (next 2 years): Awareness and Education

- Incentivise organisations to develop both offline and online **training** in the form of tutorials and courses, on existing Semantic Web standards.
- Launch preparatory studies into the **degree of coverage** that is offered by existing W3C standards and the identification of alternative standardisation activities that could be used to fill existing standardisation gaps.

Medium-term (next 4 year): Encouraging Industrial uptake

- Encourage more organisations to get involved in standardisation activities especially those that relate to the drivers and barriers identified in the technology section of the roadmap.
- Incentivise standards bodies to set up certification schemes such as those offered by multinationals such as Microsoft, Oracle, SAP etc

7.2.4. Legal and Policy

It was shown in Chapter 5 of the report, that there is high awareness among LD stakeholders of the policy and legal issues and the crucial role they play when it comes to the adoption of Linked Data by enterprises. While, Chapter 6 shows, that research and development on related topics are under represented in the research community. In this section of the roadmap we take a closer look at policy and legal issues, in the discussion that follows we differentiate between these tightly coupled concepts as follows:

- **Policy** - Refers to all phenomena and measures that may be associated with the political and/or administrative realm, including e.g. transnational or national initiatives, public private partnerships, recommendations on an EU level, connected laws and regulations, etc.
- **Legal** - Refers to all phenomena that are functionally coupled with the fulfilling of obligations coming from transnational, national law or private jurisdiction.

7.2.4.1. Description of the field

Policy

When it comes to policy and the data economy, many concrete measures and activities have already been put in place on an EU-level, for instance: several consultations (e.g. the early 2016 one on regulatory environments of platforms and online intermediaries[70]) as well as concrete activities funded under the H2020 programme, such as the European Data Portal[71] and the Big Data Value Association[72] (PPP on Data) that aim to strengthen co-operation between government and industry, or the European Data Science Academy[73] (EDSA) that aims to foster data literacy and data science education. These activities are accompanied by community building activities like the Big Data Europe project[74] or the European Data Forum[75] conferences.

At a national level several data-related activities and initiatives have also been put in place, e.g. the Digital Roadmap for Austria[76] or the Digital Agenda of the City of Vienna[77], as well as the open data activities around data.gv.at and the Open Data Portal Austria (ODP) or the community work managed by Open Knowledge Austria, the ODI Node Vienna or the Digital networked Data[78] initiative.

Although existing efforts are aligned with the European Digital Agenda[79] and the Strategy for the Digital Single Market[80], data economy initiatives on communal, national and transnational levels are: i) only loosely coupled; and ii) still in their infancy. As such, there is a lack of awareness of the importance of the data economy in general and on the opportunities presented by ELD in particular, across Austrian industry, research communities and the public.

Legal

ELD has evolved from the academic realm, whereby legal viewpoints on developments are secondary. This can be directly correlated with the fact that research has special derogations on reuse, citation, non-commerciality and scientific exploitation,

70 https://ec.europa.eu/digital-agenda/en/news/public-consultation-regulatory-environment-platforms- online-intermediaries-data-and-cloud
71 http://www.europeandataportal.eu/
72 http://www.bdva.eu/
73 http://edsa-project.eu/
74 http://www.big-data-europe.eu
75 http://www.data-forum.eu
76 https://www.digitalroadmap.gv.at/de/
77 https://www.digitaleagenda.wien/de/
78 http://networkeddata.at/en/home.html
79 https://ec.europa.eu/digital-agenda/en
80 http://ec.europa.eu/priorities/digital-single-market_en

which differ substantially from the commercial ecosystem. As a result, there is a lack of awareness and substantial lack of knowledge on the legal issues concerning Linked Data. This problem is further magnified by the fact that legal considerations are generally not even considered in more applied projects.

As a consequence, when it comes to legal matters such as: license clearance and negotiation, data protection[81] (DSG), product liability[82] (PHG), consumers protection[83] (KSchG) and connected private international and national law, ELD is not yet industry ready.

Tackling such legal issues is a time-consuming, complex and error-prone task, which:

- comes with high transaction costs because of the manual handling of legal terms and conditions;
- lacks sufficient expertise/awareness to detect all relevant legal aspects; and
- needs organisation wide effort across various departments.

Nevertheless, we are in a privileged situation given that Linked Data technologies are particularly suitable for developing technology-driven solutions to legal issues. The descriptive, open approach to metadata can be used to develop a tailor made legal descriptive layer. On a positive note, the first steps towards a technology driven solution have already been taken in the form of projects such as DALICC (Data Licenses Clearance Center) and SPECIAL (Scalable Policy-awarE Linked Data arChitecture for prIvacy, trAnsparency and compliance) that focus on licensing and regulation support.

7.2.4.2. General recommendations

Policy

We are facing a fragmented landscape of initiatives and policies that aim to support growth and vitality in data driven businesses. ELD may profit from this broad base of relevant positive impulses. But as a technology-driven data solution there is a need for technology focused support concerning policy, such as:

- Ensuring that **standardisation** is at the core of policy development.

---

81 Federal Act concerning the Protection of Personal Data, Federal Law Gazette (FLG) I 165/1999 as amended by FLG I 132/2015 through which Directive Directive 95/46/EC, OJ L 281,23/11/1995, 0031- 0050 (will be replaced by General Data Protection Regulation) was implemented.

82 Federal Act Governing the Liability for a Defective Product, FLG 99/1988, as amended by FLG I 98/2001.

83 Federal Act on Regulations for the Protection of Consumers, FLG 140/1979 as amended by FLG I 105/2015.

- The centring of policy development efforts so as to establish "**critical mass**" for ELD technologies and solutions.
- Initiate **data marketplaces** as public ELD infrastructures.
- When it comes to related data driven use-cases in public administration and/or data driven services **LD and ELD should be included in public tenders**.

Legal

To overcome the identified hurdles and barriers in the legal field, there is a need for action in all stages of the ELD development path: **research**, **applied research** and **industry integration**, so that solutions to legal challenges are an integral part of the ELD stack. The legal recommendations necessary to realise this vision are as follows:

- In order to ensure the sustainability of the sector and its future development, **personal and organisational legal expertise** need to feed into ELD projects.
- **Collaboration** between the **legal sector** (lawyers) and **technologists** should lead to models, standards and technology solutions where legal aspects are embedded into the ELD stack.
- More **research on the legal impacts** of future ELD implementations need to be initiated.
- Legal issues should be emphasised in the awareness building and educational material.
- A **set of standard**, exchangeable and machine readable **legal frameworks** for data enrichment, data fusion and data reuse are required. Standards should not only consider data copyright but must also tackle data protection, product liability, consumer protection and other connected private international and national law.
- **Legal clearing centres** need to be established in order to provide: i) personal guidance and ii) automated assistance to enterprises and legal professionals.

7.2.4.3. Austrian perspective

Legal

Strengths Good coupled ecosystem of research, government and industry allows tailormade policies to be developed	**Weaknesses** ELD relevant industry is not headquartered in Austria
Opportunities	**Threats** Lack of harmonised national policy in EC, creates competition instead of fostering specialisation

Policy

Strengths	Weaknesses ELD relevant industry is not headquartered in Austria
Opportunities • **Collaboration** between the **legal sector** (lawyers) and **technologists** can be fostered easily • **Intensive involvement into standardisation** to multiply efforts	**Threats** A dispersed European (and worldwide) legal landscape hinders generalised solutions and demands global expertise in the sector

7.2.4.4. Outlook

Key Goals:

- **Policy makers should create a positive momentum for the further development of ELD** by utilising the existing eco-system as a development and test-bed for ELD
- Promote **collaboration** between the **legal sector** (lawyers) and **technologists**, to come up with **flagship legal clearing centres** and subsequently promote these flagships as a blueprint for transnational usages.
- Incentivise the **standardisation** of legal description languages and application frameworks.

Roadmap

Short-term (next 2 years):

- Cluster existing initiatives, players and policy schemes
  - » Develop a Linked Data cluster agenda
  - » Ensure visibility in an international context
- Encourage tandem research projects between the legal sector and technologists
  - » Exploratory projects
  - » Active partnership development initiatives

Medium-term (next 4 year):

- Flagship projects on ELD legal clearing centres
  - » Demonstrators based on national use cases
  - » Broaden findings and solutions to a transnational blueprint

   » Couple outcomes of the flagship projects with the standardisation efforts

Long-term (next 7 years)

- Serving the specific niche with tailor made technological solutions and profound legal expertise
   » Establish technology driven legal clearing Linked Data in the academic research agenda
   » Encourage IT-companies to exploit commercial products

7.2.5. Funding

From the interviews and workshops conducted during the project, it is evident that ELD technologies are promising for companies of all types (startups, SMEs, large companies) and that are active in a variety of industries (*see Chapters 3-5*). While, Chapter 6 has shown that there is an active research community in the Linked Data area and distilled those research topics that require further investigation. Funding organisations and their programs can play a pivotal role in supporting further research and industrial adoption of Linked Data in Austria.

7.2.5.1. Description of the field

Austrian organisations wishing to acquire funds for ELD related research can benefit from various funding schemes, most notably FFG's IKT der Zukunft[84] call, which typically has a focus on the development of Intelligent Systems. Funding organisations that have a focus on basic research include FWF[85] and WWTF[86]. The Christian Doppler society[87] funds long term projects (up to 7 years) between universities and company partners with the goal of transitioning and adapting basic research performed in universities into industry contexts.

Going beyond funding opportunities in Austria, European level programs, such as Horizon 2020 offer additional funding sources, especially in calls related to Information and Communication Technologies[88]. Some EU funded projects, such as ODINE[89], offer funding and guidance for companies (typically startups) that use open data for creating business value.

---

84 https://www.ffg.at/iktderzukunft
85 FWF: https://www.fwf.ac.at/
86 WWTF: http://www.wwtf.at/
87 CDG: https://www.cdg.ac.at/en/
88 https://ec.europa.eu/programmes/horizon2020/en/h2020-section/information-and-communication-technologies
89 https://opendataincubator.eu/

For SMEs, dedicated acceleration programs help technology take-up (on technologies not necessarily focused on Linked Data), as well as a dedicated agency - EASME[90] (Executive Agency for SMEs). For example, the Horizon 2020 SME Instrument[91] offers funding specifically for SMEs for boosting the competitiveness of SMEs with a commercial ambition.

7.2.5.2. General recommendations

Although promising, the adoption of ELD technologies is hampered by several factors which could be mitigated through funding instruments.

Firstly, there is a lack of straightforward demonstrators that can be applied with low-cost in order to give potential adopters a quick insight into the benefits of LD technology. Therefore, ELD projects require significant investment from potential adopters with the main cost categories including skill acquisition; integration with existing software architectures; data migration. We therefore see a need to fund the creation of demonstrators that address concrete needs of potential adopters. A key player in this context is the technology transfer from universities, especially groups that perform Linked Data specific research. Funding instruments that enable **technology transfer from universities to companies** could mitigate several cases of low technology adoption.

