EXPLORING THE FACTORS THAT INFLUENCE THE DIFFUSION OF OPEN DATA FOR NEW SERVICE DEVELOPMENT: AN INTERPRETIVE CASE STUDY

Complete Research

Giovanni Maccani, Maynooth University, Ireland, giovanni.maccani.2013@nuim.ie

Brian Donnellan, Maynooth University, Ireland, brian.donnellan@nuim.ie

Markus Helfert, Dublin City University, Ireland, markus.helfert@computing.dcu.ie

Abstract

Information Systems research on Open Data has been primarily focused on its contribution to e-government inquiries, government transparency, and open government. Recently, Open Data has been explored as a catalyser for service innovation as a consequence of big claims around the potential of such initiatives in terms of additional value that can be injected into the worldwide economy. Subsequently, the Open Data Services academic conversation was structured (Lindman et al. 2013a). The research project presented in this paper is an interpretive case study that was carried out to explore the factors that influence the diffusion of Open Data Services. This paper contributes to this debate by providing both a systematic literature review study that structures research efforts available to date in this topic, and an interpretive case study (Walsham, 1995) of a successful initiative that turned several city authorities' raw open datasets into a set of valuable services. Results demonstrate that 12 factors and 56 related variables are the most relevant in the process of diffusion of open data for new service development. Furthermore, this paper demonstrates the suitability of Social Constructionism and interpretive case study research to inductively generate knowledge in this field.

Keywords: Open Data Services, Diffusion Factors, Interpretive Case Study, Systematic Literature Review.

1   Introduction

The definition of Open Data was firstly developed by the Open Knowledge Foundation (http://okfn.org/) in 2005. They defined Open Data as "data that can be freely used, shared and built on by anyone, anywhere, for any purpose". Since 2007 this definition has been governed by an Advisory Council, which is a community-run body responsible for maintaining and developing the Open Data definition and associated material. There are 3 principles behind this definition of Open Data: (1) availability and Access (i.e. people can get the data); (2) re-use and Redistribution (i.e. people can re-use and share the data); (3) universal Participation (i.e. anyone can use the data).

In their Digital Agenda (www.ec.europa.eu), European commissioners listed 4 reasons for promoting Open Data initiatives, including potential economic gains from new product and service development (estimated to be 40 billion Euros a year in the EU), addressing societal changes, fostering citizens participation, and improving internal efficiency.

So far, research on Open Data (and more generally on Public Sector Information), was mainly related to e-government inquiries, addressing aspects of democratic theory, voter participation, democratic deliberation, and open government in a broader context (Amichai et al. 2008). Later, research focuses included also Open Data seen as "vital for the vitality of the civic society" (Bertot et al. 2010). Finally,

recent research explored Open Data as a foundation and catalyser of innovation (Lakomaa and Kallberg, 2013), and particularly service innovation.

All these concepts led academic authors to introduce and structure a new research stream named Open Data Services. This novel academic conversation is giving a structure to the investigation of Open Data as a foundation of service innovation from an Information Systems perspective (see for example "Open Data Services Mini-track" at the Hawaii International Conference on System Sciences, 2013 (Lindman et al 2013a)). This research aims to contribute to this debate through the study of a successful initiative that turned a range of raw open datasets into a set of valuable services. Specifically, factors and variables that influence the diffusion of Open Data for new service development were inductively explored through an interpretive qualitative case study.

This paper is structured as follows: the next paragraph will introduce the Open Data Services academic conversation (Lindman et al. 2013a). In section two a research gap and a subsequent research question will be formulated from a Systematic Literature Review study (Okoli et al. 2010). Key research choices regarding philosophical underpinnings, decisions of reasoning and approach, and the identification of Case Study research (Walsham, 1995) as a suitable methodology will be reported in the third section. Section four includes the actual case study analysis and the outcome for this research. Finally, conclusions, limitations, and future research avenues will be presented.

1.1 Open Data Services

In information Systems (IS), data and information are conceptualized with a dichotomy of data and application, where data is used for storage and application is used for different operations based on data (Lindman et al. 2013b). In this way, it can be made a distinction between supply and demand of Open Data. Specifically, the supply perspective aims at making the data available, and the demand side builds something useful on top of the data. Hence, access to Open Data is just the first step within an infrastructure that allows end-users to consume Open Data Services. Thus, for the Open Data to become valuable there need to be a chain of steps that take the raw data, make it available to others as services, or further down the chain analyze, combine and present data in ways that make it useful for users to interpret as information. In (Lindman et al. 2013b), the authors adopted (Alter, 2010)'s work system framework to describe Open Data Services. According to Alter (2002), a work system is "a system in which human participants and/or machines perform work using information, technology, and other resources to produce products and/or services for internal or external customers". Thus, a service system can be seen as a work system that produces services for customers. These concepts were then linked to Open Data Services, and are represented in Figure 1.
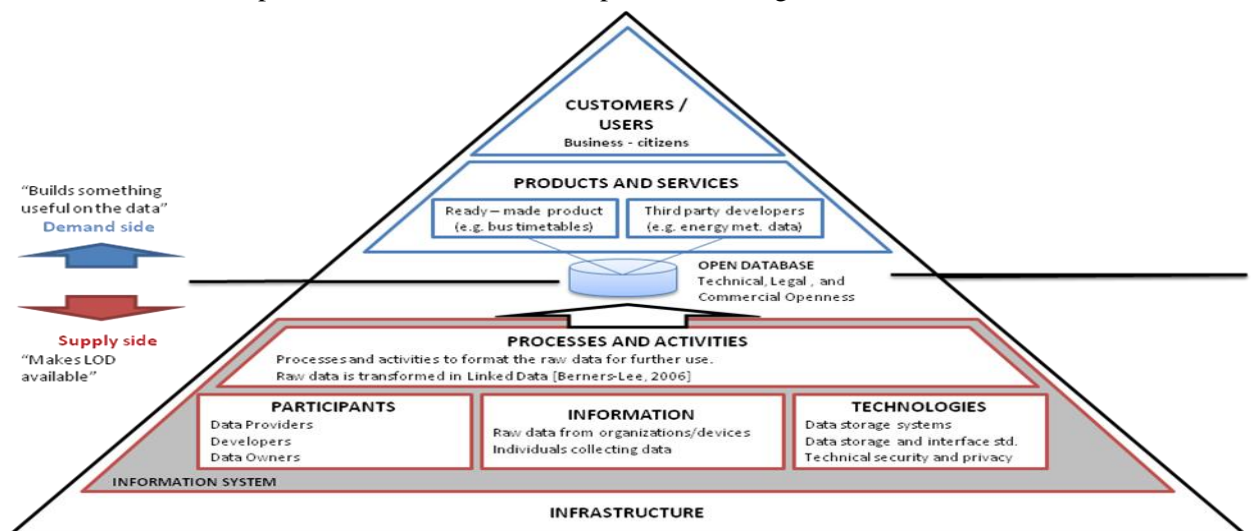

Figure 1.   Open Data Services Framework. Adapted from (Lindman et al. 2013b), (Alter, 2010).

