Acta Universitatis
Lappeenrantaensis
809

Antti Herala

BENEFITS FROM OPEN DATA: BARRIERS TO SUPPLY AND DEMAND OF OPEN DATA IN PRIVATE ORGANIZATIONS

Antti Herala

BENEFITS FROM OPEN DATA: BARRIERS TO SUPPLY AND DEMAND OF OPEN DATA IN PRIVATE ORGANIZATIONS

Thesis for the degree of Doctor of Science (Technology) to be presented with due permission for public examination and criticism in the Auditorium 2310 at Lappeenranta University of Technology, Lappeenranta, Finland on the 12th of October, 2018, at noon.

Acta Universitatis
Lappeenrantaensis 809

Supervisors Professor Jari Porras
LUT School of Engineering Science
Lappeenranta University of Technology
Finland

Professor Timo Kärri
LUT School of Engineering Science
Lappeenranta University of Technology
Finland

Reviewers Professor Tommi Mikkonen
Department of Computer Science
University of Helsinki
Finland

Professor Virpi Tuunainen
Department of Information and Service Management
Aalto University
Finland

Opponents Professor Tommi Mikkonen
Department of Computer Science
University of Helsinki
Finland

Adjunct Professor Juho Lindman
Faculty of Information Technology and Electrical Engineering
University of Oulu
Finland

ISBN 978-952-335-261-2
ISBN 978-952-335-262-9 (PDF)
ISSN-L 1456-4491
ISSN 1456-4491

Lappeenrannan teknillinen yliopisto
Yliopistopaino 2018

Abstract

Antti Herala
Benefits from Open Data: Barriers to Supply and Demand of Open Data in Private Organizations
Lappeenranta 2018
66 pages
Acta Universitatis Lappeenrantaensis 809
Diss. Lappeenranta University of Technology
ISBN 978-952-335-261-2, ISBN 978-952-335-262-9 (PDF), ISSN-L 1456-4491, ISSN 1456-4491

Open data is a concept that has been somewhat readily adopted by governments. The key ideology behind open data is to release previously restricted data with a licence that allows all forms of use, reuse, and redistribution. Governments are opening their data in order to become more transparent towards their citizens but also in order to foster individual and organizational innovations. However, the innovations become one-sided and are difficult to make based on only government data, which has led an interest towards open business data, found from privately owned organizations, who are rarely opening their data. While open data can be used to bring value to the data publisher as well as the users, the companies are not interested in these prospects.

The goal of this thesis is to illuminate issues behind the company reluctance to open their data as well as identify issues in the use of available open data. The value of opening data for companies was identified from the literature, while the interest towards open data was surveyed from companies. Afterwards, interviews were organized with software developers to further recognize the issues of supplying and demanding open data. To further highlight the issues of publishing open data, it was compared to hackathons, which are used for external innovation and crowdsourcing by companies.

Supplying open data was not seen as a viable solution for companies, exactly because the process of giving away something that could be used for value does not suit the current business environment. It is important for a company to receive benefits and/or profits from their data, and be able to have control over the data use but also the innovation process; the aspect of open data fostering innovation does not seem to apply in business. Issues were also found in the usability of open data: it is a difficult resource to use because it requires resources to find and may not even suit the necessary context. New forms of communication and intermediary business models are required, which would minimize the risks and costs of opening data to benefit the publisher as well as increase the two-way dialogue between interested parties. At the same time, new models would allow better discoverability and a usable structure for already opened data, which would make it easier to use in multiple contexts, stimulating the use of open data.

Keywords: open data, data business, business change, economic value, hackathons, data management, private organizations

Acknowledgements

A few months before starting the research for this thesis, I remember adamantly stating that I do not want to do research and I am not going to stay in the university any longer than I have to. Then, a professor in the form of Jari Porras walked into the office, where I was finalising my Master's Thesis and offered me a position on a project and a chance to do this thesis. After a careful deliberation (and some *advice* from my future colleagues), I ended up where I am now. I cannot say I do not regret this decision, since this has been the hardest and the most arduous task I have ever done, and maybe will ever do, but it was also gratifying and fun.

I'd like to offer the first thanks to Professor Porras for giving me this opportunity to do something I thought I would never do. Jari also allowed me the freedom (and funds!) to do what I thought was necessary for this topic, the freedom to go places, where I could find new people, data for my research, and a general understanding of the field. I would also like to thank my other supervisor, Professor Kärri, for his endless talks and discussions with me about this — and many other — topics. Talking about this general field allowed me to formulate the general sense of research problem and find, what was in it for me.

I also appreciate the input from my honoured reviewers, Professors Mikkonen and Tuunainen. Your feedback was valuable and most appreciated.

Aside from the official supervisors, I'd also like to thank my "shadow supervisors", Dr. Erno Vanhala and Adj. Prof. Jussi Kasurinen. Without Erno, I wouldn't have started this thesis at all and without Jussi, I most likely wouldn't have been able to finish it. So thanks for that.

I would also like to thank all of the other parties, who have been active or passive contributors to this journey. They include funders (such as DIMECC and their S4Fleet-project), data sources and collaborators (6Aika-project and Forum Virium), and all the co-authors and research participants that gave their time and energy for this research.

And finally, I need to offer my gratitude to my wonderful family. My wife, Hanna, who stood by me and listened to my continuous complaints about everything related to this research. And of course, the thanks has to be offered to my children, Juuli and Johannes, without whom this would have been so much easier, but also so much less enjoyable.

Antti Herala
October 2018
Lappeenranta, Finland

Contents

List of publications

This thesis is based on the following papers. The rights have been granted by publishers to include the articles in the dissertation.

I. Herala, A., Vanhala, E., Porras, J., and Kärri, T. “Experiences about Opening Data in Private Sector: A Systematic Literature Review.” In 2016 SAI Computing Conference (SAI), 715–724, 2016.

II. Herala, A., Kasurinen, J., and Vanhala, E. “Current Status and the Future Directions of Open Data: Perceptions from the Finnish Industry.” In Proceedings of the 20th International Academic Mindtrek Conference, 68–77. ACM, 2016.

III. Herala, A., Kasurinen, J., Vanhala, E., 2018. “Views on Open Data Business from Software Development Companies.” Journal of Theoretical and Applied Electronic Commerce Research, 13(1), pp. 91-105.

IV. Herala, A., Kokkola, J., Kasurinen, J., Vanhala, E., “Strategy for Data: Open It or Hack it?” Journal of Theoretical and Applied Electronic Commerce Research (JTAER). (Forthcoming)

In this thesis, the publications are referred to as *Publication I*, *Publication II*, *Publication III*, and *Publication IV*.

Author's contribution

I. Planned the research, collected articles for review, analysed the reviewed articles, and wrote most of the publication.

II. Made a plan for the research, analysed the collected data, and wrote most of the publication.

III. Made the research plan, participated in designing the interviews, transcribed the interviews, participated in the analysis, and wrote most of the publication.

IV. Planned the research, conducted part of the interviews, participated in the data analysis, and wrote the most of the publication.

Nomenclature

Abbreviations

API	Application Programming Interface
B2B	Business-to-business
B2C	Business-to-consumer
C2B	Consumer-to-business
CEO	Chief Executive Officer
GDPR	General Data Protection Regulation
LSC	Lean Service Creation
ODB	Open Data Barometer
PSI	Public Sector Information
SLR	Systematic Literature Review
SMS	Systematic Mapping Study

1 Introduction

Open data, a concept readily adopted by governments, is a part of a larger open movement, that consists of open source, open science, open format, open education etc. initiatives, where the goal is to provide previously paywalled or restricted material available to everyone [1]. The spreading openness is creating changes in the society [2] but also to current business models [3], [4], or even making the role of some industries difficult or redundant [5]. For example, the software development industry is already being moulded or demanded towards using open source software development and business models [6], while open science and open education are transforming the universities and scientific publishers, their roles in society [5], [7], and even their own processes [8].

Open data has not yet been adopted by larger audiences, unlike the open source initiative. Mainly open data has been realized by governments, called open government, where governments publish the data they have been collecting to the citizens and public in general and take actions to ensure transparency in government. The benefits of these actions are increased transparency and improved democratic behaviour for the government, while for citizens this means more possibilities for innovation and entrepreneurship, to mention a few [2]. Through these actions, the estimated economic benefit is said to be somewhere between over two billion [9] and three trillion [10] dollars annually, even though the economic value is difficult to estimate. The advancement of open data in public governance has raised the question, whether or not there should be commercial open data suppliers as well [11]. This would lead to accessible open data from governments, but also from businesses, as is happening in open source community [12].

Opening data does not sound like something that the businesses would jump at, giving away data they own. While it does not sound attractive at first, there are empirical examples of such actions, for example [4], [13], [14], evidence that even a business can profit from sharing their data.

On the other hand, using open data should sound interesting for a company: a free resource, which can be used in any way the company desires. However, even the use of open data has not emerged, since governments are reporting issues in the usage of their data [15].

These issues with supplying and demanding open data for companies led to the development of the main research question: "*How do private organizations supply or demand open data?*" and it is further divided into two sub-questions. The sub-questions focus on two different aspects of open data in private organizations. One of the goals is to gauge the profitability of opening data as a business, while the other goal is to determine the viability of open data as a resource for business applications. The sub-questions are:

1. How do companies perceive the profitability of opening data?
2. What kind of an asset is open data for private organizations?

To answer these questions, this research is divided into three parts. In the first part, the viability of open data is measured, the profitability of opening data for a business with a systematic review of literature and company interest towards open data with a survey. In the second phase, these two lines of research were combined through a qualitative research, interviewing software organizations on their views, and the views of their clients, towards open data and opening data, how open data initiative is executed in the current business environment. Because from this phase it was found that opening data is not done for the business, but it was highlighted that hackathons – events that incorporate some aspects of open data, especially in innovation development [16] – are used increasingly in companies. This led to the third phase, where benefits of opening data from the first phase were compared to the benefits of organizing a hackathon through a qualitative study into the hackathon organizers.

This thesis shows, that open data is a tool that can be used for profit, and there is definite interest towards it, even from industries that are not known for heavily collecting and using data, such as designers and coffee shops. However, open data has not been adopted by businesses, and this thesis illuminates multiple issues in the execution of open data initiative, that makes it unfit for business. Based on this work, the process for opening data can be and should be changed to support the business goals and profitability instead of blindly opening data because of opening data.

This thesis has been divided into two parts: the introduction and an appendix with the four scientific publications that compose the major findings and research methods. The introduction has been divided into six parts, including this chapter. In the next chapter, chapter 2, the scientific background about open data and open data business has been introduced and the key concepts are defined. In chapter 3, the research methods are presented in lieu of the research problem and position – research focus and epistemology – derived from the literature. Chapter 4 summarizes the publications, available in the appendix, shortly describing the contents of the articles and their relation to the whole thesis. Chapter 5 combines and discusses the results for both, practical and academic communities, as well as the limitations of this research. Chapter 6 finally concludes this thesis, offering a concise view of the topic and results, summarising future key points that can be derived from these results.

2 Open data for businesses

In this chapter, the current literature is presented. Open data and the value of it are explained in the light of current research, as well as the rise of open data competitions and hackathons, events that invite and encourage the use of open data [17].

2.1 Open data

For the definition of open data, the Open Definition is used: "*Open data and content can be freely used, modified, and shared by anyone for any purpose*" [18]. The concept is also further elaborated in the Open Data Handbook: "*Open data is data that can be freely used, re-used and redistributed by anyone – subject only, at most, to the requirement to attribute and sharealike*" [19]. This definition requires three main aspects of opened data:

- Availability and access
- Re-use and redistribution
- Universal participation

The goal of these aspects is to provide interoperability of data [20] and further the interoperability of systems and organizations. Through interoperability, it is possible to build more complex systems that allow participation, collaboration, and a common set of data, accessible to anyone for any reason.

The origins of open data and especially open public data can be traced back to the early years of this millennia when the European Union released a Public Sector Information (PSI) directive in 2003 [22]. This directive was designed to enable public data to be used by third parties with low costs and scarce restrictions while ensuring that all data users would have equal opportunities to use the data. This directive was later followed on the other side of Atlantic in 2009, when Barack Obama, as the president of the United States, issued the Memorandum on Transparency and Open Government [23]. These examples were imitated by multiple countries, and governments started to open their data. To this day, the Open Data Barometer (ODB) measures some form of open data activities in most of the countries in the world [21]. The countries that were measured to be the most open in 2016 are presented in Table 1.

Table 1. Top five open data leaders in the world, from [21]

Rank	Country	Score (out of 100)
1	United Kingdom	100
2	Canada	90
3	France	85
4	United States of America	82
5	South-Korea	81
5	Australia	81

The process of opening data is an iterative cycle with five steps and has essentially two main groups, data publisher and data user [24], illustrated in Figure 1. The first step in the process is the collection of the data, followed by the opening of this data. After the data is opened, the users can find and use the data in order to pursue their own needs with the accessed data. The final step is that the users and publishers offer feedback about the data, in order to make the opened data more relevant and the quality and other attributes of the data can be further improved during the first steps.

Additionally, Sieber and Johnson [25] recognized four strategies for a government to open data. These are (1) Data Over the Wall, (2) Code Exchange, (3) Civil Issue Tracker, and (4) Participatory Open Data. The first two strategies focus mostly on governments opening their data actively through portals, but in (2) guidance is offered to data users, while active engagement and even commercialization are being invited. The third strategy leans towards citizens, who are providing data from the "field" and informing the public bodies of issues; the government actively seeks feedback and participation in a form of data from the citizens. The fourth strategy is to engage citizen participation while remaining as an active operator; data is provided by governments while requesting citizen-generated data. While these strategies all vary in their execution, they can still be seen as a part of the open data process in Figure 1.

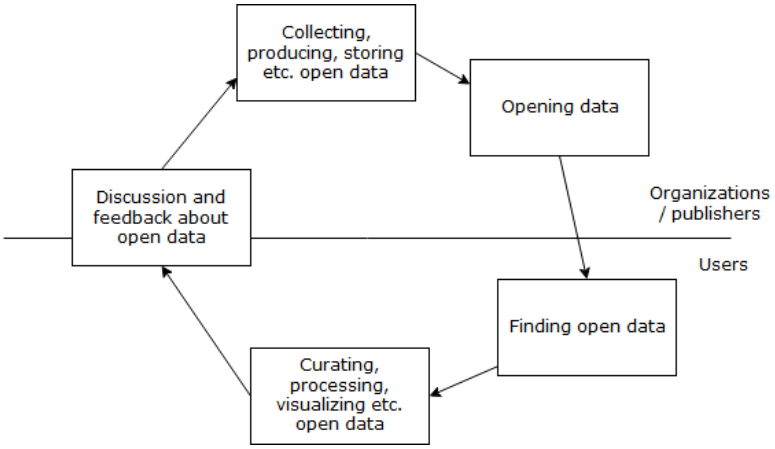

Figure 1. Open data process, adapted from (Zuiderwijk, Janssen, Choenni, Meijer, Alibaks 2012)

The perspectives that are used to view open data differ, since, on one hand, open data can be seen from the citizen or government point of view. To others, it serves a political agenda, while others see technological opportunities. Moreover, for data users, it is possible to view open data from the economic perspective, focusing on the revenue it can offer. In the open data literature, Gonzales-Zapata and Heeks [26] recognized four perspectives for open data, Bureaucratic, Technological, Political, and Economical perspectives. Additionally, other perspectives that are mentioned are Institutional

Table 2. Open data perspectives [17], [26]

Perspective	Description
Political	Political differences and developments between countries and governments, especially in relation to open data movement.
Social	Focuses on social benefits, such as transparency and accountability, from the viewpoint of cultural differences and differences in agendas of countries.
Economical	Financial benefits and profits that can be achieved with open data, including the reuse of open data and the impact on innovation.
Institutional / Bureaucratic	Describing, how institutions are enabling and constraining the publication and adoption of open data, suggesting that open data could become an integral part of data collection systems instead of being a separate step in the process.
Operational	Tools (e.g. standards) that enhances the usability of open data and focuses requirements for the data publication processes.
Technical	The viewpoint of technologies towards open data, such as platforms and infrastructures, focusing on the importance of visualizations and analyses, as well as the metadata for reuse.
Legal	Describes the importance of open data legislation, emphasizing freedom of information acts, policies, and directives.

(similar to Bureaucratic), Social, Operational, and Legal perspectives [17]. These are described in Table 2.

Depending on the perception towards open data, the perceptions towards its value can change. For example, the technical challenges of open data can be valuable for the software industry, while the political and legal agendas serve public bodies and governments. Open data can yield economic value, but there are other aspects that should be considered, which are presented in the next section.

2.2 Value of open data

The value of open data is difficult to measure since the profitability in the movement does not come from direct sales, and the openness of data is not in itself enough to sell a service [27] nor does data openness necessarily have significant value by itself [20]. Some researchers define the value of open data to be the combination and aggregation of data and the interdisciplinary user network, which allows new insights from the data, adding value to everyday life [28]. In addition to this, open data is seen to increase transparency, boost the economy, and enable external development [29]. Even a boost for creativity and innovation has been suggested, as open data allows users and entrepreneurs to explore and play with the data [30], [31].

While the value of these open data enabled actions, such as innovation and data aggregation is difficult to determine, several studies, for example [9] and [10], have been conducted in the field, estimating the economic value of open data, summarised in Table 3.

Table 3. Studies estimating the value of open data

Publication	Timespan	Scope	Estimated value
Loomis, Koontz, Miller, and Richardson [9]	2011	Global	$2.19bn
McMurren, Verhulst, and Young [32]	2005-2009	Denmark	€62m
Tong, Irshad, and Ward Revell [33]	2011-2012	UK	£1.8bn
Manyika, Chui, Groves, Farrell, Van Kuiken, and Doshi [10]	Annually	Global	$3tn

Measuring the complete value of open data is not a simple matter since the economic value does not cover all the aspect of value [34]. To measure the impact of open data, there is a definition of sustainable value, which consists of three parts: economic, environmental, and social value [34]. While the value that open data can bring is an issue of sustainable value and should be treated as such, the benefits are measured through economic value, the most common metrics of value. These different aspects of value are more closely described in Table 4.

Table 4. Aspects of sustainable value, from [34]

Value proposition	Description
Economic	Economical value is usually measured in currencies and represents the maximum amount of a selected currency that a consumer is willing to pay for an item in a free market [35].
Environmental	The environmental value represents the monetary value of environment and actions towards it, such as clean air or water and the effects of fishing and agriculture from the environmental and sustainability point of view [36].
Social	Social value is measured from the individual point of view, through experiences and subjective feelings that allow individuals to be more productive, change views towards the world and make more sustainable decisions [37], [38].

Transforming data – especially open data – into value is a difficult process. The article of Jetzek el al. [34] suggests a model, that allows the transformation of open data into sustainable value. It consists of enabling factors and value generating mechanisms, illustrated in Figure 2. The authors describe the enabling factors as methods, which offer the individual motivation, opportunity, and ability to create sustainable value from open data, while the value generation mechanisms offer tools and models to engage open data effectively and with a determined goal. The study by Zuiderwijk et al. [39]

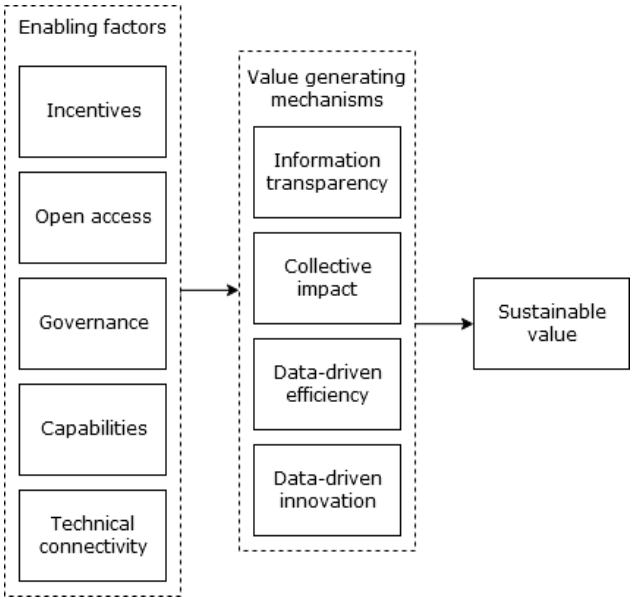

Figure 2. From data to value in sharing society, adapted from [34]

supports this model and states, for example, that the absence of open data incentives negatively affects open data usage.

This model offers possibilities for individuals to generate value from open data, but in the case of businesses, some form of even more exact business and revenue models are required. Regardless of the source, open data can be moved directly from data publishers to users, but there also exists infomediary business models between these actors [3], as is shown in Figure 3.

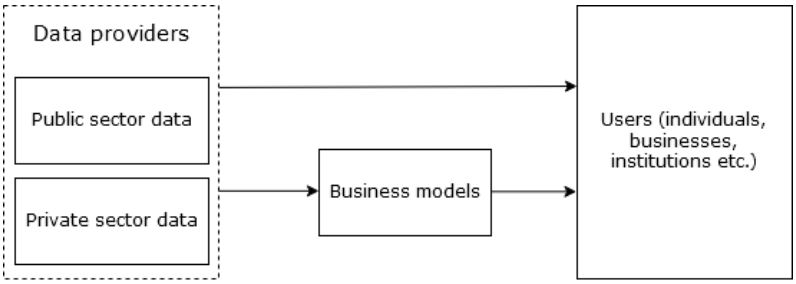

Figure 3. Position of business models in open data delivery, adapted from [3]

These open data business models can be divided into two groups focusing on a narrower concept of revenue models or a wider concept of business models; revenue model is considered as a subgroup of business models [3]. Revenue models focus on the revenue that an action brings, while business models, in addition to revenue, capture other

aspects, such as users and networks, cost structures, pricing methodologies, margins, and expected volumes [3], [40]. Ferro and Osella [41] suggest eight different revenue models and Zeleti, Ojo, and Curry [42] offer seven more, described in Table 5.

Table 5. Revenue models for open data, from [41]–[44]

Model	**Description**
Premium	Product or service in exchange for payment.
Freemium	Basic services and product are offered free, but the extended features are accessible through payment.
Open source like	Data is offered for free in exchange for assist in further development.
Infrastructural Razor and Blades	Datasets are free to use for anyone through an API, but users are charged for the amount of computing power they demand.
Demand-oriented platform	The company provides a data platform for developers and requires in exchange for advanced services and datasets.
Supply-oriented platform	Same as previous, but data providers are also charged.
Free as branded advertising	The goal is to attract attention through open data and services based on it. Revenue comes from other services after the brand is recognized by consumers.
White label development	Using open data as a marketing device and rebranding open data through own services.
Cost avoidance	Reducing the cost of publishing data with a sustainable publication tool.
Sponsorship	Giving the product for free and gaining revenue through a sponsor.
Dual licensing	Allowing the use of two separate licenses, open license for some purposes, while the closed license for others.
Support and services	Offering support for the use of data, prioritizing users and offering services for the use of data, such as visualization.
Charging for changes	The product is given free, but the changes are charged.
Increased quality through participation	The quality of the product is increased through co-creation, allowing larger margins.
Supporting primary business	Publishing data in order to support the primary business with the published data.

While there exists a multitude of revenue models for open data, they tend to focus on usefulness, process improvement, performance, and customer loyalty [45]. Ferro and Osella [41] divide the revenue models into three main categories: recurring fees or pay-per-use, advertisements, and ensuring visibility.

As it was previously mentioned, business models should take into account a larger set of actors than a revenue model does. In comparison to revenue models, the next table (Table 6.) contains the suggestions for actual business models by Janssen and Zuiderwijk [3].

Table 6. Business models for open data, from [3]

Model	**Description**
Single-purpose apps	An application that is usually based on one kind of data and offers a single function. It can provide services, such as weather information or timetables for public transportation.
Interactive apps	Extending single-purpose apps, these applications offer includes the possibility to add content, such as user ratings or complaints.
Information aggregators	This form of business takes data from multiple sources and creates new value through the new set of data. The data can be combined from multiple distinct sources, such as websites, private databases, and open data.
Comparison models	Similarly, aggregating data from multiple sources, but this form of business allows the comparison between entities.
Open data repositories	A storage for open data, where users can search the data they need or want to use; they are designed as a neutral source of open data that does not discriminate the sources.
Service platforms	Service platforms offer extended features for processing the data, such as searching, importing, cleaning, processing, and visualization. These platforms usually use different open data repositories as the source of data.

As can be seen from the different business and revenue models, it is possible to participate in the open data community and profit from it in different ways. Five roles of private – or public – organizations in open data ecosystem have been identified [46]: Suppliers, Aggregators, Developers, Enrichers, Enablers. Additionally, the literature mentions a sixth role, the Customers. The roles and their tasks are further described in Table 7. Additionally, these roles and their positions illustrated in Figure 4, that extends the view of possible business models and actors in open data delivery (Figure 3). Figure 4 also highlights the important factor in open data delivery: the role of the supplier is paramount to the success of open data based business, but at the same time the supplier is not necessary the party, that benefits the most from the release [12].

Table 7. Roles of organizations in open data ecosystem [46]

Role	Description
Supplier	Usually public, but also private organizations, who supply open data.
Aggregator	Their task is to collect and aggregate data, in order to visualize data and perform rudimentary analytics from it.
Developer	Designing and implementing applications for end-users, using the accessible open data.
Enricher	Creating added value in a form of enhanced services and products with open data.
Enabler	Facilitates the supply of open data in a form of hosting, instrumentation of data and/or products, and consultation.
Customer	End-user of the data.

The strength of open data, when compared to closed data, is that it is accessible to anyone for anything [48]. This allows open data to be used in small organizations, in start-ups, or even by individuals in any of the roles described before, albeit not necessarily supply [31]. While it may be challenging for an individual to adopt open data and communicate the results, through organizing and events open data can be used by the individuals and teams, transformed into products and services, gaining visibility. An effective medium that highlights the commercialization are hackathons and open data competitions, especially civic hackathons [17].

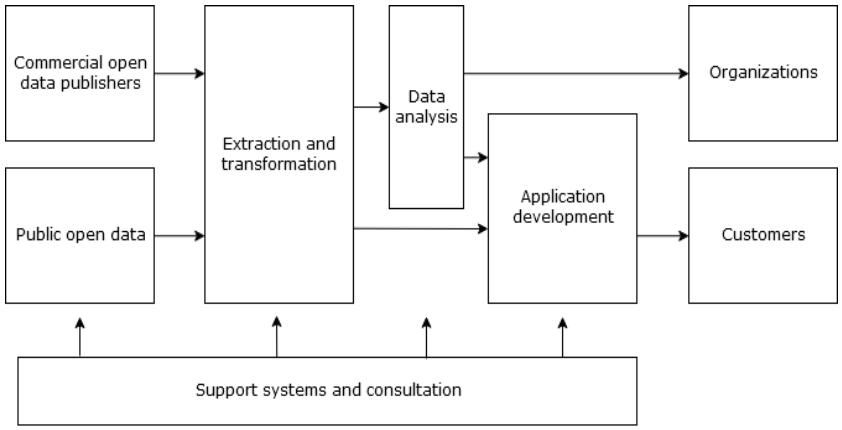

Figure 4. Roles in open data process, adapted from [47]

2.3 Issues with open data

While open data may be valuable and beneficial, there are also a multitude of issues within the initiative that does not necessarily allow companies or even public bodies to either supply open data or use open data. Even if the goals and actions are directed towards the beneficial opening of data, there can be unpredicted issues within the

organization and opening process or third parties can cause problems to the data publisher.

One of the major issues are legal constraints: local or international legislation [15]. It can be possible to open specific data in one country, while the same data is being prohibited to open in another, without mentioning changes in national laws and regulations and keeping track of them. This has been recently demonstrated with the emergence of General Data Protection Regulation (GDPR) in European Union, which implements major changes in policies related to personal data [49]. Also related to personal data, open data has a major risk of causing privacy infringements, because often the anonymization of data is a balance between revealing personal information and making the data usable [50]. Especially poorly anonymized confidential data released inadvertently can be de-anonymized, revealing groups and individuals to the public [51].

For organizations, the process of opening data can be a difficult one because of multiple reasons. It can cause issues within the organization or even conflicts between different organizations. Opening data requires changes in the data management processes, systems and technologies, employees and their mentality, division of labour [15], focus and policies [50], and even business models [17]. In addition, changes can apply to other organizations as well, especially if other companies are delivering data to the data publisher [14], [52]. In an unsuitable culture, opening data cannot necessarily reach the potential it has and may lead to unwanted behaviour in employees and processes [51] and similarly the conflicts between organizations can render opening data and harnessing benefits difficult, or even useless [52].

2.4 Hackathons and other competitions

Hackathons are relatively short events that focus on developing solutions to a problem or engaging open innovation with multiple stakeholders. The term originated from open source software developers in around the change of the millennium [53]. Because the term is somewhat new, there does not exist a common consensus on what is a hackathon and what elements differentiate it from other similar events, such as code camps [54], game jams [55] or other intensive development events. Usually, they are referred as technology-driven events, but there are also non-programming events calling themselves as hackathons [56] because they engage open innovation. Regardless of the exact goal of the hackathon, they are used as a tool to link external resources and professionals with internal innovation [57].

Hackathons are marketed and analysed based on the context and focus of the events; they vary, especially in advertising, based on the objectives and requirements [53]. They can be used in multiple contexts, for example, educational hackathons [58], industrial hackathons [59], culture hackathons [60], and civic hackathons [61]. The difference of the hackathons can be seen from the titles; educational tends to focus on improving education, industrial hackathons are further developing solutions for a company or

industry, and arts and cultures are improved and reformed through culture hackathons. Civic hackathons are not as intuitive from the name, but they are focusing on public resources and open data to improve the everyday life of citizens.

Hackathons can be grouped based on their preferred field, but they can also be grouped based on the range of what is being developed and for whom, namely tech-centric and focus-centric [53]. The main difference between these groupings is that while tech-centric aims to develop solutions for technical issues, focus-centric events aim to solve larger issues, such as social issues or business problems. These groups are also further divided into separate types, presented in Table 8.

Table 8. Tech- and focus-centric hackathons [53]

Group	Subgroup	Description
Tech-centric	Single-Application	Aims to improve one single application, such as an operating system.
	Application-Type	Focuses on single application type or genre, for example, mobile apps or web development.
	Technology-Specific	Creating solutions with one technology, that can be a programming language or environment, or source of data/API.
Focus-centric	Socially-Oriented	Addressing or contributing solutions for social issues or crisis management, many have also focused on open government.
	Demographic-Specific	Inviting participation from designated demographics (e.g. women or students) in order to offer encouragement and support, but also to create solutions for their specific demographic.
	Company-Internal	Allowing participation from within the company in order to bring the engineers together for innovations.

Hackathons are one tool to engage open innovation for governments and companies, but governments sponsor and host even larger and longer events in the form of open data competitions, which are a part of the economic instruments in their open data strategies [15]. These competitions tend to last longer than hackathons, from few months [62] to at least six months [63], that allows more time for innovation and development of the solutions, giving an equal footing for participants. Unfortunately, these competitions are not structured and they can be hosted for any period of time for any number of participants and the literature on their format is limited [64], [65]. Even though these events are not defined in an exact and uniform way, they are necessary for the stimulation of economic open data usage [66].