Secondly, the adoption of LD technologies, especially as Linked Enterprise Data across several industries, requires industry-wide action and cooperation to ensure the critical mass of adoption that can lead to a network effect and wide-scale diffusion of these technologies. Therefore, funding instruments/policy that encourage **industry wide-actions and adoption** should be offered. Note, that in this case, the issues to be solved are not country-specific but rather industry-specific and therefore international funding might be more suitable.

Thirdly, the relatively recent and short period in which LD technologies were brought to market, highlights some weaknesses of these technologies in corporate settings which need to be addressed with more (basic) research. These include aspects such as those related to security or performance which are crucial for enterprise adoption but have been only considered weakly so far in research. Therefore, funding should support basic and applied **research on topics relevant for enterprise adoption**. A study performed in PROPEL highlights those research topics that are important to companies but weakly addressed in research so far.

90 https://ec.europa.eu/easme/
91 https://ec.europa.eu/programmes/horizon2020/en/h2020-section/sme-instrument

7.2.5.3. Austrian perspective

Strengths	**Weaknesses**
• Austria is well-positioned to fund basic science in selected topics of interest • Austria is well-positioned to fund applied research in market areas/industries that are of strategic relevance to Austria (e.g., tourism) • Austria can encourage knowledge transfer from local universities to local companies • Measure can be taken to keep the right balance between funding basic and applied research (CDL)	• High access barrier to funding, especially for companies (high competition for funding, significant investment in writing proposals, lack of skills at companies for research proposal writing) • Difficulty to fund industry-wide action and adoption at an international level
Opportunities	**Threats** • Unsure access to European-level funds • Lack of control over the topics funded at European-level.

7.2.5.4. Outlook

Key Goals

Firstly, there is a need to **foster research on topics relevant for enterprise adoption**, so that various open issues that still hamper industry adoption (e.g., performance, privacy) can be sufficiently well addressed. These research topics should be investigated from a basic and applied research perspectives alike.

Second, it is important to **ensure technology transfer from universities to companies**. Universities should have the possibility to learn about the various needs of industry, while, vice-versa companies should get the opportunity to get insights into novel developments in academia.

Thirdly, going beyond technology transfer to individual companies, to take advantage of the full potential of Linked Data technologies, funding should encourage its **industry wide adoption.**

Roadmap

Short-term (next 2 years): Fostering research on topics relevant for enterprise adoption and technology transfer

In the short-term the focus should be primarily on fostering research on topics relevant for enterprise adoption while also ensuring technology transfer from universities to companies. The strategy to achieve this goal could include:

1. Funding a number of studies that go beyond PROPEL and investigate in more depth several open issues identified during the PROPEL project. In-depth studies recommended in Section 8.1.2.2 are:
    a. An in depth investigation into Linked Data security and privacy requirements and how they differ from other security and privacy requirements arising from other disciplines such as database, information systems etc.
    b. A survey of the state of the art of Linked Data security and privacy research and a study to determine the level of coverage the security and privacy community can provide with respect Linked Data security and privacy requirements.
    c. A study to examine what are the current challenges and opportunities when it comes to both determining and resolving data quality issues.
    d. An investigation into current data management challenges with a view to streamlining existing data management practices and determining the most effective means of managing large amounts of data and associated meta-data such as provenance and temporal data.
    e. An in depth investigation into specific visualisation requirements and how they differ from visualisation requirements arising from other disciplines such as database, information systems etc.
    f. A study to better understand the different systems that are referred to in academic papers, the different application areas and their maturity in terms of real world deployment.
    g. An investigation into potential industry showcases in terms of startups, projects, applications and systems based on Linked Data technologies, including concrete details of the technologies used and the lessons learned.
    h. An in depth investigation into Linked Data and recommender system models and techniques and how they differ from other disciplines such information retrieval.
    i. An investigation into current research challenges for preserving and managing dynamic linked datasets, determining the most effective way to store, synchronise, represent changes and query evolving graphs.
2. Fund research topics related to foundational technologies that have been identified as key drivers and/or barriers for Linked Data adoption in enterprises, such as (based on Section 8.1.2.4):
    a. Security and Privacy
    b. Quality
    c. Data Management
    d. Robustness, Scalability, Optimisation and Performance

    e. Analytics
    f. Human-Computer Interaction and Visualisation
    g. System Engineering
    h. Dynamic data / Streaming
    i. Computational linguistics and NLP systems

3. Provide funding to industries that are the most prone to Linked Data adoption, including: Research, Finance and Insurance, Health, Transport and Logistics, and Government.
4. Funding schemes should be considered for promoting *lightweight technology transfer*. These schemes could help in further understanding the needs of enterprises and encouraging the adoption of LD technologies at least in selected projects. For example the Irish Enterprise Partnership Scheme[92] is a funding scheme that fosters flexible knowledge transfer from universities to industries by jointly funding PhD or Post Doctoral researchers. The researchers apply their research in the context of the company needs. The funding body covers ⅔ of the researcher's costs and therefore makes such projects financially very attractive to industry partners.

Medium-term (next 4 year): Intensifying technology transfer from universities to companies

In the medium-term, the focus should be on intensifying technology transfer from universities to companies within Austria, while continuing the funding of relevant basic research. To achieve this goal, the strategy should be based on long-term funding schemes for combining basic and applied research, such as Christian Doppler Laboratories. Research topics to be prioritised for funding should be decided based on the outcome of the various studies suggested on the short-term period as well as insights gained from projects for technology transfer to industry. With Linked Data technologies becoming more robust and suitable for enterprise adoption, funding of the adoption of these technologies could also be encouraged in industries less prone to adopting Linked Data, e.g., Education, Manufacturing and Resource industries.

Long-term (next 7 years): Industry wide actions and adoption

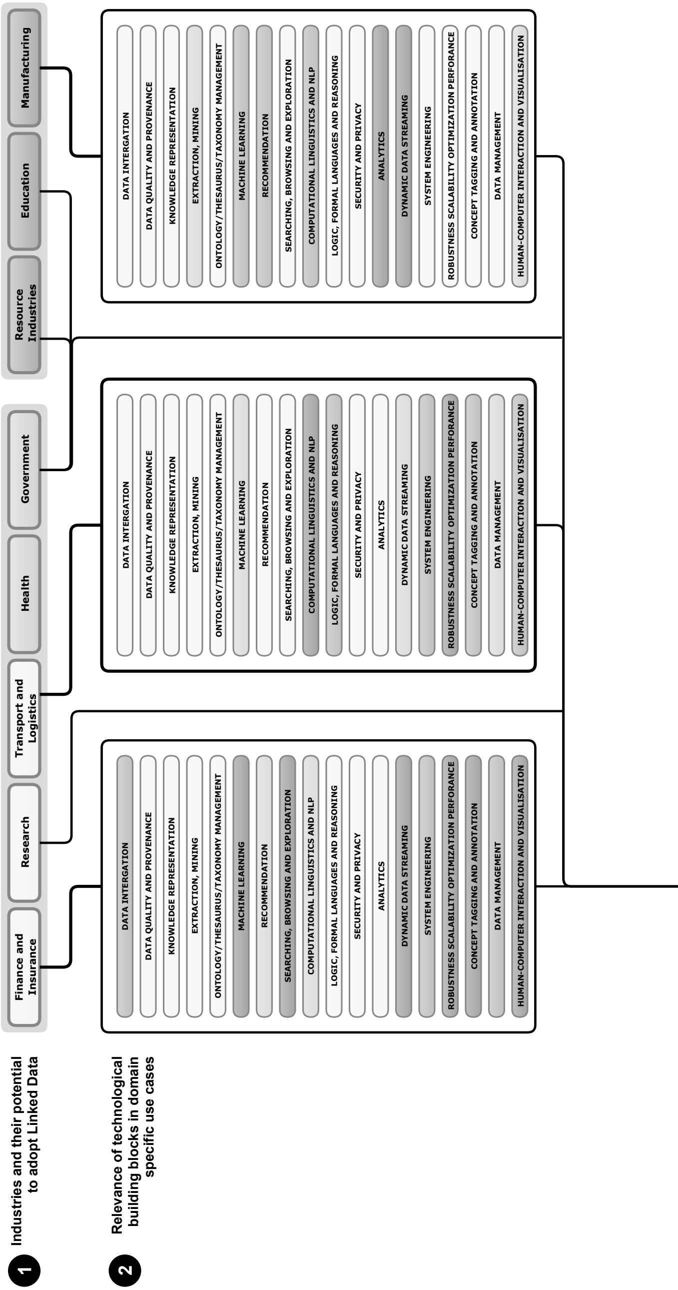
In the long-term, the funding strategy could go beyond encouraging adoption in individual enterprises and focus on industry wide actions and adoption, going beyond national borders. Such funding would particularly target industries that have an international span, such as Manufacturing. We recommend a strategy of cooperation with international funding schemes, such as CHIST-ERA to establish and foster a European level agenda and ensure access to this by Austrian companies/universities.

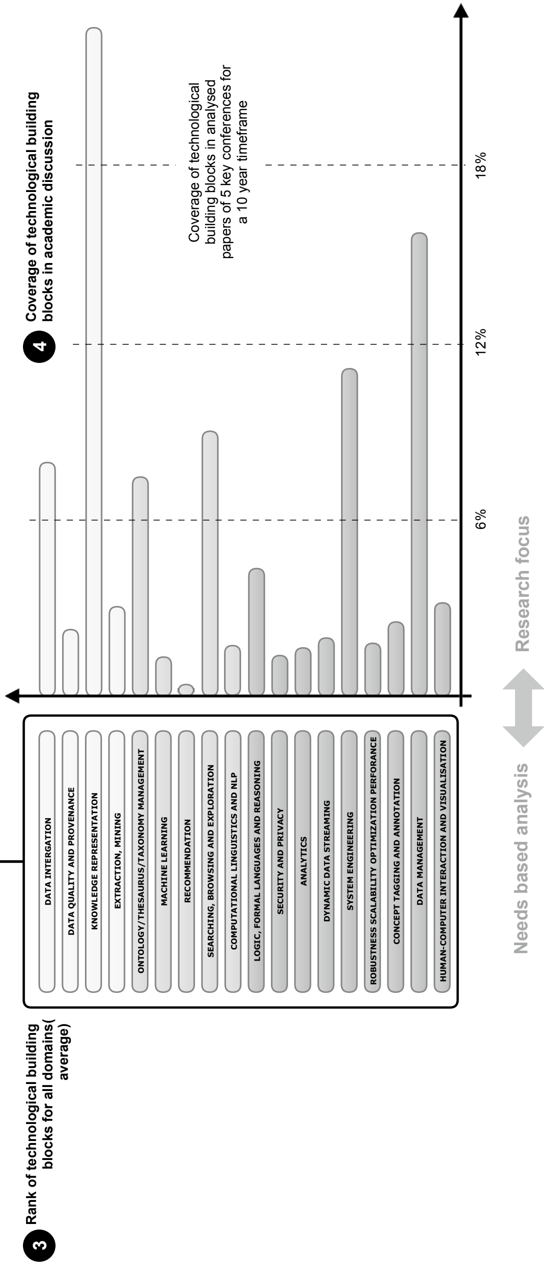
92 http://www.research.ie/enterprise/enterprise-partnership-scheme-eps

8. CONCLUSIONS

8.1. At a glance

This preparatory study takes a more concentrated look into the potential of Linked Data in Enterprises. By examining ongoing research and development efforts under the lens of concrete data and information management challenges in enterprises it is possible to assess the suitability and readiness of the technologies developed by the Linked Data community for adoption by a range of industry domains.