Within the supply side of the framework, four main factors can be detected. Regarding the technological building block, it includes data storage systems and standards for interfacing these systems. Then, the other three dimensions refer to the type of raw data that is collected and transformed (information building block), the processes and activities that are undertaken for the development of the Open Databases, and the actors involved in this work system.

Moving now to the demand side of this framework, the data is assumed to be available as Open Data (i.e. legally, commercially, and technically opened). With data openly available to the public, designers (including developers and hackers) can apply different models and/or theories to the data, and create new artifacts (Kuk and Davies, 2011). Hence, at this stage the Open Data is used to build a service. The range of actors involved in this step can vary from hackers, software consultancies, or even governments themselves (Lindman et al. 2014). Finally, to make Open Data Services sustainable, there need to be customers. According to Alter (2010), customers are recipients of a work system's services for purposes other than performing work activities within the work system. In this way, we conducted a Systematic Literature Review (SLR) in order to specifically identify research contributions to date to the Open Data Services debate.

2 Open Data Services Systematic Literature Review

This study adhered to (Okoli et al. 2010)'s 8-steps methodology. These steps are: (1) Purpose of the Literature Review, (2) Protocol and Training, (3) Searching for the Literature, (4) Practical Screen, (5) Quality Appraisal, (6) Data Extraction, (7) Synthesis of Studies, and (8) Writing the Review.

The purpose of this study was to analyze the progresses of academic research in the field of Open Data Services. The acquisition of this knowledge allowed us to precisely identify the research gaps in this field, and the subsequent opportunities for implementing a research effort. Specifically, this study is focused on reviewing the literature available on the adoption and diffusion of Open Data in terms of its usage in designing and developing new services. Four main sources for searching for the literature were considered. As suggested by (Okoli et al. 2010), Open Access Databases and Specific Subject Databases were used, including: Google Scholar (http://scholar.google.com/), Directory of Open Access Journals (http://www.doaj.org/), ACM Digital Library (http://dl.acm.org/), and IEEE Xplore (http://ieeexplore.ieee.org/). Consistently with the objectives for this study, five keywords were used as search terms: "*Open_Data Service(s)*", "*Open_Data Innovation*", "*Linked_Data Service(s)*", "*Open_Data Adoption*", and "*Open_Data Diffusion*". Finally, because of the commonality of the single words considered (i.e. "Open", "Data", "Service", "Innovation" etc.), we decided to set our search strategy on looking merely at the presence of these terms either within the title or the keywords of the papers. As a result of the searching step we found in total 262 articles that satisfied the requirements across the four sources. Due to the large amount of articles, we decided to apply exclusion criteria to both the titles and the abstracts of these papers, before proceeding with the reading of the full texts. This process allowed us to reduce the amount of papers to be read to 68. These studies were carefully analyzed and as a result 34 articles were ordered and searched and considered further. While reading these articles we selected further 10 articles from references analysis. Finally, we had an opportunity to engage with experts and researchers in the field of Open Data Services. From this early stage interaction, 2 more studies were added to this analysis. Concluding, as a result of this step, 46 articles were selected for this SLR study.

2.1 Data Extraction and Synthesis

In step 6 of the (Okoli et al. 2010)'s methodology, all the data needed for the final synthesis were collected. The output of this step was a complete list of relevant concepts from which we could synthesize our study and identify the research gap/gaps in the field of Open Data Services.

In synthesizing the data collected, we categorized the papers in three groups based on which areas of the Open Data Services framework (see Figure 2) the specific study had covered:

- Group A: these papers focused purely on the supply side of the Open Data Services framework.

- Group B: articles of this second group cover both the demand and supply side of the framework.

- Group C: the studies included here are merely focused on the demand side of the framework.

These groups and the percentages of the articles that actually belong to each of them are represented in Figure 2. Consistent with the objectives of this SLR, i.e. to systematically review the literature available on adoption/diffusion/usage of Open Data for Service Innovation, Groups B and C represent the key section of literature for identifying the research gap that this research project will aim to fill.
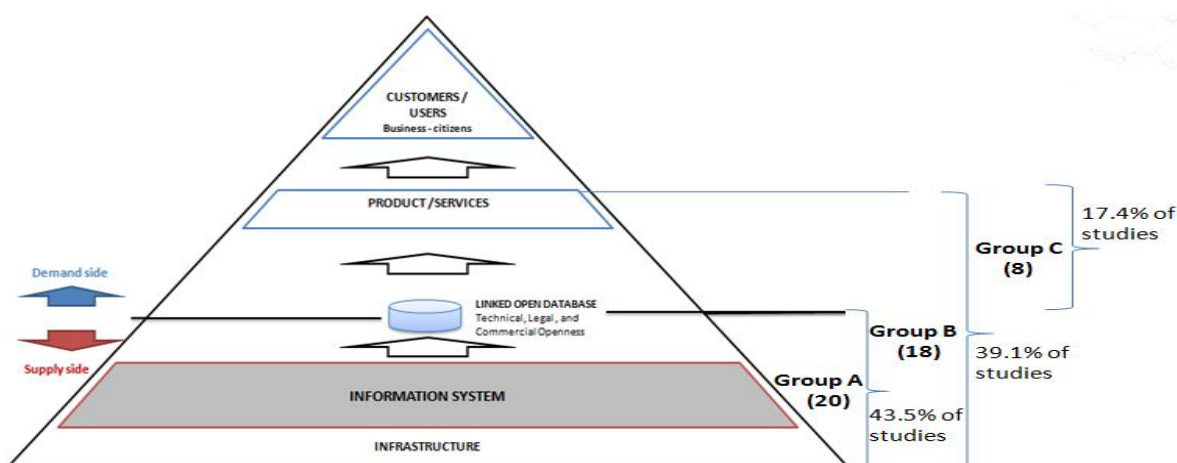

Figure 2.   Open Data Services SLR Groups.

At this stage of the research, we have systematically identified the research efforts that have been implemented in the topic of Open Data Services.

In (Lindman et al. 2013b), a literature review was presented. Furthermore, a set of research questions was outlined. The authors concluded that there is a clear research gap, and in particular "it is entirely unclear how to build a sustainable open data market and establish actors within it" (Lindman et al. 2013b, pp. 1242). However, we learnt through our SLR study how the conversation around this topic is rapidly evolving and being enriched by several studies. Firstly, we found a plethora of suggestions on how to supply Open Data to application (or more generally – service) developers (Group A). In this way, different semantic web solutions (Cifuentes-Silva et al. 2011), (Fox, 2013), (Cyganiac et al. 2010), (Wilde, 2010), new architectures (Aoyama and Kojima, 2013), institutional challenges (Felicetti et al. 2012), (Currie, 2013), and different approaches (Lorey, 2013), (King et al. 2007), (Yu et al. 2011) were investigated to enable and maintain the release of Open Data by Public Authorities. In addition, few studies focused on individual projects such as Europeana (Isaac and Haslhofer, 2011), (Haslhofer and Isaac, 2011), DataBridges (Herschel and Manolescu, 2012) and other case studies (Frosterus et al. 2012), (Oh, 2013), (Stephenson et al. 2012), to describe approaches and activities involved in the publication of Open Data.