3 Research goal and methods

In this chapter, the research problem is outlined and the reasoning behind the research questions are explained. In addition, the research perspective, as well as the selection of research and data collection methods, are presented. The goal of this chapter is to present and explain the research process, tools, and methods for this thesis.

3.1 Research problem

This thesis was initialized more from the industrial and communal point of view to answer issues in the field instead of the theoretical need for study. The initial setting for this thesis was developed based on seminars and meetings, where open data was seen as a possibility for business from the community and some companies were initially interested from the proposition. Usually, it was seen, that the open data community and competitive organizations did not see eye-to-eye; one demanded openness for the sake of transparency, while the other required profits from these endeavours. These talks and discussions between parties seemed to boil down into an argument of why against why not. Additionally, research in the field outlines economical barriers of open data, such as the profitability and accessibility of open data [17], but other scientific sources show evidence that opening data is emerging even in business [29].

The goal of this research is to understand, whether or not demanding open data from private organizations is justifiable and if opening data could become a sustainable model for profit in the future. The issue is approached from two separate but linked viewpoints: to understand the issues in the private supply-side, while simultaneously understanding the demand for open data. The simplification of this can be seen as an organization, which uses open data from other sources – public or private – and contributes back into the community by publishing enriched sets or other sets of data, illustrated in Figure 5. A similar process is already in use within the open source software development and has gained a solid foothold in the industry [12].

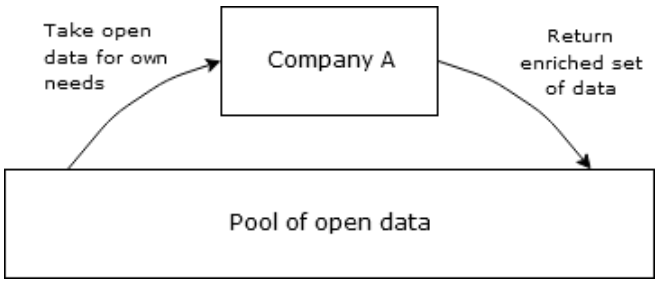

Figure 5. Illustration of open data demand and supply.

Following this setting, the main research question in this thesis is "*How do private organizations supply or demand open data?*" and it is further divided into two sub-questions and addressed in publications I-IV.

Sub-Question 1: *How do companies perceive the profitability of opening data?*

The open data initiative bases itself on the belief, that data should be usable and modifiable for anyone. This ideology does not resonate well with traditional business atmosphere, where nothing should be given away, especially not for free and for anyone for anything. To answer this question, it was necessary to first determine, whether or not opening data is profitable (Publication I), followed with studies about if open data is currently supplied by private organizations (Publication III) and how does the profitability of opening data compare to similar and more popular methods, such as hackathons (Publication IV).

Sub-Question 2: *What kind of an asset is open data for private organizations?*

Because of the ideology behind companies, they tend to take actions that are profitable to them. In order to open data, a company should be certain that the process of opening data they are investing in is profitable to them and the data is not unused and forgotten. As a major factor for supplying open data, it is necessary to evaluate the demand for open data; whether or not there is interest towards it (Publication II) and if open data is currently used for business (Publication III).

3.2 Research position

In order to properly position this thesis, it is important to define where the contribution and scope of this thesis are directed as well as the presumptions and the baseline for the research.

While open data has some popularity with start-ups [31], this research focuses on existing companies, that would benefit from changing their current practices towards supplying and demanding open data. This point of view was adopted since open data activists do not necessarily consider existing companies when determining the economic value of open data [26]. However, it is important to notice, that start-ups do not generate the same amount of value right away, that a large company could do in a short time period. This research focuses on finding whether or not companies could, or should, change their processes and practices with open data

The perceived value of open data was directed towards economic value. While the concept of sustainable value is paramount to open data, from this scope of research the social and environmental aspects do not offer a similar scale of value for companies. Technically the advancements in efficiency and organizations culture could be interpreted as social value, but from the organization's point of view, even these aspects are measured with economic value. While this point of view is adopted, open data is

seen as something that produces direct monetary value for an organization, and the process of opening data is justified in the form of profit.

As for the open data perceptions defined in the literature review, the main viewpoint towards open data in this thesis is economical, determining the profitability of open data for private organizations from supply and demand sides. The secondary views are directed towards institutional and operational perceptions. Institutional was a by-product of this thesis since during the data collection it was noticed, how open data movement is not currently recognized in the industries as a viable solution and some results are related to the changes within organizations. Similarly, operational perspective was taken into account, when in early phases the usability of open data was noticed to be an issue, requiring changes in standardization and tools in use.

3.3 Research methods

When designing research, it is important that the researcher is capable of focusing on the issues and realities in the field. But just as important are (1) the theoretical stance of the researcher, (2) the access to resources, and (3) the selection of suitable research methods that can be used for solving the issue [67]. To aid new researchers to select suitable methods for their research, there are different categorizations of research methods, such as Järvinen [68], Hevner et al. [69], and March and Smith [70]. The taxonomy by Järvinen is further presented here, because it offers a broader view into research in general, while the other guidelines are utilised more in design science research.

Järvinen [68] divides the research methods on the most fundamental level between mathematical research and research concerning reality. The studies about reality are further divided into two approaches: research stressing what is reality and stressing the utility of artifacts. The classification is further described in Figure 6.

Considering the classification in Figure 6, this research falls under theory-testing approaches. In this thesis, open data is treated as a model from literature and practice, which is then evaluated with empirical methods, such as surveys and case studies, in order to assess the usability of opening data and open data in general for business cases. While this research does not extend as far as developing a new model, it offers the initial steps by recognizing issues in the current model, allowing its improvement in practice.

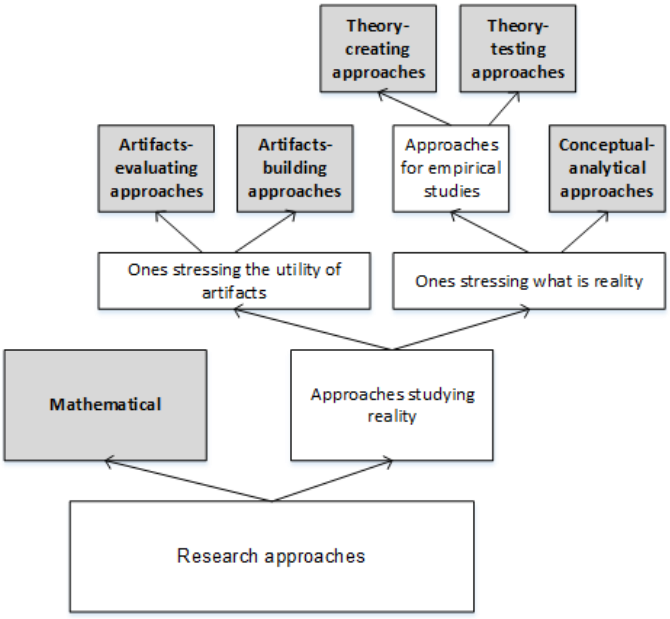

Figure 6. Approaches towards research, by Järvinen [68].

3.3.1 Research perspectives

When empirical research is conducted, a common division between used approaches are quantitative and qualitative research [71] as well as the combination of both, called mixed methods [72]. The quantitative research approach is originally developed for understanding natural phenomena, which include the use of formal methods, mathematical modelling and survey methods [73]. On the other hand, qualitative methods are designed to study cultural and social phenomena, applying action research, case studies, and ethnographies [73], [74]. While following quantitative or qualitative approaches can produce rigorous results, they both have weaknesses. Through mixing the two approaches, allows the researcher to compensate the weaknesses of each approach, since the data collected with both approaches do not compound the liabilities, but aims to neutralize them [75].

While the nature of this thesis is mainly qualitative, it still has some aspects of quantitative approach as well. This sort of mixing methods is suggested, for example, by Jick [75], who recommends surveys in addition to qualitative analysis in order to remove potential researcher and data biases, increasing the generalizability of the results.

As for the epistemology of this research, there are three epistemologies for qualitative research: positivist, interpretive, and critical [73], [76]. Additionally, Easterbrook et. al [67] suggests pragmatism as a fourth philosophical perspective.

In positivist research, the researcher conducting the study makes the assumption, that the phenomenon in question can be studied in isolation from its context [67]. Positivists tend to lean towards theory testing and development, in order to better understand and predict a phenomenon [76]. This philosophical stance expects, that the existing theories are objective and the results are independent of the researcher [73], which allows the researcher to ground their work to these theories. Controlled experiments, case studies, and surveys are usually associated to positivism [67].

If positivist stance expects everything to be objective and factual, interpretive epistemology is the opposite. In interpretive research, objective theories do not exist and generalization is not the goal [73]. Interpretive researchers study local phenomena, for example, an organization or a team, and constructs a local theory [67] in order to explain one setting that can be used to inform other, similar, settings [76]. This stance prefers abundant and rich qualitative data and is associated with ethnographies, exploratory case studies, and surveys [67].

Critical research concentrates critique towards the current practices in societies, which cause alienation and restrictions to its inhabitants [76]. Critical theorists tend to gravitate their research towards minorities and choose their topics of research based on whom it helps [67]. This epistemology treats research as a political tool with a goal to remove societal constructs that are causing alienation and restrictions [73]. These researchers often prefer case studies, but action research also reflects this philosophy [67].

Pragmatic research – as the name suggests – does not try to produce objective results but it is more on an engineering approach, where the researchers apply whatever methods suitable in order to gain practical, instead of abstract, knowledge. In pragmatism, knowledge is judged based on its usefulness for practical problems and it is acknowledged, that all knowledge is only an approximation, incomplete, and completely subjective. In pragmatic research, the preferred method is *whatever works*, but the researchers tend to prefer mixed methods research. [67]

Selections for this study

In this thesis, the applied methods are mixed methods with interpretive epistemology. Opening data is treated as an organizational or managerial issue that comes from the reluctance of companies to open their data, which allows – or even demands – the adoption of qualitative research methods [73]. The quantitative methods, namely survey, are adopted in order to strengthen the qualitative analysis [77]. This route was taken in order to gain a wider view of the current empirical evidence in the field and the interviews were conducted in order to provide context and reasons behind the survey responses.

The interpretive epistemology was chosen, because of the exploratory characteristics of this thesis, but also because this thesis aims to explain the reluctance towards open data

from the organizational and managerial contexts. However, some parts of this thesis can be thought as pragmatic, since the goal is to produce results for industrial and communal development, instead of academic endeavours. Although, while this thesis uses mixed methods – preferred in pragmatic research – the methods used are surveys and exploratory case studies, which are also valid in interpretive research. The goal of this research is to provide some local knowledge that is abstracted from multiple cases, which would direct this work towards the interpretive philosophy.

3.3.2 Selection of research methods

The issue with the first sub-question is, that the empirical evidence about open data based business in the current fields of business is limited [11]. For this reason, it was deemed necessary to use the systematic review in order to find empirical evidence from the scientific literature that would offer directions and reasons to open data (Publication I). After the evidence was found, that open data can be profitable, the empirical search was administered to software companies, who build software solutions for other fields of business. Theme-based interviews were administered and parts of grounded theory analysis were used to determine the views towards opening data from these software companies but also from their clients (Publication III). These interviews offered initial evidence for this research question, but it was decided to deepen the understanding by comparing the benefits of opening data to hackathons, which were more popular in the industrial and global setting. The data was again collected with theme-based interviews and analysed with partial grounded theory analysis. In order to minimize the bias and subjectivity one hackathon organizer could cause, each of the hackathons was treated as a case study and finally synthesized through cross-case synthesis (Publication IV).

For the second research question, the empirical evidence was easier to find, especially when the research started with the interest towards open data. There were already other actors in the field and with their help, it was possible to collect survey data from these interested organizations (Publication II). The survey was used since it was the most suitable method to gather a limited set of data from a large number of organizations simultaneously in a short period of time. The survey data was used to illustrate the interest towards open data and data business in general. This data was also used to enhance the findings from the following research into software companies with the aforementioned theme-based interviews (Publication III). The interest towards and the actual use of open data was measured, and issues in the usability of open data were raised from these studies.

The methods that were selected for the different parts of this research are presented in this section.

Systematic review

The goal for this research was based on industrial and governmental needs, but because of the lack of empirical evidence available in the field, it was deemed necessary to execute a systematic review of literature in order to systematically present, what kind of benefits and drawbacks open data can bring to private organizations. In order to execute a systematic and reproducible search into the matter, systematic mapping study (SMS) guidelines were followed [78], presented in Figure 7., while using some elements from systematic literature reviews (SLR) [79].

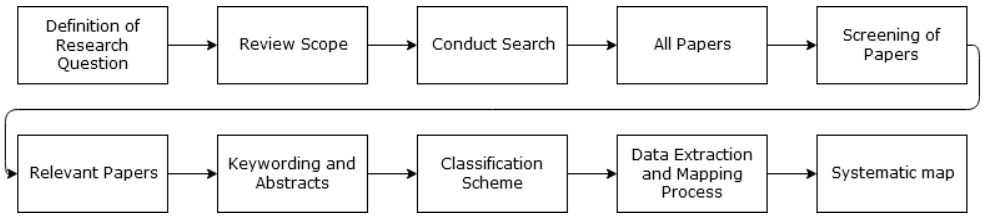

Figure 7. SMS guidelines, adapted from [78]

Following the guidelines, first, it is necessary to define a research scope from the research questions and keywords, to provide a research area and to develop search strings for the next step. After the scope has been set, a database search into the relevant search engines can be performed, which yields all the articles that match the search strings used. Additionally, it is possible to manually search specific journals and conference proceedings or to use snowballing [80] to find more relevant articles based on the references to existing literature. After all of the relevant papers have been gathered, they are screened with inclusion and exclusion criteria, which yields the number of relevant papers to the topic. Finally, the papers are classified and the map is constructed.

The difference between SMS and SLR is somewhat vague, but Petersen et al. [78] suggests, that both of the methods can be used in parallel, as is done here. In this thesis, elements from SMS were used to determine the number of empirical findings in comparison to assumptions, constructing a structured view into the topic area. SLR was used to further deepen the analysis by reading the papers and describing the empirical cases, as is suggested by Kitchenham and Charters [79].

Mixed methods: survey and interviews

Surveys are – as Fink [81] describes them – a method to collect data and analyse attributes of an individual or a society, such as behaviour, values, and knowledge. The survey is further explained as a tool, that can be completed on paper or online, with or without assistance; the data can be collected in a centralized location or the survey can be filled privately from home. Fowler [82] lists three popular uses: measuring public opinion for articles, measuring political opinions and perceptions, and understanding the

consumers through market research. Fink further extends and abstracts that a survey has three major uses: collective planning, effectiveness evaluation, and research [81].

A survey is a tool to analyse especially quantitative data and the goal is to produce objective results that are generalizable and descriptive. It can be used to guide the fieldwork of the area into a more objective and uniform direction [83]. Fink [81] states that surveys can be combined with other sources of information, especially in evaluations and research. Combining other sources of information decreases the risks of surveys, because they often do not explain the underlying meanings of the data, but can be used in order to evaluate the status of a situation and may require other, more suitable methods, to explain the phenomena [83]. In this thesis, the follow-up method was interviews and partial grounded theory analysis.

Grounded theory is a qualitative research method for qualitative data and it is used to iteratively construct a theory based on systematically collected and analysed data [73]. While the data is analysed, the theory evolves after every round of data collection, which is allowed by the interplay between the analysis and data collection [84]. The grounded theory does not use existing theories, but allows the new theory to emerge from the data; the researcher conducting the research does not necessarily know, what the theory might be until the analysis is complete [85]. The theory provided by this method is considered complete when the data collection and analysis does not provide any new insight, a point called theoretical saturation [86]. Although theoretical saturation does not mean that the theory would be complete; in grounded theory, theories are treated as evolving entities that are never complete because the resulting theory is always based on interpretations [84]. The method is commonly divided into Straussian GT and Glaserian GT. The Straussian GT focuses more on active, systematic categorization with multiple coding methods, while Glaserian GT is focusing more on discoverance of the theory instead constructing it from categorization [87]. Because of the goals of qualitative data analysis in this thesis, where the focus was directed more towards categorizations instead of emerging theory, Straussian GT seemed to be a more suitable method to follow.

In Straussian GT, the three different coding procedures are open, axial, and selective coding [87]. Because of the constraints during the research, only the first of the three was utilised and is presented here. Open coding is the first step of analysis in grounded theory, taking the unstructured and unanalysed qualitative data and it is used to interpret the phenomena reflected in the data [85]. The most important aim of open coding is constant categorization; the pieces of data are compared to each other in order to find similarities and differences. From these leads, the categories are formed.

In the process of open coding, the level of abstraction is of importance. If the categorization is done with too much detail, some of the categories would have only a few instances in it. On the other hand, if the level of abstraction is too high, the details are lost during the categorization. A fitting level is to divide categories with too many instances and combine the ones with too much detail, ensuring a balanced outcome.

The focus in this thesis was to utilise Straussian GT partially by applying open coding procedures to the collected data from the interviews. The organizations that were interviewed, were available only for one set of interviews, which made it impossible to utilise the full strength of grounded theory that would require multiple rounds of interviews with different parts of the organization. Because of this, the data was collected from interviews first and later analysed with open coding to draw inferences from the data. Such a formal method was selected, because of its structure, well-formed procedure, and general acceptance allowed the results to be acceptably communicated to the scientific community and it further enhanced the analysis by minimizing subjectivity and bias [85].

In this thesis, a survey was used to analyse the potential interest towards open data in the industry (Publication II). Those results were further analysed when the survey was combined with information from interviews, which were analysed using open coding (Publication III), where the interests gathered from the survey were compared to practice with software developers. Open coding was also applied in Publication IV, where it was used as one of the early steps of data analysis for a multiple case study.

Cross-case synthesis

A case study is a method, which aims to observe a phenomenon within its context when the impact of the context to the phenomenon is not evident [88]. The phenomena may be within individuals, organizations, processes, models, etc. that can be observed and measured. However, because of the intimate nature between the phenomena and the context, the generalizability of the case study is low [83].

Cross-case synthesis is an analysis method for studies with multiple cases [88]. It is used to combine the results of several case studies by administering the same treatment for each case and comparing them. By combining multiple similar cases, more robust findings can be gathered, where the generalizability is higher than with only one or two cases.

The qualitative data in the case studies was analysed by dividing each case into a matrix, based on the research question of the study [89]. Into the matrix, qualitative data from the interviews were collected, the corresponding topic to a corresponding cell. When the analysis for each case is completed, the findings were deconstructed into a final matrix [88] and open coding was applied for each of the cells, yielding the final categories from the interviews and cases.

The cross-case synthesis was applied in Publication IV in this thesis, in order to allow comparison between multiple hackathons. Synthesizing multiple cases allowed the researcher to draw conclusions from all the cases instead of getting results from each case individually, which eliminated some of the subjectivity and bias that analysing only one hackathon would cause.

3.4 Research process

The research process was divided into three phases, shown in Figure 8. However, before, during, and after Phase I, as the initial environmental setting, there were multiple seminars and informal meetings with different companies and networks, where their views towards open data were discussed. This also served as a search for potential cases. Some representatives of companies showed almost hostile behaviour towards open data and even more towards opening their data, which suggested that open data was seen as a radical initiative from the company perspective. These initial findings were used when constructing the structure of this thesis and they are referred as Phase 0.

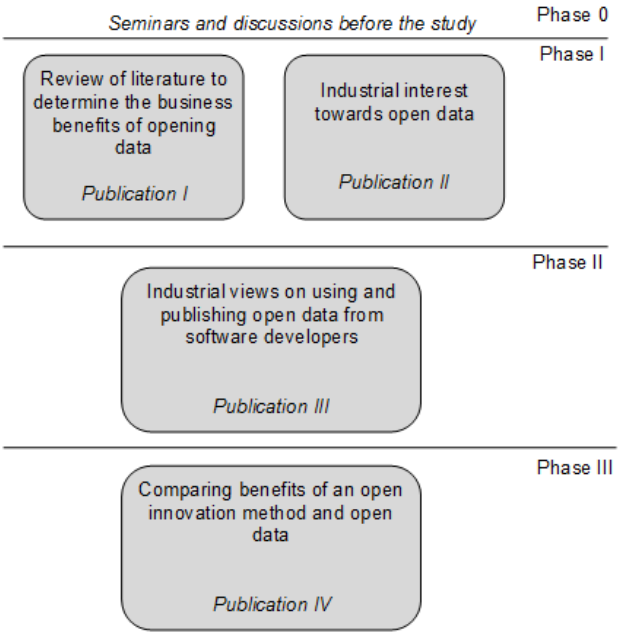

Figure 8. Research process and phases

Phase I was used to determine the environment for the thesis with a systematic mapping study (Publication I) and a survey to companies interested in open data (Publication II). While the review of the literature did not identify a clear research gap, it did show that opening data can be profitable and it could be integrated into the current business environments. For the views towards open data in general, the survey showed that companies are interested in open data and are willing to use it if there would exist a set of data they could use. These publications made a combination study in Phase II possible.

The findings from literature and survey determined the basis that required more empirical research since neither of the publications discusses the de facto state of open data in industries. The literature review discussed the previous research, while the

survey mapped the interest towards open data while it did not discuss the actual uses in practice. To further expand the research into the empirical use of open data and opening data for profit, Phase II took a deeper view towards software companies, which are either using open data for themselves or for their clients and creating solutions for opening data. The study found that open data is not used as a key business resource and data is not opened for profit. There were indicators, that pieces of open data have been used as a part of an application or data analytics, but it has not been used for a standalone business. Additionally, the study found, that companies are increasingly hosting hackathons in order to use crowdsourcing as a source for new innovations and added value from their data. This finding led to Phase III.

The goal of Phase III was to compare the methods in use to the findings from Phase I, especially against Publication I. The hackathons were seen in the companies as a valid and efficient method to engage crowdsourcing and open innovation, strengths that can be found from opening data. Currently, opening data for profit is a more hypothetical manner of approach, while it holds some similarities to hackathons, as was found in Publication III. Phase III concludes this thesis by stating, why control over the data and developers can be seen as more beneficial to an organization than opening data.

3.5 Data collection

There were two sorts of empirical data collected for this research. In the early phase, a survey was used and in the later stages, theme-based interviews were conducted.

The survey and the data collection related to that was done in two phases for Publication II. First set of data was collected, when companies joined into an industry discussion panel, where they had to fill out a registration form, that outlined the basic information about the organization participating and their interest towards data business in general. In spring 2015, a survey covering issues of open data and open data business was sent to this panel. When the survey was sent, the panel had one hundred companies and nine networks of organizations. The networks were left out of the final set of data and only one submission was accepted from each organization in order to avoid larger organizations from over-representing themselves. Out of 100 companies, 45 individual responses were collected. The responding companies and their categorizations are presented in Table 9.

Table 9. Fields and sizes of companies responding to the survey

Field of business	Responses
Software developers	19 (42.2%)
IT-related companies	10 (22.2%)
Non-IT companies	16 (36.6%)

Size	Responses
Micro	28 (62.2%)
Small	8 (17.8%)
Medium	4 (8.9%)
Large	5 (11.1%)

In Table 9, software developers are companies, which have registered their primary field of business as software development. The IT-related companies are organizations, who work directly in IT field, such as hardware providers and IT consultants. The rest are companies, which do not have a direct link to information technology, such as process design, industrial design, and security. The sizes of organizations are identified according to the European Commission's definition [90]: Micro (< 10 employees), Small (< 50 employees), Medium (<250 employees), and Large (250+ employees).

For Publication III, the survey data was used to pinpoint the views of software development companies in comparison to other industries. Additionally, semi-structured theme-based interviews were held in later half of 2016 with software company representatives in order to give depth to the findings and allow more contextual findings to emerge. Even within the software industry, organization sizes, maturities, and operation domains differ. The strategy for the interviews was to gather a heterogeneous set of organizations, which would differ within the aforementioned parameters. The organizations are presented in Table 10.

Table 10. Set of organizations for the interviews in Publication III.

Case	**Size**	**Interviewee**	**Business domain**
Case A	< 10 employees	CEO	Healthcare digitalization
Case B	< 10 employees	CEO	Data provision
Case C	200-500 employees	Project manager	Software production and life-cycle management
Case D	100-200 employees	Open source manager	Digital services
Case E	200-500 employees	Project manager	Digitalization / Digital business

The interviewees were selected based on their position in the organization that allowed as wide point of view as possible into the topic. For smaller organizations, this was the chief executive officer (CEO) of the company, while for larger organizations the interviewee came from middle management. All of the organizations were privately owned, commercial companies, but their clientele alternated from public organizations (cases A, E) to private companies (cases B, C, D).

For Publication IV, the data was also collected via semi-structured, theme-based interviews. Most of the participants were privately owned, large organizations that hosted the hackathon in their facilities and limited the number of participating teams as they saw fit. Two other organizations were publicly owned organizations that organized their hackathon challenges as a part of larger events. In total, there were ten organizations, who participated in this research and each of them were interviewed once, but updates of their status with the teams were asked during the research. Focusing on the hackathons, the organizations are not described further because of the publicity of the events, which would compromise the anonymity of the participants.

3.6 Summary

In this chapter, the research process is described, further outlining the research problem, data collection, and the analysis methods used. The research process in relation to research phases and research questions is summarised in Table 11.

Table 11. Summary of study phases, data, and analysis methods

	Phase I		Phase II	Phase III
Research questions	How do companies ...		… perceive the profitability of opening data?	
		What kind of an asset is open data for private organizations?		
Organizations involved	None	45 surveyed	45 surveyed and 5 interviews	10 organizations, 10 interviews
Instruments and protocols for data collection	Systematic review of literature	Survey	Survey and theme-based interviews	Theme-based interviews
Data analysis	Systematic mapping	Survey analysis	Survey analysis and grounded theory analysis	Cross-case synthesis
Reporting	Publication I	Publication II	Publication III	Publication IV

4 Overview of the publications

In this chapter, the overview of the publications included in this thesis is presented. The publications themselves are included in Appendix 1, containing the results and details in full. In this chapter, the results of the publications are related to the whole, their relations and findings discussed.

4.1 Publication I – Profitability of open data

4.1.1 Research objectives and results

Because of the nature of open data and the lack of opened data from private organizations, Publication I was dedicated to finding what are the main benefits and drawbacks to privately owned organizations when data is opened. The goal was to find, whether opening data brings benefits to the organization in question and if opening data would be a valid strategy to create profit from the data of the organization. The research was done as a combination of systematic mapping study and systematic literature review to scientific literature and the findings were mapped as positive and negative, and further divided as observed and assumed impacts. These findings are presented in Figure 9.

The findings were formed by reading the articles thoroughly, collecting any positive or negative impacts that open data was said to have, which touched a private organization. After all the papers were read through, the impacts were combined, creating each category with descriptions. In Figure 9, the bolded categories outside are the observed

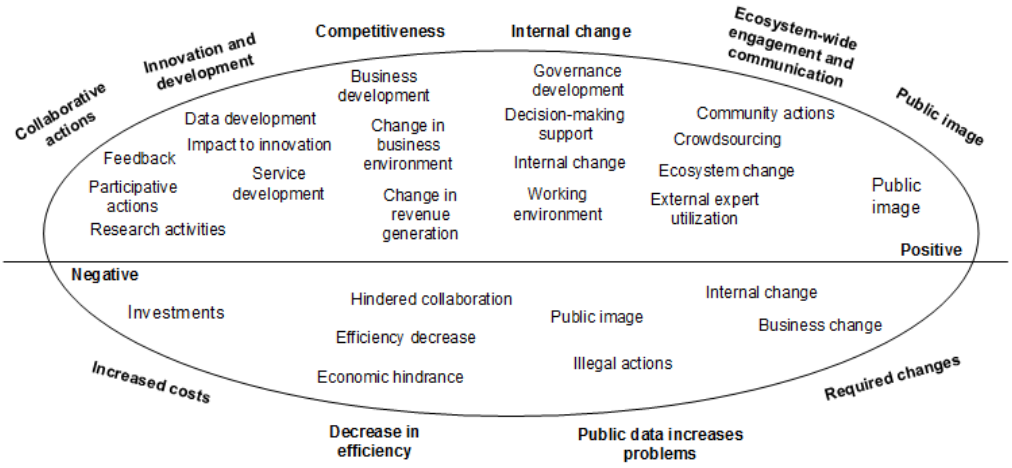

Figure 9. Positive and negative impacts of opening data, observed (out) and assumptions (in)

impacts that came from case studies, while the findings inside are assumptions, what open data is said to have an impact for. The research mainly focused on observed impacts, but the assumptions were also necessary to map out in order to form an understanding of the possibilities of open data. The observations and assumptions are described in this section.

Collaborative actions: Through opened data, the company naturally invites more collaboration from their current partners, but also from competitors and consumers. For companies, this allows efficient information sharing and for consumers, open data enables informed decisions based on real data. In addition, open data is assumed to bring more interoperability through common interest between companies (Participative actions), feedback and other external output regarding the processes, policies, products, services, and the quality of data (Feedback), and more efficient research through acceleration and better focus (Research activities).

Competitiveness: With as sufficient balance between sharing and protecting assets, the competitiveness of the company can be enhanced. Being transparent about the company actions through data also allows smaller companies to increase their publicity and consumer trust, which makes it viable to compete with larger brands in the field. Through open data, it is assumed that the business models and processes can be developed towards collaborative strategies and efficient operations through an information-driven supply chain (Business development). Open data could also be used for social and economic value through increased client commitment and efficiency gains from combined and trusted data (Change in business environment). The revenue generation could change because of open data, by decreasing transaction costs and developing credible and sustainable commerce (Change in revenue generation).

Ecosystem-wide engagement and communication: As was with collaborative actions, the ecosystem can benefit from opened data. New forms of collaboration can be achieved with existing partners, engaging in informal dialog and information sharing, but open data can attract new partners and new forms of partnerships. Through open data, the communities and communal actions could be directed towards the company technologies, allowing easier access and scrutiny of data (Community actions). Crowdsourcing is also another assumed benefit, where the company could be a part of crowd-based benchmarking and allow the recognition of problems and solutions by the public (Crowdsourcing). Transparency within the ecosystem allows accountability and data accessibility to companies and investors, which could make the communication more reliable and faster. Data could be engaged and validated within the ecosystem for constructive influences, discourses, and exchanges (Ecosystem change). Open data could also be used to engage third party experts for collective intelligence, data processing, and validation (External expert utilisation).

Innovation and development: Open data can be used in order to innovate new products by external and internal actors, but it can also be used to incrementally improve and implementing new features to existing products. The open data could also be used to

create new sets for data aggregation through the mixing of public and private data with easier movement of sustainable data, recognising data quality issues that were not noticeable before (Data development). Open data could be used to enhance and increase the collective innovation (Impact to innovation). The services that employ data could also be enhanced by improving their efficiency and performance, as well as the seamless data integration (Service development).