1 Industries and their potential to adopt Linked Data
2 Relevance of technological building blocks in domain specific use cases
Finance and Insurance
Research
Transport and Logistics
Health
Government
Resource Industries
Education
Manufacturing
DATA INTERGATION
DATA QUALITY AND PROVENANCE
KNOWLEDGE REPRESENTATION
EXTRACTION, MINING
ONTOLOGY/THESAURUS/TAXONOMY MANAGEMENT
MACHINE LEARNING
RECOMMENDATION
SEARCHING, BROWSING AND EXPLORATION
COMPUTATIONAL LINGUISTICS AND NLP
LOGIC, FORMAL LANGUAGES AND REASONING
SECURITY AND PRIVACY
ANALYTICS
DYNAMIC DATA STREAMING
SYSTEM ENGINEERING
ROBUSTNESS SCALABILITY OPTIMIZATION PERFORANCE
CONCEPT TAGGING AND ANNOTATION
DATA MANAGEMENT
HUMAN-COMPUTER INTERACTION AND VISUALISATION


3 Rank of technological building blocks for all domains(average)
DATA INTERGATION
DATA QUALITY AND PROVENANCE
KNOWLEDGE REPRESENTATION
EXTRACTION, MINING
ONTOLOGY/THESAURUS/TAXONOMY MANAGEMENT
MACHINE LEARNING
RECOMMENDATION
SEARCHING, BROWSING AND EXPLORATION
COMPUTATIONAL LINGUISTICS AND NLP
LOGIC, FORMAL LANGUAGES AND REASONING
SECURITY AND PRIVACY
ANALYTICS
DYNAMIC DATA STREAMING
SYSTEM ENGINEERING
ROBUSTNESS SCALABILITY OPTIMIZATION PERFORANCE
CONCEPT TAGGING AND ANNOTATION
DATA MANAGEMENT
HUMAN-COMPUTER INTERACTION AND VISUALISATION
4 Coverage of technological building blocks in academic discussion
Coverage of technological building blocks in analysed papers of 5 key conferences for a 10 year timeframe
6%
12%
18%
Needs based analysis
Research focus

1. **Industries and their potential to adopt Linked Data**
   The sectoral analysis presented in *Chapter 4* shows, that Energy, Retail, Finance and Insurance, Government, Health, Transport and Logistics, Telecommunications, Media, Tourism, Engineering, and Research and Development are among the industries most likely to adopt Linked Data technologies. What unites these industries is their data-drivenness and intense intra- and inter-organisational information flows. A low web-focus, however, could inhibit possible Linked Data efforts in the health industry. Similarly, the government sector, needs to be aware of its relatively low internationalisation and knowledge-intensity values.
2. **Relevance of technological building blocks in domain specific use cases**
   The demand side analysis, which was presented in Chapter 5 led to the collection of over 60 user stories and the classification of industry specific technological demands. As it is not feasible to go into detail on each and every use case, we have chosen to focus on three of the above industries: Finance and Insurance, Transport / Logistics and Manufacturing. Based on our analysis these industries have high potential from a market and industry perspective.
3. **Rank of technological building blocks for all domains (average)**
   Based on our analysis of the user stories Data Integration, Data Quality and Provenance, Extraction and Mining, and Knowledge Representation are the most important technology building blocks, followed by Taxonomy Management, Machine Learning, Recommendation, Searching/Browsing/Exploration and Computational Linguistics /NLP.
4. **Coverage of technological building blocks in academic discussion**
   Our community/academic analysis which was presented in Chapter 6, based on the same technological building blocks that were used to classify industry needs, shows that academic priorities differ from business needs. Here Knowledge Representation, Data Management, System Engineering and Searching/Browsing/Exploration, and Data Integration, Ontology/ Thesaurus/Taxonomy management receive the most attention.
5. **Gap between discussed users' needs and actual research activity**
   Based on points 3.) and 4.) it is clear that industry needs and research interests / priorities are not tightly coupled.

A strong recommendation derived from the above observations is that targeted measures are necessary in order to foster a stronger coupling between industry needs and research priorities.

8.1.1. SWOT

Strengths

Awareness & Education Few but very active players in the field. Academic and commercial organisations have a long tradition to partner in awareness raising: Vienna Semantic Web Meetup, SEMANTiCS Conference Series	**Standardisation** There are a number of Austrian organisations already actively involved in standardisation.	**Technological Innovation, Gaps and Research Question** Austria is a strong innovator, with innovation performance above or close to that of the EU average. There are several national and european projects (ADEQUATe, DALICC, SPECIAL) that are already tackling some of the open research questions.
Legal & Policy Good coupled eco-system of research, government and industry allows tailormade policies to be developed	**Funding** • well-positioned to fund basic science and applied research in market areas/industries that are of strategic relevance • encourage transfer from local universities to local companies	

Awareness & Education	**Standardisation** Setup and develop a standardisation certification scheme.	**Technological Innovation, Gaps and Research Question** General Data Protection Regulation will come into effect in May 2018, if Austria were to invest in related topics Enterprises could gain competitive advantage. Austria already has a number of relevant competency centres (i.e. STI, The TU Linked Data Lab, The WU Privacy & Sustainable Computing Lab).
Legal & Policy Liaison between legal sector (lawyers) and research and development in technology may be organized easily Intensive involvement into standardisation to multiply efforts	**Funding**	

Opportunities

Weaknesses

Awareness & Education Austrian academic and industry have to find partners and customers abroad and have no "home market" for awareness raising.	**Standardisation**	**Technological Innovation, Gaps and Research Question**
Legal & Policy For ELD relevant industry is not head-quartered in Austria	**Funding** • High access barrier to funding • Difficulty to fund industry-wide action and adoption at an international level	

Awareness & Education	**Standardisation** Standardisation bodies decide to focus their efforts elsewhere.	**Technological Innovation, Gaps and Research Question** Shortage in numbers of suitably qualified science and engineering graduates entering postgraduate studies. Cuts to core fundamental and applied research budgets.
Legal & Policy Lack of harmonized national policy in EC, creates competition instead of fostering specialisation Disperse situation on european legal landscape hinders generalized solutions	**Funding** • unsure access to European-level funds • lack of control on the topics funded at European-level	

Threats

8.1.2. Timeline over fields

Fields of activity	Outlook			Goal
	SHORT TERM →	MIDDLE TERM →	LONG TERM →	
Awareness & Education	• **Cluster existing players** » develop a Linked Data cluster agenda » ensure visibility in an international context » develop joint projects	• **Relate academic and private education** » develop a joint certification scheme » exchange of curricula, training materials and lecturers		• Ensure a solid foundation for ELD • Topic wise strengthened group of people • Sharpen the profile of LD
Technological Innovation, Gaps and Research Question	• **Capacity Building** » Incentivise organisations to develop offline and online training tailored to cater for the skills and competencies that are necessary for the adoption of Linked Data in enterprises » Preparatory studies in the foundational technologies that have been identified as key drivers and/or barriers for Linked Data adoption in enterprises	• **Collaboration** » Initiate projects in the identified key foundational technologies » The organisation of national events that provide a platform for the dissemination of research outputs, support discussion on barriers and drivers, and enable public-private and interdisciplinary matchmaking.	• **Development of centers of excellence** » Provide the supports necessary to assist in the setup and development of centers of excellence, including but not limited to financial, administrative, communication, innovation	• Develop technical competencies in the foundational technologies that have been identified as key drivers and/or barriers for Linked Data adoption in enterprises • Foster the development of multidisciplinary teams and public private partnerships such that technical research on the aforementioned technologies is informed by societal and business requirements • Provide a platform for Austrians to showcase technological results from nationally and internationally funded projects • Support the setup and development of centers, starting with Linked Data and moving beyond to Big Data Analytics and the Internet of Things

Standardisation	• **Capacity Building** » Incentivise organisations to develop both offline and online training in the form of tutorials and courses, on existing Semantic Web standards » Launch preparatory studies into the degree of coverage that is offered by existing W3C standards and the identification of alternative standardisation activities	• **Encouraging Industrial uptake** » Encourage more organisations to get involved in standardisation activities » Incentivise standards bodies to set up certification schemes such as those offered by multinationals such as Microsoft, Oracle, SAP etc.		• Identify gaps and priorities in terms of technical standards that are necessary to propel the potential of Linked Data in enterprises. • Encourage both industry and academia to get more involved in standardisation • Support the setup and development of a standardisation certification scheme
Legal & Policy	• Cluster existing initiatives, players and policy schemes » develop an own cluster agenda » ensure visibility in an international context • Encourage tandem research projects between legal sector and technology development » Sondierungsprojekte » active partnership development initiatives	• Flagship projects on ELD legal clearing centers » Demonstrators based on national use cases » Broaden findings and solutions to a transnational blueprint » Couple outcomes of the flagship projects with the standardisation efforts	• Serving the specific niche with tailor made technological solutions and profound legal expertise » Establish technology driven legal clearing linked data in the academic research agenda » Encourage IT-companies to exploit commercial products in that specific field	• Policy makers to create a positive momentum for the further development of ELD • Ensure the liaison between legal sector (lawyers) and technology development • Playing a well recognized role in standardisation and modelling of legal description language and application
Funding	• Fund most urgent research topics and most promising industries • Funding schemes for *lightweight technology transfer*	• Long-term funding schemes for combining basic and applied research, such as Christian Doppler Laboratories	• Cooperation with international schemes, such as CHIST-ERA to establish and foster a European level agenda and ensure access to this by Austrian companies/ universities	• Foster research on topics relevant for enterprise adoption • Ensure technology transfer from universities to companies • Ensure industry wide-actions and adoption
↓	↓	↓	↓	
Generic	Build on existing strength in the Austrian ELD/LD landscape, cluster players and efforts. Get momentum from new funding lines and supporting studies and pilot projects	Lay ground for institutional and technological focussing towards industry relevant research and development in dedicated niches of ELD/LD	Consolidate efforts from all fields to establish internationally recognized research centers and commercial activity in the in dedicated niches of ELD/LD	

8.2. Take home messages

The PROPEL study shows that ELD has a lot of potential in industry, and could in fact be a driver of technological, economical and procedural change. In this chapter, we present flash lights of potentials, measures and emerging ideas in relation to ELD.

8.2.1. Research and development need to be strongly coupled with industry demands

There is a general consensus that there is a need for a tighter coupling between research agendas in the academic sector and practical needs from the industry side. The success of the roadmap is dependant on buy-in by all stakeholder groups:

- More customer involvement in business processes and (open) innovation
- More industry relevance in the research agenda
- More courage in applying cutting-edge technologies in industry

8.2.2. Develop centres of excellence out of existing collaboration and expertise

Active players from academia, industry and the public are already well connected and thus capable of joining forces in order to propel the potential of Linked Data in enterprises. When it comes to basic research, applied developments and business innovation centres of excellence can leverage this potential and act as a driving force.