As represented in Figure 2, 18 of the articles considered for this SLR tackled both the demand and the supply side of the Open Data Services framework (Group B). The majority of these studies propose technological related solutions for the efficient/effective release of Open Data, and provide subsequent examples of services designed/delivered as a demonstration of the validity of the proposed solution. These include (Guo et al. 2010) (in which the application proposed facilitates the search process of Open Data), (Matheus et al. 2012) (the application developed is referred to government transparency),

(Ijima et al. 2011) (the examples are a range of mobility services designed and implemented in Japan), and (Hannemann and Kett, 2010) (integration service of Open Data across the German Library service providers). In addition, more studies focused on the design of various forms of frameworks to describe the activities and processes that are undertaken from the raw data till the adoption of Open Data for different purposes (Hartung et al. 2010), (Tcholtchev et al. 2012), (Rittenbruch et al. 2012), (Gautreau, 2012), (Chan, 2013), (Hausenblas, 2009), and (Bizer et al. 2009). Moving ahead, two more studies are included in this category: (Lindmann et al. 2013b) and (Tammisto and Lindman, 2011). These are considered the seminal works upon which the academic conversation around Open Data Services was drawn. Specifically, the former defined this new research stream and a related research agenda through (Alter, 2010)'s work system framework (see Figure 1). The latter, instead, proposes a business model for Open Data. In a similar way, (Latif et al. 2009) studied the value of the Open Data from its "Raw Data stage" till its usage.

Concerning Group C, which includes all studies focused on the demand side of the Open Data Services framework, 8 papers were analyzed. Three of them introduce innovative services designed on top of Open Data. Particularly, (Halb et al. 2010) presents a prototype for professional online editors, in (Groen et al. 2013) an innovative service for tourists in Amsterdam is proposed, and (Savelyev et al. 2011) focuses on the concept of Volunteered Geographic Services. In addition, (Hielkemma and Hongisto, 2013) adopts Porter's model (Porter, 1990) to study how Living Lab methodologies can drive forward the urban competition for Open Data. An original contribution to this field is also given in (Ferro and Osella, 2012, 2013) which propose a set of business models for private entrepreneurs that aim at harnessing and using public datasets for profit oriented businesses. Finally, the last article included in Group C, refers to an analysis that was conducted in 2012 and aimed at including Open Data as a fundamental factor for innovation in web-based applications and information services (Lakomaa and Kallberg, 2013).

2.2   Research Gap

The SLR study showed that the demand side of the Open Data Services framework remains unexplored in literature. In addition, we searched across municipal Open Data websites of European capital cities and of the top 50 cities in Europe in terms of population. Our findings, confirmed the statement made in (Kuk and Davies, 2011), i.e. in the open data context the most significant emphasis towards service innovation has been placed to catalyze "civic hacking" taking through weekend-hack days and competitions. In fact, high profile competitions such as "Apps4Fingal" (http://data.fingal.ie/Apps4Fingal/) in Ireland, "Apps4BCN" in Barcelona (http://opendata.bcn.cat/opendata/ca), "App4MI" in Milan (http://dati.comune.milano.it/), and many more, have emphasized the potential benefits for innovation, of releasing data to developers, and allowing actors outside of government to build services off the back of it. However, as shown in (Kuk and Davies, 2011), out of 130 distinct projects that were proposed, only 10 remained active one month later. Then, the authors made a very strong statement as a conclusion of their study: "grand claims for the service revolutions that open data may bring about are overstated; though more modest claims can be grounded in evidence" (Kuk and Davies, 2011, pp.15). As a result we see that the potential for Open Data for service innovation has been demonstrated in several ways, but in practice there are some challenges that need to be overcome in order to make it happen. In our SLR study we found two papers that specifically focused on diffusion and adoption of Open Data. However, they are related to the supply-side of the Open Data Services framework (Janssen et al. 2012) and (Barry and Bannister, 2014). Both of them used the lens of Institutional Theory (Scott, 1995) to investigate the benefits and the barriers to the diffusion of Open Data, in terms of obstacles that governments face and need to overcome to release their data and make them "openly" available.

Thus, we conclude that there is a clear research gap concerning the investigation on how Open Data diffuses for new service development, in order to achieve the benefits that have been strongly claimed

in literature, and surprisingly not yet achieved in practice. As a consequence the following research question was formulated for this study:

What are the factors that influence the diffusion of Open Data for new service development?

3 Research Approach and Methodology

This section is dedicated to both the key research decisions and the methodological choices that have been taken to guide this study.

The purpose of this research project is to investigate the factors that influence the diffusion of Open Data for Service Innovation. Despite the well acknowledged Diffusion of Innovations Theory (Rogers, 2005), our SLR showed that there is a substantial lack of theoretical insights specifically related to the recently introduced academic conversation of Open Data Services (Lindman et al. 2013b). As a consequence, an interpretivist perspective is considered as the most suitable for this project.

The resulting paradigm assumed for this project is Social Constructionism. Ontologically, the basic premise that is made here is that the Open Data released by cities authorities diffuses among service developers through an apparently natural process, but that in fact is complex, and contingent on several social actors and activities. From the epistemological point of view, the knowledge that will be developed assumes a correspondence of meaning of subjects constructing the diffusion process of Open Data (i.e. the reality to be studied). Hence, this study, consistent with its Social Constructionist philosophical underpinnings, involves an inductive reasoning. Generally, inductive studies can be conceptualized as a set of steps (Glaser and Strauss, 1967). Huff (2008) summarizes them as follows: (1) extensively describe an interesting situation without use specialized vocabulary from existing academic literature; (2) create a first level of substantive categories by coding these descriptions; (3) modify and improve codes as additional data are collected, categorized and compared; (4) expect theoretic insights to emerge as categories stabilize and their relationships become apparent (5) conclude empirical observation when new categories are not required to account for further observations (Huff, 2008).

Finally, we decided to adopt a qualitative approach (as opposed to a quantitative one) for two main reasons. First, the emerging concept of Open Data for Service Innovation, due to its novelty, must be investigated through an exploratory-oriented study. Secondly, as the reality of Diffusion of Open Data for new service development will be analyzed from the meaning given by the people that are actively involved in constructing such reality, qualitative research methods are appropriate as they are "designed to help researchers understand people and the social and cultural contexts within which they live" (Myers and Avison, 1997).

3.1 Case Study Research

As a result of a comparative analysis between different methodologies employed in interpretive/constructionist qualitative inductive research (with a specific focus on IS), Case Study research is found to be the most suitable for the purposes of this research. Case study research is the most common qualitative method used in IS (Orlikowski and Baroudi, 1991). Although there are numerous definitions, Yin (1994) defines the scope of a case study as follows: "a case study is an empirical inquiry that:

- Investigates a contemporary phenomenon within its real-life context, especially when…
- the boundaries between phenomenon and context are not clearly evident" (Yin 1994, p. 13).

In relation to this research, Case Study:

- Is a methodology consistent with Social Constructionism, Inductive Reasoning, and Qualitative Approaches (Stake, 2006), (Walsham, 1995), (Eisenhardt, 1989), (Eisenhardt and Graebner, 2007), and (Lauckner et al. 2012).