Internal change: By publishing data, the company exposes itself to public scrutiny, which can reflect poorly on their own processes and data quality, increasing motivation to develop systems and processes to better reflect the expectations. Opening data could modernize the governance of the company, optimizing the administration and unties the development from the administration (Governance development). Legitimate decisions could be made based on available data and opinions, opening the process and making the decisions visible, comprehensible and discussable (Decision-making support). Open data could be used to decrease the duplication of data and effort, which allows the cost reductions and productivity boost. By removing duplication, the information asymmetry could be reduced and the support could be directed towards new insights, knowledge, and technologies (Internal change). In the culture of openness that open data could bring, the corruption and internal criticism could be reduced, while empowering the employees to affect their own working conditions and job descriptions (Working environment).

Public image: Instead of third party certification, the company can showcase the sustainability and ethicality of their products directly to the public. These actions can increase the positive image of the company and strengthen their brand. The assumptions are on the same page with the observations in a sense that the brand, reputation, and public profile could be improved, adding trust towards the publisher and conveying competence and integrity (Public image).

Decrease in efficiency: Because opening data requires new processes, it can case a conflict between old and new systems, dividing the available resources and making the management ineffective. The efficiency is also affected by existing conflicts between organizations, which can make the publication of data difficult or impossible. It was also noted that data has to have a clear ownership in order to be able to publish the data, it can become an ineffective process if the data is owned by a third party. The assumptions noted that open data has issues from the economic point of view since the profits may be affected by unclear impact, outcomes, and costs. The company can also lose market share and confidentiality if they publish critical data (Economic hindrance). The assumptions do agree with the observations about the efficiency decrease, adding multiple different possible sources, where it could stem from, such as technical issues, human resources, and the lack of interoperability, to name a few (Efficiency decrease). The collaboration can also be affected, if the companies would not have matching systems or there is a distinct lack of technical expertise, hindering interoperability (Hindered collaboration).

Increased costs: Opening data requires investments, as was mentioned before. There need to be new investments into the new systems to support the process of opening data, as well as the training of staff is needed to efficiently use the systems. Data alone might not always be enough by itself and it could require other materials and documents, such as metadata and records. The assumptions agree especially on the technological investments and it was also found that opening data could bring costs from tying the management as well as the other resources, such as staff, budget, and time (Investments).

Public data increases problems: While the publicity of the data is one strength of open data, it can also be quite problematic. Depending on the nature and the quality of the data, the release can negatively affect the company brand, or even lead to privacy infringements through de-anonymizations. The key point is that open data is not controlled, which can lead to misinterpretations and misunderstandings. The assumptions support these observations since open data could be used to cause potential threats to the company and individuals, leaving the company open for malpractices, such as data fraud, hacking, and data manipulation (Illegal activities). The public image could also be affected by opening data through negative publicity, or legal actions, critical and unwanted questions, and increased confusion about the data (Public image).

Required change: Opening data requires multiple changes inside and outside the company, affecting even the whole ecosystem, through policies, processes, employee mentality, and business models. On the ecosystem level, if other companies are managing or supplying the data for the company, who is opening data, the change is extended to them as well, including the requirements for data quality and risks. Additionally, when opening data the company culture has to change, which can cause issues with the staff, who are not comfortable with openness. The assumptions point out that the open data usage cannot be controlled outside the company and this causes changes in business and operational models as well as the culture of the organization (Business change). The cultural shift could bring a harmful culture with self-censorship, informal procedures, non-recoding, and excessive caution; an environment, which does not support honesty and openness (Internal change).

The findings offered evidence that in scientific literature opening data is seen to provide more positive than negative impacts for the data publisher. Out of the 75 observed impacts from the literature, 50 of them were positive, yielding a positive/negative ratio of 2.0. From the assumptions, the ratio of positive and negative impacts was less, only 1.64, but the number of assumed positive impacts was 243, out of 391 assumed impacts. This offered evidence, that it is possible to benefit – as a privately owned organization – from opening data, while there was a definite lack of practical cases, where direct income was achieved.

4.1.2 Relation to the whole

Publication I offered insight, that opening data does not automatically translate to financial or operational losses to private organizations. Opening data can be beneficial for a company when measuring collaboration and interoperability, or even as an instrument for competitiveness. While open data, by multiple definitions, is still mainly popular by governments and public organizations, this research offers a research gap to start an investigation, why privately owned organizations have not started to open their data, while apparently there are benefits to reap.

4.2 Publication II – Interest towards open data in Finnish scope

4.2.1 Research objectives and results

The goal of Publication II was to clarify the interest towards the usage of open data; whether or not open data is seen as an actual resource for business. In this research, the main question was divided into separate sub-questions in order to determine, if different aspects of an organization affected the perceptions. These aspects were namely the size of the organization as well as their role within open data ecosystem. The data was collected with a designated survey and a registration form within an industry discussion panel. This yielded a sample size of 45 organizations varying from software developers to designers and coffee shops. What was noticeable, was the high number of software developers in the panel, 19 organizations out of 45 reported their key area of business as software development.

The results of Publication II can be summarised into two main aspects. First, the size of the organization does not affect the interest towards data business, but it does change the views towards open data. Smaller organizations are considering open data as a resource for direct software development and applications, while larger organizations prefer to use open data as an additional resource for data analysis. Overall, the applicability of open data is seen as an aggregating set of data instead of the direct value in applications.

The second aspect of Publication II was the necessity of communication. There seems to be a mismatch between data publishers and data users when it comes to informing one another about the needs and requirements of the data. Open data is being published by public organizations, but they rarely take feedback from third parties, who might be interested in this data. The data can also be published through channels, which are not visible to the potential users. Meanwhile, the users who are interested in the use of open data do not communicate their needs to the data publisher, neither before the opening of data nor when the data is accessible. It would seem that the field of open data has an issue that the data publisher does not know, what data is wanted, and the data users do not know what data is being published.

Through these results, it can be argued that the field of open data is still young but starting to mature. The data is being made open for the sake of being open, and there has not been much discussion with third parties about the interest and applicability of data. From the industrial size, there is a definite interest towards open data as a free and usable resource for software and analysis, but communications are proving to be an issue.

4.2.2 Relation to the whole

To understand, whether or not there is potential demand for open data, the goal was to determine the interest towards open data and to get a view to the current, empirical status of open data environment in Finnish scope. The findings determined that the environment is still developing from an early stage, but also illustrated that there is definite interest towards open data for industrial use. While the environment is still in the early stages, there might exist some instances that are already applying open data in their operations.

4.3 Publication III – Practices of open data from the field

4.3.1 Research objectives and results

The survey data in Publication II showed that open data is especially interesting for software developers, and in Publication III, these claims were observed more closely. The research was done with a two-fold-method, first, the environment of the relation of software developers and open data was set using the survey data from Publication II and the second step was done with interviews and qualitative analysis from actual software companies. It is also important to notice, that the steps were executed first and the analysis was done after both sets of data – qualitative and quantitative – were collected, so the interviews were not affected by the survey results in this scope. The goal of this research was to inquire, whether the interest towards open data has materialized enough to be used in any industry on some scale. Additionally the findings from Publication I were a part of the agenda. The research questions were formulated to allow the interviews to focus on using open data but also publishing open data from the industrial setting.

First, the survey results were analysed. Out of 19 software developers, the results show that the added value of open data comes from direct software development (84%) and data combination for analysis (79%). Open data in an application was seen more important in software industry than in others, and vice versa with data aggregation, where other industries saw that combining data might bring more added value. Another comparison was made to dictate the overall interest towards data-based business, not just open data. Amongst the software developers, the highest ranked were data-based applications and services (95%), interface development (74%), and data enhancement

(58%). In addition to data management, these aspects were seen as more important in software industry than in other industries.

The second, qualitative, step of this research delved deeper into software developers and their views towards using and publishing open data as well as the trends of open initiatives. Publication III offered five major findings presented in Table 12.

Based on one of the findings regarding different uses of data and data sharing, in Publication III a strategy classification was described to further emphasize the differences between strategies and actors. This classification is presented in Figure 10.

The metrics in the classification are collected from the interviews, where control over the data was mentioned as one of the key points when dealing with data. This relates on how much control does the organization need to exert over the data in order to share it; personal data is required to be controlled more because of legislations and possible privacy infringements, while open data – by definition – cannot or should not be controlled. Another measurement selected for the classification was publicity of the data. As mentioned, business critical data cannot be public, while competitions and hackathons are being marketed with the access to data and availability to innovate with it.

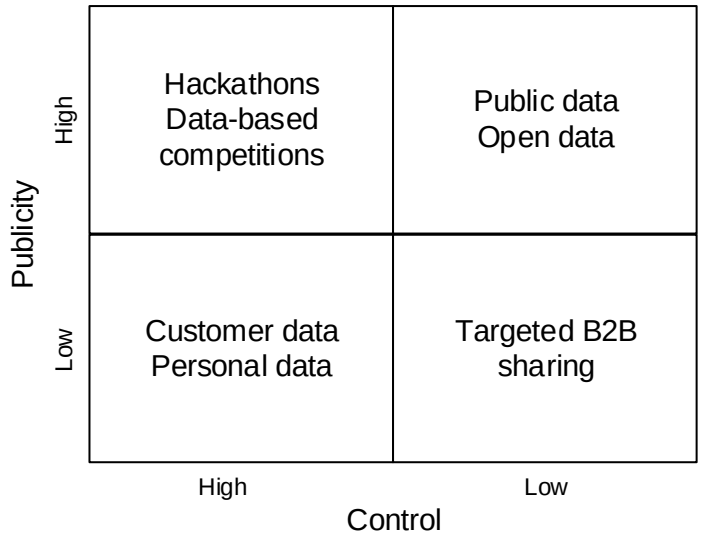

Figure 10. Strategy classification for data sharing from Publication III.

Table 12. Findings and explanations in Publication III

Finding	Justification
Open data is not considered as a key business asset in industries	The interviewees stated, that it is difficult to make a business directly from open data. They did point out that their clients had had open data as a basis for their product, in order to have some data before the publication to increase functionality.
	The software developers pointed out, organizations currently have multiple issues with internal data, which is usually collected by or for them and for their case specifically. These issues multiply with external data and open data, since the structure or level of granularity do not necessarily match.
	Open data may be accessible as a set of data through a website, which is not an optimal medium to deliver data. While data may be declared open, there can exist legal and organizational issues which makes the usage difficult.
Opening data is not done for business initiatives	Some organizations are currently opening data, but this is done solely because of legislative needs or done as publicity stunts. The organizations mentioned come from media, banking sector, and electricity production.
	While opening data is mainly an action for public organizations, even they are having issues with opening data. There are legislative needs to open data, but they are not executed efficiently in practice.
	A major hindrance to open data is the profitability of data. Both private and public organizations can have a revenue stream from the data and opening data would disrupt this income. This leads to the opening of non-critical data, which may not be usable for business.
	Ownership and control over data are raised as one of the most important factors, especially in private organizations. Data is not even shared between partners without strict contracts and predetermined uses.
Open data is considered as a potential added value for software products	Open data is recognized in software development, and sometimes even used to support start-ups and clients, who might need such data. The developers are also making plans for the use external data, if there happens to be a set of data that can be applied in their cases.
	Communication about the data was raised as an issue, since software developers do not have the resources to keep track on what data is being opened and where it can be accessed, which leaves them uninformed about the data.
Open data is not a popular open initiative trend	The most popular open initiative that was mentioned was open source initiative. This initiative has progressed to a point that the clients are already inquiring about the use of open source solutions for their needs. Especially in public organizations the drive towards open source software has strengthened significantly.
	Another popular initiative is the development of open APIs. There appears to be two definitions for this, either open-for-all API or following-open-standards API. The first relates to open data by offering open data efficiently, while the latter makes development and integration easier through a common interface.
Open and closed data are not the only forms of data management	Organizations are sharing data between each other (B2B). In some cases a client shares data to the software developer and in some cases the developer acts as an intermediary, visualizing data from one client to another client. These methods are considered rather rare and the main reason for this is the desire to control data, even if it does not have immediate value to the owner.
	Personalizing a service for a consumer requires personal data to be inputted into the service as well as the actions within the service are logged and used to further develop the personalized service. Smart devices are also used to collect data from everyday life and this data could be used for the development of new or existing services, for example in healthcare. This is currently one of the key operations, where software developers are investing, collecting and using consumer data for personalization and services.
	Hackathons and other competitions were mentioned as a method to share data to developers without losing control over it. The views towards these events and their results were sceptical, while they are currently being organized as public or restricted competitions by multiple actors.

4.3.2 Relation to the whole

The goals of Publication III were two-fold. One goal was to determine the use of open data in the current industrial landscape and the second was to determine the willingness to publish open data from private organizations. This publication was used to empirically gauge the use and publication of open data, in order to empirically validate previous findings. To the whole, the findings determine that the use of open data is limited because of the nature of the data from the business scope, issues in the data, and the delivery of the data. The publication of open data was not seen relevant for private businesses, especially because of the ideology – giving something away for free – does not fit into the current business models in private or public organizations; the ownership and control of the data are seen as a necessity in order to harvest the added value of the data. While the view towards data may be developing towards a more open direction, it is still uncertain, whether open data would be the go-to method for crowdsourcing and open innovation or not. Based on the interviews, it was seen that companies are engaging open innovation through hackathons instead of opening their resources.

4.4 Publication IV – Engaging open innovation based on data

4.4.1 Research objectives and results

Since there were practically no empirical cases in Phase I to determine, what benefits and drawbacks does a company have when they are opening their data, Publication IV compares the effects from the literature review to similar, empirical cases. Hackathons were selected to serve as these cases since in Publication III they were recognized as a somewhat popular method to crowdsource and share data in a controlled environment. Since the benefits and drawbacks of opening data had been determined from the literature in Publication I, this research focuses on the benefits and challenges of hackathons in order to provide similar and comparable results.

For the research, ten organizations who had hosted a hackathon within the past few years were interviewed. Most of these cases were industrial hackathons, events purposefully designed in order to apply open innovation for the needs of the hosting company. Two of the cases were public hackathons, where organizations submitted challenges and data for larger events, where different types of hackathons have been combined. From each organization, the interviewee came from the organizing team, who had been part of the team from beginning to end. The interviews consisted of three major parts: (1) the motivations behind hosting the event, (2) the benefits and challenges they posed and lastly (3) the future sights how the events could be improved. Each case was analysed by using open coding and after the individual cases had been finalised, the cases were combined into one case through cross-case synthesis, where the results were drawn.

In Publication III, the hackathons and other competitions were determined to be on the same level of publicity with opening data and the key difference between the strategies was the control of data. Controlling and owning data was seen as the main factor behind not opening data in Publication III and because of this, the following research is used to determine, if there are additional benefits from hackathons when compared to opening data.

The data collected through interviews was analysed qualitatively. First, each of the organized hackathons was treated as an individual case and the interviews were codified into a matrix based on the interview structure. All mentions of motivations, benefits, and challenges were recorded into the matrix and then codified into a shorter description, making sure that the context and meaning were not lost. After all of the cases were recorded and codified in this manner, they were combined into one case through cross-case synthesis.

It was found that the hackathons offered multiple benefits. The main benefits of hackathons were found to be the products, new projects, and research. This category was mentioned the most, in ten cases out of ten organizations. Other major benefits were the organizational development (9/10) and ecosystem development (8/10). The hackathons were seen as a method, which allowed the organization to gain new products but also allowed the organization to develop internally, giving them new tools, culture, and views towards agile development in their innovation process. Similarly, the ecosystem of the organization was allowed to develop through new contact from the event and the visibility of the event; new partnerships were initialized and old ones were expanded. All of the benefits are presented in Table 13.

The challenges in hackathons, unlike the benefits, concentrated on the event itself, but there were also challenges for the organization. The category that was mentioned the most was the issues in processes and organization; it was mentioned eight out of ten times. Many of the organizations noticed that in order to take some prototypes forward from the hackathon, the current innovation processes and project offices were too rigid and monolithic to allow agile development and fast prototyping. There were also issues in the culture of the organization: the hackathon was not seen to have value for the organization and it was difficult to engage individuals in the event. Even if the employees were excited about the event, it was found that it is still difficult to engage them into the hackathon because their day-to-day responsibilities do not allow variations: the employees did not have the time and the organization did not necessary have the funds ready to take prototypes forward. All of the challenges are shown in Table 13.

Table 13. Benefits and challenges of hackathons

Benefits	Sum	Challenges	Sum
New ventures: products, projects, and research	10	Process and organizational issues	8
Internal / organizational development	9	Resource and cost management	6
Ecosystem development	8	Difficulty of communication	5
Feedback and communication	6	Technological issues	5
Motivated system development	5	Challenges in preparation	5
Accelerated R&D	5	Team commitment	4
Visibility	4	Schedule	3
Archive of ideas and concepts	3	Ownership	3
Recruitment	2	Lack of follow-up	2

After these findings, the benefits and challenges were compared to the observed benefits and drawbacks of open data from Publication I. The comparison was done by comparing both sets of descriptions and categories to each other through identical or similar categories and keywords. Through these comparisons, it was possible to draw conclusions as to why hackathons are seen as the more beneficial method to engage external professionals for a company.

The major findings of this research were that hackathons offer:

- More control over the data and developers for the organization
- Practical business cases and culture improvements inside the organization
- Meaningful communication and exchange of ideas with external professionals
- Required changes were limited in day-to-day processes and non-existent in ecosystem level

In comparison, opening data has the ability to attract more visibility to the data and the organization from the public. While publicity is necessary for an organization, it can be argued that opening data may not be the most efficient method to market an organization to the public.

When an organization invests into external development, control over the data and the process are seen as of paramount importance, in order to ensure the sufficient return of investment. With hackathons, it is possible to control everything from beginning to end, while allowing external development. This control ensures that the organization can gain something, if they want. With open data, this control is lost, since the data is given outside of the company. The business cases do not necessarily come back to the data publisher, but with hackathons, the organizer has the initial option to buy, or invest into, the prototype.

The changes within the organization are also seen differently. When adopting to opening data, it is important that the whole organization changes as well, so the change to processes, management, employees, and company culture has to be changed so that opening data would be beneficial. With hackathons, it was seen that the change comes from the inside; agile development is introduced to the organization by external professionals, changing the culture and the way in which people think, enabling change from the individual level. Change becomes something that does not have to be managed since it is voluntary. In addition, while opening data also might require changes in the company ecosystem and partners, with hackathons the changes are bound within the organization.

Lastly, data is not an efficient medium of communication. Through hackathon events, it is possible to meet new people and talk with the developers who are building the prototypes. Through direct communication, it is easier to engage new partnerships or recruit individuals and even direct the whole innovation process and prototype development. This way, any misconceptions and misunderstanding can be cleared and the data is not used in a way that it should not be used. With open data, external individuals have all the control over the data and can do anything with it, which can expose the organization to malicious attacks.

4.4.2 **Relation to the whole**

In order to understand, why data is not opened by companies in reality, it is necessary to understand how companies currently engage open innovation with their data and what are they getting out of it. The goal of Publication IV was to demonstrate, what benefits hackathons bring in comparison to opening data and it was also mapped, what negative aspects that might have for the company. Overall, opening data was deemed to attract more publicity and visibility, but hackathons offer practical benefits, that can be measured and their profitability determined. From this publication, it would seem, that open data does not suit private organizations in its current form, but requires new and more suitable strategies to be implemented into a business process. Then again, when compared to the hackathons, opening data did not offer anything more than hackathons in a sense of direct profits, which can raise the question if open data should be a viable concept for an organization that concentrates on profit.

5 Discussion

In this chapter, the results of this thesis are summarised and the contributions to practice and academia are presented. After the contributions, the limitations of this study are discussed and the value of this thesis evaluated. At the beginning of this research, the research question was set: "*How do private organizations supply or demand open data?*". This chapter aims to offer answers to that question.

5.1 Theoretical contributions and implications for further research

Open data and its benefits have been a popular area of research for a relatively short time: the number of publications featuring open data in title, abstract, or keywords started to increase sharply in 2009 [17]. It is also mentioned, that the aspects of the research have been directed differently and open data can be viewed from multiple different perspectives, which changes the needs and measurable elements of the initiative. In the open data communities the economic perspective is sometimes regarded the least popular or least necessary point of view [26] and even then the focus is directed towards start-ups and new innovations instead of existing companies. Even academic research suggests that open data should not only bring economic value, but it requires a sustainable value from social and environmental points of view [34].

In their review of literature about open data, Zuiderwijk et al. [17] list multiple issues in open data from the economical perspective. The major barriers are (1) difficulty to profit from open data, (2) costs of acquiring open data, and (3) the changes in revenue models when previously sold data has to be opened. This research contributes to the first and third barrier, since the problem is not to profit from the data, but to *control the opened data* as was seen in Publications III and IV. Control does not mean only to control the access to data, but also controlling the marketing of the data and other methods to stimulate the use of the opened data.

Through control, it is possible to profit from the data and it does not necessarily require any changes in the current business models, if a sufficient intermediary business exists to run the process. Open business data should be considered as a specific venue for research, instead of lumping it together with all open data research. When considering the economical perspective of open data and especially open business data, it should not be thought to be the same as open government data. This is because business data is not usually opened because of legislation and transparency, excluding government-owned businesses, but because of potential profits, which should makes the fundamental requirements different. Although these statements raise a question if a company can produce open data, if it requires control over different aspects of the process transforming data to a product. Answering this dilemma has to be left for future research, but with sufficient intermediary businesses and smart communications it should be possible.

In their article, Zuiderwijk et al. [17] also outline issues that have to be researched about open data. Two of these directions are the stimulation of innovation through open data and stimulating the use of open data. This research contributes to both of these directions.

In this thesis, Publication IV compares hackathons and open data as a tool for innovation, trying to show the differences of now-popular hackathons and open data in business. It became clear, that open data is not as sufficient tool as hackathons in the process of innovating and creating products from raw data. In hackathons, the participants and prototypes are controlled and the whole process is directed by the hosting organization, which makes the outcomes more predictable and the profits can be estimated more closely. Because of this, open data is not necessarily a viable solution when stimulating innovation *in business*.

The use of open data is said to be stimulated through hackathons and other competitions [66]. However, in this research it became evident, that the key factor in stimulating open data usage is sufficient communication. Many companies are not aware of what data is available as open data, because they do not have resources to find it. Competitions are a tool to increase the discoverability of open data, but this targets technologically oriented companies, such as software developers. In this thesis, it was found that while software developers are a majority of the companies, who are interested in open data, they are not the only ones (Publication II). Many, more traditional, companies are also interested in open data and they are not necessarily in the segment, who can participate to hackathons or are even interested in them. Disregarding the issues of discoverability, the opened data may also have issues that make it difficult to use after discovering it (Publication III). These issues require future research about how open data can be sufficiently communicated to everyone with sufficient quality, who might have an interest in it, without the need to search the data by themselves.

To summarise, these findings would suggest that supplying and demanding open data by companies is not a viable strategy currently. Companies want control over their data and also a voice within the innovation process, in order to direct the process into a beneficial direction, since the open data supplier does not always benefit the most [12]. There also exists an issue with the need for data and the fact that open data, while a suitable set of data, does not fit the necessary context. A larger scope of future research into this matter should be directed towards the partnerships, how public organizations and private companies could use their different perspectives and bring benefits to both sides. A new process model that would solve an issue of communication about the data and also an easy method of data delivery, being open data, but still answering specific needs. Further research should be conducted in order to determine the possible venues how open data ideology and movement can be integrated into current business fields.

5.2 Contribution to practice

While the nature of this research is exploratory, the results can be used to produce some inferences towards the local and global open data community. This thesis illuminates several reasons, why companies are not opening data similarly to governments and sharing their data gratuitously. While the major focus is towards opening data, this thesis also answers the question of why open data is not used for profit, while it is free and accessible to anyone.

5.2.1 Supplying open data

In Publication I, it was determined from the scientific literature, that opening company data can become a profitable venture. However, it was also highlighted that if the opening process is done carelessly, none of the benefits can materialize. The results showed, that the empirical evidence was somewhat scarce, but still leaned towards the positive impacts for business. It is possible to argue, that this shows clear indicators that open data initiative does not conform to the current business environment and implementing it would be too radical for existing businesses.

The issues were highlighted in Publication III, where software development companies were in question. The main issue of opening data is that it lacks profit and clear business models that could be implemented without radical changes. The companies also mentioned how important factors the control over the data and the ownership are, since data is currently kept within organizations, even if they do not need it at the moment. The owners can become quite jealous about their data and its uses. Opening data is seen as a marketing stunt in industries, it is used to produce a new form of advertisements for interested citizens, but also as a tool to show developers, what data does the company have.

The visibility of data is also seen as the benefit of open data in Publication IV, while the other uses, such as open innovation and organizational change, are not engaged efficiently. In the publication, it was evident, that controlled environment and face-to-face discussions are more important than access to data. Companies are not willing to open their data since they cannot control the innovation process and do not necessarily benefit from supplying open data, which is against the goals of a company. Industrial hackathons and other events were more popular and currently in use since with sufficient planning, the organization can benefit from it.

These results would indicate that while opening data is not a beneficial proposition for businesses in practice – at least currently – the initiative is not a lost cause. The open data community should direct their views towards different methods to introduce new methods to publish data, and ensure that the data supplier can benefit from opening this data. The intermediary businesses should also be supported more, in order to provide services that allow companies to publish their data without radical changes or massive costs, minimizing the hurdle of first steps to be taken, offering low-cost solutions to

open data for interested parties. The time of opening data for the sake of opening data is over, and a more controlled supply chain is necessary to make opening data a beneficial business action.

5.2.2 Issues in open data usage

While there is a definite interest towards open data in a wide variety of industries (Publication II), the usage of open data is low, because of the usability, accessibility, and discoverability of open data (Publication III). Companies are currently struggling with their own internal data and using external data seems to be somewhere in the distant future (Publication III). In Publication II, it was found that the preferable actions for data usage are application development and data analytics. However, when creating an application or analysis, it is important to notice, that there should exist a problem or a question that can be solved with data. Using data, which has not been collected for the exact purpose, may lack a few or multiple elements, such as context and granularity (Publication III). These issues clearly show, that with open data, the communications about specifics are an issue and something to be improved (Publication II). Data, open or otherwise, should never be the goal of an action since data should be collected for something, not for the sake of having data.

The usability of open data is a difficult issue and requires close communications between suppliers and users, as was noticed in Publication IV about data and the innovation related to it. Data in itself is not a sufficient medium for communication (Publication II), and the practitioners should be encouraged to improve the visibility and discoverability of open data through current contacts and ecosystems. Additionally, there should be new ways to publish data, which would make it easy to find for interested parties, who do not have the resources to actively search for the data.

5.2.3 Summary

This study would suggest that if open data is seen as a possible resource in the future, the publication process has to be changed in a way that would somehow compliment businesses from participating in the community. The issue with communications is difficult to address since data itself does not serve as a sufficient medium for communication and gathering all interested parties within the same communication channels is cumbersome and costly, especially because one set of open data may not be usable for any of the parties.

As a result of this thesis, it was found that opening data for the sake of opening data is not a sustainable strategy when the data publisher is not aware of the contexts where the data is needed. Neither it is automatically usable for data consumers since the structure of the data or the level of granularity does not necessarily match the required uses, or the data does not always contain all of the necessary elements.

The results of this thesis would suggest, that the open data community should not ask companies to open data without a sufficient tool for easy publishing. Companies do not want to invest in opening their data since the payoff is not guaranteed and may even be non-existent. Neither do they want to use open data, which they cannot find without resources directed towards it and which does not necessarily fit their needs. These issues would invite new intermediary businesses and business models, which would actually make data publishing cheap and profitable, and structures the data to a form that is wanted and needed in other contexts. However, it would seem that open source business models cannot necessarily be utilised, possibly because of the infancy of open data movement or because of the tight relationship between data and data publisher compared to open source development. No matter the reason, the field would require some form of an intermediary business model that combines opening data and other tools, such as hackathon events or similar innovation methods, or even new technologies, such as blockchain [91], on a regular basis.

5.3 Limitation and validity of this thesis

When conducting research, it is impossible to execute research plan in an empirical environment without some limitations and threats to validity [92]. The task of the researcher is to recognize the threats beforehand when designing the research, as was done in this thesis. Interpretive and especially qualitative research is vulnerable to validity threats and pitfalls, including lack of replicability, controllability, generalizability, and deductibility, the risk of misinterpretation, the difficulty of manipulating individual variables, and the inability to randomize [67], [75], [83]. Because of a large number of the threats to validity, guidelines have been developed for inexperienced researchers to avoid the basic pitfalls (e.g. [92], [93]).

Maxwell [93] defines five key threats to research validity regarding qualitative research, that should be addressed in qualitative works. While the validity threats are addressed from qualitative literature, a part of this research was one in a quantitative setting. However, only a small part of this thesis draws the conclusions from the quantitative findings and the issues of validity are addressed in the research article (Publication II), and so, this thesis will be validated mostly from a qualitative perspective. These threats to validity are descriptive validity, interpretive validity, theoretical validity, generalizability, and evaluative validity. Each of these threats is addressed in this section.

Descriptive validity: Maxwell [93] describes descriptive validity as the first concern of a qualitative researcher; the researcher should have precise factual accuracy from their account, not distorting or inventing their perceptions. The author continues that while inventing results might not be deliberate, and there are issues such as misunderstandings, mishearing, mis-transcriptions, and omissions of tone and pitch.

The descriptive validity was addressed in this study by using multiple researchers to construct the interviews and having at least two researchers conduct the interviews for

Publication III and IV. The interviews were also recorded in order to preserve as much information and minimizing the issues of misunderstanding and omission during the transcription process. By using these steps, the results were reported to be as factual as possible, minimizing the influence of one researcher.

Interpretive validity: In qualitative research, it is important that the researcher does not impose his or her own concept and vocabulary into the interview. The interview should yield data that matches the views of the interviewee and not the views of the researcher, a key issue in interpretive research [93]. Closely related to interpretive validity is what Kitchenham et al. [92] describe as 'cherry picking', over-analysing the data in order to reach the desirable results. This form of validity threat emerges in both quantitative and qualitative research, where data is being analysed.

In this thesis, this threat to validity was responded by using multiple researchers to validate the analysis and results. The quantitative data in Publication II was analysed online in order to allow all three researchers to perform and validate the analysis simultaneously. On the other hand, the qualitative data in Publications III and IV were first transcribed from recordings and then analysed by multiple researchers who worked separately and agreed on the results after individual analysis.

Theoretical validity: According to Maxwell [93], theoretical validity goes beyond the two previous types of validation. The descriptive and interpretive validity focus on the accuracy of the terms used in the study, while theoretical validity concerns itself with issues that do not disappear with the agreement of "facts" of the situation. The issue lies within the legitimacy of the terms or the application of them; whether or not the terms are properly used in the community or if the application of the terms is appropriate to study the given phenomena.

In this thesis, there were two communities, where the applied research was reviewed: the practical and academic communities. Theoretical validity was addressed constantly when doing the research, through discussions with practitioners, collecting feedback and comments from interviewees about the topic and possible directions, and via reviews of different publication venues. Practitioners offered viewpoints that are important to their community and took some of the results into practice, while simultaneously the academic community saw the relevance of the terms and the fit of their application for research in their community

Generalizability: Maxwell [93] describes generalizability as to how effectively the experiences from one population or situation can be transferred to another, similar population or situation. This issue of validity is highly separated from the issues in quantitative or experimentation, where generalizability is natural, while qualitative research has not been designed for generalizations in the wider population. Because of the rift between qualitative and quantitative generalizability, Lee and Baskerville [94] proposed a framework that can be used to distinguish the separate, and equally important, notions of generalizability.