"Key stakeholders who can further drive such approaches are universities and the research sector as long as they (still) function as open information systems. Therefore, it is necessary to guarantee open/free spaces in the academic environment".

Founder of Austrian Start-up on Semantic Technology

8.2.3. Subsidise the ELD eco-system

Considering the fact that impact can be measured as a function of the number of adopters, ELD shows exemplary economic network effects. However, flagship implementations and pioneering projects are key to furthering the growth of ELD in Austria. Both financial and infrastructural support are necessary in order to accelerate the development of the sector. Core preparatory steps include:

- Base infrastructures (stores, services, data) to build solutions on top
- Project related funding

8.2.4. Use the power of ELD for disruptive industries

Many industries face disruptive changes that force them to come up with new business models. Traditional solutions don't provide the results necessary to meet the demands of the digital transformation. Even conservative industries such as financial services or steel production see the need for a "two speed IT" that incorporates the ability to respond to disruptive threats. Linked Data is highly developed and has proved successful in multiple data and information management use cases. As such, organisations that are serious about digital transformation could potentially leverage Linked Data. The redevelopment of companies coupled with disruptive developments, present companies with the opportunity to adopt ELD, facilitating new business narratives, new processes and new products.

8.2.5. ELD is the backbone for the upcoming content industry

Linked Data is especially relevant for online businesses (media, e-commerce, etc). Given that ELD has its strength as an underpinning technology for new business narratives, ELD provides tools and approaches that could be used to build on the internet typical network effects on which the upcoming content industry is based. Aside from Linked Data principles, the community has also been actively involved in the development of tools (ontologies, RDF, NLP) that enrich digital products and makes them exchangeable within a broader digital environment.

8.2.6. ELD has to convince for a change

While, well established industries have their own home grown IT environments, ELD does not have this level of specialisation and thus has to be adapted beforehand. Therefore, the integration of ELD in running environments means that technological, behavioural and cultural hurdles need to be overcome. For instance, it took a long time to convince industry to use SQL for data management. To the same extent there is now resistance against "non-relational" models like Linked Data, where new standards and new skills (like SPARQL) are required. In order to instigate change, ELD has to:

- show, that ELD addresses the general demands from business perspective: present clear business cases, fast returns, tangible and quantifiable benefits.
- be piloted in low-barrier experimental/open source/freemium solutions that can be used in an experimental/try-out phase of industry implementation.
- provide answers to security/privacy/compliance issues which are still a weakness of ELD.

9. REFERENCES

- Abulafia, D. (2014). Thalassocracies. In P. Horden and S. Kinoshita (eds.). A Companion to Mediterranean History, Oxford.
- Al-Ghassani, A. M. (2003). Improving the structural design process: a knowledge management approach (Doctoral dissertation, © Ahmed M. Al-Ghassani). Chicago. Available at: https://dspace.lboro.ac.uk/dspace-jspui/bitstream/2134/7605/3/Thesis-2003-Al-Ghassani.pdf (accessed on 22.10.2016).
- Altbach, P. G., and J. Knight (2007). The internationalization of higher education: Motivations and realities. Journal of studies in international education, 11(3-4), 290-305. Available at: http://www.pef.uni-lj.si/ceps/dejavnosti/sp/2011-06-02/Internac.%20HE%20-%20motivations%20and%20realities%20Altback,%20Knight%2007.pdf (accessed on 22.10.2016).
- Arias, J., C. Severino, J. A. Tortosa and E. Manso (2014). The cost of Complexity for Telecoms and how Companies can reduce it. Strategyand. Available at: http://www.strategyand.pwc.com/media/file/Strategyand_The-cost-of-complexity-for-telecoms.pdf (accessed on 22.10.2016).
- Auer, S., J. Lehmann, A.-C. N. Ngomo and A. Zaveri (2013). Introduction to Linked Data and its Lifecycle on the Web. In Reasoning Web.
- Austrian Chamber of Commerce (2016). Wirtschaftslage und Prognose. Available at: http://wko.at/statistik/prognose/arbeitslose.pdf (accessed on 17.11.2016).
- Bauer, F., D. Recheis and M. Kaltenböck (2011). Data. reegle. info–A new key portal for Open Energy Data. In Environmental Software Systems. Frameworks of eEnvironment (pp. 189-194). Springer Berlin Heidelberg. Available at: https://reeep.org/sites/default/files/reegle%20Open%20Data%20paper%20ISESS.pdf (accessed on 22.10.2016).
- Bauernhansl, T., M. ten Hompel and B. Vogel-Heuser (2014). Industrie 4.0 in Produktion, Automatisierung und Logistik, Springer.
- Bealey, F., and A. G. Johnson (1999). The Blackwell Dictionary of Political Science: A User's Guide to Its Terms. Blackwell Publishing Ltd. Available at: http://bayanbox.ir/view/1259615488552050843/Blackwell-Dictionary-of-Politics.pdf (accessed on 22.10.2016).
- Bentley, J. (1993). Old World Encounters: Cross-Cultural Contacts and Exchanges in Pre-Modern Times. New York: Oxford University Press.
- Berners-Lee, T., J. Hendler, and O. Lassila (2001). The Semantic Web. *Scientific American* 284.5: 28-37.
- Bernstein, A., J. Hendler, and N. Noy (2016). A new look at the Semantic Web *Communications of the ACM* 59.9: 35-37.
- Bertelsen, S. (2003). Construction as a complex system. In proceedings of IGLC (Vol. 11). Available at: http://leanconstruction.dk/media/16764/Construction%20as%20a%20Complex%20System.pdf (accessed on 22.10.2016).
- Biffl, S. and M. Sabou (2016). Semantic Web for Intelligent Engineering Applications. Springer

- Bizer, C., T. Heath and T. Berners-Lee (2009). Linked Data—The Story So Far. International Journal on Semantic Web and Information Systems. 5 (3): 1–22.
- Brandl, K. (2010). Perspectives on Globalization and the Danish Architecture and Engineering Industry. Available at: http://www.danskeark.dk/Medlemsservice/Internationalt/~/media/Dark/Medlemsservice-dokumenter/Internationalt/Globaliseringsundersoegelse/Globaliseringsrapporten(UK)-final.ashx (accessed on 22.10.2016).
- Brinkley, I. (2006). Defining the knowledge economy - Knowledge economy programme report. The work foundation. Available at: http://www.theworkfoundation.com/DownloadPublication/Report/65_65_defining%20knowledge%20economy.pdf (accessed on 22.10.2016).
- Brown, D. (2015). IDC Predicts GDPR Will Create a $3.5B Market Opportunity for Security and Storage Vendors. Press release on 3.11.2015.
- Caracciolo, C., A. Stellato, A. Morshed, G. Johannsen, S. Rajbhandari, Y. Jaques and J. Keizer (2013). The AGROVOC Linked Dataset. Semantic Web, 4(3), 341-348. Available at: http://www.semantic-web-journal.net/system/files/swj274_1.pdf (accessed on 22.10.2016).
- Carolan, L., F. Smith, V. Protonotarios, B.Schaap, E. Broad, J. Hardinges, and W. Gerry (2015). How Can We Improve Agriculture, Food, and Nutrition with Open Data? The Open Data Institute (ODI). Available at: http://www.godan.info/wp-content/uploads/2015/04/ODI-GODAN-paper-27-05-20152.pdf (accessed on 22.10.2016).
- Cata, T., and S. M. Lee (2006). Adoption of web-based applications in the financial sector: The case of online insurance. Journal of Internet commerce, 5(2), 41-61. Available at: http://www.tandfonline.com/doi/abs/10.1300/J179v05n02_03?journalCode=wico20 (accessed on 22.10.2016).
- Chatti, M. A., V. Lukarov, H. Thüs, A. Muslim, A. M. F. Yousef, U. Wahid, C. Greven, A. Chakrabarti and U. Schroeder (2014). Learning Analytics: Challenges and Future Research Directions. eleed, Iss. 10. Available at: https://eleed.campussource.de/archive/10/4035 (accessed on 22.10.2016).
- Christiansen, A.C. et al. (2015). IDC FutureScape: Worldwide IT Security Products and Services 2016 Predictions, Doc # 259836.
- Chua, C.K., D. Schubmehl and Q. Li (2016). IDC PeerScape: Practices for Cognitive Systems Initiatives, May 2016, #AP41156315.
- Claessen, F. N., and H. La Poutré (2014). Towards a European Smart Energy System. Available at: https://www.eitdigital.eu/fileadmin/files/docs/Action_Lines/SES/Towards_a_European_Smart_Energy_System.pdf (accessed on 22.10.2016).
- Clegg, J., and S. Kamall (1998). The internationalization of telecommunications services firms in the European Union. Transnational Corporations, 7, 39-96. Available at: http://unctad.org/en/docs/iteiit9v7n2_en.pdf#page=44 (accessed on 22.10.2016).
- Cortada, J. W., D. Gordon and B. Lenihan (2012). The value of analytics in healthcare. IBM Institute for Business Value IBM, Global Business Service. Available at: https://www.ibm.com/smarterplanet/global/files/the_value_of_analytics_in_healthcare.pdf (accessed on 22.10.2016).

- Cottle, M., W. Hoover, S. Kanwal, M. Kohn, T. Strome and N. W. Treister (2013). Transforming Health Care Through Big Data. Institute for Health Technology Transformation, Washington DC, USA. Available at: http://c4fd63cb482ce6861463-bc6183f1c18e748a49b87a25911a0555.r93.cf2.rackcdn.com/iHT2_BigData_2013.pdf (accessed on 22.10.2016).
- Coviello, N. E. (1994). Internationalizing the Entrepreneurial High Technology, Knowledge-intensive Firm'. PhD Thesis. Department of Marketing, University of Auckland, New Zealand.
- Creswell, J. W. (2009). Research design: Qualitative, quantitative, and mixed methods approaches (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.
- Dapp, T. F. (2014). Fintech–The digital (r) evolution in the financial sector. Deutsche Bank Research. Available at: https://www.dbresearch.com/PROD/DBR_INTERNET_EN-PROD/PROD0000000000345837.pdf (accessed on 22.10.2016).
- Deloitte (2015). Global powers of retailing. Available at: https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Consumer-Business/gx-cb-global-powers-of-retailing-2016.pdf (accessed on 22.10.2016).
- Efthymiou, K., A. Pagoropoulos, N. Papakostas, D. Mourtzis and G. Chryssolouris (2012). Manufacturing systems complexity review: challenges and outlook. Procedia CIRP, 3, 644-649. Available at: http://www.sciencedirect.com/science/article/pii/S221282711200282X (accessed on 22.10.2016).
- Egbu, C. O., and H. S. Robinson (2005). Construction as a knowledge-based industry. Knowledge management in construction, 4, 31-49. Available at: http://onlinelibrary.wiley.com/doi/10.1002/9780470759554.ch3/summary (accessed on 22.10.2016).
- Elisseeff, V. (2001). The Silk Roads: Highways of Culture and Commerce. UNESCO Publishing / Berghahn Books.
- ElMaraghy, W., H. ElMaraghy, T. Tomiyama and L. Monostori (2012). Complexity in engineering design and manufacturing. CIRP Annals-Manufacturing Technology, 61(2), 793-814. Available at: http://www.sciencedirect.com/science/article/pii/S0007850612002004 (accessed on 22.10.2016).
- Elsner, S. (2013). Retail Internationalization: Analysis of Market Entry Modes, Format Transfer and Coordination of Retail Activities. Springer Science and Business Media.
- Ericsson (2014). ICT and The Future of Financial Services. Available at: http://www.ericsson.com/industry-transformation/wp-content/uploads/sites/6/2014/11/ict-and-the-future-of-financial-services.pdf (accessed on 22.10.2016).
- Ericsson (2015). ICT and The Future of Retail. Available at: http://www.ericsson.com/id/res/docs/2015/ict-and-the-future-of-retail.pdf (accessed on 22.10.2016).
- Eröcal, D. (2005). Case Studies of Successful Companies in the Services Sector and Lessons for Public Policy, OECD Science, Technology and Industry Working Papers, 2005/07, OECD Publishing. Available at: http://dx.doi.org/10.1787/863338840653 (accessed on 22.10.2016).