- Ensures richness and depth in order to understand the phenomenon of interest (Anaf et al. 2007), (Flyvbjerg, 2006), and (Stake, 2006).
- Enables the exploration of complex situations allowing for the gathering of multiple perspectives, from a range of sources, including contextual information (Stake, 2006), (Flyvbjerg, 2006), and (Lauckner et al. 2012).
- Is particularly useful when the unit of analysis is a process, which is compatible with the research question of this study (Stake, 2006), (Walsham, 1995), and (Lauckner et al. 2012).

In the context of the constructionist paradigm in which this research is situated, Yin (2003) and Stake (2006) emphasize the importance of establishing a specific framework that structures data collection in a case study. Inductive qualitative case study researchers usually combine multiple data collection methods (Eisenhardt, 1989) and keep the design of the process flexible. Yin (2003) argues that evidence for case studies may come from six different sources: documents, archival records, direct observation, participant observation, interviews and physical artifacts.

While observation was a natural source for collecting data during the time we spent within the case, document analysis and interviews were chosen as the other main sources for the data collection process. In interpretive IS case studies, as an outside observer, Walsham (1995) argues that interviews are the primary data source, "since it is through this method that the researcher can best access the interpretations that participants have regarding the actions and events which have or are taking place, and the views and aspirations of themselves and other participants" (Walsham, 1995, pp. 78). In relation to document analysis, Yin (2003) highlights five main documents that are often used in case study research. These are: (1) organizational website; (2) newspaper articles; (3) company reports and regulations; (4) policies; (5) e-mail correspondences. We believed that all these types of documents could enrich the understanding of the case from several angles.

3.2 Role of Theory

The central notion is to use a case study as the basis from which to develop theory inductively. "The theory is emergent in the sense that it is situated in and developed by recognizing patterns of relationships among constructs within cases and their underlying logical arguments" (Eisenhardt and Graebner, 2007, pp. 25). In this way, it is important to reflect upon the use of theory. Our research question can be classified as phenomenon-driven, i.e. that focuses more on the importance of the phenomenon and emphasizes the lack of explanations given by existing theory. Here the research question is broadly scoped to give the researcher more flexibility. "The justification rests on the phenomenon's importance and the lack of viable theory and empirical evidence" (Eisenhardt and Graebner, 2007, pp. 26).

Diffusion was generally defined as "the process by which an innovation is communicated through certain channels over time among the members of a social system" (Rogers, 2005). In the context of this study, the innovation is represented by Open Data released by cities' authorities, and the social system is the group of users of such innovation for Service Innovation purposes. According to the innovation diffusion research, the user first has to make a decision on whether to adopt an innovation. To this purpose information is collected about the innovation and it leads to the formation of perceptions about the innovation. In line with these perceptions a decision to adopt or reject the innovation is made (Rogers, 2005). IS literature on Diffusion of Innovations is plentiful (Nakicenovic and Grubler, 1991) and several IS studies have focused on identifying factors that influence the diffusion of particular innovations (Quaddus and Xu, 2005), (Norton and Bass, 1987). Most of these studies used the models proposed by Ajzen and Fishbein (1980) and Davis (1989). In general terms, these authors suggested that some external factors affect the perceptions about an innovation, which in turn impact the diffusion of such innovation. This model is generic in nature and is likely to be applicable in most IS innovation diffusion processes (Quaddus and Xu, 2005). As a consequence, the term diffusion in the definition of the focus for this study is structured as follows: "External Factors" →"Perceptions" → "Dif-

fusion". This simple scheme will allow us to bind the data collection and analysis from the case that will be studied within the diffusion process without constraining too much the research and subsequently limit both the flexibility of the study and the quality of the findings.

4 The Case Study

The case study that was carried out in this research was in a company that offers a set of Open Data Services which main goal is to visualize planning and building permit data from open datasets to allow users to see what projects are being proposed in their area or city. This particular case was selected for several reasons:

- This is one of the few successful initiatives in terms of profitable business that developed a range of services from local authorities' open datasets that satisfy a want or need within a market.
- As a consequence of the previous point, the case is relevant for studying the factors that influence diffusion of Open Data for service innovation.
- The company made their data accessible from the beginning. In other words, a commitment for collaboration was ensured with the company as a first step. In particular, the CEO of the company agreed to share documents and experiences related to the case, and to allocate some of his time to be interviewed.
- The location of the business' offices, people and operations allowed the possibility of collecting data in person, and to actually observe activities when/if needed. This enabled more flexibility in scheduling the various meetings required to execute this research.
- The CEO of the company is also actively involved in other Open Data initiatives. This created the opportunity to engage with a subject with deep knowledge in the field of Open Data.

The company proposes a set of Open Data Services that analyze building permit data for the home remodeling and construction market. After its launch in Europe, it expanded its activities in United States. Currently, the service is available in eight major cities. A team of six people is working within the company. According to CEO, these people include a business developer based in US, three software developers, and a person responsible for front-end activities and marketing initiatives.

All in all, the services provided aim at making urban planning information easier to find. The company visualizes planning and building permit data to allow users to see what projects are being proposed in their area or city. Hence, it helps anyone who owns property to know what permits are being requested near or on their property, in real time. The service charges retailers, property owners and construction companies monthly fees to get leads and understand what's happening in the building environment around them or the one of their interest. This allows a user to see what is being planned in their area, and at what stage in the planning process the application is at. The service has additional features too which allow users to be notified of new applications in their area, a 3D interactive house to identify what works require planning permission, planning leads for suppliers, a directory of professionals, and a news feed of planning related information from local area plans to strategic infrastructure.

4.1 Data Collection

The case study research process proceeded through five main different stages. In first place, the CEO and founder of the company was contacted through emails to define the details of the potential collaboration, and to explain the objectives for this study. After receiving positive feedback, three more face to face meetings were agreed in the company's offices. The first was formulated as an informal meeting that would have allowed us to both collect data around the general information about the company, and the services offered, and to clarify ethical aspects for this study. A further meeting was also scheduled to undertake the Semi-Structured Interview (SSI) process. Ultimately, a last meeting was

agreed with the goal of ensuring that correct interpretation of the data collected has been achieved (Stake, 2006).

Data about the company was also gathered from all the information publicly available. We found several documents reporting the company's activities and interviews to the key people within it. From the information gathered through this process, additional elements to be considered and to be tackled more in depth through the semi-structured interview process emerged. Hence, after the first meetings, a clear picture around the company, its services, its people, and the activities it is involved in was achieved. The contextual data were also collected, analyzed and structured. At this stage what has been defined in (Stake, 2006) as the "situationality of the quintain" (i.e. the specific case-related context around the phenomenon to be studied), was explored. The next step was the conducting of the semi-structured interview process. The interviewee that was chosen was the CEO, as he was found as the most knowledgeable person within the company. In total the interview lasted for 55 minutes. Approximately 40 minutes were recorded, while the rest of the time was employed in the Entry and Exit stages (Myers and Newman, 2007). The interview was transcribed from the same day in which it was conducted. When transcribing the data, further concepts emerged. As a consequence another interview session was implemented to clarify and deepen on those aspects.