When considering the framework [94], it consists of four categories of generalizability. There four types are: (1) from Data to Description, (2) from Description to Theory, (3) from Theory to Description, and (4) from Concepts to Theory. Because of the explorative nature of this thesis, and its initiation from practice, it can be argued that this thesis generalized from data to description. The data, which was collected during the research, is used to understand the nature of open data business and the viewpoints from the industries towards it, offering observations and descriptions instead of contributions to a theory.

The results of this thesis are also difficult to widely generalise. The focus is towards open data business and the results are relevant in that context, but the extent of the relevance is discussable. The data was collected from Finnish setting, which can be considered to be a somewhat more suitable environment for open movements, both in business, for example, Lean Service Creation (LSC) [95], and in public acquisitions [96]. Because of this, the results cannot be generalised to other countries for certain, but common elements can be drawn into practice, especially if these results are already hinted at in other research concerning other environments.

While there were multiple organizations in this thesis, representing different fields of industry, it is noteworthy that traditional industries do not have the capabilities to use open data. Because of this, the results concerning open data usage in this thesis mainly comes from the software industry and this might bias the results, which affects the generalizability.

When discussing generalizability, it is important to note that this thesis was not conceived with the goal of generalizable results. Instead, the purpose was to explore the issues in the current Finnish open data environment and helping to solve how open data can be made more usable and if opening data could be a viable business strategy.

Evaluative validity: The fifth and final item in Maxwell's [93] list is evaluative validity, which is not considered to be in the centre of qualitative research similarly as the descriptive, interpretive, and theoretical validities. The researchers do not tend to make a claim to evaluate their choices in studies.

In this thesis, the evaluative validity was addressed through scientific peer-review and academic community, as well as through self-analysis and group evaluation within the research team. The findings in this thesis are descriptions from a real-life phenomenon, offering guidelines to practitioners in order to advance their processes and methods in the field, but also for academics to further understand the phenomena and what would be required in order to solve the current issues.

6 Conclusions

Nowadays, data is seen as an important asset in practically any industry. Data, whatever data it happens to be, can be used to improve, to predict, to enhance, and to create value for business and government, to name a few uses. Data is an incredibly versatile asset and a key resource for relatively new and exponentially grown businesses, such as Facebook and Google.

In this thesis, the focus is directed towards open data. Data, which is published for anyone to be used for anything. Because of such a monumental impact that data is said to have on business and the world, open data as a concept should be an interesting movement and companies should be thrilled to use data that does not require investments to collect. Also, by following a common saying, that "one man's trash is another man's treasure", it would be reasonable to expect that companies could try to benefit from this movement by offering their data as open, for anyone to use. Not critical business data, but instead something they collect, but do not need or cannot process.

As the results of this thesis show, such ideology does not match the current business atmosphere. It was apparent that traditional thinking was still prevailing, especially in brick and mortar industries, while the use of data has increased even there. However, younger industries, namely software industry, treated open data as something usable, but even they do not pay special attention to it. Data has been opened by companies and open data has been also used, but these are marginalized events that are not directed towards revenue.

From the viewpoint of a company, opening data is not a viable strategy, because there is a definite lack of practical business models that can be integrated into the models in use. This is not an issue with start-ups who can create their business models in a way, that allows the use of data and open data, but the issue lies with existing companies, who were in the focus in this research. While companies are interested in openness and crowdsourcing, there are issues that prevent them from opening their data: the lack of revenue, difficulties in communication, and the loss of control and ownership. These issues would invite more intermediary businesses that would make opening data profitable and a usable solution for data publisher and user.

Creating more intermediary businesses and a viable publishing process would also invite more use of open data. The issues in opening data reflect on some level the usage of open data, since the issues in communication affect the usability of open data, causing the data to be inefficient or unsuitable for a use, where it would seem to be a perfect fit. Open data through websites and portals, or other methods that require manual labour, are not the most optimal methods of data delivery since, for example, the software companies would require an effective medium, such as an API and assurances that the data will be accessible when needed.

Contrary to the assumption about the value of open data, it is not seen as a viable resource for companies, because it is difficult to find and use. The same goes for supplying open data that does not necessarily allow profits to be made, while some supportive evidence has been provided in the literature. If these issues were to be taken under further assessment, it would be justified to claim, that the first stage of open data should be coming to a close and there is a clear indication to reform the movement, taking into account the business side and not only the governments and their citizens. Opening and sharing data should always benefit the provider, as is happening with hackathons, which would ensure profits for the provider, sufficient resources to further innovations, and transparency for the citizens and consumers. Simply pushing open data available somewhere in the Internet does not allow this. New models would be needed to reform the initiative, while staying true to the key aspects of open data: accessibility, reuse and redistribution, and universal participation.

References

[1] F. J. García-Peñalvo, C. García de Figuerola, and J. A. Merlo, 'Open knowledge: Challenges and facts', *Online Information Review*, vol. 34, no. 4, pp. 520–539, 2010.

[2] M. Janssen, Y. Charalabidis, and A. Zuiderwijk, 'Benefits, Adoption Barriers and Myths of Open Data and Open Government', *Information Systems Management*, vol. 29, no. 4, pp. 258–268, Sep. 2012.

[3] M. Janssen and A. Zuiderwijk, 'Infomediary Business Models for Connecting Open Data Providers and Users', *Social Science Computer Review*, vol. 32, no. 5, pp. 694–711, Oct. 2014.

[4] T. Aitamurto and S. C. Lewis, 'Open APIs and news organizations: A study of open innovation in online journalism', presented at the International Symposium on Online Journalism, Austin, TX, 2011.

[5] G. Eysenbach, 'Citation Advantage of Open Access Articles', *PLoS Biology*, vol. 4, no. 5, p. e157, May 2006.

[6] A. Bonaccorsi, S. Giannangeli, and C. Rossi, 'Entry strategies under competing standards: Hybrid business models in the open source software industry', *Management Science*, vol. 52, no. 7, pp. 1085–1098, 2006.

[7] S. Harnad and T. Brody, 'Comparing the impact of open access (OA) vs. non-OA articles in the same journals', *D-lib Magazine*, vol. 10, no. 6, 2004.

[8] J. C. Molloy, 'The Open Knowledge Foundation: Open Data Means Better Science', *PLoS Biology*, vol. 9, no. 12, p. e1001195, Dec. 2011.

[9] J. Loomis, S. Koontz, H. Miller, and L. Richardson, 'Valuing Geospatial Information: Using the Contingent Valuation Method to Estimate the Economic Benefits of Landsat Satellite Imagery', *Photogrammetric Engineering & Remote Sensing*, vol. 81, no. 8, pp. 647–656, Aug. 2015.

[10] J. Manyika, M. Chui, P. Groves, D. Farrell, S. Van Kuiken, and E. A. Doshi, 'Open data: Unlocking innovation and performance with liquid information', *McKinsey Global Institute*, p. 21, 2013.

[11] A. Immonen, M. Palviainen, and E. Ovaska, 'Towards open data based business: Survey on usage of open data in digital services', *International Journal of Research in Business and Technology*, vol. 4, no. 1, pp. 286–295, 2014.

[12] J. Lindman and L. Nyman, 'The businesses of open data and open source: Some key similarities and differences', *Technology Innovation Management Review*, vol. 4, no. 1, p. 12, 2014.

[13] D. S. Sayogo and T. A. Pardo, 'Understanding Smart Data Disclosure Policy Success: The Case of Green Button', in *Proceedings of the 14th Annual International Conference on Digital Government Research*, New York, NY, USA, 2013, pp. 72–81.

[14] D. S. Sayogo *et al.*, 'Going Beyond Open Data: Challenges and Motivations for Smart Disclosure in Ethical Consumption', *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 9, no. 2, pp. 1–16, 2014.

[15] N. Huijboom and T. Van den Broek, 'Open data: an international comparison of strategies', *European journal of ePractice*, vol. 12, no. 1, pp. 4–16, 2011.
[16] I. Susha, M. Janssen, and S. Verhulst, 'Data collaboratives as "bazaars"?: A review of coordination problems and mechanisms to match demand for data with supply', *Transforming Government: People, Process and Policy*, vol. 11, no. 1, pp. 157–172, Mar. 2017.
[17] A. Zuiderwijk, N. Helbig, J. R. Gil-García, and M. Janssen, 'Special Issue on Innovation through Open Data: Guest Editors' Introduction', *Journal of theoretical and applied electronic commerce research*, vol. 9, no. 2, pp. i–xiii, 2014.
[18] Open Definition, 'The Open Definition - Open Definition - Defining Open in Open Data, Open Content and Open Knowledge'. [Online]. Available: http://opendefinition.org/. [Accessed: 20-Sep-2017].
[19] Open Data Handbook, 'What is Open Data?' [Online]. Available: http://opendatahandbook.org/guide/en/what-is-open-data/. [Accessed: 20-Sep-2017].
[20] A. Halonen, 'Being open about data', 2012. [Online]. Available: http://www.fininst.uk/wp-content/uploads/2017/09/being-open-about-data.pdf. [Accessed: 26-Nov-2017].
[21] Open Data Barometer, 'Open Data Barometer'. [Online]. Available: http://opendatabarometer.org/?_year=2016&indicator=ODB. [Accessed: 20-Sep-2017].
[22] P. Cox and G. Alemanno, 'Directive 2003/98/EC of the european parliament and of the council of 17 november 2003 on the re-use of public sector information', *Official Journal of the European Union*, vol. 46, pp. 1–156, 2003.
[23] B. Obama, 'Memorandum for the heads of executive departments and agencies', *Presidential Studies Quarterly*, vol. 39, no. 3, p. 429, 2009.
[24] A. Zuiderwijk, M. Janssen, S. Choenni, R. Meijer, and R. S. Alibaks, 'Socio-technical Impediments of Open Data.', *Electronic Journal of e-Government*, vol. 10, no. 2, 2012.
[25] R. E. Sieber and P. A. Johnson, 'Civic open data at a crossroads: Dominant models and current challenges', *Government Information Quarterly*, vol. 32, no. 3, pp. 308–315, Jul. 2015.
[26] F. Gonzalez-Zapata and R. Heeks, 'The multiple meanings of open government data: Understanding different stakeholders and their perspectives', *Government Information Quarterly*, vol. 32, no. 4, pp. 441–452, Oct. 2015.
[27] J. Lindman, 'Similarities of Open Data and Open Source: Impacts on Business', *Journal of theoretical and applied electronic commerce research*, vol. 9, no. 3, pp. 46–70, Sep. 2014.
[28] S. Ekins, A. M. Clark, and A. J. Williams, 'Open Drug Discovery Teams: A Chemistry Mobile App for Collaboration', *Molecular Informatics*, vol. 31, no. 8, pp. 585–597, Aug. 2012.
[29] Y. Tammisto and J. Lindman, 'Definition of open data services in software business', *Software Business*, pp. 297–303, 2012.

[30] J. Cranefield, O. Robertson, and G. Oliver, 'Value in the mash: Exploring the benefits, barriers and enablers of open data apps', presented at the 22nd European Conference on Information Systems, Tel Aviv, Israel, 2014.
[31] E. Lakomaa and J. Kallberg, 'Open data as a foundation for innovation: The enabling effect of free public sector information for entrepreneurs', *IEEE Access*, vol. 1, pp. 558–563, 2013.
[32] J. McMurren, S. Verhulst, and A. Young, 'DENMARK'S OPEN ADDRESS DATA SET - Consolidating and Freeing-Up Address Data', Jan. 2016.
[33] A. Tong, H. Irshad, and D. W. Revell, *Market assessment of public sector information*. UK Government-Department for Business Innovation and Skills, 2013.
[34] T. Jetzek, M. Avital, and N. Bjørn-Andersen, 'Generating sustainable value from open data in a sharing society', in *International Working Conference on Transfer and Diffusion of IT*, 2014, pp. 62–82.
[35] A. Fontinelle, 'Economic Value', *Investopedia*, 21-Nov-2013. [Online]. Available: http://www.investopedia.com/terms/e/economic-value.asp. [Accessed: 27-Sep-2017].
[36] R. Brouwer, 'Environmental value transfer: state of the art and future prospects', *Ecological economics*, vol. 32, no. 1, pp. 137–152, 2000.
[37] R. Haenfler, 'Social value of public information', *The American Economic Review*, vol. 92, no. 5, pp. 1521–1534, 2002.
[38] Social Value UK, 'What is Social Value?', *Social Value UK*. [Online]. Available: http://www.socialvalueuk.org/what-is-social-value/. [Accessed: 26-Sep-2017].
[39] A. Zuiderwijk, M. Janssen, K. Poulis, and G. van de Kaa, 'Open data for competitive advantage: insights from open data use by companies', in *Proceedings of the 16th Annual International Conference on Digital Government Research*, 2015, pp. 79–88.
[40] M. Morris, M. Schindehutte, and J. Allen, 'The entrepreneur's business model: toward a unified perspective', *Journal of business research*, vol. 58, no. 6, pp. 726–735, 2005.
[41] E. Ferro and M. Osella, 'Eight business model archetypes for PSI re-use', *Open Data on the Web*, 2013.
[42] F. A. Zeleti, A. Ojo, and E. Curry, 'Exploring the economic value of open government data', *Government Information Quarterly*, vol. 33, no. 3, pp. 535–551, 2016.
[43] J. Musings, 'Open data business models', 2012. [Online]. Available: http://www.jenitennison.com/2012/08/20/open-data-business-models.html. [Accessed: 19-Oct-2017].
[44] K. Suhaka and J. Tauberer, 'Business Models for Reuse of Open Legislative Data', 2012. [Online]. Available: https://legalinformatics.wordpress.com/2012/07/09/suhaka-and-tauberer-on-business-models-for-reuse-of-open-legislative-data/. [Accessed: 19-Oct-2017].
[45] F. A. Zeleti, A. Ojo, and E. Curry, 'Emerging Business Models for the Open Data Industry: Characterization and Analysis', in *Proceedings of the 15th Annual*

International Conference on Digital Government Research, New York, NY, USA, 2014, pp. 215–226.

[46] H. Jaakkola, T. Mäkinen, and A. Eteläaho, 'Open Data: Opportunities and Challenges', in *Proceedings of the 15th International Conference on Computer Systems and Technologies*, New York, NY, USA, 2014, pp. 25–39.

[47] J. Lindman, T. Kinnari, and M. Rossi, 'Industrial open data: case studies of early open data entrepreneurs', in *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, 2014, pp. 739–748.

[48] P. F. Uhlir and P. Schröder, 'Open Data for Global Science', *Data Science Journal*, vol. 6, pp. OD36–OD53, 2007.

[49] C. Tankard, 'What the GDPR means for businesses', *Network Security*, vol. 2016, no. 6, pp. 5–8, 2016.

[50] A. Zuiderwijk, M. Janssen, R. Meijer, S. Choenni, Y. Charalabidis, and K. Jeffery, 'Issues and Guiding Principles for Opening Governmental Judicial Research Data', in *Electronic Government*, Berlin, Heidelberg, 2012, vol. 7443, pp. 90–101.

[51] F. Bannister and R. Connolly, 'The Trouble with Transparency: A Critical Review of Openness in e-Government', *Policy & Internet*, vol. 3, no. 1, pp. 1–30, 2011.

[52] E. Almirall, M. Lee, and A. Majchrzak, 'Open innovation requires integrated competition-community ecosystems: Lessons learned from civic open innovation', *Business Horizons*, vol. 57, no. 3, pp. 391–400, May 2014.

[53] G. Briscoe and C. Mulligan, 'Digital innovation: The hackathon phenomenon', *London: Creativeworks London Work Paper*, vol. 6, 2014.

[54] J. Porras, K. Heikkinen, and J. Ikonen, 'Code camp: a setting for collaborative learning of programming', *Advanced Technology for Learning*, vol. 4, no. 1, pp. 43–52, 2007.

[55] A. Fowler, F. Khosmood, and A. Arya, 'The evolution and significance of the Global Game Jam', in *Proc. of the Foundations of Digital Games Conference*, 2013, vol. 2013.

[56] M. Komssi, D. Pichlis, M. Raatikainen, K. Kindström, and J. Järvinen, 'What are Hackathons for?', *IEEE Software*, vol. 32, no. 5, pp. 60–67, Sep. 2015.

[57] H. W. Chesbrough, *Open innovation: The new imperative for creating and profiting from technology*. Harvard Business Press, 2006.

[58] A. Nandi and M. Mandernach, 'Hackathons As an Informal Learning Platform', in *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, New York, NY, USA, 2016, pp. 346–351.

[59] M. Raatikainen, M. Komssi, V. d Bianco, K. Kindstöm, and J. Järvinen, 'Industrial Experiences of Organizing a Hackathon to Assess a Device-centric Cloud Ecosystem', in *2013 IEEE 37th Annual Computer Software and Applications Conference*, 2013, pp. 790–799.

[60] G. Briscoe and J. X. Hon, 'Choreograthons: hackathons for dance composition', 2014. [Online]. Available: http://webspace.qmul.ac.uk/gbriscoe-disabled/site/Publications_files/Choreograthons-%20Hackathons%20for%20Dance.pdf. [Accessed: 17-Oct-2017].

[61] P. Johnson and P. Robinson, 'Civic hackathons: Innovation, procurement, or civic engagement?', *Review of Policy Research*, vol. 31, no. 4, pp. 349–357, 2014.
[62] Z. Yang and A. Kankanhalli, 'Innovation in government services: The case of open data', in *International Working Conference on Transfer and Diffusion of IT*, 2013, pp. 644–651.
[63] H. Hielkema and P. Hongisto, 'Developing the Helsinki smart city: The role of competitions for open data applications', *Journal of the Knowledge Economy*, vol. 4, no. 2, pp. 190–204, 2013.
[64] A. Hjalmarsson, N. Johansson, and D. Rudmark, 'Mind the gap: exploring stakeholders' value with open data assessment', in *System Sciences (HICSS), 2015 48th Hawaii International Conference on*, 2015, pp. 1314–1323.
[65] A.-S. Hellberg and K. Hedström, 'The story of the sixth myth of open data and open government', *Transforming Government: People, Process and Policy*, vol. 9, no. 1, pp. 35–51, 2015.
[66] A. Zuiderwijk and M. Janssen, 'Open data policies, their implementation and impact: A framework for comparison', *Government Information Quarterly*, vol. 31, no. 1, pp. 17–29, 2014.
[67] S. Easterbrook, J. Singer, M.-A. Storey, and D. Damian, 'Selecting empirical methods for software engineering research', in *Guide to advanced empirical software engineering*, Springer, 2008, pp. 285–311.
[68] P. Järvinen, 'On a variety of research output types', 2004. [Online]. Available: http://urn.fi/urn:isbn:951-44-6197-5.
[69] A. R. Hevner, S. T. March, J. Park, and S. Ram, 'Design science in information systems research', *MIS quarterly*, vol. 28, no. 1, pp. 75–105, 2004.
[70] S. T. March and G. F. Smith, 'Design and natural science research on information technology', *Decision support systems*, vol. 15, no. 4, pp. 251–266, 1995.
[71] C. Wohlin, M. Höst, and K. Henningsson, 'Empirical research methods in Web and software Engineering', in *Web engineering*, Springer, 2006, pp. 409–430.
[72] A. Bryman, 'Integrating quantitative and qualitative research: how is it done?', *Qualitative research*, vol. 6, no. 1, pp. 97–113, 2006.
[73] M. D. Myers, 'Qualitative research in information systems', *Management Information Systems Quarterly*, vol. 21, no. 2, pp. 241–242, 1997.
[74] B. Kaplan and J. Maxwell, 'Qualitative research methods for evaluating computer information systems', *Evaluating the organizational impact of healthcare information systems*, pp. 30–55, 2005.
[75] T. D. Jick, 'Mixing qualitative and quantitative methods: Triangulation in action', *Administrative science quarterly*, vol. 24, no. 4, pp. 602–611, 1979.
[76] W. J. Orlikowski and J. J. Baroudi, 'Studying information technology in organizations: Research approaches and assumptions', *Information systems research*, vol. 2, no. 1, pp. 1–28, 1991.
[77] A. J. Vidich and G. Shapiro, 'A comparison of participant observation and survey data', *American Sociological Review*, vol. 20, no. 1, pp. 28–33, 1955.
[78] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, 'Systematic Mapping Studies in Software Engineering.', in *EASE*, 2008, vol. 8, pp. 68–77.

[79] B. A. Kitchenham and S. Charters, ‘Guidelines for performing systematic literature reviews in software engineering’, in *Technical report, Ver. 2.3 EBSE Technical Report. EBSE*, sn, 2007.
[80] C. Wohlin, ‘Guidelines for snowballing in systematic literature studies and a replication in software engineering’, in *Proceedings of the 18th international conference on evaluation and assessment in software engineering*, 2014, p. 38.
[81] A. Fink, *How to conduct surveys: a step-by-step guide*, Sixth edition. Los Angeles: SAGE, 2017.
[82] F. J. Fowler, *Survey research methods*, 4th ed. Thousand Oaks: Sage Publications, 2009.
[83] G. G. Gable, ‘Integrating case study and survey research methods: an example in information systems’, *European journal of information systems*, vol. 3, no. 2, pp. 112–126, 1994.
[84] A. Strauss and J. Corbin, ‘Grounded theory methodology’, *Handbook of qualitative research*, vol. 17, pp. 273–85, 1994.
[85] J. Corbin and A. Strauss, ‘Grounded theory research: Procedures, canons and evaluative criteria’, *Zeitschrift für Soziologie*, vol. 19, no. 6, pp. 418–427, 1990.
[86] A. Strauss and J. Corbin, *Basics of qualitative research techniques*. Sage publications, 1998.
[87] H. Heath and S. Cowley, ‘Developing a grounded theory approach: a comparison of Glaser and Strauss’, *International journal of nursing studies*, vol. 41, no. 2, pp. 141–150, 2004.
[88] R. K. Yin, *Case study research: Design and methods*. Sage publications, 2013.
[89] M. B. Miles and A. M. Huberman, *Qualitative data analysis: An expanded sourcebook*. Sage, 1994.
[90] European Commission, ‘What is an SME? - European Commission’. [Online]. Available: http://ec.europa.eu/growth/smes/business-friendly-environment/sme-definition/index_en.htm. [Accessed: 12-Apr-2016].
[91] J. Yli-Huumo, D. Ko, S. Choi, S. Park, and K. Smolander, ‘Where Is Current Research on Blockchain Technology?—A Systematic Review’, *PLOS ONE*, vol. 11, no. 10, p. e0163477, Oct. 2016.
[92] B. A. Kitchenham *et al.*, ‘Preliminary guidelines for empirical research in software engineering’, *IEEE Transactions on Software Engineering*, vol. 28, no. 8, pp. 721–734, Aug. 2002.
[93] J. Maxwell, ‘Understanding and validity in qualitative research’, *Harvard educational review*, vol. 62, no. 3, pp. 279–301, 1992.
[94] A. S. Lee and R. L. Baskerville, ‘Generalizing generalizability in information systems research’, *Information systems research*, vol. 14, no. 3, pp. 221–243, 2003.
[95] Futurice, ‘Lean Service Creation’, 2017. [Online]. Available: https://www.leanservicecreation.com/. [Accessed: 14-Nov-2017].
[96] JUHTA, ‘JHS-suositukset - JHS 169’, 2012. [Online]. Available: http://www.jhs-suositukset.fi/suomi/jhs169. [Accessed: 14-Nov-2017].

Publication I

Herala, A., Vanhala, E., Porras, J., and Kärri, T.
Experiences about Opening Data in Private Sector: A Systematic Literature Review

Proceedings of 2016 SAI Computing Conference (SAI)
pp. 715-724, 2016

SAI Computing Conference 2016
July 13-15, 2016 | London, UK

Experiences About Opening Data In Private Sector: A Systematic Literature Review

Antti Herala, Erno Vanhala, Jari Porras and Timo Krri
Lappeenranta University of Technology
Skinnarilankatu 34
53850 Lappeenranta, Finland
Email: {antti.herala; erno.vanhala; jari.porras; timo.karri}@lut.fi

***Abstract*—Open data has been described to, for example, reduce bureaucracy, remove corruption and lower the cost of a product. The discussion has mainly been over the data that has been provided by public organizations: governments, states, and municipalities. In this systematic literature review the aim is to see why private organizations would want to participate to open data initiative and how it would impact their operations. This study presents observed benefits and challenges as well as assumed implications. It was found out that the studied research is mainly speculative and observed results are shallow compared to the assumed and speculated benefits and problems.**

***Keywords*—Open data; Private organizations; Business change; Business impacts**

I. Introduction

Open data is gaining popularity as an initiative, especially in governmental organizations [1]. According to Masip-Bruin, open data is "*an approach to managing data so that it enables the structured free flow of non-sensitive information to those who have a need or interest in using it, both within and across government agencies and to the public. It allows different types of users to access, organize and use data in ways that make sense to them*" [2, p. 331]. With open data and open movement developing, there resides potential for changes in the ecosystems on how the data is processed, shared, and used [3]. In their study Geiger and von Lucke present three major changes of data handling [4, p. 186-187]:

- "*Everything is secret, if is not explicitly marked as public.*" → "*Everything is public, if it's not explicitly marked as secret.*"
- "*Range and time of publication are determined by public authority. Often, inspection of files is on application, based on the Freedom of Information Act.*" → "*All data not determined by qualified data privacy protection or data security are fully published, proactive and contemporary.*"
- "*Published data are permitted to be inspected for private use. Further usage is reserved and can be allowed on demand.*" → "*Published data are usable by everybody for everything including commercial usage without any restrictions exempt from charges. This contains the possibility of editing and distributing of the public data.*"

While the public bodies hold vast amount of data, open data is required from private organizations in order to realize the full benefits [5]. For example in Earth observations open data is a de facto standard to use, since most of the data is applied by and for the public [6]. While the major discussion about open data is currently concentrated on governmental level [7], the private companies also need to have strategies and policies to harness the advantages of disclosure [5]. This phenomenon is partially explained by Jetzek when talking about open government data (OGD): *When governments become open, the mechanisms that affect value generation and appropriation move beyond the traditional buyer-seller relationships; thus connections between the public and the private, as well as the social and the economic dimensions begin to emerge* [3, p. 3]. The main objective of this study is to survey the existing literature on how the companies have benefited from open data and what negative effects they have suffered. The impacting segments are separated into two groups based on the perspective: assumptions and observations. The main focus are the observed impacts and assumptions are used to support and explain the nuances of the observations. This study found six beneficial and four detrimental impacts to companies from opening their data.

After the introduction, in the second section the research process is presented, followed by the research questions and the setting of the study in section three. The fourth section outlines the search: used databases, search terms, and the yield of the search. In the fifth section the findings are presented by starting from positive impacts and followed by negative impacts. Lastly the findings are discussed and the limitations of the study are outlined in the sixth section. The study is concluded in the seventh section.

II. Research process

This study follows the systematic mapping study guidelines introduced by Kitchenham and Charters [8]. This method was selected in order to gather and present a phenomenon in a thorough manner, especially since the topic has been a target of a biased conversation [7]. The systematic mapping research process has been illustrated by Petersen et al. [9] and presented in Figure 1.

The first step in the process of a systematic mapping study is to define and outline the topic. From the clear topic, a review scope can be created that will be used to conduct the research and categorize the search, which yields all the articles. The papers are then screened against the criteria set at the beginning of the research and only the relevant studies are included for the next step. When all the non-relevant articles are excluded

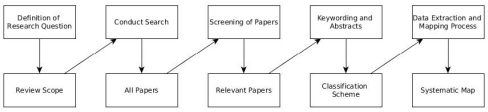

Fig. 1: Systematic mapping research process [9]

from the search, the remaining papers are keyworded from abstracts. Finally, after the classification scheme is placed, the selected studies are read carefully, the data concerning the research problem is extracted, and the systematic map can be constructed with the results in a suitable format.

III. Research questions

The landscape of open data currently requires scientific evidence about the benefits and negative effects for companies and their businesses to increase the understanding and participation in open data initiative [10]. The discussion about the topic has been biased between company representatives and open data experts, with little scientific evidence [11]. This mapping concentrates on two research questions:

- "*What are the effects of opening their data to private sector?*"
- "*What are the open data perspectives of private sector in scientific literature?*"

For the both technological and management goals of this study, the following databases were used: ACM Digital Library, IEEE Xplore Digital Library, Science Direct, ProQuest, EBSCO, and additionally Google Scholar. The aim is to find as many articles as possible from all the scientific areas, where open data has been published or is being considered for publishing. These databases offer articles from both governmental and private open data, focusing on the effects that open data can bring and has brought already.

In addition, to the organizational effects the papers were selected with additional requirements: 1) the article has to cover observed or assumed effects of open data instead of a list of possible barriers, 2) the article has to be peer-reviewed, 3) the language of the article has to be English and 4) full-text articles are required.

Two initial searches were conducted from February 2015 to March 2015 and an additional search was done in April 2015.

IV. Search

The search was done in three individual stages. Round 1 (R1) focused on the term "*open data*" AND "*private sector*" AND ("*impact*" OR "*effects*") from all search fields. In round 2 (R2) the search focused on the title-abstracts with more limited search term ("*private sector*" OR "*effects*") AND "*open data*") in order to limit the search results. In addition to these two search rounds, the selected articles were collected and their references were mapped for additional information. The additional, reference-based search was conducted by extracting data from articles and if the data was referenced from a non-analyzed article, that paper was selected for further analysis. Those articles were searched from the predefined databases and additionally from Google Scholar, presented as Other in TableI The additional search was used in order to cover as much previous research as possible. The results of the three stages of the search can be found on Table Search. Total of 347 articles were searched and it yielded 74 papers having few duplicates. The results are presented in Table I.

TABLE I: R1: "*open data*" AND "*private sector*" AND ("*impact*" OR "*effects*") from all fields, R2: ("*private sector*" OR "*effects*") AND "*open data*" from title-abstract, additional search from previous references. Numbers are presented as Accepted / Found.

	ACM DL	IEEE Xplore	Science Direct	ProQuest	EBSCO	Other	
(R1)	17/29	0/0	12/122	6/83	1/2	0/0	36/236
(R2)	3/37	6/8	4/11	5/11	4/10	0/0	22/77
Ref. search	0/2	0/0	1/7	7/11	0/0	8/14	16/34
	20/68	6/8	17/140	18/105	5/12	8/14	74/347

The 347 title-abstracts of the papers were read in the search and 74 were accepted for further studying. Five duplicates were found in R2, after R1 was completed. The duplicates were dropped from the total count, leaving the final count to 69 individual articles. Out of the 69 articles, 21 were discarded after reading the full article. The major reason for the exclusion was the focus on technological aspects, such as implementations and effective delivery, instead of the effects of open data to the organization. In these papers the authors either analyzed or constructed systems to utilize open data from the technological point of view. Another reason was the focus on single governmental unit that constructed strategies for opening the data. This shows only possibilities for a singular entity without any concrete studies, evidence-based hypotheses, or necessary insights. In the end, out of 347 articles, 48 papers were fully read and selected for data extraction. The findings are presented in the next section.