- European Commission (2010). A digital agenda for Europe. Publications Office of the European Union. Available at: http://europa.eu/pol/pdf/flipbook/en/digital_agenda_en.pdf (accessed on 22.10.2016).
- European Commission (2014). Enforcement of the Third Internal Energy Market Package. Commission Staff Working Document. Available at: https://ec.europa.eu/energy/sites/ener/files/documents/2014_iem_communication_annex6_0.pdf (accessed on 22.10.2016).
- Eurostat (2013). Science, technology and innovation in Europe. European Commission. Available at: http://ec.europa.eu/eurostat/documents/3930297/5969406/KS-GN-13-001-EN.PDF (accessed on 22.10.2016).
- Feigenbaum, L., et al. (2007). The Semantic Web in action. *Scientific American* 297.6: 90-97.
- Feldmann, S., D. Schubmehl and H. Reynolds (2012). White Paper: Content Analytics and the High Performing Enterprise.
- Feldmann, S., S.J.I. Herzig, K. Kernschmidt, T. Wolfenstetter, D. Kammerl, A. Qamar, U. Lindemann, H. Krcmar, C.J.J. Paredis, and B.Vogel-Heuser (2015). Towards Effective Management of Inconsistencies in Model-Based Engineering of Automated Production Systems. In: Proc. of IFAC Symposium on Information Control in Manufacturing.
- Forbes, R. J. (1993). Studies in ancient technology, Volume 2. Brill.
- Forza, C., and F. Salvador (2001). Information flows for high-performance manufacturing. International Journal of Production Economics, 70(1), 21-36. Available at: http://www.sciencedirect.com/science/article/pii/S0925527300000384 (accessed on 22.10.2016).
- Fox, B., R. van den Dam, and R. Shockley (2013). Analytics: Real-World Use of Big Data in Telecommunications. IBM Institute for Business Value. Available at: http://www-935.ibm.com/services/multimedia/Anaytics.pdf (accessed on 22.10.2016).
- Fraunhofer (2012). Information Technology for the Mechanical Engineering Industry of the Future. Available at: https://www.iosb.fraunhofer.de/servlet/is/25994/e_Fraunhofer_Mechanical%20Engineering_Web.pdf?command=downloadContent&filename=e_Fraunhofer_Mechanical%20Engineering_Web.pdf (accessed on 22.10.2016).
- Fromm, J. A., and T. V. Kern (2000). Investment Opportunities in Education: Making a Profit While Making a Difference. The Journal of Private Equity, 3(4), 38-51.
- Gajendran, T., G. Brewer and M. Marimuthu (2013). Internationalisation of construction business and e-commerce: Innovation, integration and dynamic capabilities. Construction Economics and Building, 13(2), 1-17. Available at: https://epress.lib.uts.edu.au/journals/index.php/AJCEB/article/viewFile/3321/3489 (accessed on 22.10.2016).
- Gens, F. (2015). IDC's Worldwide IT Cloud Services Taxonomy.
- Gens, F. (2015a). IDC FutureScape: Worldwide IT Industry 2016 Predictions — Leading Digital Transformation to Scale.

- George, T., et al. (2011). ICT in agriculture : connecting smallholders to knowledge, networks, and institutions. Washington, DC: World Bank. Available at: http://www.ictinagriculture.org/sites/ictinagriculture.org/files/ICT%20in%20Agriculture%20e-Sourcebook_0.pdf (accessed on 22.10.2016).
- Groth, P., A. Loizou, A. J. G Gray, C. Goble, L. Harland, and S. Pettifer (2014). API-centric Linked Data integration: The Open PHACTS Discovery Platform case study, Web Semantics: Science, Services and Agents on the World Wide Web, Volume 29, Pages 12-18.
- Groves, P., B. Kayyali, D. Knott and S. Van Kuiken (2013). The 'big data' revolution in healthcare. McKinsey Quarterly, 2. Available at: http://www.pharmatalents.es/assets/files/Big_Data_Revolution.pdf (accessed on 22.10.2016).
- Hagenhoff, S. (2006). Current Challenges in the Media Industry and ICT based solutions. Institut für Wirtschaftsinformatik. Working Paper No 14/2006. Available at: http://webdoc.sub.gwdg.de/ebook/serien/lm/arbeitsberichte_wi2/2006_14.pdf (accessed on 22.10.2016).
- Hannoun, H. (2006). Internationalisation of financial services: implications and challenges for central banks. Available at: http://www.bis.org/speeches/sp060304.htm (accessed on 22.10.2016).
- Havens, T., A.D. Lotz and S. Tinic (2009). Critical media industry studies: A research approach. Communication, Culture and Critique, 2(2), 234-253. Available at: http://sites.lsa.umich.edu/lotz/wp-content/uploads/sites/45/2014/02/CCC-CMIS.pdf (accessed on 22.10.2016).
- Heath, T. and C. Bizer (2011). Linked Data: Evolving the Web into a Global Data Space (1st Edition), Volume 1 of Synthesis Lectures on the Semantic Web: Theory and Technology.
- Heath, T. and C. Bizer (2011). Linked Data: Evolving the web into a global data space. Synthesis lectures on the Semantic Web: theory and technology, 1(1), 1-136. Available at: http://info.slis.indiana.edu/~dingying/Teaching/S604/LODBook.pdf (accessed on 22.10.2016).
- Hey, T., S. Tansley, and K. Tolle, eds. (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research.
- Högselius, P. (2009). The internationalization of the European electricity industry: The case of Vattenfall. Utilities Policy, 17(3), 258-266. Available at: www.sciencedirect.com/science/article/pii/S0957178709000022 (accessed on 22.10.2016).
- Holden, C. (2003). Actors and motives in the internationalization of health businesses. Business and Politics, 5(3), 287-302. Available at: http://www.tandfonline.com/doi/abs/10.1080/1369525042000189410 (accessed on 22.10.2016).
- Holster, H. C., S. Horakova, A. H. Ipema, B. Fusai, G. Giannerini, F. Teye and O. Schmid (2012). Current situation on data exchange in agriculture in the EU27 and Switzerland (No. D2. 4, p. 84). AgriXchange. Available at: http://www.efita.net/apps/accesbase/bindocload.asp?d=6913&t=0&identobj=VpugfXnk&uid=57305290&sid=57&idk=1 (accessed on 22.10.2016).
- Houghton, J. (2002). Information technology and the revolution in healthcare. Pharmaceutical Industry Project. Equity, Sustainability and Industry Development. Working Paper Series. Available at: http://www.eldis.org/vfile/upload/1/document/1201/Information%20Technology%20and%20Revolution%20in%20Healthcare.pdf (accessed on 22.10.2016).

- Hult, G.T.M., S. T. Cavusgil, and R. J. Calantone (2003). Knowledge as a Strategic Resource in Logistics and Purchasing. Working Paper Series. Marketing Science Institute Reports. Available at: https://global.broad.msu.edu/hult/publications/msi03.pdf. (accessed on 22.10.2016).
- ICLEI (2011). The Contribution of ICT to Energy Efficiency: Local and Regional Initiatives Implementing energy efficiency initiatives harnessing ICT: A toolkit for local and regional authorities. Available at: http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=4266 (accessed on 22.10.2016).
- ICTC (2012). ICT in the Financial Services Sector. Available at: http://www.ictc-ctic.ca/wp-content/uploads/2012/09/ICTC_ICT_in_FinancialService_EN_9_12.pdf (accessed on 22.10.2016).
- IDC (2014). Market Analysis: Worldwide Content Analytics 2014–2018 Forecast, IDC #249185.
- IDC (2015). IT-Spending Survey Austria.
- IDC (2016a). IT-Security Umfrage, Mai 2016.
- IDC (2016b). Digital Transformation Austria 2016, Online Survey, Juli 2016.
- IDC (2016c). Austria IT Services Market 2016–2020 Forecast and 2015 Analysis.
- IDC (2016d). Big Data Survey Austria. Update zum Big Data und Analytics Markt für Österreich.
- Jespersen, L. M., J. P. Hansen, G. Brunori, A. L.Jensen, K. Holst, C. Mathiesen and I. A. Rasmussen (2014). ICT and social media as drivers of multi-actor innovation in agriculture–barriers, recommendations and potentials. European Commission, Directorate-General for Research. Available at: http://icrof.eu/pdf/2013%20SCAR_CWG_AKIS_SocialmediaICTandinnovationprocesses%20report%2031%2005%202013.pdf (accessed on 22.10.2016).
- Jung, W., S. H. Han, H. Park and D. Y. Kim (2010). Empirical assessment of internationalization strategies for small and medium construction companies. Journal of construction engineering and management, 136(12), 1306-1316.
- Kamrani, A., A. Adat and M. Azimi (2010). A methodology for analysis of manufacturing operations due to complexity. International Journal of Industrial Engineering: Theory, Applications and Practice, 18(2). Available at: http://journals.sfu.ca/ijietap/index.php/ijie/article/download/325/195 (accessed on 22.10.2016).
- Kanagarajah, S. (2005). Business Dynamics in Canada. Statistics Canada. Chicago.
- Kayyali, B., D. Knott and S. Van Kuiken (2013). The big-data revolution in US health care: Accelerating value and innovation. Mc Kinsey and Company, 1-13. Available at: http://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-big-data-revolution-in-us-health-care (accessed on 22.10.2016).
- Kleis, L., P. Chwelos, R. V. Ramirez and I. Cockburn (2012). Information technology and intangible output: The impact of IT investment on innovation productivity. Information Systems Research, 23(1), 42-59. Available at: http://www.ucdenver.edu/academics/colleges/business/about/Faculty-Research/workingPapers/test/Ramirez_2011_ITandInnovation.pdf (accessed on 22.10.2016).