4.2 Analysis

After transcribing the SSI, summarizing field notes and collecting other documents, the steps that were implemented were (Miles and Huberman, 1994), (Darke et al. 1998), (Berg, 2001), (Patton, 2002): (1) manually review the transcripts, line by line and sentence by sentence, and all the data collected to uncover key patterns/themes and produce key words/phrases (indicated as "key statements" in this document); (2) produce labels/categories of these key words/phrases. Identify high level factors and corresponding variables and formulate tentative assertions about those for which strong evidence is found; (3) look for relationships among the factors; (4) develop raw tables of factors variables and their links for each case study; (5) ensure correct interpretation of the data has been achieved.

As a first stage, SSI transcripts were reviewed line by line. The choice of starting with this source of data was made as this is the main data collection method employed. In addition, a review of the documents collected and of the field notes taken was already implemented and it informed the SSI process. Thus, the objective was to extract all the key notions and subsequently to categorize them in first approximation in relation to the three main concepts of Diffusion of Innovations (see paragraph 3.2). In particular, we distinguished concepts referred to external factors (E), those related to perceptions (P) and ultimately those concepts that referred merely to the diffusion process (D). As a result, 186 statements fell in the "E" (External Factors) category, while for perception and diffusion 118 and 70 statements arose respectively. Due to the high number of concepts, these needed to be divided into more focused units of analysis. After reading several times these statements, we classified them in ten different clusters that are presented in Table 1.

Cluster	Description	Number of key statements included
A	External Factors related to Open Data and dataset characteristics	23
B	External Factors related to the local authority (i.e. the data publisher)	97
C	External Factors related to the company	67
D	External Factors related to the interaction between the Open Data supplier and the company	55
E	Other External Factors	56

F	Perceptions about usefulness of Open Data	60
G	Perceptions about Open Data Services business opportunity	32
H	Perceptions about local authorities opportunities	56
I	Other Perceptions	7
J	Diffusion	70

Table 1.   Clusters in Case Study Analysis

For each of these clusters, tentative assertions about the case were formulated. In particular, similar statements across different sources were grouped together. When strong evidence (estimated as evidence from at least two sources of information) was found, tentative assertions were formulated about the factors. Related variables and links were explored and defined subsequently.

4.3   Results

In conclusion, from this Case Study 12 general assertions were formulated in relation to 12 different factors (7 External Factors and 5 Perceptions Factors) influencing the Diffusion of Open Data for new service development. Furthermore, in relation to the Diffusion Factor, 6 additional sub-factors emerged. Each of these factors is described by a set of variables, and links between factors were derived as well. The resulting diffusion factors model is depicted in Figure 3.
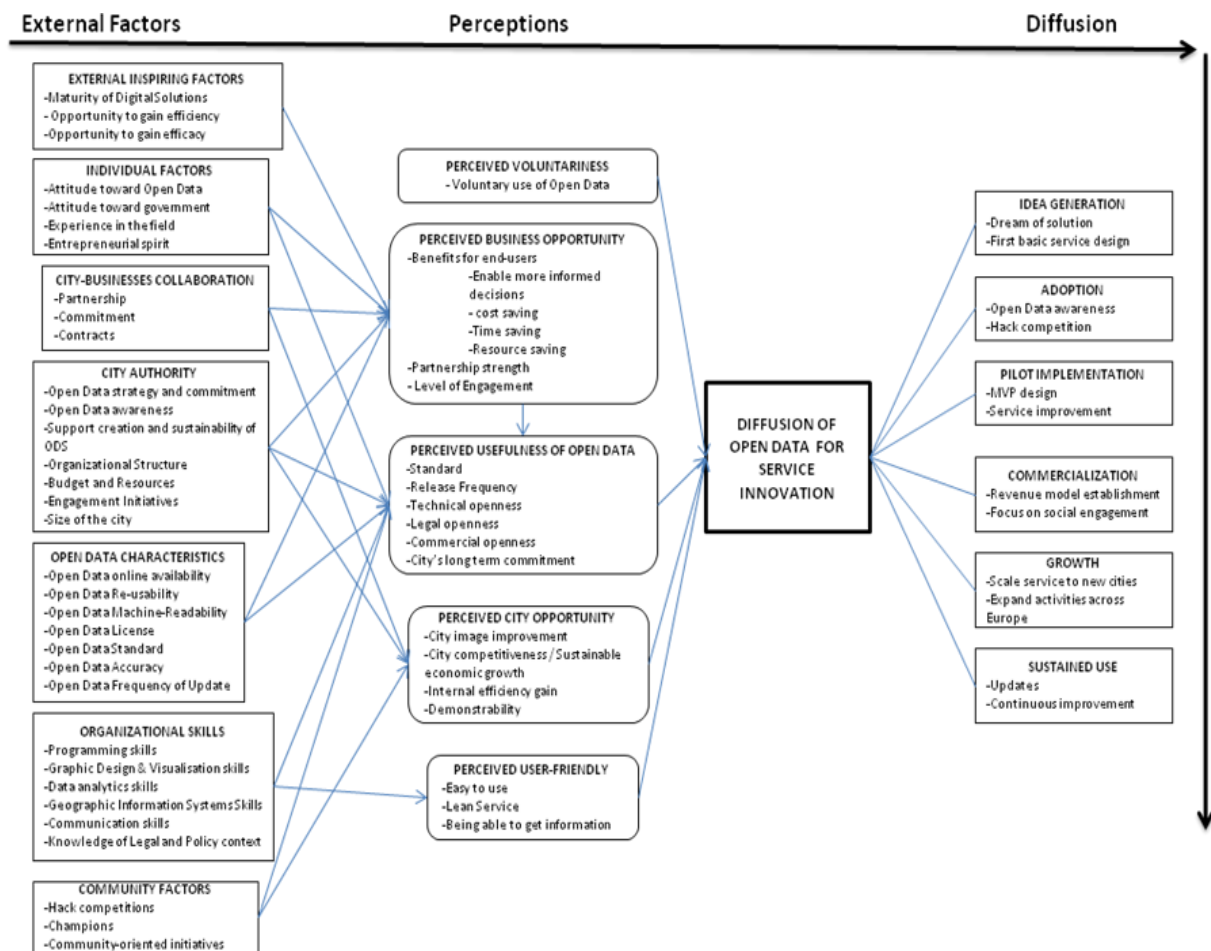

Figure 3.   Diffusion Model of Open Data for Service Innovation.

4.3.1 External Factors

As shown in Figure 3, seven External Factors that influence the diffusion of Open Data for new service development emerged from this case study. In first place the characteristics of the open dataset arose to be critical for fostering its re-usage. This factor includes aspects such as *online availability, machine readability, re-usability,* and availability of a commercial re-usage *license* that are already ingrained in the definition of Open Data (see introductory section of this document). However, *frequency of update* ("having near real-time data is a basic requirement for the delivery of our services" – the CEO stated), data quality dimensions and the presence of a common *standard* were also found to be critical variables in order for the dataset to comply with the company's aims, systems, and operations.