V. Findings

The articles qualified for this study rely heavily onto the assumed propositions, while the studies showing concrete implications were considerably fewer. This is evident from Figure 2, where the papers are categorized by the number of impacts per article and the emphasis on the nature of the impacts: assumed or observed and positive or negative. In Figure 2 the assumed impacts are on the left side of y-axis and the observed impacts on the right side while the positive impacts are over the x-axis and the negative impacts below it. The position for x-axis was calculated by subtracting the number of assumed impacts from observed impacts, leaving the mainly assuming papers on the negative side of the axis and the mainly observing articles on the positive side. The position for y-axis was calculated by subtracting the number of negative impacts from positive impacts. The size of the bubble correlates the total number of mentioned impacts in the article.

The Figure 2 shows the current bias about open data: most of the papers are concentrating on the assumptions.

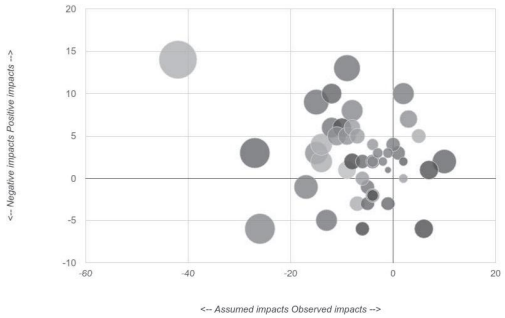

Fig. 2: Emphasis on positive and negative impacts (different levels of gray are for clarification).

Only a handful of studies present cases, where opening the data has brought any concrete impacts to the industry. Such examples are the cases of Nike and Levi-Strauss [12] and the industrial revolution in British isles [13]. Most of the positive impacts are assumed by various actors without actual references or research on how the use open data improves business. The negative impacts rely to assumptions as well by showing even fewer observed cases about the negative impacts of open data. Most of those papers suggest and warn about the negative effects hypothetically without experience, for example concerns about privacy infringements.

In Figure 3, the articles are presented individually to show the different emphasis they had. The figure confirms that the number of assumed impacts is higher than observed and some of the papers concentrate purely on assumptions. Some researchers, such as Nuvolari [13] and Tjoa [14] present only few observed impacts while for example, Janssen [7] present only assumptions as well as the largest number of impacts in an individual article in this study.

The collected impacts were analyzed in order to extract compact information from the initial data. The process that was used is as follows:

1) Impacts are divided into four categories: observed positive/negative, assumed positive/negative.
2) Positive and negative impacts are separated and the following steps are applied to both sets separately.
3) All of the impacts are clustered based on their attributes and the clusters are named with an abstracted topic that explains them with few words.
4) The clusters of assumed impacts are then clustered based on the clustered observed impacts and the result is presented in Table II (positive) and Table IX (negative).
5) The clustered observed impacts are illustrated with the articles and the clustered assumed impacts are abstracted into tables.

In Table II the clustered observed impacts are presented in the first row and in the parenthesis the corresponding table, where the assumptions are explained. Beneath the observed impacts in the columns are the clustered assumed impacts

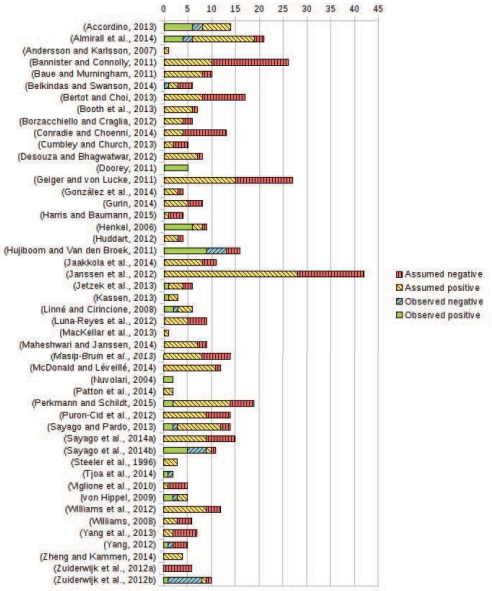

Fig. 3: Impact categories in mapped papers.

and after the assumed impacts in the parenthesis, the number of references in the articles. In some articles the collected individual impacts were categorized similarly and thus, the total number of impacts here and the number of all impacts in Discussion differ. In the next section, the positive impacts are presented and negative impacts follow afterwards.

A. Findings providing positive impacts

1) Collaborative actions: As an example about opening data to collaboration, from the first industrial revolution Nuvolari [13] presents a case of Cornish pumping engines from 19th century. The study speaks about collaborative actions by mine managers, who released the information about their mining engines: technical characteristics, operating procedures and performance. The motivation for the publication was the current monopoly of one engine provider, who was able to block other steam technology development with a patent. After the patent expired, the market opened for other developers, which caused the current engines to be discarded and new versions developed by individual engineers. By providing data and information to engineers, the amount of trial-and-error development decreased, which helped the efficiency of the engines in use to improve steadily.

Since the first industrial revolution, the technology for data and information sharing has become more effective. Through Internet, sharing, collaborating, and participation are made easier continuously and the standing problem is how to harness the available technologies [15]. For example, major cities in

TABLE II: Positive assumed impacts categorized with observed impacts.

Collaborative actions (Table III.)	Innovation and development (Table VI.)	Competitiveness (Table IV.)	Ecosystem-wide engagement and communication (Table V.)	Internal change (Table VII.)	Public image (Table VIII.)
Participative actions (12)	Data development (5)	Business development (14)	Community actions (5)	Governance development (3)	Public image (11)
Feedback (11)	Impact to innovation (14)	Change in business environment (10)	Crowdsourcing (10)	Decision-making support (7)	
Research activities (9)	Service development (10)	Change in revenue generation (10)	Ecosystem change (20)	Internal change (21)	
			External expert utilization (9)	Working environment (8)	

different countries are using technologies to share the data they have collected from their functions to enhance civic open innovation to accelerate the development of tools for citizens [16]. The strategy is to use external resources and collective actions to increase the added value to citizens [17], for example through combination of data in a new way [14]. According to Geiger and von Lucke [4], added value is created when unconnected datasets are combined and new conclusions achieved. After the data is published, the publisher should be actively searching the knowledge that has emerged from the data [7]. The cities also use open data initiative to engage the citizens to the policymaking through collectivity [15], making the dialog and participation possible between policymakers and citizens [3].

Private organizations are also a part of opening the data. In 2005 Nike and Levi-Strauss disclosed their factory data after multiple inquiries from student unions [12]. While the companies had initial fears about the disclosure, the long-term benefit was enhanced industry collaboration and information sharing. In addition, the private companies can benefit from the increased dialog and participation also from its customers: Sayogo et al. [5] report smart disclosure by companies, who open their data to the consumers about their production and supply chain. The goal is to increase public understanding about the sustainability of the products and empower the consumers to take actions the companies provide. The smart disclosure by companies can decrease consumer confusion in complex market, empower consumers by informing them about their decision and increase data-driven products and services and thus improve economy [5].

2) Competitiveness: The increase in competitiveness influences both the data providers and the enablers, who enable the opening and use of data [1], [39]. An example of enabler business is the targeted advertisement, where the social media data is used [15]. The study of Huijboom and Van den Broek [1] found that by gaining access to more data, the ICT companies can create new programs and enhance old applications by adding new features and functionalities that uses data from multiple sources. In addition, to the increased actions of ICT companies, the applications produced from the published data can increase the awareness about the provider. Open data leaves the aspect of ownership and focuses on the developers and companies, who provide products and services tailored to larger audiences for greater financial gains [16]. The companies can transform to become more transparent about their actions and processes, increasing publicity and consumer trust. According to the interviews done by Sayogo et al. [5], becoming transparent and open about the company's activities allows smaller enterprises to compete with larger players, who have the advantages in scale, resources, visibility and brand, and consumer base.

TABLE III: Assumed impacts that refer to collaborative actions.

Term	Description
Participative actions	**Collaboration and participation** [18], [19], [7], [4], [20], [2], [15], [21] **in topics of interest** [22] **and in application development** [23], [4] **increases interoperability** [24] **with academia and other companies** [25], [21], [10], [2], [20], [26], [4], [22]**.**
Feedback	**Contribution through effective feedback** [20], [15], [27], [24] **channels** [7], [11] **from data consumers to improve processes, policies** [28]**, products, services** [7] **and data quality** [29] **increases company's understanding about customers** [30], [27], [31] **and generates external input** [32]**.**
Research activities	**Acceleration** [33], [34], [35], [10] **and focus** [36] **to transparent research process** [37] **allows multiple goals** [22] **in efficient research** [38], [10], [16] **with academia** [36] **and reduces the high failure rates** [22]**.**

While the opening of data can provide benefits, the companies have to strictly limit the amount and quality of published data. In a case by Yang [17], the releasing organization showed less concern of losing valuable assets, since the data was updated daily; it "went bad" fast. Henkel [23] presents a case about embedded Linux, where the actors in the development share their code to public use through open source software. The report found that companies do share their code with strict rules: either generic code or highly specialized code designed only for their hardware. The study also reports that some companies reveal their code, because of its importance to the competitive position and differentiation. From the answers a conclusion is drawn that for some companies the revealing may even support their competitive advantage. The major reason for this stand comes from the fact that by revealing their code they gather technical benefits, such as external development and reduce their maintenance support. The study concludes that the companies can collect benefits by creating a right balance between protecting and sharing their products. The enhancement to competitiveness through revealing may come from shaping collaborative behavior of others or discouraging direct competition. [22].

While the companies lack cases in the literature Borzacchiello and Craglia [40] found that in governmental sector the benefits have been found to outweigh costs in most of the cases. A case from the literature shows that the initial costs were recovered in six months due to increased efficiency [41]. Another study found that by opening data, the average saving of costs and time decreased 11% and 17% respectively [42].

TABLE IV: Assumptions that affect competitiveness of the company.

Term	Description
Business development	**Developing new, more effective business models and processes** [43], [44], [34], [32], [27], [39] **with collaborative and competitive strategies** [16] **and operational efficiency** [45] **with new information, methods, and business intelligence** [21] **about trends, issues, and challenges** [15]**. Support to economic, knowledge** [7], [32] **and business development** [23], [32] **and sustainable production** [5] **by understanding possibilities of economic growth in every point of the information-driven supply chain** [46], [5]**.**
Change in business environment	**Achieving and maximizing new economic opportunities** [32], [16] **by stimulating competitiveness** [25], [5], [7], [2] **and enhancing the competitive position and advantage** [22], [32], [25] **by efficiency gains from combined and trusted data** [31], [39] **and increased client commitment** [39] **by releasing social and commercial value** [18]**.**
Change in revenue generation	**Increased revenue** [6], [16] **and lowered transactions costs** [19] **through credible and sustainable commerce** [31] **and added value through augmentation** [27]**, reuse** [4] **and combination** [7] **of data. Boosted economic development** [35] **leads to profitability** [45]**, widening the company portfolio** [16] **and discouraging competition** [22]**.**

3) Ecosystem-wide engagement and communication: The roles in business ecosystems vary from direct to indirect roles, including the value chain and also other roles, such as outsourcing companies and regulatory agencies [47]. Since opening the data is meant to affect mostly on the public, it can affect every actor in business ecosystem; open data ties together governments, businesses, and citizens [25]. Henkel's [23] study about selective revealing in a case of embedded Linux shows that companies are having extended dialog through revealed artifact. Doorey [12] also reports similar activities in the case of Nike and Levi-Strauss, where increased transparency and factory disclosure added the informal dialog as well as information sharing and collaboration in monitoring, training, and remediation between competitors. Companies have also been able to harness the knowledge of private actors and communities [15], [12].

Openness in ecosystems is a complicated matter, since it requires clear data ownership and clear policies about the disclosure; one company opening the data can easily affect the whole ecosystem [5]. However, the ownership of the data is increasingly moving back to the consumer, who can use the data as they see fit. This would suggest that the customers themselves decide, based on their own data, which advertisements they want to see and what services they want to use [25]. The value of data should be calculated from the financial viewpoint as well as the amount of reuse when compared to the effort needed to publish it [19].

4) Innovation and development: In the era of information-driven development, where even one user can innovate and create their own products, the free and public open data makes a powerful business resource [35]. Von Hippel [51] presents a study about user-based innovation, an innovative action by the users instead of the product manufacturers. In the paper he underlines the importance of user-based innovation in software business that requires extensive amount of data from users over time. Acquiring and applying this data becomes possible through open data, as have been seen from the example of cities and public bodies [30]. Accordino [15] presents an example of innovative software that is connected to open data: the Futurium Platform, a software created solely on decision-making. The data is collected from open data providers, social media APIs and from stakeholders and policymakers. All this data is pooled together and Futurium is capable of using this data to derive knowledge for future use when making decisions. Gurin [35] proposes a demand-driven data disclosure; the opening process is driven by the needs of users by involving the stakeholders.

TABLE V: Assumptions related to the positive change in ecosystem.

Term	Description
Community actions	**Expand the community** [31] **and increase communal activities around the company and their technologies** [28]**, allowing easier access to possible customers** [31] **for comparison** [24] **and scrutinization of data** [7]**. Users can also submit their own data sources and analyzes through the communities** [24]**.**
Crowdsourcing	**Accessibility to crowdsourcing** [20], [35], [21], [10], [16], [22] **encourages crowd-based benchmarking** [24] **and involvement in data collection, analysis, and application** [2]**. Crowdsourcing and public engagement** [7] **increases awareness of problems and solutions** [48]**.**
Ecosystem change	**Transparency** [40], [49], [32], [7], [4], [3], [20], [2], [21], [5], [24], [18] **in ecosystem allows accountability** [50], [29], [20], [21], [25] **and accessibility to investors and companies** [48], [7], [16]**, making the two-way communication between stakeholders and companies faster and more reliable** [19], [28], [15], [4]**. The data can be engaged and validated through external resources** [28], [16], [7] **with transparent process that allows tracking and makes explanations possible** [29] **in the ecosystem. Through the external engagement, the publisher remains ready for constructive influences, discourses, and exchanges** [7]**.**
External expert utilization	**Enhancing performance** [28] **through increased communications with third parties** [51]**, allowing the use of collective intelligence** [4], [7] **and involving third parties to data processing** [7] **and validation** [22]**. Third parties include global expert community** [10] **that can be accessed** [20], [16] **through the Internet** [34]**.**

TABLE VI: Assumptions about the enhancement of innovation and development.

Term	Description
Data development	**The mix of public and private data** [7] **adds cross-data interactions and aggregations** [2] **that are made possible with easier data movement** [22] **and sustainable data** [7]**. The technologically independent access** [2] **provides new challenges to third party data mining** [29] **and unearths underlying data quality issues** [43]**.**
Impact to innovation	**Enhanced** [21], [10]**, accelerated** [33]**, promoted** [49]**, increased** [26], [44], [25], [22], [39], [19]**, improved** [50] **and stimulated** [7] **cumulative** [22] **and collective** [4] **innovation** [3]**.**
Service development	**New** [7]**, improved** [16], [7] **and enhanced** [32] **services with improved service provision** [24]**. The performance** [10] **and efficiency** [48] **of the services can be increased through better standards** [50] **and seamless integration** [2]**, increasing the quality** [39], [45] **and outreach** [32] **of the services.**

Linne and Cirincione [36] studied the effects of open data in real estate, appraisal, and mortgage banking. This is only one example of developing industries, who have embraced open data, renewing the real estate analytics. According to the study, the benefit of open data primarily comes from cost savings and increased efficiency in the real estate industry, since the data can be typed only once. Also opening the data records to commercial vendors, academics, and practitioners the business analytics and metrics are required to change in order to analyze millions of aggregated data points [36]. Through open data the market's movements and patterns are now visible and exploitable, enhancing the quality of services

and products available for consumers [5].

5) Internal change: The reported benefits from internal change emerging from open data is rather scarce in the literature. Huijboom and Van den Broek [1] report budget cuts as a driver to open data in United Kingdom. The government wanted to create savings through publishing data about the public expenditures and they involved citizens in the process of cutting, asking where the cuts should be made. Doorey [12] reports that open data is thought to provide better performance through publicity, when corporations open the data that reflects poorly on their requirements and it encourages managers to improve the performance of the processes where the data is gathered. The decisions of Nike and Levi-Strauss that lead to the disclosure of their supplier factories pressured the suppliers to increase the transparency on their working environment and processes [12].

TABLE VII: Assumptions about the changes in internal processes.

Term	Description
Governance development	**The culture of administration changes to welcome opposing views and inputs** [7] **and the process of administration is optimized** [7] **and modernized through openness** [4]**. Administration can receive supportive data from multiple sources to leverage investments** [48] **and development does not necessary need to be tied to administration** [4]**.**
Decision-making support	**Legitimate decisions** [15] **and policies can be based on** [40], [29] **and improved with** [7] **a rich set of analyzes and opinions** [33]**, making the process** [7] **transparent** [52] **and the consequences visible, comprehensible and discussable** [4]**.**
Internal change	**Creating cost efficiency, decreasing costs** [16], [22], [20], [30], [25], [36], [33], [39] **and boosting productivity** [3], [32], [5], [40], [27] **by countering unnecessary duplication of data and work** [10], [7], [39]**. By illuminating problems in the processes** [29] **the information asymmetry is reduced** [49]**, support is given to the creation of new insights, knowledge, and technologies** [48], [28], [7] **in agile and adaptable environment** [16], [15] **and the information and data can be easily maintained and extended** [52]**.**
Working environment	**Internal criticism** [29] **and corruption** [35] **can be reduced through easier monitoring** [28], [29] **and transparent guidance** [5]**, when openness becomes a standard procedure** [21] **and employees are empowered** [20] **to change their own working conditions and future within the company. By opening the data, employees can be moved to more interesting jobs** [40] **and the stress and workload can be reduced on individual level** [48]**.**

6) Public image: Companies are increasingly being inquired about their actions by the governments and the information about the sustainability of companies products and practices is being increasingly demanded by the customers [49]. This drives them to change their working methods towards transparent actions [12]. According to Doorey [12] Nike and Levi-Strauss disclosed their supplier factories because of the public demand; they deemed the potential benefits of being the first to market with transparency to be more beneficial than the associated risks. The companies can also provide their data directly to the public, but can use data intermediaries to validate and publish the data to the public [49]. While the companies use third party certifiers to calm the public about the sustainability and ethicality of their products, social pressure and government enactment are guiding the companies from certifications to transparent actions, where each individual consumer can monitor the product throughout the supply chain [5]. Even charities have found that donors are actually expecting accountable and transparent actions instead of assurances [50].

TABLE VIII: Assumed enhancement of public image.

Term	Description
Public image	**Brand, reputation** [31], [25], [4], [7] **and public profile** [10] **is improved through positive publicity, visibility, and transparent actions that broadens understanding** [7], [28]**, adds trust towards the publisher** [35], [29], [32], [25], [5], [39], [7] **and conveys competence and integrity** [5]**.**

B. Findings providing negative impacts

The negative impacts in Table IX are presented similarly to the positive impacts in the previous section. The first row of the table presents the topic (observed impact) and in the parenthesis the table, where the assumed impacts are explained after the description. In the parenthesis after the assumed impact is a number that indicates how many articles mentioned the said impact.

TABLE IX: Negative assumed impacts clustered with observed impacts.

Decrease in efficiency (Table X.)	Increased costs (Table XI.)	Public data increases problems (Table XII.)	Required changes (Table XIII.)
Economic hindrance (20)	Investments (17)	Illegal activities (21)	Business change (5)
Efficiency decrease (18)		Public image (9)	Internal change (7)
Hindered collaboration (9)			

1) Decrease in efficiency: Yang [17] found in a case study about the Taiwanese e-government that when an organization is striving towards openness, the usually forgotten fact is that setting up openness and creating new processes parallel to the old ones increase the workload on existing, limited resources. Even if the open processes are used to substitute the old ones, the preparations and modification take up substantial amount of time and new resources are needed to maintain the systems. While the preparations and management could be handled, Tjoa [14] found a challenge that the technical experience of the publisher and especially users are usually lacking. In other words, even if the data is published, the ones who would want to use the data lack the skill to access it. This would require significant investments into the user interfaces, which would conflict with the target of the publisher.

In addition, Huijboom and Van den Broek [1] found that the scattered strategies of the organization are a major hindrance, if the organizations cannot decide which processes they should use, thus dividing the available resources. They also found that national laws and ethics can be a significant factor in opening the data, since in some countries the publishing of the same datasets are illegal, even if they would be legal in others. Such datasets that hold any data that can be classified as private will have multiple problems from legislative regulations. Almirall [16] also found that existing conflicts between different organizations can prevent opening the data. While the study by Doorey [12] discredited the necessity to hide the data from competitors, some companies still refuse to collaborate with each others business, limiting the amount of data to be released.

TABLE X: Assumptions about the decreasing efficiency.

Term	Description
Economic hindrance	**The income can be reduced** [11], [49], [17], [34] **because of the uncertain economic shift** [34]**, impact** [1]**, outcomes** [19], [7] **and costs** [31], [19] **or because of regulations** [44], [22]**, lack of intellectual rights** [23] **or the lack of use of the data** [7], [2], [3], [5], [37], [20]**. The company can also lose market share** [22] **and commercial confidentiality** [32], [5], [29], [4], [21], [44], [22] **by accidentally publishing critical data.**
Efficiency decrease	**The efficiency can be decreased due to legal restrictions (licensing and copyright** [43], [11], [7], [4], [10], [39], [6] **and opaque data ownership** [19]**), technical issues (information asymmetry** [26], [4]**, lack of standardization** [4], [26] **and poor data quality** [35], [50]**), human reasons (lack of technical expertise** [39], [20], [21]**, misinterpretations** [24] **and the lack of users** [37], [48], [21]**) and the lack of collaborative actions and interoperability** [22], [21]**. Opening the data may cause the processes to be obfuscated** [29] **to the point where the effectiveness suffers.**
Hindered collaboration	**The collaborative companies have mismatched technologies** [31] **or business models** [16], [4]**, lacking the interoperability between systems** [37], [2], [26], [10] **and the created technologies will be unused due to the lack of technological expertise** [7], [37], [11] **and linguistic problems** [31]**.**

2) Increased costs: According to Accordino [15], the first costs about the data comes from the collection. The instantaneous data collection technology and the use of ICT is largely untapped, demanding technological investments to simply gather data. Accordino continues that after the data has been successfully collected, the next problem is that the data has to be rendered useful to stakeholders and policymakers through visualization, which requires more tools. In order to make the use of technology effective, the users of the new systems also require training to the systems, which increases the technology related costs [1].

If the company has existing technologies and systems they use, then changing to openness may require more investments into the systems. The interviews done by Sayogo [5] point that there exists multiple costs, when striving towards openness. First, ensuring openness throughout the supply chain requires significant investments due to the possibility of suppliers being unable to use the latest technology. Usually the suppliers use remote technological resources, which requires additional staff. To publish the data online, if not completely automated, also requires more time and dedication from the existing staff, a time they could use to complete their original jobs. Opening process data might also require extensive explanations and background material available [29]. The data maintenance, such as records and documentations, is also a significant cost to data producers because of the collection, accuracy, and credibility of the data, especially when the data is published to the public who scrutinize the data [29].

TABLE XI: Assumed aims for new investments that are required to realize openness.

Term	Description
Investments	**Significant financial** [27] **and technological costs** [26]**, because of tying management** [21] **and time** [43], [29] **during the opening process. The technological tasks are unclear** [19] **until the whole system has been implemented, requiring resource, budget, and time** [2] **as well as technological investments** [7], [4], [37], [10], [20], [2], [21], [25], [39], [6], [18], [35], [19], [43]**. After the new systems have been implemented, the management requires new mechanisms, capabilities, and processes for governance** [7]**.**

3) Public data increases problems: The study of Zuiderwijk [53] concentrates on the case of the Research and Documentation Centre (WODC) located in Netherlands. The WODC collects, stores, enhances, and provides information to the Dutch Ministry of Security and Justice. The data the WODC processes is partially sensitive and private information; the study creates guidance for organizations with sensitive data on how and what to publish. When mapping 45 datasets, the most commonly recognized issues by Zuiderwijk [53] were privacy-sensitivity and anonymization, a lack of metadata, and a lack of data quality. The privacy-sensitive data can be misused and misunderstood and it can be used to trigger spurious findings that may affect the provider's publicity. For example, inadvertent release of confidential data may lead to privacy infringements by de-anonymizing data, identifying groups and individuals [29]. According to Zuiderwijk [53], the lack of metadata and data quality makes the dataset more difficult to search through and the usability of data is lowered by confusing the users about the source and function of the data.

TABLE XII: Assumed problems that opening the data may bring.

Term	Description
Illegal activities	**The opened data is exposed to misuses and misinterpretations** [40], [4], [28], [21], [5], [19], [11], [24], [29]**, it can cause potential threats** [29] **to the company** [1] **and to individuals** [37], [1], [29], [4], [7], [31], [26], [2], [21], [44], [25], [6], [19] **and can be available for malpractices** [20]**, such as data fraud** [49]**, hacking and data manipulation** [29]**.**
Public image	**By opening the data, publisher may receive negative publicity** [4] **which can lead to loss of trust** [49] **and brand** [21]**, including other socially undesirable outcomes** [29] **such as critique towards individuals** [28] **instead the company. By opening the datasets, company may face legal actions** [29]**, critical and unwanted questions** [7] **or increased discussion and confusion about the data** [7]**, especially if the data lacks validity, completeness, or exhaustiveness** [3], [2], [4], [7]**.**

4) Required changes: Open data and openness brings changes into organizations, ecosystems, and cultures. According to Zuiderwijk [53], the organizations face changes in the form of new focus and policies, such as funding and reward systems, and time consumption structures. The interviews conducted by Sayogo [5] indicate that the decision of one company opening their data affects the supply chain directly. The suppliers are required to provide exact and complete data to their customers, who want to publish the data, requiring changes in the ecosystem and both organizations. The closed cultures of organizations are also required to change, when the data is published outside the company. Huijboom and Van den Broek [1] report that the closed culture of an organization is a major hurdle for opening the data, requiring changes in policies and processes but as well in the employees and their mentality. According to Bannister and Connolly [29], transparency should be used in constructive climate and a culture of quality improvement, or it may lead to cover-ups and blame assignment.

Aside from the changes in the organizations, the old business models and sources of value can become ineffective. Belkindas and Swanson [43] point out that especially to companies who make revenue by selling data, launching open data requires more financial resources to cover the expanded offerings, while the revenue from data has disappeared.

TABLE XIII: Assumptions about the negative impacts that the required changes bring to organizations.

Term	Description
Business change	**The control over the data of the publisher cannot be predefined but only guided** [7], [22]**, causing changes in existing culture** [7], [5], [22]**, business models** [34] **and operational models** [26]**.**
Internal change	**Cultural shift in administration** [4]**, organization** [20], [16], [19] **and processes** [19], [4] **may lead to biased overview** [7] **and into new culture, where excessive caution and conformity, non-recording, informal procedures, and self-censorship increase** [29] **due to destruction of jobs** [40]**. Such environment does not support radical thinking, honesty, and openness** [29]**.**

VI. Discussion

In the beginning of this article, two research questions were outlined: RQ1 – "*What are the effects of opening their data to private sector?*" and RQ2 – "*What are the open data perspectives of private sector in scientific literature?*".

This study outlines multiple impacts that affect organizations, both positive and negative. The positive impacts comprise from increasing collaborative actions and competitiveness, addition to ecosystem-wide engagement and communication, enhanced innovation and development, internal change within the company processes and methods, and positive public image. The negative impacts consist of decrease in efficiency, increasing costs, problems caused by public access to data and the changes that are required from the organizations.

In the search, 48 articles were included into the reading process and 466 impacts were collected from the papers. The number of impacts and the ratio of positive and negative impacts in a cluster are presented in Table XIV. The majority of the papers are making assumptions but still both clusters, observed and assumed, have more positive impacts than negative. The ratio of positive impacts over negative impacts in the observed cluster was 2.0 and in the assumptions cluster it was 1.64. The articles had 50 positive observed impacts and only 25 negative impacts. The positive assumed impacts was counted to be 243 in the articles, while the number of negative assumptions was 148. Both observed and assumed impacts would suggest that the opening of data has more positive than negative impacts to businesses but the number of assumptions show that openness is still seen as a disadvantage. This finding suggests that open data is perceived as a threat in current businesses, while the empirical experience suggests otherwise.

TABLE XIV: The division of impacts found from the selected articles.

	Observed	Assumed
Positive	50	243
Negative	25	148
Ratio [positive/negative]	2.0	1.64
Sum	75	391

A. Limitations

A major limitation is the search through selected databases that does not guarantee that every article about the topic is presented. The lack in the search functionalities of the databases were partially covered with the reference-based search. The databases were selected to cover the viewpoint of both engineering and management to get as wide a sampling as possible. In the search, the selection also excluded books and non-peer-reviewed articles by focusing on peer-reviewed publications. The research is partially covered in non-scientific publications and reports, especially since this study concentrates on the organizations, who usually do their own research. The main concentration was also the articles that were written in English, while the open data initiative covers multiple countries and governments and the research is written in multiple languages. This has possibly ruled out reports, where the effects of open data are being discussed.

Open data is also currently being published by public organizations. While they are fundamentally different from private organizations, they share similar attributes. The possible effects that are influencing changes in the public organizations can also remain true for private organizations as well. By excluding the governmental open data from the search can have left out valuable impacts that can affect an organization.

B. Future work

The data gathered from the 48 articles indicate that the private organization have not opened their data storages with the same speed and interest as the state owned organizations. It shows a clear need for more research about the lack of eagerness to open the data on private organizations. The next step in the research is to study the actual private organizations and their reasons to open their data and keep their data closed.

VII. Conclusion

Open data is a rising ideology that can change the current business landscape through sharing the data. Multiple governments have already started releasing their data systematically and only few companies are yet following the trend.

This study highlights reasons, what could happen to private organizations if they opened their data. The impacts were found by mapping existing literature systematically from scientific databases. The search was made using five databases and 48 articles were selected for further analysis. The impacts from the articles were clustered as observed and assumed as well as positive and negative impacts.

The impacts lean heavily on assumptions; 391 of 466 of the found impacts were assumed. However, the literature also suggests that the impacts are mostly positive to the organizations, as long as the opening is done systematically and carefully. The observed impacts also show a higher ratio of positive impacts per negative impacts, totaling a ratio of 2.0 while the assumed impacts total only 1.64 ratio. This would suggest that when the data is opened, the results are beneficial but for some reasons private organizations have not opened their data; they may see it as a loss of business advantage.

While the scientific literature shows that opening the data brings benefits to the provider, the practice does not follow the examples. While the governments are opening their data, private organizations are waiting to see, how the new ideology can enhance the existing markets and businesses. Based on what has been found in this paper, the organizations have multiple opportunities to benefit from openness.

ACKNOWLEGMENT

We'd like to thank Fimecc and S4Fleet-project for funding this study.

REFERENCES

[1] N. Huijboom and T. Van den Broek, "Open data: an international comparison of strategies," *European journal of ePractice*, vol. 12, no. 1, pp. 4–16, 2011.

[2] X. Masip-Bruin, G.-J. Ren, R. Serral-Gracia, and M. Yannuzzi, "Unlocking the Value of Open Data with a Process-Based Information Platform," in *2013 IEEE 15th Conference on Business Informatics (CBI)*, Jul. 2013, pp. 331–337.