- Koczanski, A. and M. Sabou (2015). Open Government Data for Business and Social Innovation. Project report for Wirtschaftskammer Wien. Available at: http://data.ifs.tuwien.ac.at/ogda/observatory/Final_Report_Sabou.pdf (accessed on 17.11.2016).
- Koczanski, A. and M. Sabou (2015a). Sustainability Implications of Open Government Data: A Cross-Regional Study. Proc. of the ACM Web Science Conference.
- Kok, W. (2004). Facing the challenge. The Lisbon Strategy for Growth and Employment. Report from the High Level Group Chaired by Wim Kok. Luxembourg: Office for Official Publications of the European Communities. Available at: https://ec.europa.eu/research/evaluations/pdf/archive/fp6-evidence-base/evaluation_studies_and_reports/evaluation_studies_and_reports_2004/the_lisbon_strategy_for_growth_and_employment__report_from_the_high_level_group.pdf (accessed on 22.10.2016).
- Kunisch, M., J. Frisch, D. Martini and S. Böttinger (2007). agroXML – a standardized language for data exchange in agriculture. In Proceedings of 6th Biennial Conference of European Federation of IT in Agriculture Glasgow. Available at: http://www.informatique-agricole.org/download/efita-conference/Congres_EFITA_2005/PA111-%20Reiner%20Doluschitz.pdf (accessed on 22.10.2016).
- Kvavik, R. B., P. J. Goldstein and J. Voloudakis (2005). Good Enough!: IT Investment and Business Process Performance in Higher Education. EDUCAUSE. Available at: http://er.educause.edu/~/media/files/articles/2008/1/ers0504w.pdf?la=en (accessed on 22.10.2016).
- Layne, K., and J. Lee (2001). Developing fully functional E-government: A four stage model. Government information quarterly, 18(2), 122-136. Available at: http://www.sciencedirect.com/science/article/pii/S0740624X01000661 (accessed on 22.10.2016).
- Lee, F. C. and H. Has (1996). A quantitative assessment of high-knowledge industries versus low-knowledge industries. In: Howitt, P. (ed.): The implications of knowledge-based growth for micro-economic policies. Calgary: University of Calgary Press, S. 39-77.
- Levinson, M. (2008). The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger. Princeton University Press.
- Lund, S., T. Daruvala, R. Dobbs, P. Härle, J. Kwek and R. Falcón (2013). Financial globalization: Retreat or reset? McKinsey Global Institute. Available at: http://www.mckinsey.com/global-themes/employment-and-growth/financial-globalisation (accessed on 22.10.2016).
- Lutes, T. (2015). Data-driven government: Challenges and a path forward. IBM Analytics white-paper. Available at: http://www.smartresilient.com/system/files/IBM_Data-driven_Government_Challenges_WP.pdf (accessed on 22.10.2016).
- Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh and A. H. Byers (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. Available at: http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation (accessed on 22.10.2016).

- Mattsson, S., P. Gullander and A. Davidsson (2011). Method for measuring production complexity. In 28th International Manufacturing Conference. Available at: http://publications.lib.chalmers.se/records/fulltext/local_146662.pdf (accessed on 22.10.2016).
- McDermott, P. (2010). Building open government. Government Information Quarterly, 27(4), 401-413. Available at: http://igup.urfu.ru/docs/Bank%20English_Transleted%20Articles/English/HR%20Management/Building%20open%20government.pdf (accessed on 22.10.2016).
- Miles, I. (2003). Knowledge intensive services' suppliers and clients (p. 81). Ministry of Trade and Industry Finland. Available at: http://ktm.elinar.fi/ktm_jur/ktmjur.nsf/12b74ae4d1122aadc22565fa003211a6/f0ed7947073a4ac3c2256de4003cf376/$FILE/tura15teoeng.pdf (accessed on 22.10.2016).
- Mohanty, R. P. and A. K. Behera (1996). TQM in the service sector. Work Study, 45(3), 13-17. Available at: http://www.emeraldinsight.com/doi/full/10.1108/00438029610115479 (accessed on 22.10.2016).
- Nadkarni, A. (2015). Press Release: New IDC Forecast Sees Worldwide Big Data Technology and Services Market Growing to $48.6 Billion in 2019, Driven by Wide Adoption Across Industries, 09 Nov 2015.
- Nadkarni, A. and D. Vesset (2015). Market Forecast: Worldwide Big Data Technology and Services Forecast, 2015–2019, IDC #259532.
- Naidoo, V. (2006). International education A tertiary-level industry update. Journal of research in international education, 5(3), 323-345. Available at: http://bern.library.nenu.edu.cn/upload/soft/Internationaleducation.pdf (accessed on 22.10.2016).
- National Research Council (1990). The Internationalization of U.S. Manufacturing: Causes and Consequences. Washington, DC: The National Academies Press, 1990. doi:10.17226/1573.
- Ndou, V. (2004). E-government for developing countries: opportunities and challenges. The electronic journal of information systems in developing countries, 18. Available at: http://www.ejisdc.org/Ojs2/index.php/ejisdc/article/view/110 (accessed on 22.10.2016).
- Nitithamyong, P., and M. J. Skibniewski (2004). Web-based construction project management systems: how to make them successful?. Automation in construction, 13(4), 491-506. Available at: http://www.sciencedirect.com/science/article/pii/S0926580504000184 (accessed on 22.10.2016).
- OECD (2015). Frascati Manual 2015, Guidelines for Collecting and Reporting Data on Research and Experimental Development, DOI: 10.1787/9789264239012-en.
- Open Data Institute (2016). Open enterprise: how three big businesses create value with open innovation. London, UK. Available at http://theodi.org/open-enterprise-big-business (accessed on 22.10.2016).
- Osman, I. H., A.L. Anouze, Z. Irani, B. Al-Ayoubi, H. Lee, A. Balcı, T. Medeni and V. Weerakkody (2014). COBRA framework to evaluate e-government services: A citizen-centric perspective. Government Information Quarterly, 31(2), 243-256. Available at: http://www.sciencedirect.com/science/article/pii/S0740624X14000380 (accessed on 22.10.2016).

- Pardo, A., and G. Siemens (2014). Ethical and privacy principles for learning analytics. British Journal of Educational Technology, 45(3), 438-450. Available at: https://www.researchgate.net/profile/Abelardo_Pardo/publication/261331318_Ethical_and_privacy_principles_for_learning_analytics/links/548159c10cf263ee1adfb165.pdf (accessed on 22.10.2016).
- Passant, A. (2010). Semantic Web Technologies for Enterprise 2.0.
- Porter, M. E. (1985). The Competitive Advantage: Creating and Sustaining Superior Performance. NY: Free Press, 1985.
- Powell, W., and K. Snellman. "The Knowledge Economy." *Annual Review of Sociology* 30, no. 1 (August 2004): 199–220. doi:10.1146/annurev.soc.29.010202.100037.
- Press, G. (2016). Internet Of Things By The Numbers: IDC Survey Finds It's All About The Data. Blog entry on Sep 22, 2016 on Forbes.com. Available at: http://www.forbes.com/sites/gilpress/2016/09/22/internet-of-things-by-the-numbers-idc-survey-finds-its-all-about-the-data/#451ddb6e2856 (accessed on 08.10.2016).
- Queensland Government (2013). Department of Health Open Data Strategy 2013–2016. Brisbane. Available at: https://www.health.qld.gov.au/performance/docs/opendata/open-data-strategy.pdf (accessed on 22.10.2016).
- Rehse, O., A. Rauchfuss and B. Roughan (2010). Lean Advantage in Telcos - Reducing Complexity and Transforming Culture. The Boston Consulting Group. Available at: http://www.bcg.at/documents/file51742.pdf (accessed on 22.10.2016).
- Reina, P. and G. J. Tulacz (2015a). The top 225 international design firms, ENR, New York. Available at: http://www.mairetecnimont.com/it/media/rassegna-stampa/repository-it/27-07-2015-engineering-news-record-the-top-225-international-design-firms (accessed on 22.10.2016).
- Reina, P. and G. J. Tulacz (2015b). The top 250 international contractors, ENR, New York. Available at: http://www.ym.com.tr/Files/Content/News/ENR_Top250_International_Contractors_2015.pdf (accessed on 22.10.2016).
- Ridgway, K., C.W. Clegg and D.J. Williams (2013). The Factory of the Future, Government Office for Science, Evidence Paper 29.
- Right2info.org (2013). Access to Information Laws: Overview and Statutory Goals. Available at: http://www.right2info.org/access-to-information-laws/access-to-information-laws (accessed on 22.10.2016).
- Sabharwal, R. (2016). IDC's Worldwide Semiannual Big Data and Analytics Spending Guide Taxonomy, 2016, Doc # US41207615.
- Schmid, O., J. Hecht, H. Holster, M. Holpp and D. Martini (2013). agriXchange – an analysis of the data exchange in agriculture in Europe. LANDTECHNIK–Agricultural Engineering, 68(1), 14-17. Available at: https://www.landtechnik-online.eu/ojs-2.4.5/index.php/landtechnik/article/viewFile/2013-1-014-017/174 (accessed on 22.10.2016).

- Schmidt, N., A. Luder, S. Biffl, and H. Steininger (2014). Analyzing Requirements on Software Tools According to Functional Engineering Phase in the Technical Systems Engineering Process. In: Proc. of the Int. Conf. on Emerging Technologies and Factory Automation.
- Schoenmaker, D. (2014). Cross-Border Insurance in Europe. Duisenberg School of Finance Policy Paper, (45). Available at: http://www.dsf.nl/wp-content/uploads/2014/09/DSF-Policy-Paper-No-45-Cross-border-Insurance-in-Europe.pdf (accessed on 22.10.2016).
- Schubmehl, D. (2013). Is the Semantic Web and Linked Data ready for Primetime? IDC Community IT Agenda Group Blog post on March 23, 2013. Available at: https://idc-community.com/groups/it_agenda/business-analytics-big-data/isthesemanticwebandlinkeddatareadyforprimetime (accessed on 29.12.2015).
- Serber, S., and E. Wise (2010). Internationalization of Research and Innovation - new policy developments", In: Corporate R&D: An Engine for Growth, a Challenge for European Policy, Conference on corporate R&D (CONCORD 2010), European Commission
- Shao, G., S. J. Shin and S. Jain (2014). Data analytics using simulation for smart manufacturing. In Proceedings of the 2014 Winter Simulation Conference (pp. 2192-2203). IEEE Press. Available at: https://www.researchgate.net/profile/Guodong_Shao/publication/271197235_Data_analytics_using_simulation_for_smart_manufacturing/links/54c6801c0cf219bbe4f86e09.pdf (accessed on 22.10.2016).
- Shaw, M. J. (2001). Information-based manufacturing with the web. In Information-Based Manufacturing (pp. 7-21). Springer US. Available at: http://link.springer.com/chapter/10.1007/978-1-4615-1599-9_1 (accessed on 22.10.2016).
- Simmhan, Y., S. Aman, A. Kumbhare, R. Liu, S. Stevens, Q. Zhou and V. Prasanna (2013). Cloud-based software platform for big data analytics in smart grids. Computing in Science and Engineering, 15(4), 38-47. Available at: https://www.computer.org/csdl/mags/cs/2013/04/mcs2013040038.pdf (accessed on 22.10.2016).
- Smith, D. (2007). An Introduction to Building Information Modeling (BIM). Journal of Building Information Modeling: 12–4. Available at: http://www.wbdg.org/pdfs/jbim_fall07.pdf (accessed on 22.10.2016).
- Smith, Keith (2000). What Is The 'knowledge Economy'? Knowledge-Intensive Industries and Distributed Knowledge Bases. DRUID Summer Conference on The Learning Economy, Rebild, Denmark: AEGIS, University of Western Sydney, 2000. Available at: http://www.druid.dk/uploads/tx_picturedb/ds2000-123 (accessed on 23.11.2016).
- Sonka, S., and Y.-T. Cheng. (2015). Precision Agriculture: Not the Same as Big Data But...Farmdoc daily (5):206, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign. Available at: http://farmdocdaily.illinois.edu/2015/11/precision-agriculture-not-the-same-as-big-data.html (accessed on 22.10.2016).
- Southon, G. (2003). Advancing knowledge in health: a knowledge-based health system. In HIC 2003 RACGP12CC [combined conference]: Proceedings (p. 377). Health Informatics Society of Australia (HISA); Royal Australian College of General Practitioners (RACGP).