In addition, given that the case delivers its Open Data Services in different cities, the process on how different cities are targeted and selected was investigated. The findings showed that two main factors should be added to the model. These are *City Authority* and *City-Business Collaboration.* In relation to the former, in order to foster Open Data adoption by the case, cities needed to have *Awareness* about Open Data and a subsequent *Strategy* in place. Furthermore, from the continuous interactions that the case is having to "convince cities to come on board", it emerged that is crucial for cities to have a CIO, an established IT department, and people exclusively devoted to Open Data initiatives within the councils. The latter, instead, sheds light on the importance of establishing relationships between Open Data Suppliers (i.e. city authorities) and the Open Data Services business. In particular, quoting the CEO, having cities that "come on board" and "establish win-win relationships" with them are believed to be cornerstones for the diffusion of Open Data among service developers. In some cities, the case also stipulated *contracts* where in exchange of the service provided, the city authority commits to supply Open Data on a daily basis. Connected to this aspect, it could be interpreted that the low maturity (or often the absence) of digital solutions in the context strongly stimulated the case to design and deliver its Open Data Services. In other words, the opportunities of gaining both efficiency and efficacy in the urban planning operations were critical for the successful adoption of Open Data for new service development. Moreover, the importance of community-driven initiatives confirms the findings presented in (Kuk and Davies, 2011). General *Organizational Skills* (referred mainly to IT skills) needed to design and deliver a user-friendly Open Data Service were also outlined and structured in Figure 3. Finally, it emerged that the individual attitude towards Open Data and, more in general towards working with government bodies in the specific context, are also critical factors that stimulated the case to adopt the innovation.

4.3.2 Perceptions

As explained in the theoretical section (paragraph 3.2) the presence of a number of External Factors generates Perceptions about the innovation, and depending on those the innovation is (or is not) adopted. In first place, *Perceived Usefulness of Open Data* derived mainly from: (1) the characteristics of the datasets outlined above, (2) the access to community-oriented initiatives (to spread awareness about Open Data and its potential), (3) the presence of a long-term Open Data program within the council, (4) the availability of appropriate internal skills to deal with transforming the dataset into a valuable service, and (5) the individual attitude towards governments and Open Data. In other words, if these External Factors are in place, Open Data is perceived as useful for service innovation purposes, in terms of the achievement of a sustainable and frequently updated provision of accurate content free of charge for the design of the final service. Another perception that was found to influence the diffusion process of Open Data is referred to the related *Perceived Business Opportunity*. Once again, this perception is generated from the characteristics of the open datasets. In addition, the *status-quo* of solutions in the context, the previous experiences in the specific urban planning field, and the establishment of long-term relationships with the data suppliers were found to influence the creation of these perceptions. Within this cluster several aspects are encompassed, including opportunity to enable: more informed decisions, and cost time and resource savings from the previous way of conducting

these sets of operations. Clearly, if a business opportunity in the context from the availability of Open Data is achieved, it leads to perceived Usefulness of Open Data to implement such opportunity. Interestingly, another cluster of perceptions arose regarding the *Perceived City Opportunity*. In fact, as underlined several times, having cities that "come on board" was the cornerstone for the success of the case. Hence, if perceptions of opportunities from the city council are in place, sustainable provision of valuable Open Data is believed to be fostered. These perceptions include the will for cities to state that they "embraced Open Data" (which in turn is perceived as an improved image of the city itself). In addition it comprises also increased *City's Competitiveness* in terms of providing new sources (i.e. Open Data) for new businesses creation and so to attract new forms of capital and in turn foster sustainable local economic growth. Finally the last two clusters that emerged are referred to the characteristic of the service itself and the generations of perceptions about voluntary use of Open Data. In relation to the first, the importance of providing a service that is *lean, easy-to-use,* and that is able to provide accurate information was highlighted as a fundamental aspect. The second factor, instead, emerged as a standalone set of perceptions related to the voluntary use of Open Data (this aspect was mainly referred to ordinary people that join community initiatives voluntarily to develop new services for the community from Open Data).

4.3.3   Diffusion

The last category refers to the actual Open Data diffusion observed for the Open Data Services' company analyzed in this research. In this cluster, 6 sub-factors emerged as an outline of the diffusion process undertaken by the case. In first place, the CEO and founder of the company designed a very basic solution from a clearly identified problem derived from his working experience in the field of public urban planning processes. At the time of internally designing a solution for the city, the current CEO of the case gained awareness about the existence of Open Data through the participation to an Open Data Hack Competition. At this stage, Open Data was adopted for efficiently and effectively provide the content needed for the proposed solution. From there the first pilot of the application was designed and delivered in the local context. As part of this process, a detailed knowledge of the potential market was achieved. This contributed to the subsequent implementation of the *Commercialization* phase, in which that knowledge was leveraged into the establishment of revenue models connected to the service. At this stage the solution was available in a major European city and it could be "easily scaled" (according to the CEO) to other urban environments. Therefore, the company started its *Growth* by engaging with other city councils and subsequently expanding its services. Finally, *Sustained Use* of Open Data is achieved through the establishment of contracts with the partner cities in which the provision process of Open Data under certain characteristics is ensured. Furthermore, continuous improvement of the service and the underlying technology contribute to the actual implementation of this diffusion stage.

5   Conclusions and Future Works

This paper presented an interpretive case study on the diffusion factors of public sector's Open Data for new service development. The identification of 12 factors, the links between them, and 56 related variables that influence the diffusion of Open Data Services, represent the main contribution of this paper. Among other factors, we underlined the importance of collaboration, partnerships, and contracts between Open Data suppliers and users for fostering the sustainable delivery of an Open Data Service. The key point here is that cities need to be aware of the benefits that might derive from undertaking an Open Data program. These benefits include improving city image, increasing competitiveness in terms of ability of attracting investments and so foster sustainable economic growth, and internal efficiency gain opportunities. If this happens, cities are ready to ensure long-term commitment (some through contracts) in terms of provision of accurate data on a daily basis, in exchange for the delivery of a service that adds value to their data and increases substantially the efficiency of internal operations. In this regard, *Communication Skills* from the company to engage with new cities are crucial. Further-

more, the well acknowledged role of champions in the Diffusion of Innovations theory (Rogers, 2005) also emerged as fundamental in this process.

Overall, the identified factors led the case to overcome many of the barriers identified in (Janssen et al. 2012) and (Barry and Bannister, 2014). Finally, the company that was subject of this case study won one of the European hack-competitions for the re-usage of Open Data for new service development. Since this way of catalyzing Open Data Services innovation was highlighted as crucial in such ecosystem by previous studies (see section 2.2), some clarifications are needed. In detail, such initiative, instead of being the starting point from which the service was then refined and improved, was leveraged for increasing Demonstrability of the potential of Open Data Services, and so as a mean to engage with other cities (and therefore to enable business growth). In other words, the award is being shown to targeted city authorities in order to convince them in fostering an Open Data Services ecosystem.