[3] T. Jetzek, M. Avital, and N. Bjørn-Andersen, "Generating Value from Open Government Data," in *Proceedings of the 34th International Conference on Information Systems. ICIS 2013*. Atlanta, GA: Association for Information Systems. AIS Electronic Library (AISeL), 2013.

[4] C. P. Geiger and J. von Lucke, "Open Government Data," in *Conference for E-Democracy and Open Government*, 2011, pp. 183–194.

[5] D. S. Sayogo, J. Zhang, T. A. Pardo, G. K. Tayi, J. Hrdinova, D. F. Andersen, and L. F. Luna-Reyes, "Going Beyond Open Data: Challenges and Motivations for Smart Disclosure in Ethical Consumption," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 9, no. 2, pp. 1–16, 2014.

[6] R. Harris and I. Baumann, "Open data policies and satellite Earth observation," *Space Policy*, vol. 32, pp. 44–53, 2015.

[7] M. Janssen, Y. Charalabidis, and A. Zuiderwijk, "Benefits, Adoption Barriers and Myths of Open Data and Open Government," *Information Systems Management*, vol. 29, no. 4, pp. 258–268, Sep. 2012.

[8] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," Technical report, EBSE Technical Report EBSE-2007-01, Tech. Rep., 2007.

[9] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic Mapping Studies in Software engineering," in *Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering*. Italy: British Computer Society, 2008.

[10] A. J. Williams, L. Harland, P. Groth, S. Pettifer, C. Chichester, E. L. Willighagen, C. T. Evelo, N. Blomberg, G. Ecker, C. Goble, and B. Mons, "Open PHACTS: semantic interoperability for drug discovery," *Drug Discovery Today*, vol. 17, no. 21–22, pp. 1188–1198, Nov. 2012.

[11] A. Viglione, M. Borga, P. Balabanis, and G. Blöschl, "Barriers to the exchange of hydrometeorological data in Europe: Results from a survey and implications for data policy," *Journal of Hydrology*, vol. 394, no. 1–2, pp. 63–77, Nov. 2010.

[12] D. J. Doorey, "The transparent supply chain: From resistance to implementation at Nike and Levi-Strauss," *Journal of Business Ethics*, vol. 103, no. 4, pp. 587–603, 2011.

[13] A. Nuvolari, "Collective invention during the British Industrial Revolution: the case of the Cornish pumping engine," *Cambridge Journal of Economics*, vol. 28, no. 3, pp. 347–363, 2004.

[14] A. M. Tjoa, B.-L. Do, A. Anjomshoaa, T.-D. Trinh, P. Wetz, and E. Kiesling, "Exploring Linked Statistical Data Using Linked Widgets," in *Proceedings of the Fifth Symposium on Information and Communication Technology*, ser. SoICT '14. New York, NY, USA: ACM, 2014, pp. 1–3.

[15] F. Accordino, "The Futurium–a Foresight Platform for Evidence-Based and Participatory Policymaking," *Philosophy & Technology*, vol. 26, no. 3, pp. 321–332, Sep. 2013.

[16] E. Almirall, M. Lee, and A. Majchrzak, "Open innovation requires integrated competition-community ecosystems: Lessons learned from civic open innovation," *Business Horizons*, vol. 57, no. 3, pp. 391–400, May 2014.

[17] T.-M. Yang, "The Complexity of Cross-boundary Information Sharing: An Organizational Perspective on Taiwan e-Government," in *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance*, ser. ICEGOV '12. New York, NY, USA: ACM, 2012, pp. 143–145.

[18] J. C. González, J. Garcia, F. Cortés, and D. Carpy, "Government 2.0: A Conceptual Framework and a Case Study Using Mexican Data for Assessing the Evolution Towards Open Governments," in *Proceedings of the 15th Annual International Conference on Digital Government Research*, ser. dg.o '14. New York, NY, USA: ACM, 2014, pp. 124–136.

[19] P. Conradie and S. Choenni, "On the barriers for local government releasing open data," *Government Information Quarterly*, vol. 31, Supplement 1, pp. S10–S17, Jun. 2014.

[20] G. Puron-Cid, J. R. Gil-Garcia, and L. F. Luna-Reyes, "IT-enabled Policy Analysis: New Technologies, Sophisticated Analysis and Open Data for Better Government Decisions," in *Proceedings of the 13th Annual International Conference on Digital Government Research*, ser. dg.o '12. New York, NY, USA: ACM, 2012, pp. 97–106.

[21] J. C. Bertot and H. Choi, "Big Data and e-Government: Issues, Policies, and Recommendations," in *Proceedings of the 14th Annual International Conference on Digital Government Research*, ser. dg.o '13. New York, NY, USA: ACM, 2013, pp. 1–10.

[22] M. Perkmann and H. Schildt, "Open data partnerships between firms and universities: The role of boundary organizations," *Research Policy*, vol. 44, no. 5, pp. 1133–1143, Jun. 2015.

[23] J. Henkel, "Selective revealing in open innovation processes: The case of embedded Linux," *Research Policy*, vol. 35, no. 7, pp. 953–969, Sep. 2006.

[24] D. Maheshwari and M. Janssen, "Reconceptualizing measuring, benchmarking for improving interoperability in smart ecosystems: The effect of ubiquitous data and crowdsourcing," *Government Information Quarterly*, vol. 31, no. 1, pp. S84–S92, Jun. 2014.

[25] D. S. Sayogo and T. A. Pardo, "Understanding Smart Data Disclosure Policy Success: The Case of Green Button," in *Proceedings of the 14th Annual International Conference on Digital Government Research*, ser. dg.o '13. New York, NY, USA: ACM, 2013, pp. 72–81.

[26] T.-M. Yang, J. Lo, H.-J. Wang, and J. Shiang, "Open Data Development and Value-added Government Information: Case Studies of Taiwan e-Government," in *Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance*, ser. ICEGOV '13. New York, NY, USA: ACM, 2013, pp. 238–241.

[27] P. Booth, P. Gaskell, and C. Hughes, "The Economics of Data: Quality, Value & Exchange in Web Observatories," in *Proceedings of the 22nd International Conference on World Wide Web Companion*, ser. WWW '13 Companion. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2013, pp. 1309–1316.

[28] B. Baue and M. Murninghan, "The accountability web: Weaving corporate accountability and interactive technology," *Journal of Corporate Citizenship*, vol. 41, no. 2, pp. 27–49, 2011.

[29] F. Bannister and R. Connolly, "The Trouble with Transparency: A Critical Review of Openness in e-Government," *Policy & Internet*, vol. 3, no. 1, pp. 1–30, 2011.

[30] M. Kassen, "A promising phenomenon of open data: A case study of the Chicago open data project," *Government Information Quarterly*, vol. 30, no. 4, pp. 508–513, Oct. 2013.

[31] L. F. Luna-Reyes, D. F. Andersen, D. L. Andersen, D. Derrick, and H. Jarman, "Full information product pricing regimes: Policy implications for US-Mexico sustainable commerce," in *Puentes Symposium, Houston, TX*. Citeseer, 2012.

[32] J. McDonald and V. Léveillé, "Whither the retention schedule in the era of big data and open data?" *Records Management Journal*, vol. 24, no. 2, pp. 121–99, 2014.

[33] C. Zheng and D. M. Kammen, "An innovation-focused roadmap for a sustainable global photovoltaic industry," *Energy Policy*, vol. 67, pp. 159–169, Apr. 2014.

[34] A. J. Williams, "A perspective of publicly accessible/open-access chemistry databases," *Drug Discovery Today*, vol. 13, no. 11–12, pp. 495–501, Jun. 2008.

[35] J. Gurin, "Open Governments, Open Data: A New Lever for Transparency, Citizen Engagement, and Economic Growth," *The SAIS Review of International Affairs*, vol. 34, no. 1, pp. 71–82, 2014.

[36] M. R. Linné and J. Cirincione, "Creating Open Data Standards for

Real Estate, Appraisal, and Mortgage Banking," *Journal of Real Estate Literature*, vol. 16, no. 1, pp. 75–81, 2008.

[37] A. Zuiderwijk, M. Janssen, S. Choenni, R. Meijer, and R. S. Alibaks, "Socio-technical Impediments of Open Data," *Electronic Journal of E-Government*, vol. 10, no. 2, pp. 156–172, 2012.

[38] B. MacKellar, C. Schweikert, and S. A. Chun, "Patient-oriented clinical trials search through semantic integration of Linked Open Data," in *2013 12th IEEE International Conference on Cognitive Informatics Cognitive Computing (ICCI*CC)*, Jul. 2013, pp. 218–225.

[39] H. Jaakkola, T. Mäkinen, and A. Eteläaho, "Open Data: Opportunities and Challenges," in *Proceedings of the 15th International Conference on Computer Systems and Technologies*, ser. CompSysTech '14. New York, NY, USA: ACM, 2014, pp. 25–39.

[40] M. T. Borzacchiello and M. Craglia, "The impact on innovation of open access to spatial environmental information: a research strategy," *International Journal of Technology Management*, vol. 60, no. 1/2, pp. 114–129, Sep. 2012.

[41] P. G. Almirall, P. Q. Ros, M. Craglia, and M. Moix, *The Socio-economic Impact of Spatial Data Infrastructure of Catalonia*. Office for Official Publications of the European Communities, 2008.

[42] M. Craglia and M. Campagna, "Advanced regional SDIs in Europe: Comparative cost-benefit evaluation and impact assessment perspectives," *International Journal of Spatial Data Infrastructures Research*, vol. 5, pp. 145–167, 2010.

[43] M. V. Belkindas and E. V. Swanson, "International support for data openness and transparency," *Statistical Journal of the IAOS*, vol. 30, no. 2, pp. 109–112, Jun. 2014.

[44] R. Cumbley and P. Church, "Is "Big Data" creepy?" *Computer Law & Security Review*, vol. 29, no. 5, pp. 601–609, Oct. 2013.

[45] L. A. Streeter, R. E. Kraut, H. C. Lucas, and L. Caby, "How open data networks influence business performance and market structure," *Association for Computing Machinery. Communications of the ACM*, vol. 39, no. 7, p. 62, Jul. 1996.

[46] M. Andersson and C. Karlsson, "Knowledge in Regional Economic Growth—The Role of Knowledge Accessibility," *Industry & Innovation*, vol. 14, no. 2, pp. 129–149, May 2007.

[47] M. F. Baghbadorani and A. Harandi, "A Conceptual Model for Business Ecosystem and Implications for Future Research," *DOI*, vol. 10, pp. 82–86, 2012.

[48] K. C. Desouza and A. Bhagwatwar, "Citizen Apps to Solve Complex Urban Problems," *Journal of Urban Technology*, vol. 19, no. 3, pp. 107–136, Jul. 2012.

[49] D. S. Sayogo, J. Zhang, H. Liu, S. Picazo-Vela, and L. Luna-Reyes, "Examining Trust As Key Drivers in Smart Disclosure for Sustainable Consumption: The Case of I-choose," in *Proceedings of the 15th Annual International Conference on Digital Government Research*, ser. dg.o '14. New York, NY, USA: ACM, 2014, pp. 137–146.

[50] S. Huddart, "Renewing the Future: Social Innovation Systems, Sector Shift, and Innoweave," *Technology Innovation Management Review*, vol. 2, no. 7, pp. 5–9, Jul. 2012.

[51] E. von Hippel, "Democratizing Innovation: The Evolving Phenomenon of User Innovation," *International Journal of Innovation Science*, vol. 1, no. 1, pp. 29–40, Mar. 2009.

[52] E. W. Patton, P. Seyed, P. Wang, L. Fu, F. J. Dein, R. S. Bristol, and D. L. McGuinness, "SemantEco: A semantically powered modular architecture for integrating distributed environmental and ecological data," *Future Generation Computer Systems*, vol. 36, pp. 430–440, Jul. 2014.

[53] A. Zuiderwijk, M. Janssen, R. Meijer, S. Choenni, Y. Charalabidis, and K. Jeffery, "Issues and Guiding Principles for Opening Governmental Judicial Research Data," in *Electronic Government*, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, H. J. Scholl, M. Janssen, M. A. Wimmer, C. E. Moe, and L. S. Flak, Eds., vol. 7443. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 90–101.

Publication II

Herala, A., Kasurinen, J., and Vanhala, E.
Current Status and the Future Directions of Open Data: Perceptions from the Finnish Industry

In Proceedings of the 20th International Academic Mindtrek Conference
pp. 68-77, 2016

Current Status and the Future Directions of Open Data: Perceptions from the Finnish Industry

Antti Herala
Lappeenranta University of Technology
Skinnarilankatu 34
53850 Lappeenranta, Finland
antti.herala@lut.fi

Jussi Kasurinen
Lappeenranta University of Technology
Skinnarilankatu 34
53850 Lappeenranta, Finland
jussi.kasurinen@lut.f

Erno Vanhala
University of Tampere
Kalevantie 4
33100 Tampere, Finland
erno.vanhala@staff.uta.fi

ABSTRACT

Open data has been a hot topic of the decade, as even the president of the United States has given input to the deployment of open data practices. Besides politics, open data has been discussed in the scientific literature. So far the big breakthrough has not happened, although several cases have proved that the open data concept works, and provides positive results in several ways. In this article, we study and discuss the perceptions towards open data in business. Our data consisted of 45 survey responses gathered from various Finnish companies. The results indicate that companies want to use open data in different ways — ranging from application development to process efficiency — but at the moment open data in Finland is scattered and most of the companies do not find what they need. These findings are in line with the previous research executed on the field and argue that open data still needs more positive examples of application, but the potential is real, and companies consider it as a useful concept.

CCS Concepts

•**Applied computing** → **Information integration and interoperability;** *Enterprise data management;* IT governance;

Keywords

Open data, Economic value, Finnish industry

1. INTRODUCTION

Open data — data that can be used, reused, and shared by anyone — has been estimated to offer a boost to the economic and social value between three and five trillion dollars annually, as calculated by McKinsey Global in 2013 [15]. They estimated that the value would be unlocked across seven sectors, for example with 210-280 billion in finance and 300-450 billion in the healthcare business. However, unlike some other concepts to increase revenue, the data itself does not hold any value [2] since the data is shared without costs, but the added value stems from the use of data, where anyone can innovate and develop new business from the data. In order to realize this goal, the most critical requirement is that the governments need to release their data sources for its citizens and private sector, and the opened data has to satisfy specific needs in a business context. This mainly requires three aspects: high quality, high update frequency, and ease of reuse [21].

There currently exists multiple business models around open data [1]. These models allow companies to build viable businesses solely based on open data, or enhance their existing products and services with new data [22].

However, in order to create new innovations and develop new businesses around open data it is necessary to know what the companies expect from the current and future open data and what do they want to do with it.

This study aims to answer to a question "*How do companies perceive open data?*". This question is divided into sub-questions: "*How does the company size affect the actions in open data ecosystem?*", "*How do the perceptions differ within open data ecosystem?*", and "*What affects the perceived added value?*". To understand these considerations in the business, our research group conducted a survey of 45 business organizations, collecting information on themes such as interest towards data business in general and application of open data in the current business strategies. Based on our observations, the smaller organizations are more eager to adopt open data into their business, but there also seems to be confusion with the strategy that successful application of open data practices would require. In general, the business needs more success stories and positive example cases to invest in the open data practices.

Rest of the paper is structured as follows: In the next section the related research about similar studies are covered and their relevant findings are described. The third section outlines the research process: protocol, participants, and data analysis. In the fourth section, the results are summarized and discussed further with the statistical significance and implications from the hypotheses. The fifth section offers discussion about the most important findings and implications. We conclude this paper in the sixth section.

2. RELATED RESEARCH

The entire worth of the software industry is approximately

AcademicMindtrek'16, October 17 - 18, 2016, Tampere, Finland

ISBN 978-1-4503-4367-1/16/10...$15.00

DOI: http://dx.doi.org/10.1145/2994310.2994312

407 billion USD [17] if taking into account only the companies, which provide software as a product. If every area of industry, which applies some form of a software component is taken into account, there is a question of what product or service industry provider doesn't fit the bill, since almost every new product or service has some form of a software component. In this sense, the application areas of open data are not limited by the area of industry, since almost every business domain has some service, product or analysis method, which could apply open data or open data-related services. Open data has a potential to become a major component for business [21], even to the extent that it should be possible for companies to exist solely on the business based on the open data resources — as has already happened in the U.S. [22].

Gonzalez-Zapata and Heeks [7] identified four different perspectives of open government data (OGD): Bureaucratic, Technological, Political, and Economical. The Chilean case study [7] found that in the analysis of stakeholders of OGD, the private companies and economic perspective were expressed the least as stakeholders from all of the four perspectives. Their results are indicating that companies do not have any specific strategies for the use of open data. However, at the same time, the open data activists from the supply side see the fostered innovation and entrepreneurship as a sufficient economic impact, disregarding the enhancement of existing businesses.

Jaakkola et al. [11] used a survey to determine, how Finnish organizations are interested towards open data and its applications. Jaakkola et al. targeted private and public organizations with a survey in the Satakunta region, receiving 43 responses. They measured the knowledge about open data, the business opportunities, and information related to open data. The study discovered three main results: 1) companies do not have experience about open data, 2) companies can see the business potential of open data but not the business opportunities, and 3) information about open data is lacking. In their article Jaakkola et al. also determine different roles in open data ecosystem that are applied in this article. However, the concept of open data ecosystem is much grander and divided than the roles, as is presented in the literature review by Zuiderwijk et al. in 2014 [23].

3. RESEARCH PROCESS

In order to find out what perceptions companies have concerning open data, a quantitative survey was conducted in the Spring of 2015. The objective of this survey was to determine the corporate opinion towards published open data, and the current methods of publishing. The survey method for data collection was selected in order to engage multiple companies within a short period of time.

Protocol: In order to collect data from the companies, a five-minute online survey was conducted. The survey comprised sixteen multiple choice and open-ended questions and it was constructed for Finnish companies only, so the survey was conducted only in Finnish to ensure that all respondents understood the question items correctly. The survey was structured into three topics: basic information, open data in business, and future contacting. The goal of the survey was to assess, how companies currently perceive open data, how their actions affect their views, and where does the added value come from in their business domain. In addition to the main survey, the industry discussion panel was formed as a professional discussion forum and has been open for registration continuously. The registration form for industry panel contained questions about basic information concerning the companies and networks, as well as their existing interest towards data business. Both the survey instrument and the panel registration form are described in the appendix.

Participants: The survey was sent to an industry discussion panel — in which each participant voluntarily joined per their interest — that consisted of one hundred companies and nine networks of organizations. In order to conduct a business-focused analysis, the networks were left out from the analysis. From each business organization, only one submitted answer was accepted to maintain a balance between the different sized organizations and to ensure that the data does not over-represent the larger organizations of the industry panel. Out of the 100 companies, 45 individual answers were collected. 19 out of the 45 respondents (42.2%) were identified directly as software development companies. 10 respondents (22.2%) were identified as companies working directly with information technology, such as IT consultation or hardware provider. The rest of the companies represent different business domains which by themselves do not have a direct link to information technology, such as machinery and process design, industrial design or security. The companies are identified as Micro (< 10 employees), Small (< 50 employees), Medium (< 250 employees), and Large (250+ employees), according to the European Commission's definition [5]. For the survey respondents, the distribution of sizes was as follows: Micro 62.2%, Small 17.8%, Medium 8.9%, and Large 11.1%.

Data analysis: The hypotheses of this research are divided into two categories: size-related and role-related. The first three hypotheses H1-H3 are used to establish if the company size affects their actions related to the data business and open data ecosystem. Sayogo et al. [20] determined that smaller companies can benefit more from publishing open data, than the larger companies. With the first three hypotheses, the goal is to understand if similar phenomenon in the open data consumption is possible from the company point of view. The hypotheses H4, H5 and H6 take into account the roles in open data ecosystem and try to establish their impact on the perceptions about open data and data business. The article by Jaakkola et al. [11] determined that definite roles exist in the open data ecosystem: in our research, the objective is to test if these roles have an impact on the business of different companies. In the final hypothesis H7, the aim is to find out how do the interests towards specific civil sectors affect the added value from data. As stated by Manyika [15], open data can bring a different amount of economic value to different sectors. This hypothesis brings additional insight about the impact of current trends on the sources of added value of open data.

The following hypotheses were developed and assessed via the survey instrument:

- Hypothesis 1: *The target of interest towards data business is determined by human-based resources.* (Applies registration form Q10 and Q11)
- Hypothesis 2: *Size of the company determines the key source of added value.* (Registration form Q10 and Survey Q10)
- Hypothesis 3: *The availability of human resources increases actions in every role of open data ecosystem.*

(Registration form Q10 and Survey Q11)

- Hypothesis 4: *Roles in open data ecosystem defines the necessary information about open data.* (Survey Q11 and Q12)
- Hypothesis 5: *Role in open data ecosystem determines the sources of added value.* (Survey Q11 and Q10)
- Hypothesis 6: *Role in open data ecosystem defines the interest towards data business.* (Registration form Q11 and Survey Q11)
- Hypothesis 7: *Source of added value correlates directly with the interest towards a specific data topic.* (Survey Q10 and Q6)

The hypotheses H1, H2, and H3 were analyzed using Least Squares-method. The method was selected because the data fills the four recommendations offered by Gray and MacDonell [8]: multiple degrees of freedom, homogeneous dataset without drastic outliers, linear variables, and harmonious data. With the sample size of 45, the analysis cannot be used for predictivity, but it offers insight towards the trends. The hypotheses with non-linear variables, H4, H5, H6, and H7, are analyzed with descriptive analysis and their statistical significance is tested with Pearson's χ^2 test [3].

The organizations were divided into different role groups based on their open data practices. Jaakkola et al. [11] identified five roles, which were defined as follows:

- *Suppliers* are organizations that supply data through an interface or a portal.
- *Aggregators* are organizations, who collect and aggregate data for visualization and analysis.
- *Developers* create applications based on the available data to benefit consumers as well as other organizations in their ecosystem.
- *Enrichers* are actively seeking added value from open data to their existing products and/or services.
- *Enablers* facilitate the opening and usage of open data by offering hosting services, instrumentation, and consultation and advisory services.

The article also presents a sixth role: a consumer. However, the role of the consumer is usually understood as a citizen or citizens in the ecosystem, which is not in the scope of this study. Excluding the supplier and consumer, other roles can be described as intermediaries between the supplier and consumer. The roles and their interactions are presented in Figure 1.

4. RESULTS

In this section, each of the hypothesis is presented separately and reflected to the available data. Before the hypotheses, the statistical significance is addressed and the implications of these results are discussed in the last part of this section.

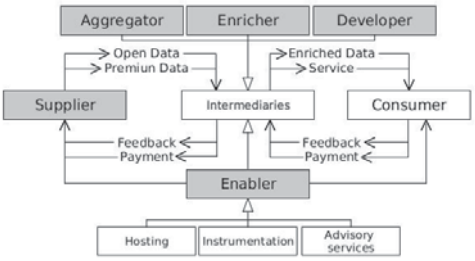

Figure 1: Roles in open data ecosystem, adapted from [11].

Table 1: Pearson's χ^2 test results

Hyp.	Degrees of freedom	χ^2 value	χ^2 critical (5%)
H4	28	9.7467	41.34
H5	16	6.4189	26.3
H6	40	4.6832	55.76
H7	72	26.8800	92.81

4.1 Statistical significance

The hypotheses from H4 to H7 were tested for statistical significance with Pearson's χ^2 test. The test measured if the measured variables correlate with each other; the tests were measured with the limiting value of 5% and the results are presented in Table 1.

Pearson's statistics is used to determine if there is a correlation between two variables: the experimental values are being compared with the critical value [9]. Greenwood and Nikulin continue that with Pearson's χ^2 test there are always two hypotheses: the hypothesis that variables correlate or the null hypothesis that the variables do not correlate. When the χ^2 value is higher than the critical value, the null hypothesis is rejected and similarly if the critical value is not exceeded, the null hypothesis is accepted. The critical value is determined by the level of significance and available degrees of freedom [18]. The level of significance was determined to be the least acceptable level of 0.05 [4]. In this study, our hypotheses were — for every separate hypothesis from H4 to H7 — that the variables in the analysis correlate. However the χ^2 value does not exceed the critical value in any set of variables, which dictates that the variables do not correlate and there cannot be any definite outcomes. In other words, our findings are not predictive, they are descriptive.

4.2 Hypotheses

In this article, we tested seven different hypotheses; the results are summarized in Table 2 and presented further in this subsection. The implications of these results are discussed after the hypotheses.

Hypothesis 1: ***The target of interest towards data business is determined by human-based resources.***

The first hypothesis was tested with the data from registration form items Q10 and Q11. The data analysis indicates that the interest towards any section of data business is not

Table 2: Summary of the hypotheses and results.

Hyp.	Result	Justification
H1	Rejected	The analysis indicates that larger companies are more interested in every section of the data business.
H2	Accepted	The analysis indicates that the smaller companies prefer application development and specialized use of open data, while the larger companies tend to lean more towards data aggregation, data usage in research and development, and working process improvements.
H3	Rejected	Smaller companies are more interested in the roles of developer and enricher than larger organizations, who prefer the roles of aggregator, enabler, and supplier.
H4	Rejected	The interest towards information about the open data is unanimous and not related to these roles in this set of organizations. The most important information are 'lists the about open data' and 'lists about the prospective dataset'.
H5	Rejected	The sources of added value are very similar between different roles and since the data does not correlate, any conclusions from small differences would be meaningless. The most potential sources of added value were data usage in an application and data aggregation.
H6	Rejected	All of the groups are interested in the same aspects of data business. The most interest is pointed towards data-based products and data enhancement.
H7	Inconclusive	All of the sources of added value lean towards specific sectors, but it does not provide enough insight to make conclusive deductions. The most popular data topics are the geographical, traffic, and IT sectors.

determined by the size of the company. The analysis, presented in Figure 2, clearly indicates that the interest towards different sections rise when the size of the company grows. Some of the data business categories are rising faster than others, but with this data sample, statistical significance cannot be found. This leads us to discard the hypothesis H1.

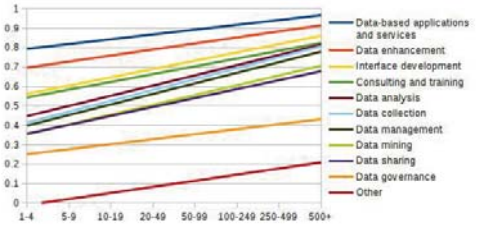

Figure 2: Least Squares-analysis on company size correlation to interest in data business.

Hypothesis 2: ***Size of the company determines the key source of added value.***

This hypothesis is tested with an item from the registration form (Q10) and an item from Survey (Q10). The analysis in Figure 3 shows that using open data in an application is far more popular for small companies than it is for the large ones. Also the category Other is more popular in smaller companies, dictating that smaller companies use open data for specific, specialized goals not listed in this survey. Based on the analysis, larger companies tend to be more interested in the research and development, data combination and intensifying their own working processes. This result leads us to accept hypothesis H2.

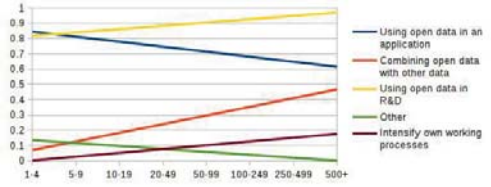

Figure 3: Least Squares-analysis on company size correlation to the source of added value.

Hypothesis 3: ***The availability of human resources increases actions in every role of open data ecosystem.***

In order to find out if the hypothesis is acceptable, questions Q10 from registration form and Q11 from the survey are used. Based on the analysis seen in Figure 4, the role of the aggregator is rising rapidly with the size of the company. Also, the role of enabler and supplier are more popular with larger organizations. Only the roles of developer and enricher — direct manipulators of data — are more popular with smaller companies, which would indicate similar results as in hypothesis H2. This result is rejecting the hypothesis H3.

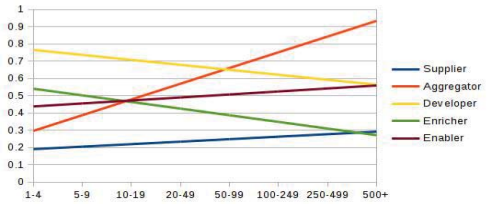

Figure 4: Company size correlation to roles in open data ecosystem analysed with Least Squares-analysis.

Hypothesis 4: Roles in open data ecosystem defines the necessary information about open data.

This hypothesis was tested with the survey items Q11 and Q12. As based on the results illustrated in the Figure 5, the different roles of the open data users correlate between the different groups. In all groups, the application types list about open data and list about prospective dataset are the two most important information sources while open data service-level agreement was the least important factor. Even if there is really no strong correlation between the roles and the usages (see Table 1), there are some minor differences: for example, the enablers are relatively most interested in the applications that apply open data sources. The data would indicate that all companies are fairly interested about two main aspects: what data is currently available and what data will be available in the future. All of the groups are unanimous in their preferences, which leads us to reject hypothesis H4.

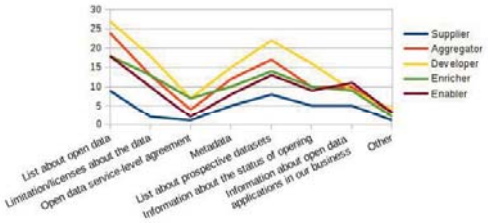

Figure 5: Necessary information about open data for different roles.

Hypothesis 5: ***Role in open data ecosystem determines the sources of added value.***

This hypothesis was tested with survey questions Q11 and Q10. It was expected that different roles would perceive the added value from open data differently, but as can be seen from Figure 6, the ratio of answers were practically unanimous. The ratio was calculated by dividing the number of answers of one role about one source of added value with the sum of all the answers in that group. There is a small rise of added value from open data in applications for the group of developers, but the rise is a lot smaller than initially expected and it takes space from the R&D, which is logical. Added value from combining open data with other data is similar to every role, while it was expected that aggregators would have more interest towards that than the rest of the groups. The differences are significantly small and since the data does not correlate based on Pearson's χ^2 test, drawing definite conclusions from the data is not feasible. However, the data would suggest that we have to reject H5.

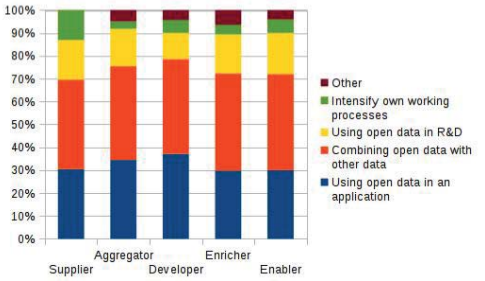

Figure 6: The source of added value for different roles.

Hypothesis 6: ***Role in open data ecosystem defines the interest towards data business.***

For this hypothesis, the questions from registration form and survey were used, registration form Q11 and survey Q11. Looking at Figure 7, it is quite evident that every company, independently from the role, is interested in the same aspects of data business. The ratios in the figure are calculated by summing all of the answers per role and using that as the divisor for the number of answers in each interest. All of the groups raise the data-based applications and services as the most interesting topic. Otherwise, the results seem more uniform, only the interest towards data governance and data collection drop from the others. While the data is not statistically significant (Table 1), the participating companies do seem to be interested in the same fields of open data and data business. This result would suggest rejecting the hypothesis H6.