- Statistic Austria (2016). Available at: https://www.statistik.gv.at (accessed on 17.11.2016).
- Statistik Austria – Bundesanstalt Statistik Österreich (2016). Statistisches Jahrbuch Österreichs 2016. Kapitel Wissenschaft und Technologie. Wien. Available at: http://www.statistik.at/wcm/idc/idcplg?IdcService=GET_PDF_FILE&dDocName=054404 (accessed on 22.10.2016).
- Sundström, M., and J. Reynolds (2014). Final Report from the Expert Group on Retail Sector Innovation. Available at: http://ec.europa.eu/research/innovation-union/pdf/Report_from_EG_on_Retail_Sector_Innovation_A4_FINAL_2.pdf (accessed on 22.10.2016).
- Tayal, M. (2016). Examining New Factory Concepts Enabled by Industry 4.0. Available at: http://www.happiestminds.com/blogs/examining-new-factory-concepts-enabled-by-industry-4-0/ (accessed on 22.10.2016).
- Tene, O. and J. Polonetsky (2013). Big Data for All: Privacy and User Control in the Age of Analytics, 11 Nw. J. Tech. and Intell. Prop. 239. Available at: http://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=1191&context=njtip (accessed on 22.10.2016).
- Tian, G. Y., G. Yin and D. Taylor (2002). Internet-based manufacturing: A review and a new infrastructure for distributed intelligent manufacturing. Journal of intelligent manufacturing, 13(5), 323-338. Available at: http://link.springer.com/article/10.1023%2FA%3A1019907906158 (accessed on 22.10.2016).
- Turner, Vernon, Bob Kraus, Carrie MacGillivray, Monika Kumar (2015). IDC's Worldwide Internet of Things Taxonomy, 2015, Doc # 256186.
- United Nations (2008). International Standard Industrial Classification of All Economic Activities. Revision 4. Statistical papers. Available at: http://unstats.un.org/unsd/publication/seriesM/seriesm_4rev4e.pdf (accessed on 22.10.2016).
- Uppenberg, K. (2009). R&D in Europe: Expenditures across sectors, regions and firm sizes. Available at: http://aei.pitt.edu/32615/1/54._R%26D_in_Europe.pdf (accessed on 22.10.2016).
- W3C (n.d.). Linguistic Linked Data for Content Analytics: a Roadmap. Available at: https://www.w3.org/community/ld4lt/wiki/Linguistic_Linked_Data_for_Content_Analytics:_a_Roadmap (accessed on 29.12.2015).
- Waling, L. J. (2006). Managing tacit knowledge in hospitals: successfully delivering the intangible with IT (Doctoral dissertation, TU Delft, Delft University of Technology). Available at: http://repository.tudelft.nl/assets/uuid:81ba327d-bd95-409e-a83b-846078dbb72a/waling_2006.pdf (accessed on 22.10.2016).
- Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. International Journal of Forecasting, 30(4), 1030-1081. Available at: http://www.sciencedirect.com/science/article/pii/S0169207014001083 (accessed on 22.10.2016).
- Willis, H. L. and L. Philipson (2005). Understanding electric utilities and de-regulation (Vol. 27). CRC Press. Available at: http://119.235.48.163/library/Ebooks/EEE/Electrical%20and%20Electron-

ics%20Engineering/Understanding%20Electric%20Utilities%20and%20De-Regulation/Understanding%20Electric%20Utilities%20and%20De-Reguration%202ed%20-%20Lor.pdf (accessed on 22.10.2016).

- Willoughby, K. and P. Galvin (2005). Interorganizational Collaboration, Knowledge Intensity, and the Sources of Innovation in the Bioscience-Technology Industries. Knowledge, Technology, and Policy, 18(3), 56-73. Available at: http://link.springer.com/article/10.1007%2Fs12130-005-1005-z (accessed on 22.10.2016).
- Windrum, P. and M. Tomlinson (1999). Knowledge-intensive services and international competitiveness: a four country comparison. Technology Analysis and Strategic Management 11.3. 391-408. Available at: https://www.researchgate.net/profile/Paul_Windrum/publication/263567827_Knowledge-intensive_Services_and_International_Competitiveness_A_Four_Country_Comparison/links/540968600cf2822fb738cee8.pdf (accessed on 22.10.2016).
- Wood, David (ed.) (2010). Linking Enterprise Data.
- Wood, David, Marsha Zaidman, Luke Ruth and Michael Hausenblas (2013). Linked Data: Structured data on the Web.
- Wood, H., and K. Gidado (2008). Project complexity in construction. In Proceedings of The construction and building research conference of the Royal Institution of Chartered Surveyors–COBRA-2008, Dublin. Available at: https://core.ac.uk/download/files/122/9552740.pdf (accessed on 22.10.2016).
- World Wide Web Consortium (2014). Best practices for publishing Linked Data. Available at: https://www.w3.org/TR/ld-bp/ (accessed on 22.10.2016).
- Wyatt, J. C. (2001). Management of explicit and tacit knowledge. Journal of the Royal Society of Medicine, 94(1), 6. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1280081/pdf/11220080.pdf (accessed on 22.10.2016).
- Wyman, O. (2015). Managing Complexity – The State of the Financial Services Industry 2015. Available at: http://www.oliverwyman.com/content/dam/oliver-wyman/global/en/2015/jan/SoFS-2015/State_of_the_Financial_Services_Industry_Report_2015.pdf (accessed on 22.10.2016).
- Zeng, S. X., G.X. Lou and V.W.Y. Tam (2007). Managing information flows for quality improvement of projects. Measuring Business Excellence, Vol. 11 Iss 3 pp. 30 - 40. Available at: http://dx.doi.org/10.1108/13683040710820737 (accessed on 22.10.2016).

10. ABBREVIATIONS

ELD - Enterprise Linked Data
LD - Linked Data
ISWC - International Semantic Web Conference
ESWC - Extended Semantic Web Conference
JWS - Journal of Web Semantics
SWJ - Semantic Web Journal
W3C - World Wide Web Consortium

11. ANNEXES

11.1. Annex A: Interview Guide

Date: DD/MM/YYYY
Name: XXX
Brief job description and functional role: (e.g., CXO, Head of Communication, IT Project Manager, etc.)
Organization: XXX
Industry: according to the PROPEL industry categorisation
Employees: < 50 | 50-200 | > 200

Organization

- What is the main focus of your organisation's business?
- Can you describe your organisation's operational and organisational structure? [e.g., functional, matrix, projectized]
- Is there much collaboration within your organisation, both in terms of business processes and software systems? Please provide some examples that demonstrate this. [e.g., across departments / functional units]

Data integration / exchange / linking use cases

- Considering your existing business processes and software systems, do you currently have any specific organisational or technical challenges with respect to data exchange/integration within the organisation and with third-party organisations? Please provide some examples. [e.g., integrating your suppliers or customer's data and/or integrating data within your organisation with publicly available data]

- Where has data been shared successfully? Do you have examples for good practices and existing solutions for data integration / exchange in your organisation?
- What data would generate an added value if it was integrated/linked in your company? [e.g., customer data and spatial data, weather data and transportation data, etc.]

Linked Data (LD)

- Are you familiar with the concept of LD?

[in case of yes:]

- What experience do you have with LD?
- What is your understanding of the value of this technology from a business and/or technical perspective?
- Has your organisation adopted LD or is planning to do so? Describe past/current/planned projects/initiatives/trials of your organisation.
- What are the main difficulties encountered in the adoption of LD concepts and technologies generally speaking and for your organisation?
- Which organisational conditions positively impact LD adoption (generally speaking and for your organisation)?

Note: For those, who are new to the subject, we can provide a short insight on what LD is: *Linked Data is a method of publishing structured data so that it can be interlinked and become more useful through semantic queries. It builds upon standard Web technologies such as HTTP, RDF and URIs, but rather than using them to serve web pages for human readers, it extends them to share information in a way that can be read automatically by computers. This enables data from different sources to be connected and queried (Wikipedia)* to achieve the grand vision of a global information space. *Optionally you can send this video after the interview: Manu Sporny: What is Linked Data (12:09)*

Technology

- Do you have a technology roadmap within your organisation? If yes, what are your short term and medium term objectives?

- Which stakeholders are involved when it comes to strategic technology decisions in your organisation?
- What are drivers and barriers in getting experienced with emerging technologies for your organisation?

11.2. Annex B: Workshop Agenda

Agenda

15:00–15:10	**Begrüßung und Einführung in das Projekt PROPEL** Julia Neuschmid, Research Analyst, IDC Central Europe GmbH
15:10–15:20	**Eine kurze Einführung in Linked Data** Elmar Kiesling, Head of Linked Data Lab, Technische Universität Wien
15:20–15:30	**Technology gaps and research requirements for effective data integration solutions** Sabrina Kirrane, Postdoctoral Researcher, Wirtschaftsuniversität Wien
15:30–15:40	**Use cases für den Linked Enterprise Data Einsatz** Thomas Thurner, Head of Transfer, Semantic Web Company
15:40–16:00	Pause
16:00–17:00	Expertendiskussionen
17:00–17:30	Zusammenfassung der Ergebnisse
17:30–18:00	Get-together

Die Veranstaltung ist **kostenlos** und findet in Deutscher und Englischer Sprache statt. Da die Teilnahmeanzahl beschränkt ist, bitten wir Sie um Anmeldung bis zum 2. Mai 2016.

Hier geht es zur Anmeldung

Der Workshop findet im Rahmen des Projektes PROPEL – Propelling the Potential of Linked Data in Austrian Enterprises **statt. PROPEL** erhebt technologische Herausforderungen, neue unternehmerische Chancen und offene Forschungsfragen des Datenaustauschen und der Datenintegration in Unternehmen. Ziel ist die Erstellung einer Research Roadmap, die den Pfad zukünftiger Forschung und Entwicklung im Bereich Linked Data beschreibt. Gefördert durch Österreichische Forschungsförderungsgesellschaft FFG und Bundesministerium für Verkehr, Innovation und Technologie BMVIT.

Wir freuen uns auf Ihre Teilnahme!