These findings represent in our opinion several aspects to be further investigated as a potential contribution to the broad Diffusion of IS Innovations literature. In addition, a Systematic Literature Review study was presented in order to detail and structure the research efforts to date in this emerging field. Finally, the suitability of interpretive case study to inductively generate theory was demonstrated as a suitable way of conducting empirical research in this field. Nevertheless, we also acknowledge a potential limitation for this study. Despite generalized outcomes from one case study are achievable (Flyvbjerg, 2006), (Yin, 2004), (Eisenhardt, 1989), this debate is still ongoing. In this research, by assuming a Social Constructionist philosophical underpinning the process was kept coherent and consistent from its problem definition to the formulation of the final outcome. However, in order to provide a stronger contribution to existing theory, some case study methodologists argue that Multiple Case Study is a more viable option (Stake, 2006). Thus, future research will involve more cases with context-related diversities. As a result, additional aspects and stronger assertions are expected to emerge when conducting analysis across cases.

Acknowledgements

References

Ajzen, I., and Fishbein, M., (1980), *Understanding Attitude and Predicting Social Behavior*, Prentice-Hall, Englewood Cliffs, NJ.

Alter, S. (2002). *The work system method for understanding information systems and information systems research*. Communications of the Association for Information Systems, 9(1), 6.

Alter, S. (2010). *Viewing systems as services: a fresh approach in the IS field*. Communications of the association for information systems, 26(1), 11.

Amichai-Hamburger, Y., McKenna, K. Y., & Tal, S. A. (2008). *E-empowerment: Empowerment by the Internet*. Computers in Human Behavior, 24(5), 1776-1789.

Anaf, S., Drummond, C., Sheppard, L. A. (2007), *Combining case study research and systems theory as a heuristic model*. Qualitative Health Research, 17(10), 1309-1315.

Aoyama, M., Kojima, H. (2013), *LISA: Linked Services Architecture Based on the Linked Data and Service Broker*, IEEE 20th International Conference on Web Services.

Astley, W. G., & Van de Ven, A. H. (1983). *Central perspectives and debates in organization theory*. Administrative science quarterly, 245-273.

Barry, E., & Bannister, F. (2014). *Barriers to open data release: A view from the top*. Information Polity, 19(1), 129-152.

Berg, B. L. (2001), *Qualitative Research Methods for the Social Sciences*, Allyn and Bacon, Boston.

Bertot, J. C., Jaeger, P. T., & Grimes, J. M. (2010). *Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies*. Government Information Quarterly, 27(3), 264-271.

Bizer, C., Heath, T., & Berners-Lee, T. (2009). *Linked data-the story so far*. International journal on semantic web and information systems, 5(3), 1-22.

Chan, C. M. (2013, January). *From Open Data to Open Innovation Strategies: Creating E-Services Using Open Government Data*. In System Sciences (HICSS), 2013 46th Hawaii International Conference on (pp. 1890-1899). IEEE.

Cifuentes – Silva, F., Sifaqui, C., Labra – Gayo, J. E. (2011), *Towards an architecture and adoption process for Linked Data Technologies in Open Government Contexts: A Case Study For the Library of Congress of Chile*, 7th Conference on Semantic Systems.

Currie, J. L. (2013), *The Role of Canadian Municipal Open Data Initiatives: a Multi-City Evaluation*, Department of Geography at Queens University, Kingston, Canada.

Cyganiak, R., Maali, F., Peristeras, V. (2010), *Self Service Linked Government Data with dcat and Gridworks*, I-SEMANTICS 2010, Graz, Austria.

Darke, P., Shanks, G., & Broadbent, M. (1998). *Successfully completing case study research: combining rigour, relevance and pragmatism*. Information systems journal, 8(4), 273-289.

Davis, F. D., (1989), *Perceived usefulness, perceived ease of use, and user acceptance of information technology*, MIS Quarterly, 319-340.

Eisenhardt, K. M. (1989), *Building theories from case study research*, Academy of management review, 14(4), 532-550.

Eisenhardt, K. M., & Graebner, M. E. (2007), *Theory building from cases: opportunities and challenges*. Academy of management journal, 50(1), 25-32.

Felicetti, A., d'Andrea, A., & Niccolucci, F. (2012). *The Open Data Semantics and the (re) use of Open Information in Cultural Heritage*. ECLAP, International Conference on Information Technologies for performing Arts, Media Access and Entertainment, 2012.

Ferro, E., Osella, M. (2012), *Business Models for PSI Reuse: a Multidimensional Framework*, European Commission Workshop on Using Open Data: Policy Modelling, Citizens Empowerment, Data Journalism, Brussels.

Ferro, E., Osella, M. (2013), *Eight Business Model Archetypes for PSI Reuse*, Open Data on the Web Workshop, Google Campus, London.

Flyvbjerg, B. (2006), *Five misunderstandings about case-study research*. Qualitative inquiry, 12(2), 219-245.

Fox, M. S. (2013), *City Data: Big, Open and Linked*, Enterprise Integration Laboratory, working paper, available at: http://www.eil.utoronto.ca/smartcities/papers/City-Data-v5.pdf.

Frosterus, M., Tuominen, J., Wahlroos, M., & Hyvönen, E. (2013). *The Finnish Law as a Linked Data Service*. In The Semantic Web: ESWC 2013 Satellite Events (pp. 289-290). Springer Berlin Heidelberg.

Gautreau, P., Hasenack, H., Lerch, L., Merlinsky, G., Noucher, M., Severo, M. (2012), *A Comparison of the Open Environmental Data Diffusion in Argentina, Bolivia, and Brazil: some preliminary results from the Baguala Project for methodological discussion*, UFRGS, Brazil.

Glaser, B., Strauss, A., 1967, *The discovery of grounded theory*, Aldin, New York.

Groen, M., Meys, W., & Veenstra, M. (2013, September). *Creating smart information services for tourists by means of dynamic open data*. In Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication (pp. 1329-1330). ACM.

Guo, W., & Kraines, S. B. (2010, October). *Enriching city entities in the EKOSS failure cases knowledge base with Linked Open Data*. In Computer Information Systems and Industrial Management Applications (CISIM), 2010 International Conference on (pp. 404-409). IEEE.

Halb, W., Stocker, A., Mayer, H., Mülner, H., & Ademi, I. (2010, September). *Towards a commercial adoption of linked open data for online content providers*. In Proceedings of the 6th International Conference on Semantic Systems (p. 16). ACM.

Hannemann, J., & Kett, J. (2010, August). *Linked data for libraries*. In Proc of the world library and information congress of the Int'l Federation of Library Associations and Institutions *(IFLA)*.

Hartung, C., Lerer, A., Anokwa, Y., Tseng, C., Brunette, W., & Borriello, G. (2010, December). *Open data kit: Tools to build information services for developing regions*. In Proceedings of the 4th ACM/IEEE International Conference on Information and Communication Technologies and Development (p. 18). ACM.

Haslhofer, B., Isaac, A. (2011), *data.europeana.eu: the Europeana Linked Open Data Pilot*, International Conference on Dublin Core and Metadata Applications.

Hausenblas, M. (2009), *Exploiting Linked Data to Build Web Applications*, IEEE Internet Computing.