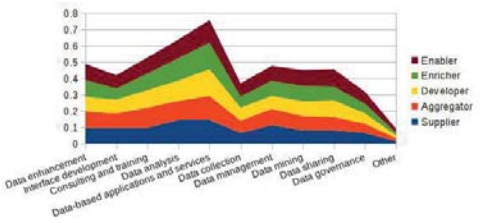

Figure 7: Roles per interest towards data business.

Hypothesis 7: ***Source of added value correlates directly with the interest towards a specific data topic.***

Figure 8 illustrates the division between the application methods with the most common application domains identified from the industry. In all of these domains minimum of ten organizations indicated that they have some business-related open data application in mind. To specify an answer for this hypothesis, survey questions Q10 and Q6 were used. The scale in the figure is the number of answers per data topic.

From Figure 8 the sectors where open data could bring in the most added value are IT and traffic sectors as well as the geographical applications. In every field, the value of combining open data exceeds the interest towards direct

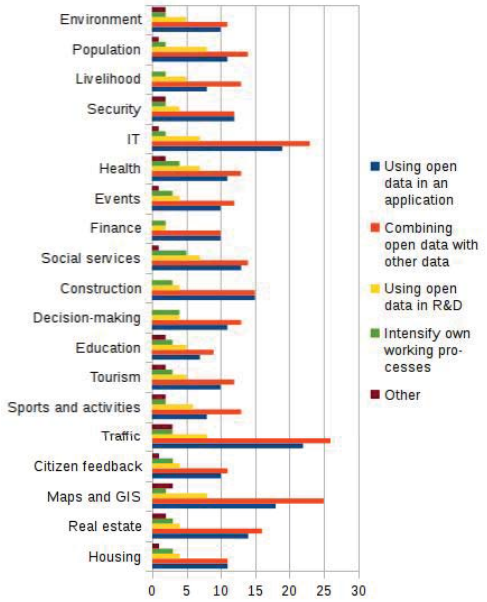

Figure 8: Added value from different data topics

applications. Only fields where the interest towards direct applications are at least on par with data aggregation are housing, construction, finance, and security.

From this data, the acceptance or rejection of the hypothesis H7 is inconclusive. On one hand, the interest peaks definitely towards few fields, but on the other hand, it includes all of the sources of added value. From the data, it is impossible to ascertain that the specific fields lean towards a specific source of added value. In a dataset that is not correlating, according to Pearson's χ^2 test, the differences are not that remarkable. However, the resulting figure does raise interesting insights on the mindset of the companies about what they want to do with the available data. Even with the differences in popularity, the sources of added value have approximately the same shape in every field. The popularity of these topics is similar to the results of the study analyzing the commercial interest towards Spanish open data domains [21], where the geographic and financial data were clearly the most popular topics, 51.1% and 46.8% respectively. In our data traffic and IT are also popular topics while the financial data is not in the spotlight. However, both datasets agree on the popularity of geographic information.

4.3 Implications of the results

As stated earlier, due to the constrained dataset, the results cannot be used to predict and present definite correlations. However, some implications can be drawn from the data and they are discussed in the following. The acceptance of H2 is not that surprising; it offers insight to the fact that smaller companies are trying to create applications based on open data; they are trying to create direct value with products. Some products and even companies can be built essentially on top of open data. Larger companies are more interested in indirect value, such as research and efficiency improvements.

In the hypothesis H3, the implications are fairly similar to H2: smaller companies prefer the role of developer and enricher, which would indicate that smaller companies are trying to benefit directly from open data and transform the freely accessible data to a new or improved product. The rapidly increasing interest towards the role of aggregator by larger firms is in line between the two hypotheses. Larger companies, who are interested in data aggregation and R&D with open data are heavily leaning towards the role of aggregator, disregarding the other roles.

H4 would indicate that what companies really need is information about the sources of open data, where it is and how can it be accessed. The problem with this is the fragmented nature of the open data release since open data in Finland is available in a governmental portal while some cities are providing their data through websites.

The analysis of H6 brings an interesting contrast to the other hypotheses, namely H2 and H3. Those two hypotheses stated that smaller companies are more interested in application development based on open data, but at the same time, H6 states that most of the companies still regard applications as the data-based core business. This shows the difference for a data business as a whole and the open data business: larger companies are regarding open data as something more than just a business resource for direct value, while smaller companies use it as a resource in order to grow.

In H7 the most selected fields were traffic, IT, and maps and GIS (Geographic Information System). Since the geographic data is practically ready for publishing and does not require any anonymization, it is an easy solution for data suppliers to share. The use is also universal since multiple different applications can be improved or innovated based on the available data about locations and other spatial data. The traffic sector has been discussed a lot in the open data community since it could bring multiple benefits to the society. Such like the spatial data, traffic data can be used to satisfy multiple aims, from which the smart city viewpoint is one of the grandest. The interest towards open data in IT sector can be used directly to improve their existing products, benefiting the existing clientele and society, and also to increase the portfolio of services. The analysis also shows the differentiation of the interest towards applications versus the combination of data, where data aggregation is more popular in any field. This would confirm that the added value of open data does not come from the data in itself but through shared value from the social and economical standpoint [13].

What strikes most from the data is the fact that while the sample is not statistically significant, the results are looking very similar, nearly identical in H4, H5, and H6. Even when the companies are all working in their own, respective fields and extending their own portfolio, this sort of uniformity seems impossible. The interest towards specific data topics is also biased towards few topics, which are easy to use. It would seem that while companies already have experience about open data and they see the possibilities of it, they still follow the directions set by data supplier. This would suggest that companies do have the basic knowledge about open data, but they are unable to follow the data suppliers

as much as would be required, so they follow the recommendations made by the publisher. This limits the reuse of open data since the interested parties cannot find the data they need. It also creates a problem, since the most preferred way of communication and learning about the open data possibilities (see Figure 9) is direct conversations with the suppliers and workshops, which would require knowledge and experience in the domain.

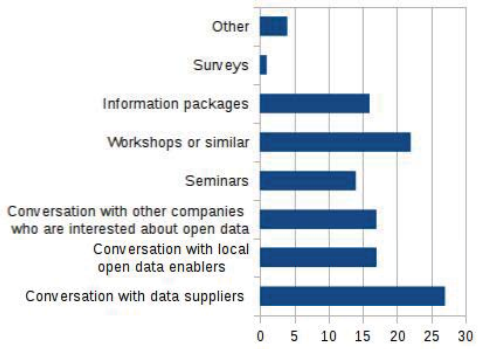

Figure 9: Preferred communications methods from the company perspective (Survey Q13)

The hypotheses of this survey discuss the different applicabilities of the open data. These observations can be summarized as follows:

- Size of a company does not change the interest towards data business, but it changes how the company sees open data. Smaller companies are directing their attentions towards open data based application development, while larger companies prefer indirect applications and benefits of open data.
- Participating in the open data ecosystem does not cause any differentiations about what knowledge is needed about open data, where the added value comes from, nor the views about data business.
- The added value in open data ecosystems origins from the open data based application development for smaller companies and data aggregation for larger firms. It would seem that in every field of society, the interest towards combining data exceeds the interest towards data-based applications.
- Companies do not have the capability to follow the data supplier as much as they are required to in order to utilize open data to the full extent. It seems that companies, in general, do not know what they want, or what is offered.

5. DISCUSSION

Based on the results, it can be argued that companies are currently following the open data trend, but it is not clear for them how and where they could apply the open data effectively. In the case of applicability of the open data, the case study by Gonzalez-Zapata and Heeks [7] found that the companies are not that interested in open data and governments are not that interested in cooperating with private firms. Our findings in this article support some of the findings, but are not as definite; the companies are interested in the possibilities of open data, but they do not know what data could be applied and when. Our data and our findings are closer to the findings of Jaakkola et al. [11]: companies see and understand the business potential in the open data activities, but they are lacking the means such as knowledge and experience to tap into the business opportunities. In addition, the companies lack the communication channels concerning the open data possibilities, as seen in Figure 9. In any case, the research done by Jaakkola concentrates on one provincial region, while our study has a national focus. Comparison between our findings and Jaakkola would suggest, that the results do not differ much between the different Finnish provinces and at least in the Finnish domain, these concerns are valid and need to be addressed.

Blame about the lack of communication can be shifted towards data supplier, who should be able to discuss with companies when they are opening their data. Some governments and municipalities nowadays are opening their data based on the reason that it has to be open [12]; they do not think about the opportunities that can emerge from the data. Therefore there should be more dialogue between data openers and users, whatever the role in the ecosystem or society. Currently, both sides are concentrating on their own data policies and they practically disregard each other in the process. In order to find new economic development from open data, the publisher should know what data is necessary to provide for a company to use it in their product or service. This works both ways: the companies who are interested in the government's data should actively research and discuss with the supplier about what they want and how. This would enable the publisher more freedom and flexibility in their processes and decision about which set of data will be opened.

These issues are not in any way new and they are being actively tackled by organizations such as OGP with other similar issues [16]. The most distressing fact is that these aspects we found in this study have been addressed for years, but still it seems to be a problem. So it begs a question: are the governments doing enough and are they doing the right decisions and actions to further the economic impact and value of open data?

5.1 Validity

First of all, in the survey the sample size of 45 organizations may seem somewhat limited. However, similarly, as in [10], the sample size is small but sufficient if analyzed correctly. In our study, the threat of overfitting the data — over-representing certain sub-groups of participants — was addressed by selecting the organizations to represent different software domains and types of organizations and only allowing one responded per organization to prevent larger organizations from over-representing themselves in the data. Also related to the number of organizations, a paper by Sackett [19] discusses the conceptualization of signal-to-noise-ratio in statistical research. Their approach to define confidence as based in practicality of observations: *confidence = (signal / noise) * square root of sample size*. In practice, this indicates that the confidence for the result

being non-random weakens if the amount of noise increases while signal decreases. In the Sackett model, the attributes are abstracted, meaning that the noise can be considered to be any uncertainty on the data. In this study, the sample population was first screened with a participation survey and then ensured with the industry discussion board that the operating domain and the intentions of the answers were understood correctly. Since the concept of the Sackett study is that the confidence in the survey data increases the validity of the study, our study addressed this problem by screening the sample for wanted types of organizations. Therefore it can be argued that our signal was very good and noise low, so the overall confidence should be good. Finally, the Pearson's χ^2 test is considered a reliable tool for assessing goodness of fit between two groups (e.g. [3]).

Considering the technical structure of the survey, the responses were collected with established methods, for example by applying the 'like best' (LB) technique with roles established in the prior research, which is common survey data collection strategy (e.g. [6]). In addition, Kitchenham et al. [14] divides comparable survey studies into exploratory studies from which only weak conclusions can be drawn, and confirmatory studies from which strong conclusions can be drawn. This study is an exploratory, observational, and cross-sectional study that explores the phenomenon of applying open data and the expectations of the open data in practice, and provides more information and understanding to both researchers and practitioners to refine their future work into the topic.

The validity can be also questioned with the biased size of participating organizations as well as their individual fields of business. Only 11.1 percent of the respondents were large organizations, while the rest of them were SMEs. This issue of validity was noticed by Verhulst and Caplan, where the smaller organizations were recognized as the main beneficiary because of the equalizing effects of open data. The positive effects of open data are most likely felt by smaller organizations first because they have limited access to data, information, and analytical tools, a trait not shared with larger companies [22]. 42.2 percent of the survey participants are doing business directly in the software development field and 22.2 percent of the companies are doing business in an area closely related to information technology. It causes a bias towards the software development and software business in general. However, this threat to validity would be larger, if the companies were randomly selected and non-screened since the amount of irrelevant noise would also increase. For this study, only companies which are currently doing business with the open data or within the open data ecosystem are participating in the survey since they have tangible viewpoints.

6. CONCLUSION

In this paper, we have presented our results on a national survey regarding the application and adoption practices of open data in Finland. The survey collected 45 responses from several different business domains, such as construction, software development, tourism and even sports event organizers. In general, the results confirmed our prior expectations knowledge based on the prior research and another regional survey on the application of open data in Finland.

Across the different industry domains and company sizes, the survey implicates similar results: the application of open data offers potential for economic value in several sectors, but so far the companies have had difficulties finding a suitable strategy for implementing open data resources. The most important observation was that the companies cannot find the open data they need for their business, even though they are willing to use the data. The scattered supply of open data requires more resources from companies than they want to deploy. In general, the results were in line with the earlier regional studies, and confirm that the observations and recommended actions for promoting open data initiative needs to address these issues on the national level.

In the future, the results of this study can be used to guide the decision-makers and open data suppliers to further optimize their strategies of publishing data. One possible line of inquiry would be to focus on determining the possibilities and most promising venues, which could be used to supply open data more efficiently to citizens and businesses with reasonable costs to the publisher.

7. ACKNOWLEDGMENTS

We are grateful to 6Aika-project for their cooperation. We would also like to thank DIMECC S4Fleet-project and the network of companies associated with it.

8. REFERENCES

[1] F. Ahmadi Zeleti, A. Ojo, and E. Curry. Exploring the economic value of open government data. *Government Information Quarterly*, Feb. 2016.

[2] P. Booth, P. Gaskell, and C. Hughes. The economics of data: quality, value & exchange in web observatories. pages 1309–1316. ACM Press, 2013.

[3] H. Chernoff and E. L. Lehmann. The Use of Maximum Likelihood Estimates in χ^2 Tests for Goodness of Fit. *The Annals of Mathematical Statistics*, 25(3):579–586, Sept. 1954.

[4] M. Cowles and C. Davis. On the origins of the. 05 level of statistical significance. *American Psychologist*, 37(5):553, 1982.

[5] European Commission. What is an SME? - European Commission.

[6] A. Fink and J. Kosecoff. *How to Conduct Surveys: A Step by Step Guide*. Sage Publications, Inc. California, 1985.

[7] F. Gonzalez-Zapata and R. Heeks. The multiple meanings of open government data: Understanding different stakeholders and their perspectives. *Government Information Quarterly*, 32(4):441–452, Oct. 2015.

[8] A. R. Gray and S. G. MacDonell. A comparison of techniques for developing predictive models of software metrics. *Information and Software Technology*, 39(6):425–437, Jan. 1997.

[9] P. E. Greenwood and M. S. Nikulin. *A guide to chi-squared testing*, volume 280. John Wiley & Sons, 1996.

[10] J. Iivari. Why are CASE tools not used? *Communications of the ACM*, 39(10):94–103, Oct. 1996.

[11] H. Jaakkola, T. Mäkinen, and A. Eteläaho. Open Data: Opportunities and Challenges. In *Proceedings of*

the 15th International Conference on Computer Systems and Technologies, CompSysTech '14, pages 25–39, New York, NY, USA, 2014. ACM.

[12] M. Janssen, Y. Charalabidis, and A. Zuiderwijk. Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, 29(4):258–268, Sept. 2012.

[13] T. Jetzek, M. Avital, and N. Bjorn-Andersen. Generating Value from Open Government Data. In *Proceedings of the 34th International Conference on Information Systems. ICIS 2013*, Atlanta, GA, 2013. Association for Information Systems. AIS Electronic Library (AISeL).

[14] B. Kitchenham, S. Pfleeger, L. Pickard, P. Jones D. Hoaglin, K. El Emam, and J. Rosenberg. Preliminary guidelines for empirical research in software engineering. *IEEE Transactions on Software Engineering*, 28(8):721–734, Aug. 2002.

[15] J. Manyika. *Open data: Unlocking innovation and performance with liquid information*. McKinsey, 2013.

[16] Open Government Partnership. From commitment to action — Annual Report 2015. Technical report, 2015.

[17] C. Pettey. Gartner Says Worldwide Software Market Grew 4.8 Percent in 2013, 2014.

[18] R. L. Plackett. Karl Pearson and the Chi-Squared Test. *International Statistical Review / Revue Internationale de Statistique*, 51(1):59, Apr. 1983.

[19] D. L. Sackett. Why randomized controlled trials fail but needn't: 2. Failure to employ physiological statistics, or the only formula a clinician-trialist is ever likely to need (or understand!). *Canadian Medical Association Journal*, 165(9):1226–1237, 2001.

[20] D. S. Sayogo, J. Zhang, T. A. Pardo, G. K. Tayi, J. Hrdinova, D. F. Andersen, and L. F. Luna-Reyes. Going Beyond Open Data: Challenges and Motivations for Smart Disclosure in Ethical Consumption. *Journal of Theoretical and Applied Electronic Commerce Research*, 9(2):1–16, 2014.

[21] D. Tinholt. The Open Data Economy: Unlocking Economic Value by Opening Government and Public Data. *Capgemini Consulting Analysis*, 2013.

[22] S. Verhulst and R. Caplan. Open data: A twenty-first century asset for small and medium-sized enterprises. Technical report, 2015.

[23] A. Zuiderwijk, M. Janssen, and C. Davis. Innovation with open data: Essential elements of open data ecosystems. *Information Polity*, 19(1, 2):17–33, 2014.

APPENDIX

Survey

Background information

1. I answer to this survey as:
   (a) Company representative/Network representative
2. First name
3. Last name
4. Name of the company/network
5. City

Open data in business

1. (Q6) Which data or programming interfaces are the most interesting to you from the viewpoint of your company's/network's business?
   (a) Housing / Administration / Real estate / Maps and GIS / Citizen feedback / Culture / Law and regulation / Traffic / Sports and activities / Tourism / Counselling / Education / Decision-making / Construction / Social services / Finance / Events / Health / IT / Security / Livelihood / Population / Environment / Other
2. (Q7) Does your company/network have services or products in use or in design, where you apply open data or open interfaces?
   (a) Yes / No
3. (Q8) If yes, what kind of product/service and what data/interface it is using or could use?
4. (Q9) Do you think that using open data could bring added value to your company/network in the future?
   (a) Yes / No
5. (Q10) If yes, where does the added value come from?
   (a) Using open data in an application / Combining open data with other data / Using open data in R&D / Intensify own working processes / Other
6. (Q11) What are the roles of your company/network in open data ecosystem?
   (a) Supplier (provides data) / Aggregator (collects data from different sources and visualizes it) / Developer (develops applications based on the data) / Enricher (combines open data with their own data) / Enabler (raises awareness and interest towards open data) / Other
7. (Q12) What are the most important sets of information for your company/network that the cities should provide about open interfaces and data (top 1-3 most important)?
   (a) List about open data / Limitations or licenses about the data / Open data service-level agreement / Metadata / List about prospective datasets / Information about the status of opening / Information about open data applications in our business / Other
8. (Q13) What are the preferred methods of communication your company/network would be most interested about?
   (a) Conversation with data suppliers / Conversation with local open data enablers / Conversation with other companies who are interested about open data / Seminars / Workshops or similar / Information packages / Surveys / Other
9. Comments about the industry discussion panel?

Future communications

1. I want to be contacted by a representative about utilizing open data.
2. Which city representative you want to have communications with?
   (a) Helsinki / Espoo / Vantaa / Tampere / Turku / Oulu

Registration form (for companies)

Contact information

1. Name
2. Appellation
3. Phone number
4. E-mail address
5. Organization
   (a) Company/Network

Company information

1. Company name
2. Company website
3. Business ID
4. Office locations
5. (Q10) Company size
   (a) 1-4 / 5-9 / 10-19 / 20-49 / 50-99 / 100-249 / 250-499 / 500+
6. (Q11) Interest towards data business
   (a) Data-based applications and services / Data enhancement / Interface development / Consulting and training / Data analysis / Data collection / Data management / Data mining / Data sharing / Data governance / Other
7. Expectations from the panel

Publication III

Herala, A., Kasurinen, J., and Vanhala, E.
Views on Open Data Business from Software Development Companies

Journal of Theoretical and Applied Electronic Commerce Research
Vol. 13(1), pp. 91-105, 2018

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 13 / ISSUE 1 / JANUARY 2018 / 91-105

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762018000100106

Views on Open Data Business from Software Development Companies

Antti Herala[1], Jussi Kasurinen[2], and Erno Vanhala[3]

[1] Lappeenranta University of Technology, School of Business and Management, Lappeenranta, Finland, antti.herala@lut.fi
[2] Southern-Eastern Finland University of Applied Sciences, Department of Digital Economies, Kotka, Finland, jussi.kasurinen@xamk.fi
[3] University of Tampere, Department of IT Administration, Tampere, Finland, erno.vanhala@staff.uta.fi

Received 15 March 2017; received in revised form 14 July 2017; accepted 25 August 2017

Abstract

The interest towards the concept of open data has increased during the last ten years, as governments and municipalities have decided to open their data repositories. This has led to a new generation of mobile apps, which utilize this data to improve the feature richness and the overall user experience for the customers. In this study, we interviewed representatives of five software organization and discussed their views towards opening data - private and public - and also using the open data in practice. Based on our observations, the companies see very limited scope for the use of open data as a business asset: the main applications seem to gravitate towards function as an additional feature for an existing product, not a source of new innovations or business ventures. The results also illustrate on how little benefit the organizations consider to gain from opening their private data, and what alternatives there are for sharing data in a profitable manner. Additionally, as based on the observations, a strategy classification on the different data sharing methods is formulated and presented.

Keywords: Open data, Data business, Data management, Software development, Private data

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 13 / ISSUE 1 / JANUARY 2018 / 91-105

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762018000100106

1 Introduction

The main concept of open data and its application is simple; access to the publicly-funded data provides greater returns from the public investment and can generate wealth through the downstream use of outputs, such as traffic information or weather forecast services [11]. However, even though open data and data sharing as concepts are forty years old with the open data initiative reaching ten, the practical actions and applications have tended to stay on the superficial level, (e.g. [29]), and significant progress or success stories are hard to find. The current trend is that the governments and municipalities are opening their data, but the impact and usefulness of raw open data repositories to citizens - and even to businesses - can be questioned [7]. Besides the governments, a handful of private organizations are opening their data in an attempt to unlock the economic value of open data [10], but even they have difficulties finding innovative usage, let alone generate additional profit [14].

In a previous study [8] it was found that companies are interested in open data and that this mindset spans over different industries, from both publicly available data to the private business-to-business data access. Open data is not only a resource for software companies, but also for traditional engineering industries and even for small, non-franchised local markets and shops. In our previous study, it was established that there is evidence [9] on recognizing the applicability of open data, and opening the data to the clients by private organizations leads to business opportunities, creating new value. However, while there is interest towards open data in a wide variety of businesses, the question still remains whether or not open data is actually used to generate income or are there some other sharing methods in use that are more efficient and more profitable.

For this study, four research questions were formulated. The first three are concentrating on the usage of open data as well as the interest towards opening or sharing data and the fourth research question revolves around the different types of openness:

- How do new clients express interest towards open data?
- What kind of open data-based solutions is the existing clientele expecting?
- How does the product portfolio of a software company respond to open data?
- What are the current trends of open initiatives?

To gain insight into these aspects, we conducted a qualitative interview study with five software organizations applying different strategies for opening and sharing data between organizations. In addition, the interest towards open data in different areas of expertise was measured with a quantitative survey, to establish the initial view. In this study, we introduce our findings and discuss the open data practices applied in the industries. The goal of this research is not to represent the software industry as a whole but to highlight some ideas why open data has not been used by professional software developers.

The rest of the paper is structured as follows: Chapter two discusses the related research works on open data and how it relates to business. The third chapter further explains the methods applied in this research, from data collection to analysis as well as the organizations for the interview. In chapter four the results are presented and further discussed in the fifth chapter. The study is concluded in chapter six.

2 Related Research

Open data is defined as *data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike* [21]. Open data is something that is usually published by public organizations, such as governments and municipalities, referred as Open Government Data or OGD [18]. There are also other publically funded bodies, such as World Bank and United Nations that are not linked to only one country but are still funded by taxpayers and provide a source of open information and datasets. Another-and much anticipated- venue of published open data are the privately funded organizations, companies that are not funded through taxes, which could find sources of profit, innovation, and revenue through publishing open data. These commercial open data publishers are a rare occurrence in practice due to various reasons [10]. Lindman et al. [14] argue that open data publishing still lacks practical revenue models, which can prevent players entering the field. Contradicting this, Zeleti et al. [34] present multiple business models for open data publishers, which are viable and applied in practice. While the field may or may not lack practical examples and literature gravitating towards the lack of practices, a systematic literature review into the matter revealed multiple opportunities, how the private sector can profit from opening their data [9].

Open data has been referred to have a significant value in the economic sense, especially for small and medium-sized enterprises [31]. While the economic impact of opening data has not yet emerged, open data is treated as a valuable business resource especially because of its characterization of being free to use - even commercially [6].

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 13 / ISSUE 1 / JANUARY 2018 / 91-105

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762018000100106

There has been a definite interest towards open data as a business resource from various points of view [10]. However, even the usage of open data has not found its way into the estimate of economic value. For instance, the study of Gonzales-Zapata and Heeks [5] into Chilean open data environment and stakeholders recognized the economic perspective of open data as the least regarded, behind political, technological, and bureaucratic perspectives. While there are only a few scientific pieces of evidence about the economic impact of open data, the study of Lindman et al. [14] found, that companies utilizing-or wanting to utilize- open data tend to lean towards data analysis and data-based applications, a finding which was also supported by Herala et al. [8]. Additionally, the study by Zuiderwijk et al. [36] outlines propositions for a company, for the creation of value from open data.

To further engage the business opportunities of open data, there already exists a market for open source business. Lindman [13] studied open source and open data business back to back in Finnish scope in order to detect similarities between them. In the research, it was found that the open data and open source businesses are rather identical when it comes to community management and developer motivation, and open data practitioners could draw valuable lessons from open source communities and research.

Open data is not the only method to share data (Figure 1). It is considered even as a rare method for data sharing since open data initiative is a rather unorthodox method in business systems. While the data sharing is usually done between two actors, or in some cases between multiple actors, open data can be seen as an omnidirectional data sharing technique, where the data is shared to everyone at the same time. Before open data, there have been connections between companies in terms of sharing resources or data in supply chains [15], [29] or even in larger consortiums and ecosystems [27]. This is referred to as business-to-business (B2B) sharing.

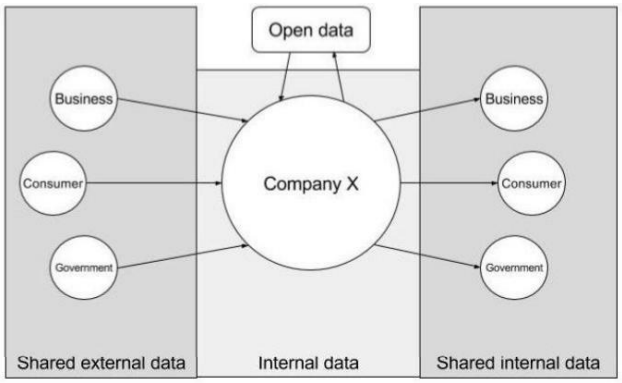

Figure 1: Available external data (L) and internal data (R), based on [15], [25], [28]

In addition to B2B context, there are also companies who have their transactions with consumers instead of businesses. In this environment, transactions were easier and faster when a customer bought an item from a store. Nowadays the store is an online store and the transaction leaves a digital imprint on the system and the company knows automatically who the client is and what they have bought. Creating analysis from this can bring value to the owner of the online store. In a sense, consumers are sharing their data with the company as consumer-to-business sharing (C2B). The environment has started to change since the data ownership is being challenged and consumers are demanding this data to themselves [17], [24], and some companies are even starting to share this data (e.g. [28]) creating an option for business-to-consumer (B2C) sharing.

While the transactions between businesses and consumers exist, governments also have roles in data sharing, especially with sensitive information about citizens [22]. Governments can supply data to businesses but they can also request data from a company, ensuring a flow of data in both directions [25], which are usually referred as G2B and B2G. The data sharing between these instances can be beneficial to both parties [3], [32].

To summarize, a company may receive data from another company, from their customers and also from governments, and it is possible to use open data as a resource. A company may then share their data with other companies and governments, or create services to share their data with their consumers or in extreme cases, publish their data as open. While the actual methods of data sharing may be more complex, for example having multiple actors in a transfer of just one set of data, these methods are simplified in order to illustrate the different strategies in the field of collaborative data management. All of the different methods of data sharing are on some level discussed in this study. The most weight is given to any open data related solution that may be presented during the interviews, but the scope of other related methods is also held relevant.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 13 / ISSUE 1 / JANUARY 2018 / 91-105

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762018000100106

3 Methods

In this study, the focus was directed towards software companies, while not limiting their amount of resources or the scope of operations, in order to determine viewpoints from organizations with different clientele. In the study population, all of the organizations were privately owned software providers, with software or service development being their main source of revenue. The focus was placed on software companies because, in our previous study [8], the voluntarily participating set of survey respondents consisted of 45 companies, where 19 of them were directly involved in software development. This suggested a popularity of open data in software development and it was decided to examine closer, what is going on in the industry regarding open data, especially considering the similarities between open source and open data activities [13].

3.1 Data Collection

For the quantitative data in this study, we used the survey data already gathered for the previous study. That set of data was collected during spring 2015 and it was a five-minute online survey to an industry discussion panel, consisting voluntary participants interested in open data. The participating companies were also asked to register for the panel and a separate set of data was collected from that registration form. From these companies, 19 out of 45 respondents were doing business in software development. The rest acted in other areas of business, varying from consultation and hardware providers to coffee shops and designers. This set of data was used to gain the initial view, how the software business-where open data is supposed to be more popular- varies from other areas of business.

In addition to this survey data, a set of qualitative data was collected via interviews (Figure 2). This *zoom in* was applied to gain more insight into the observations made on the statistical data; the sample selection was a polar sample, collecting different types and sizes of volunteering organizations from the original survey. Overall, five interview sessions were held during the latter half of 2016. The initial strategy for the population criteria and selection were based on our prior experiences on conducting industry-wide studies on software industry in general, made by our research group (for example [12]) and on our prior knowledge concerning the application of open data concepts, which were reported [8]. Our sample strategy included the application of polar examples of different operating domains and company sizes along with different viewpoints into the open data and the software industry to gain a wide perspective on the open data practices. The sample of the interview rounds consisted of three development organizations selected from our research partners and supplemented with additional volunteering organizations to achieve a heterogeneous group of different organization sizes, maturities, and operating domains.

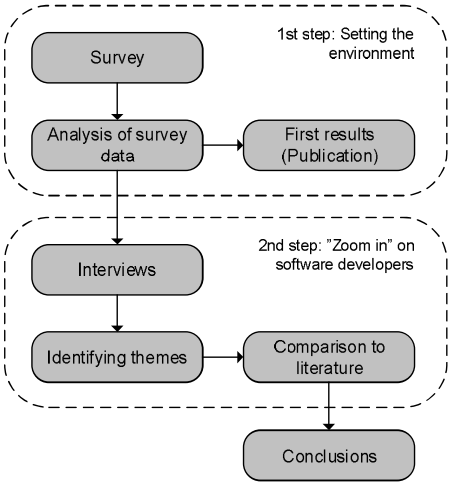

Figure 2: Data collection and research process

The five organizations in the study group were professional software developers, which were either recent technology adopters or companies already developing open data solutions, as identified by the previous survey study. The organizations varied (Table 1) from small software developers working on healthcare, to business-to-business solution services and service digitalization. The smallest organization in the focus group was a software company with three employees; the largest organization employed several hundred people that contributed to the

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 13 / ISSUE 1 / JANUARY 2018 / 91-105

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762018000100106

product and service development or deployment. All of the participating organizations were commercial, privately owned, companies.