Das PROPEL-Konsortium

propel TU WIEN WU SEMANTIC WEB COMPANY bmvit FFG

http://idc-austria.at/de/events/63608-linked-data-datenintegration-der-zukunft/7-overview

11.3. Annex C: User Stories

Media

Theme	As a	I want to	so that
Automatic and dynamic content production and display	creator of marketing content for the tourism industry	translate my message for various markets and cultural contexts	design more clearly understandable and finely targeted messages for various target markets
Automatic and dynamic content production and display	media house	recycle my content frequently and generate automatically different variations of the content	content production costs decrease and content efficiency increases
Customization and CRM	media house	adapt the paywall for my content individually to customer profiles	I increase revenues
Customization and CRM	media house	enrich my content offering with e-commerce based on user profiles	I can build a solid media business model
Customization and CRM	media house	display personalized content as precise as possible	my readers stay as long as possible on my website
Data and information integration	journalist	enrich my articles with facts and figures	I can decrease my research effort
Performance measurement and reporting	media controller	relate changing KPI`s to each other that are referring to similar actions	I can compare media consumption in the long-term
Product management	journalist	link my articles that are published on the web	I can track it and follow which websites make a reference to my article
Search and information retrieval	news agency (newspaper)	provide own articles on specific topics and provide additional topic-related data and information from other sources (e.g. Wikipedia) to the newspaper readers	my readers can get a comprehensive picture on the topic
Search and information retrieval	video editor	find visual materials based on specific characteristics (e.g. 1 man and 1 child talking to each other)	I can save time and work more efficiently
Search and information retrieval	media house	detect online users that could be interested in my content	I can first establish a community and provide content in a next step, when I reach a critical mass
Tracking and transparency	media house	keep track of where my content is displayed and how it is consumed centrally	I don`t have my content analytics distributed on many external platforms

Professional Services

Theme	As a	I want to	so that
Automatic and dynamic content production and display	real estate broker	display my rental objects (picture, description, address) at different real estate platforms, but I would like to maintain changes centrally	so that I don`t have to update the content in different places
Automatic and dynamic content production and display	real estate broker	include legal requirement changes automatically in all documents where I have them included	my contracts are up-to-date with low manual effort
Data and information exchange, coordination and collaboration	consulting company	share information across subsidiaries	we make use of existing know-how and content
Data and information integration	real estate platform provider	offer additional geographic, statistic and qualitative data about an apartment's neighborhood (open government data, wikipedia, public statistics, newspaper articles, etc.)	the people who use my platform to search for apartments can get a holistic picture of the appartment's surrounding and can make a decision if the location fulfills their requirements
Data and information integration	civil engineer	receive additional geographic, statistic and qualitative data about a location (e.g. urban renewal site, construction site, brownfield area, etc.)	I can get a holistic picture of the location, analyse data that comes from different sources, and can make a plan (urban development plan, construction plan, etc.)
Product management	caregiver for elderly working at an NGO	search for assistive products and services on the Internet that support elderly people in a specific situation / use case (e.g. health monitoring at home)	I can consult and recommend specific services and products to elderly people
Product management	event company	group my contacts by attributes (industry, company name, position, location, etc.)	I can manage the contacts of my event participants more efficiently
Tracking and transparency	real estate broker	know which objects belong to which owners and which intermediaries are involved	I can simplify the administration process of contract management, and revenue sharing

Various

Theme	As a	I want to	so that
Customization and CRM	sales representative	find out about customer-related prior communication across various channels (E-mail, CRM)	I can provide a better customer experience; make an informed offer, cross/up-selling etc.
Data and information exchange, coordination and collaboration	IT system modeler	work on a model with multiple tools and exchange models with other software/system developers that use a different set of tools	I can collaborate more efficiently with other software/system modelers
Data and information exchange, coordination and collaboration	participant in a value chain of physical products	better exchange information and coordinate production and logistics with suppliers and customers	I can improve efficiency, effectiveness and flexibility of my inventory management and operations
Data and information integration	business user	be able to access all relevant information (ERP, CRM, mails, cloud services etc.) from a single point	do my job more efficiently and effectively
Data and information integration	project manager	get new employees up to speed quickly and reuse deliverables and experiences from prior projects	I can cope with high employee turnover and avoid duplication of effort
Data and information integration	employee	know which roles, responsibilities, know-how my colleagues have	I can ask the right people for information and support
Data and information integration	HR department	include external data as on the labour market in my analytic tool	I can do better workforce estimations
Data and information integration	HR department	include financial data into my analytics tool	I can be more prepared for budget discussions with the finance department
Data and information integration	information architect	to merge different data sets and store meaningful metadata for different data sources	I can perform value added analytics and sell this data to 3rd parties
Data and information integration	data scientist	capture tacit knowledge in the company	reduce the time it takes to get familiar with who's who within an organisation
Data consistency and elimination of duplication	HR department	ensure that employee data are consistent across multiple systems	
Other automation of processes	information architect	automate rights management, which is a very manual process that is unable to keep pace with big data requirements	I can be legally sure what datasets I can use

Search and information retrieval	strategic management, key account manager	find people with relevant competences or get a high-level overview of the organization's strategic capabilities	I can make informed decisions and leverage the knowledge and experiences available within the firm
Search and information retrieval	data scientist	identify expertise within a large organisation and be able to pinpoint the relevant experts	I can identify top trends within the organisation and expertise for the organisation as a whole
Tracking and transparency	consumer	find out about the origin of products that I consume	ensure that the circumstances of their production (workers’ rights, product safety and integrity, environmental responsibility) are safe and consistent with my personal ethical framework
User integration/ open innovation	product developer	quickly obtain information on how customers use my product	I can evolve and adapt them
Using ELD ecosystem	CIO, CTO or similar	use a infrastructure, which provides me all basic LD functions	so that i have not to build up own facilities and ease the start of Linked Enterprise Data in my company

Other

Industry	**Theme**	**As a**	**I want to**	**so that**
Education	Product management	teacher	link my training material (data, documents, etc.) to material on the same topic	I can provide my students an overview on all available training material
Finance	Benchmarking	bank	analyse financial data of customers (companies) and show customers industry trends	my customers can benchmark their company against industry data (where am I standing in comparison to the overall industry / in comparison to my region / in comparison to other companies with my size, etc.)
Finance	Customization and CRM	provider of payment solutions	automatically analyse customer data such as who buys when, what, where, how often, for how much, etc.	I can develop and sell customer profiles to companies who want to sell products and services by addressing each customer with individual offers that might interest the customer the most
Finance	Customization and CRM	credit institute	collect data about customers from different sources on the web (social media, public websites, etc.)	I can analyse their behaviour and financial standing/credit rating

Finance	Customization and CRM	insurance company	offer my customers highly customized products	can improve customer satisfaction and gain competitive advantages in the marketplace
Government	Other automated analysis of data	secret service agent/police officer	extract relevant information from social media streams	
Government	Performance measurement and reporting	public company	be able to continuously and (semi-) automatically create financial reports	I can comply to regulatory requirements and evolve my organizational structure flexibly
Government	Search and information retrieval	public servant	be able to easily find relevant files/ cases (e.g. court ruling) based on their context	more quickly gather relevant information
Health	Data and information integration	pharmaceutical company	integrate disparate systems that are: hard to integrate, widespread, contain the same data that contradicts each other	I can gain insights from other clinical trials
Health	Other automated analysis of data	analytics company	automatically analyse drugs and component substances	
Insurance	Product management	insurance company	link attributes of product data (client name, insurance price, insurance type, etc.)	I can keep an overview of the wide range of available products
Manufacturing	Benchmarking	manufacturing company	share production data (where is when how much produced) from different branch offices within the company / with management	I have an overview on what happens on the different production sites
Manufacturing	Data and information integration	HR department	include manufacturing and ordering data into my analytic tool	I can be more precise in my workforce planning
Research	Search and information retrieval	market analyst	find articles about specific IT vendors based on their names in Austria on the Internet in different newspapers and magazines	I can track and monitor IT provider's business development on the market (e.g. growths, decline, new products)
Research	Search and information retrieval	market analyst	link existing reports on specific technologies that have been published by my organization based on their attributes/metadata (keywords, technology segment, region covered, publishing date, author, etc.)	I can find existing reports that have been published by my company

Retail	Benchmarking	retailer	share sales data (when is where what and how much sold) from different shops within the company (sales team, management)	I have an overview on current sales number in each shop
Retail	Customization and CRM	online market place provider	analyse user data (who buys when, where, what, for how much, etc.)	I can recommend each user individual products and services that based on their profile interests them the most
Retail	Product management	online market place provider	group products and services with similar attributes (e.g. function, colour, size)	users can find products and services that interests them and that are similar
Retail	User integration / open innovation	designer of automated warehouse solutions	elicit specific customer requirements and make sure that a designed solution fulfills them	I can efficiently and effectively design warehouse solutions for clients
Tourism	Customization and CRM	tourism manager/service provider	reach customers directly and offer integrated, seamless services	I can create a better "user experience" for visitors (and manage visitor streams)
Tourism	Performance measurement and reporting	tourism manager	collect tourism data (e.g., revenues, arrivals, bednights) from the various stakeholders involved	I can efficiently create and publish reports on detailed performance indicators in near-real time
Transport	Data and information exchange, coordination and collaboration	national postal service	exchange information with my international counterparts	we can work together on innovating our industry
Transport	Data consistency and elimination of duplication	logistics operator	align internal and external delivery ID's	distribution processes are more transparent in case of inquiries
Transport	Resource scheduling and logistics	motorway operator	manage data about staff, locations on the road that need maintenance, availability of working equipment and vehicles and other data, in combination with traffic and weather data	I can plan resources more efficiently and save costs

Transport	Resource scheduling and logistics	logistics operator	decrease delivery time by providing my drivers with driving routes that take traffic information into consideration	distribution costs decrease
Transport	Resource scheduling and logistics	logistics operator	gain insights when parcels were picked up	I can optimize my delivery process.

Image credits and copyrights

The pictures used for this book were obtained from commercial photography libraries.

Sponsoring

PROPEL is funded by the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT) and the Austrian Research Promotion Agency (FFG) under the program "ICT of the Future" between November 2015 and December 2016.

bmvit
Bundesministerium für Verkehr, Innovation und Technologie

FFG
Austrian Research Promotion Agency

Acknowledgments

This book could not have been made without the help of those who have contributed ideas, discussions, demos and texts. We would like to thank everyone who helped us:

Andreas Blumauer, Martin Kaltenböck, A Min Tjoa, Tassilo Pellegrini, Ioana Hulpus, The Semantic Web Science Association (SWSA), STI International, the Semantic Web Journal, the Journal of Web Semantics

The PROPEL project – Propelling the Potential of Enterprise Linked Data in Austria – surveyed technological challenges, entrepreneurial opportunities, and open research questions on the use of Linked Data in a business context and developed a roadmap and a set of recommendations for policy makers, industry, and the research community.

SEMANTIC WEB COMPANY linking data to knowledge
WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS
TECHNISCHE UNIVERSITÄT WIEN Vienna University of Technology
IDC

edition mono monochrom

ISBN: 978-3-902796-55-4