Herschel, M., & Manolescu, I. (2012, October). *DataBridges: data integration for digital cities.* In Proceedings of the 2012 ACM workshop on City data management workshop (pp. 1-4). ACM.

Hielkema, H., & Hongisto, P. (2013). *Developing the Helsinki smart city: the role of competitions for open data applications*. Journal of the Knowledge Economy, 4(2), 190-204.

Huff, A. S. (2008), *Designing Research for Publication*, Thousand, Oaks: Sage.

Ijima, C., Morita, T., Enomoto, Y., & Yamaguchi, T. (2011). *A mobility service based on Japanese linked data*. In Knowledge-Based and Intelligent Information and Engineering Systems (pp. 246-255). Springer Berlin Heidelberg.

Isaac, A., & Haslhofer, B. (2013). *Europeana linked open data–data. europeana. eu*. Semantic Web, 4(3), 291-297.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). *Benefits, adoption barriers and myths of open data and open government*. Information Systems Management, 29(4), 258-268.

Kuk, G., Davies, T. (2011), *The Roles of Agency and Artifacts in Assembling Open Data Complementarities*, 32nd International Conference on Information Systems, ICIS, Shanghai.

Lakomaa, E., Kallberg, J. (2013), *Open Data as a Foundation for Innovation: The Enabling Effect of Free Public Sector Information for Entrepreneurs*, IEEE Access, September 2013.

Latif, A., Saeed, A. U., Hoefler, P., Stocker, A., & Wagner, C. (2009, September). *The Linked Data Value Chain: A Lightweight Model for Business Engineers*. In I-SEMANTICS (pp. 568-575).

Lauckner, H., Paterson, M., Krupa, T. (2012), *Using Constructivist Case Study Methodology to Understand Community Development Processes: Proposed Methodological Questions to Guide the Research Process*, The Qualitative Report 2012 Volume 17, Article 25, 1-22.

Lindman, J., Rossi, M., Tuunainen, V. K. (2013), *Introduction to Open Data Services Minitrack*, In Proceedings of the 46th Hawaii International Conference on System Sciences (a).

Lindman, J., Rossi, M., Tuunainen, V. K., (2013), *Open Data Services: a Research Agenda*, In Proceedings of the 46th Hawaii International Conference on System Sciences (b).

Lindman, J., Tomi Kinnari, T., Rossi. M. (2014), *Industrial open data: Case studies of early open data entrepreneurs*, 47th Hawaii International Conference on System Sciences (HICSS), IEEE.

Lorey, J. (2013). *Storing and Provisioning Linked Data as a Service*. In The Semantic Web: Semantics and Big Data (pp. 666-670). Springer Berlin Heidelberg.

Matheus, R., Ribeiro, M. M., & Vaz, J. C. (2012, October). *New perspectives for electronic government in Brazil: the adoption of open government data in national and subnational governments of Brazil*. In Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance (pp. 22-29). ACM.

Miles, M. B., & Huberman, A. M. (1994), *Qualitative data analysis: An expanded sourcebook*, (2nd ed.). Newbury Park, CA: Sage.

Myers, M. D., Avison, D. (1997), *Qualitative research in information systems,* Management Information Systems Quarterly 21, 241-242.

Myers, M. D., & Newman, M. (2007). *The qualitative interview in IS research: Examining the craft*, Information and Organization, 17(1), 2−26.

Nakicenovic, N., Grubler, A. P., (Eds.) (1991), *Diffusion of Technologies and Social Behaviour*, Springer, Berlin.

Norton, J. A., Bass, F. M., (1987), *A diffusion theory model of adoption and substitution for successive generations of high-technology products*, Management Science 33 (9) 1069–1086.

Oh, J. (2013), *Literature Review – Open Data in New Zealand*, University of Auckland Department of Statistics, available at: https://www.stat.auckland.ac.nz/~joh024/LitReviews/LitReview_NZOpenData.pdf.

Okoli, C., Schabram, K. (2010), *A Guide to Conducting a Systematic Literature Review of Information Systems Research*, Sprouts.

Orlikowski, W., Baroudi, J. (1991), *Studying Information Technology in Organizations: Research Approaches and Assumptions*, Information Systems Research (2) 1991, pp. 1-28.

Patton, M.Q. (2002), *Qualitative Research and Evaluation Methods*, Thousand Oaks, CA: Sage.

Porter, M. E. (1990), *The competitive advantage of nations*, Harvard Business Review.

Rogers, E. M., (2005), *Diffusion of Innovations*, fifth edition, Free Press, NY.

Quaddus, M., & Xu, J. (2005), *Adoption and diffusion of knowledge management systems: Field studies of factors and variables*, Knowledge-Based Systems, 18(2-3), 107-115.

Rittenbruch, M., Foth, M., Robinson, R., & Filonik, D. (2012). *Program your city: designing an urban integrated open data API*. In Proceedings of Cumulus Helsinki 2012 Conference: Open Helsinki–Embedding Design in Life (pp. 24-28). Cumulus International Association of Universities and Colleges of Art, Design and Media.

Savelyev, A., Xu, S., Janowicz, K., Mülligann, C., Thatcher, J., & Luo, W. (2011, November). *Volunteered geographic services: Developing a linked data driven location-based service*. In Proceedings of the 1st ACM SIGSPATIAL International Workshop on Spatial Semantics and Ontologies (pp. 25-31). ACM.

Scott, W. R. (1995), *Institutions and Organizations,* Thousand Oak, CA: Sage.

Stake, R. E., (2006), *Multiple Case Study Analysis*, NY: The Guilford Press.

Stephenson, M., Di Lorenzo, G., Aonghusa, P. M. (2012), *Open Innovation Portal: a Collaborative Platform for Open Data Sharing*, PerCom Demos, Lugano, Switzerland.

Tammisto, Y., & Lindman, J. (2011). *Open data business models*. In The 34th Information Systems Seminar in Scandinavia, Turku, Finland (pp. 762-777).

Tcholtchev, N., Farid, L., Marienfeld, F., Schieferdecker, I., Dittwald, B., Lapi, E. (2012), *On the Interplay of Open Data, Cloud Services and Network Providers Towards Electric Mobility in Smart Cities*, IEEE GoSmart 2012, Global Trends in Smart Cities.

Walsham, G. (1995), Interpretive *Case Studies in IS research: Nature and Method*, European Journal of Information Systems (4) (pp. 74-81).

Wilde, E. (2010). *Linked Data and Service Orientation.* In Service-Oriented Computing (pp. 61-76). Springer Berlin Heidelberg.

Yin, R. K. (1994), *Discovering the future of the case study method in evaluation research*, Evaluation Practice, 15(3), 283-290.

Yin, R. K. (2003), *Case study research*, (Vol. 5). Thousand Oaks, California.

Yu, H. Q., Dietze, S., Li, N., Pedrinaci, C., Taibi, D., Dovrolls, N., ... & Domingue, J. (2011). *A linked data-driven & service-oriented architecture for sharing educational resources*, In: 1st International Workshop on eLearning Approaches for Linked Data , 8th Extended Semantic Web Conference