Table 1: Participating organizations, their size, and their business domains

Company	Size	Business domain
Case A	Less than 10 employees	Digitalization for healthcare
Case B	Less than 10 employees	Data provision
Case C	200-500 employees	Software production and life-cycle management
Case D	100-200 employees	Digital services
Case E	200-500 employees	Digital business / digitalization

The objective of this approach was to gain a broader understanding of the practice of opening data and to identify the general factors that affect the open data solutions these organizations apply and provide for their customers. To achieve this, our research team developed a questionnaire based on four research questions. At first, the participants were asked about the 1) new clientele and their interest towards open data, with angles from usage to sharing. While in the topic of their clientele, the following questions regarded 2) the existing clientele about the same matters. The third set of questions was about 3) the software company itself; have they reacted to the usage or sharing of open data through their portfolio or any other form of sharing data. The final part of the interview was about 4) the different trends of openness as well as the company's activity with current trends. When constructing the interview, it was anticipated that not all companies use-or especially share- open data, so there was also a number of questions about different sorts of data sharing strategies to identify different methods to share data instead of open data. These questions revolve around the business to business (B2B) sharing as well as business to customer (B2C) sharing, since business data or customer data are private data protected by legislation. The reason to include these questions was to validate if the software company has knowledge and expertise in any form of data sharing, which then affects the impact of their responses to the results.

The interviews were constructed as semi-structured interviews with a list of questions (Site 1), and the whole sessions were tape-recorded for qualitative analysis. Typically, an interview lasted for approximately half an hour and they were arranged as face-to-face interviews with one or two organization participant and one or two researchers or conducted via video call over the Internet. During the research, we collected approximately 134 minutes of interview data for further qualitative analysis. The interviews were organized as semi-structured, where the interviewer was mainly listening to the interviewee and interrupted the flow only by keeping the direction of the discussion within the research questions. The pre-determined questions served as a tool for the researcher to make sure that all of the topics had been discussed.

The decision on who to interview mainly came from the companies. However, the researchers set requirements on what qualifications they should have and what they should know about the topic. From the small and medium-sized enterprises, the interviewee was the chief executive officer of the company and from large companies, there was an interviewee from middle management, usually a project manager managing one or multiple projects directly or indirectly. This was based on our aim to gain a better understanding of the operational level of software development and their considerations of different general themes related to their business activities. It was also necessary to recognize if our findings in the previous study could be validated with these organizations.

The collected qualitative interview data was classified and codified following the principles of the open coding method from Grounded Theory [4], [30]. The open coding and case analysis were done to collect observations and identify repeated themes from the data; the number of observations did not warrant a full Grounded Theory analysis but used to pinpoint major themes and understand the different observations.

4 Results

The results are collected in two steps: first, the initial survey is analyzed and its conjectures to the environment of open data explained from the quantitative data. The second step, the qualitative data and the results from the interviews are presented, with the conjectures towards opening data and the usage of open data, based on the research questions.

4.1 Interest Towards Open Data

The first insights were gained through the survey responses, which are not an integral part of this research but serve as the first step of the research method to provide insight into the environment. 19 companies-out of 45- were software development companies and the rest of them range from consultation and hardware providers to electrical and mechanical process designers. Before the results from the interviews, the survey responses are compared in terms of added value from open data (Figure 3) and interest towards open data (Figure 4).

When the companies were asked *Do you think that using open data could bring added value to your company/network in the future?* and *If yes, where does the added value come from?*, they selected either using open

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 13 / ISSUE 1 / JANUARY 2018 / 91-105

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762018000100106

data in an application or combining sets of data (Figure 3). For instance, out of 19 software development companies, 16 (84%) selected *Using open data in an application*. The ratio in the figure is calculated by dividing the number of selections with the total number of responses by respective fields.

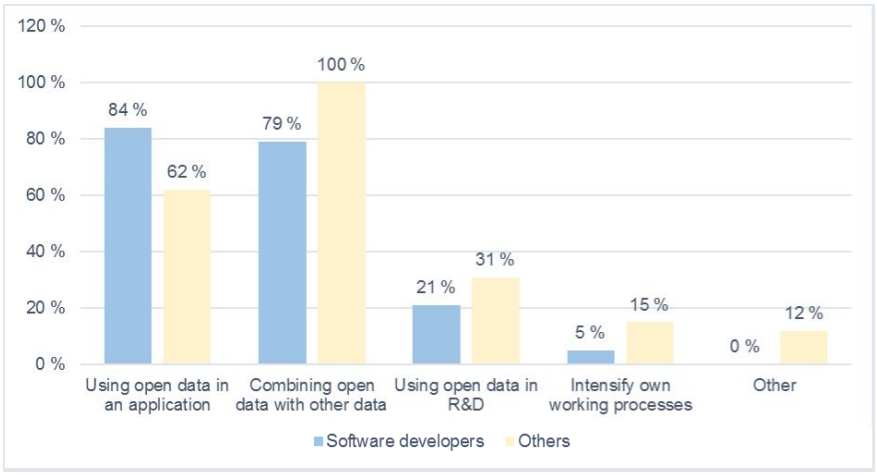

Figure 3: The added value of open data for software developers and other areas of business

The research data shows, that the perceived added value of open data by software companies exceeds the perceptions only in application development and the other fields exceed software development the most in data combination. This is in line with our prior observations [8], where the most common types of open data application were similar.

The other measured entity was collected from the industry discussion panel registration form, where the participants were asked their interest towards data business (Figure 4). The ratios are calculated in the same manner as in the previous figure, dividing the number of selections with the total number of responses from the respective fields of industry.

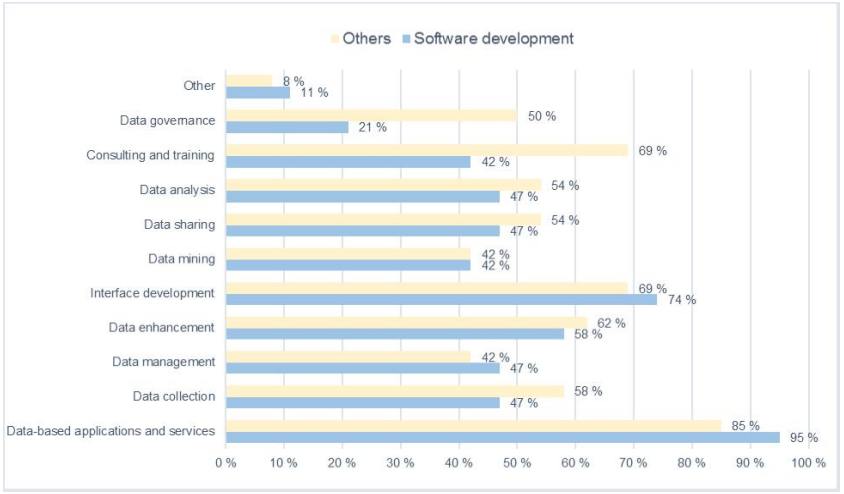

Figure 4: Company interest towards different segments of data business

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 13 / ISSUE 1 / JANUARY 2018 / 91-105

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762018000100106

These result support each other; when comparing to the other fields of industry, the software development companies seem to be more interested in data-based application development, data management, and interface development. This is understandable since these activities also represent their core business activities.

4.2 Open Data in Practice

The second step of the research method, the qualitative interviews with five companies, were used to assess the software company aspects further, with the observations from analysis divided into three different categories (Table 2). The *Using open data* summarizes the viewpoints of software developers and also their clients on whether or not open data is used or will be used in their business. The *Opening data for value* presents the willingness to open data in order to create business or other activities with that data. The final category *Trends of openness* implies the current methods on how to be open in the software enterprise. This focuses towards the actions of software companies, but also the insights they have, or may have, about other industries. The first two categories described here are addressing research questions 1-3, first from the side of open data usage and then from the viewpoint of opening data. The third category offers views for research question 4 only. The questions in the interviews are linked to the research questions (Site 1), available online.

Table 2: Summarization of the most important observations

Company	Fields of clients in scope of study	Using open data		Opening data for value	Trends of openness
		Client	Developer		
Case A	Mainly public healthcare sector	No	No	No	Open source
Case B	Media research and market monitoring	Rare	Yes	No	Open data
Case C	Media, retail, communications, manufacturing	Rare	No	No	Open source
Case D	Private sector	Rare	No	Rare (public clients)	Open source, open API, open governance
Case E	Public sector	No	Yes	Yes (public clients)	Open source, open API

From the interviews, a total of five conjectures were found through the qualitative theme identification described in the third section. These conjectures emerged from the qualitative data and they could serve as a base for future research in this field. The first two are critical towards the agenda of open data initiative, using open data and opening data are not considered as viable methods for profit, while some positive considerations were noted. In general, software companies in this study are interested in open data, if some data, that can be found applicable to their cases, is found. The trends of openness also lean towards open source development and open APIs instead of open data, and even data management has other methods to share data instead of opening it.

4.2.1 Open Data is not Considered as a Key Business Asset in Industries

Contrary to the findings from the survey, through the interview, there was a theme that open data is not considered as a key business asset by software developers or their clients. In the more general view, open data is seen as a difficult asset to use for a business. Companies do have knowledge about open data, but there has not been a case, where open data would have been used as the sole base for a business.

Open data projects are very challenging in a company sense since it is difficult to create business out of it. - Case A
If anyone is doing actual business with open data, I would assume that all [of our client's open data projects] have been regulative needs or marketing stunts. - Case D.

All of the companies had the same view, that open data cannot be used as the only source of a standalone business, but a few cases were reported, where open or accessible data was used as a basis. An example was offered from Case C, where the data was used for a start-up, so there would be some data in their product before deployment. However, this open data has been only one part of the database and scarcely used after new data was collected. Other companies did also have some examples from public projects or from data aggregation, enhancing a client's product.

We have used some library data and at least some events data from public application programming interface (API) for a customer [...]. When they started a start-up from nothing, we were able to get real data from the beginning and then added to it from other sources. - Case C.

We are providing data from social media as a part of [other company's] tool. - Case B.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 13 / ISSUE 1 / JANUARY 2018 / 91-105

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762018000100106

One of the main issues that was mentioned in the interviews was the quality of data. Not necessarily just for open data but for data usage in general. Currently, there are issues of data quality in internal data-even one source of data- and open data may not be compatible or even in the same context as would be needed.

The problem is that when the client's service uses some kind of data and [in the external data] there might be mismatches and shortages. - Case C.

Companies are-at least for now- still wrestling with the application of internal data, so open data and public data, or external data, in general, may be used only in the future. - Case B.

While there are issues with external data, it is possible to take some data from external sources and many software companies do it already when requested. These are usually done because a client is requesting a specific set of data.

We would want that as many clients as possible would use the data we already have, but we do listen very carefully if a client or partner says, that they need some kind of data. - Case B.
Usually we get some propositions from clients, that there are these kinds of sources and how much it would cost to integrate this data with their service. And then we have started to analyze, what it is. - Case C.

Open data in the sense of accessibility is a special case since the data is distributed to anyone and it usually comes from a public source. There are still issues since some open data is published as raw data through a website and the focus is more of getting the data open instead of accessible.

Instead of open data, the main concern should be to get open APIs-no matter if the data is open or not- but in order to create something, APIs are necessary. - Case E.

And even if the data is open and accessible through a sensible API, there still can be legal and organizational issues that make the data inaccessible.

And when there are these open APIs, it still might take from four to five months of negotiations to use it. - Case E

4.2.2 Opening Data is not Done for Business Initiatives

Open data and especially opening data is an initiative that targets mainly governments and their funded institutions. This was evident from the interviews, where companies targeting the private market tended to lean more towards privatization of data instead of opening it. But even in the more reluctant cases, there had been some evidence of opening or sharing data, either as open data or through a pilot, such as a hackathon or even through a private API.

We have organized a hackathon with [a private organization], that based on their data, also the same with [their partner]. They have been our clients and wanted us to help them organize something to get developers interested in their organization and the data they have. - Case D.

There has been sort of API-platform building, where someone could build a service that uses [the client's] data. - Case C.

Through the interviews, there were some examples on what kinds of companies are currently opening data and the reasons behind them. From Case D there were mentions of media companies, banks, and electricity producers who are either fulfilling regulatory requirements or publicity stunts.

There are some media companies, who probably do not have obligations to open anything but they want to do so because of publicity. - Case D.

We've had requests [from publically funded companies], that some data has to be opened and can you design an API for us based on these requirements. - Case D.

In Case E, the focus is directed towards public organizations, such as municipalities, who are also opening data because of regulations. And so, mostly data is being opened because it belongs to the public strategy.

There exist this INSPIRE directive that we have used and are using for developing [this tool]. - Case E.

When discussing open data, public organizations are more in favor of opening data than private organizations. A common theme from the interviewees was, that the current open data is something that is easy to open, which is not business critical and does not contain information about individuals or organizations.

If it's something like [the location of] the public trash cans that are not so business critical, that sort of data is being opened and published. - Case E.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 13 / ISSUE 1 / JANUARY 2018 / 91-105

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762018000100106

Since its initiation in 2009, open data initiative has been developed and it is being adopted by governments around the world. However, one aspect that was found from the interviews is that while a governmental legislation would be in favor of open data, it is not done on ground level as effectively as possible. There were issues of availability, quality, and context that were raised in the interviews.

We are more the cause than the effect of what [data] is opened [in municipalities]. - Case E.

While this has been recognized in non-critical data, such as geolocation, open data is an even more difficult topic in critical sets of data, for example, individual's data and health care.

Often in health care, there is ignorance, a great ignorance towards information technology and suspicion towards everything. [...] And because of it, this kind of data publishing and open data sounds even more terrifying. - Case A.

An issue about open data that many critics often raise is the lack of profit from the data. Currently, some companies and even municipalities are gaining profit from data by selling data sets and access to data. From the interviews, it became clear, that the ownership is a big issue and also the lack of profitable revenue streams for the owner of data.

What increases the control over data is the fear, that they [data owner] will lose the profit from it. -Case E.

The ownership and control of data is an issue, which is highlighted in private organizations. This was evident in Case B, where the interviewee engages in combinative data analytics between public data, open data, and organization's data. Also in Case A, it was mentioned that they rarely see any data from their clients and are forced to test their systems solely on generated dummy test data.

We get data from our clients if agreed on a contract, but usually we do not take the data but are trying to integrate everything into their systems. - Case B.

When we do the systems, we generate the data for ourselves and use that, so we do not have access to patient information or data whatsoever. - Case A.

4.2.3 Open Data is Considered as a Potential Added Value for Software Products

Similarly, what was noticed from the survey, software companies are interested in open data at least on a hypothetical level. Some of the companies-especially the larger ones- already have examples and cases, where open data has been used as a part of their product.

If there is an example… basically [open data is used] to support decision making. - Case D.

It's mostly geographic information that we use [in this case]. - Case E.

This has also been a case in smaller, specialized companies, in this study Case B, where open data is used on some level in their day-to-day business.

Our idea is to collect [open and public] data once and then partner up with multiple stakeholders, and we'd like that as many as possible would use our existing sets of data. - Case B.

In other cases, such as Case A and Case C, open and public data has been thought of and discussed to be a part of future solutions they are creating.

We do have plans to use some open or public data in the future. [...] It would be like something that helps our customers to engage services easier, or something like this has been a part of our conversations. - Case A.

We are currently planning some IoT-projects with our clients and it would be possible or necessary to use some external data in those. - Case C.

4.2.4 Open Data is not a Popular Open Initiative Trend

When the software developers were asked about trends of openness, open source code was the first among responses. Open source is seen as an important trend in software development and on some level, it is even demanded by the clients.

In the past 5-10 years the use of open source code has become much easier, there are no fears directed towards it that were felt before. - Case C.

It [the project] was specified to be executed with the open source principle. - Case E.

Another point that was raised by the interviewees are open interfaces (APIs) and also semi-open APIs, that can be connected easily with permission. It seems that there are varying definitions of open APIs in the industry, from open-

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 13 / ISSUE 1 / JANUARY 2018 / 91-105

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762018000100106

for-all APIs to following-open-standards APIs. In domains which handle critical data, such as health care, the movement of data between actors is important. However, there have been issues with the APIs and access to them in those scenarios.

What I have understood that open APIs are something that should already exist [in institutions] and there are some health care institutions have negotiated access [to their APIs]. - Case A.

There has been some more discussion about APIs, but open APIs are currently mainly for public organizations. - Case D.

Creating APIs is one direction of development software companies can take, but they also require them as an effective method to deliver data from one system to another. In some cases, the effective use of an API-open or not-can be paramount for a project to success. This is also necessary for an effective delivery of open data, as was mentioned before.

I think us as completely dependent of open APIs since we connect to the public systems. - Case E.

4.2.5 Open and Closed Data are not the Only Forms of Data Management

In this article, the main focus has been open data and external data and its opposite closed data. Between these two extreme scenarios exists shared data, which was also mentioned by the interviewed organizations. These forms to share data are mentioned in the interviews are B2B data sharing, user generated (C2B and B2C) data sharing, and hackathons. Hackathons were not included in the initial scope of data sharing, however, they were mentioned multiple times in the interviews and included here.

The most common of these three is B2B sharing, which means that data is shared between two or more companies based on agreements and contracts on how and where the data can be used. Software companies in these cases either had two clients who shared data between each other or a client that shared data to the software company. For example in Case C, the software company worked as an intermediate, who directed collected data to another client.

At least in one direction there are cases, where we collect data with one company's machinery and then the data is moved into our service, where we visualize that data to another client. - Case C.

However, mostly the interviewees considered this method of sharing rather rare in their environment. Either the data is highly controlled by the owner or it is not necessary to deliver this data anywhere from the system.

Based on contracts, usually we do not take data from our clients to ourselves but try to do the integration in the client's systems. - Case B.

Based on the cases, these problems seem to stem from the desire to control the data. System owners, where the data is stored, tend to treat data as *theirs* unless this ownership has been specified in agreements. In this scope, companies seem to think, that the data has some sort of value, and this value is not something that they want to share, even though the data owner cannot necessarily define the value of data.

We do get inquiries that is it possible to get this and this data, but that rarely is possible, because some external actor owns the system and does not allow access. - Case A.

Currently an understanding has been changed that intellectual property rights (IPR) do not necessarily contain that much value, but the thinking has been shifted towards the ideology, that the collected data has a lot of value. - Case D.

As the popularity of personalized services has risen, the companies are collecting more and more data on their users. This includes traditional services, where users log in and the service remembers their information and also usage data in the service. This data is highly personal and should never be used without the user's permission and never published.

In majority [of our services] there is some form of login and after logging in the user can see their information. - Case C.

I think currently the government offers this service, where you can see your own medical information and what information has been modified and who has looked at those. - Case A.

Another form of user generated data has become from the use of smart devices, such as smart wristbands and other devices that are used in everyday life. This data is collected by a user and transmitted to the manufacturer and their service, or given to the user to do what they want. A form of this was mentioned in Case A, about health care systems, where such data could be used, if the user-or in this case the patient- allows this. This has not materialized yet, but some of the software companies are aware of it and are also developing and planning something in this field.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 13 / ISSUE 1 / JANUARY 2018 / 91-105

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762018000100106

One future idea is this sort of quantified self, that when individuals have these smart devices that can be used as for self-diagnosis. - Case A.

There have been talks about those [personal data systems], but to my knowledge, there are not any concrete projects where they have been implemented. - Case D.

There were also some skeptical mentions of different competitions and hackathons, where the data owner gives access to defined group of developers, who then innovate new services and products based on the challenge offered by the organizer. In the case of open data, these competitions are public for everyone but there are also more limited events, where the participants are selected from applicants.

It feels like currently it is believed that hackathon is the solution, putting a lot of different actors to the same space and allow them to innovate new services. - Case D.

5 Discussion

In this study, four separate research question were defined. These were *How do new clients express interest towards open data*?, *How much open data-based solutions are the existing clientele expecting?*, *How does the product portfolio of a software company respond to open data?*, and *What are the current trends of open initiatives?*. From these questions, it was determined after the interviews that the first two do not hold any separate value, so those will be discussed together. So there are no differences between new clients or already existing clients in terms of open data. The research questions 1-3 served two purposes: to determine the use of open data but also the practical business of opening data.

RQ 1-3, using open data: The first step of this research determined, that while the data collected from the industry panel survey is not a large set, it does imply a definite interest towards open data, which exceeds business domains. While the data is not statistically significant [8], it does define clear themes from the companies interested in open data. Software developers are interested in creating software on top of open data and other industries are interested in using open data in addition to other sets of data in order to do more efficient data analysis.

However, through the interviews, it was possible to gain additional insight concerning the open data and software companies. While the survey shows that there is interest, it does not define if there are any companies operating in the field of open data. Through the interviews it was clear, that open data is not used as a key business resource, while it may be used on some level to initiate a business or enhance data analysis. It would seem that the use of open data as a key resource is avoided because of its unreliability. The data does not necessarily match the context where it is needed and the quality of data does not necessarily satisfy the need in terms of granularity, similarly as observed by Immonen et al. [10]. Using contracts or buying data seems to be a much more reliable source of raw data for business operations, as it enforces the third party to provide and maintain their services.

The general trend is that the open data is an interesting resource for software companies, who are the forefront actors to utilize this data. From the interviews it was noticed, that software companies do keep an eye out for external sets of data and they do think of ways to use the data to their advantage. For instance, a data broker would use open data as a free resource, which they could sell through their service. Depending on how the company does their business, open data could be used to enhance existing or new solutions through geolocation or another easy-to-use set of data. What is interesting that similar opportunities were already reported in 2014 [10] and propositions for value creation in 2015 [36], but practical examples are still small scale projects at best. In general, there is an undertone that the open data possibilities are not that well-known [8], so the practical applications and processes to adopt open data are limited.

RQ 1-3, opening data: It would seem that open data is rarely produced as a business, or by private companies. This stems from the belief that giving away something-especially for free- is bad business practice even if the given thing has no value to the owner. While companies may not open data because of this, the legislative pressure from governmental level forces municipalities and public organizations to open data, which may cause them to lose revenue streams they used to get from selling that same data. Because of protecting these revenue sources, the data that is opened tends to be somewhat useless for the publisher, which usually also means that it is not relevant or interesting for any independent business, although it may provide information which can be used as an additional service. The additional legislative problem is the general ignorance of rights; the data owners are not necessarily informed enough to know, what kind of data is possible to open and therefore nothing is released.

In this study, it was found that some companies open their data and their motivations were identified. Publically funded companies are required to do so through legislation, while some companies open data as a form of marketing, increasing the visibility of data. Other forms of data sharing in a more controlled fashion were also identified from the interviews, such as directly between businesses [29], between businesses and consumers [17], [24], or through hackathons [1] and competitions, as is the suggestion for stimulating use of open data [35]. Hackathons and competitions were not identified as a method to share data before when creating a scope of different sharing

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 13 / ISSUE 1 / JANUARY 2018 / 91-105

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762018000100106

environments. This may be because of the nature of the events is different than usual data transactions with clear actors on both sides.

RQ 4, trends of open initiatives: Open data is a form of openness, that still has not gained popularity like other open trends like open source or open science [19], [23]. From the interviews, open source code was deemed as the most popular form of openness-at least in software business- also from the client side. Open source has become so popular, even clients are demanding it and on the governmental level, it is used more and more. Another popular form of openness, which could provide solutions to the data ownership problems, are the open APIs and APIs in general. APIs can be used to simplify the data usage and application development [33], and to gain profits and possibilities [2]. They are seen as necessary tools to distribute data, no matter if the data is open for everyone or only shared to a restricted group of individuals.

5.1 Strategy Classification for Data Sharing

Besides the research questions, the interviews provided an additional theme for this study. It was clear, that open and closed data are not the only methods to manage data between organizations, as was expected, and the participating companies were aware of these different methods. In the world of data, it would seem that the level of control-such as IPR and data ownership- administered to the data is of utmost importance, as in who can use the data, for what and why. Also moving hand in hand with control, the publicity of data is a factor that is usually thought of. For instance, accidentally showing a patient's medical records, even to the patient in question, may be seen as a breach of privacy. Based on these two factors-control and publicity- and the results of these interviews and research, a strategy classification of different data management strategies has been defined (Figure 5).

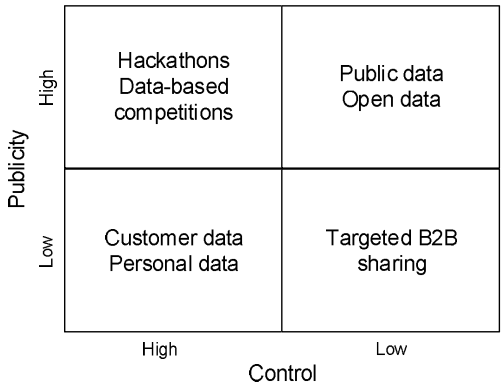

Figure 5: Strategies to share data

From the interviews, it became clear that data can be shared in other ways than between companies. Data could be also shared with a strict group of developers through a hackathon or it could be open for a definite time for any developer through a competition. In hackathons, the IPRs are usually left to the developers, but the data can be used only per contract outside the hackathon, which leaves the control of data to the owning company, while it is accessible to developers.

There have also been various discussions whether the customer data and customer collected data is owned by the company where the customer has given their data or if the data is owned by the customers, for example [24]. Both definitions still place multiple constraints on how a company may use the customer data since they usually must have a permission from the customer to use this data whatsoever. The data is also a target for extreme privacy- because of consumer trust- and that data should never be given or lost outside a company.

Meanwhile, public data and open data are different concepts, since public data is not necessarily available for reuse, as per the license. Some researchers and companies lump these two definitions under open data [14], but considering how companies are willing to give public data through websites, but not as open data, they are separate definitions in the scope of this research.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 13 / ISSUE 1 / JANUARY 2018 / 91-105

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762018000100106

5.2 Validity of the Research

In any case in this type of research project, several threats to study validity exist [20]. For example, in the codification of the observations, the researcher bias can be troublesome, skewing the results of data analysis and further on the refined implications. Similarly, design issues on questionnaire could have steered the collected data towards certain viewpoints. In our study, the threats to validity were addressed by taking certain measurements to ensure neutrality. For example, the questionnaire was designed by a group of three researchers, with the feedback and adjustment ideas collected from other empirical software engineering researchers from the laboratory. In addition, the interviews were conducted by the questionnaire designers, to ensure that the interviewees understood the questions correctly and in all cases, in the native language of the interviewee to catch the indirect undertones and allow informal discussion. The researchers encouraged informal discussion and allowed the interviewee to control the narrative of the interview instead of reflecting their own views through the pre-determined questions, minimizing the researcher bias that may come from poorly designed questions or the views of the researcher. Finally, the codification process was conducted by two researchers to ensure minimal interference of personal opinions or individual preferences. These actions also address most of the common problems of qualitative studies, which besides Onwuegbuzie [20] are also identified by for example Robson & McCartan [26] and Miles & Huberman [16].

Other concern was the number of interviewed organizations. First of all, it is important to note that the goal of this research is not to describe software industry as a whole but to highlight issues behind publishing and using open data. This research was limited to studying a selected sample of five organizations. This was taken into account while writing this article and the authors were trying to avoid any generalizing undertones. The reader should notice, that these results are not exhaustive in describing the views towards open data. Another limitation of the results was that in the studied organizations, only one organization applied open data in their day-to-day business. However, while the other organizations did not apply open data in their daily operations, most of them had the knowledge or even experience with open data and other forms of openness. This can be seen as an indication that open data is not applied as widely as is suggested by the scale and amount of open data in the literature, but that many organizations have adjusted their processes to use open data if it is necessary for the context.

6 Conclusion

In this study, our aim was to determine the views and experiences of industrial use of open data and its applications. We interviewed five project managers or upper management representatives from five organizations to understand and assess how these organizations are applying open data concepts currently or going to apply in the future. From our earlier industry survey, we already knew that the general application levels of open data techniques are not very high, and in this qualitative study we aimed to understand how and why the state of the art is what it is.

Overall, our results from the interviews confirmed the observations from the survey; the companies do not have very strong strategies towards the application of open data. The organizations do have the knowledge and even experience with open data, but it is rarely used in the day-to-day operations. Even an organization that handles open data daily does not recognize it as a critical form of data, especially when compared to available public data. The research data would suggest that open data is not used because of the lack of success stories and the required format and quality of data. On the other hand, data is not opened because of the lack of revenue models and the negative views towards opening data in an organization. These practical reasons, while not conclusive, seem to hinder the creation of commercial open data applications. The results of this study do raise the question whether or not open data should be regarded as a key business resource at all, but an additional part of a company's operations, that can be used when necessary to enhance and engage. Does such a set of data actually exist that can simultaneously be open and profitable?

As this study considered only software development companies it would be critical to direct future research towards other sectors, using the findings in this research as initial hypotheses. It could also be beneficial to get in touch with organizations that are slowly opening their data to be used by some other organizations. The trend has been strong on public sector but the private sector is still a mystery and it would require more research. The following work should concentrate on how the currently applied data sharing methods are being used and why - remaining in the scope of open data in order to determine, how much of the open data initiative can be realized in business.

Acknowledgements

The authors of this study would like to thank Digital, Internet, Materials & Engineering Co-Creation (DIMECC) and their project, Service Solutions for Fleet Management (S4Fleet) for funding this study.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 13 / ISSUE 1 / JANUARY 2018 / 91-105

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762018000100106

Websites List

Site 1: Full list of interview questions
http://www2.it.lut.fi/GRIP/datatools/opendata/Interviewtable.pdf

References

[1] J. Aboab et al., A datathon model to support cross-disciplinary collaboration, Science Translational Medicine, vol. 8, no. 333, p. 333ps8, 2016.
[2] T. Aitamurto and S. C. Lewis, Open APIs and news organizations: A study of open innovation in online journalism, presented at the International Symposium on Online Journalism, Austin, TX, April 1, 2011.
[3] J. E. Fountain, The virtual state: Transforming American government?, National Civic Review, vol. 90, no. 3, pp. 241-52, 2001.
[4] B. G. Glaser and A. L. Strauss, The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago: Aldine Publishing Company, 1967.
[5] F. Gonzalez-Zapata and R. Heeks, The multiple meanings of open government data: Understanding different stakeholders and their perspectives, Government Information Quarterly, vol. 32, no. 4, pp. 441-452, 2015.
[6] J. Gurin, Open governments, open data: A new lever for transparency, citizen engagement, and economic growth, The SAIS Review of International Affairs, vol. 34, no. 1, pp. 71-82, 2014.
[7] M. B. Gurstein, Open data: Empowering the empowered or effective data use for everyone?, First Monday, vol. 16, no. 2, 2011.
[8] A. Herala, J. Kasurinen and E. Vanhala, Current status and the future directions of open data: Perceptions from the finnish industry, in Proceedings of the 20th International Academic Mindtrek Conference, Finland, 2016, pp. 68-77.
[9] A. Herala, E. Vanhala, J. Porras, and T. Kärri, Experiences about opening data in private sector: A systematic literature review, in Proceedings 2016 SAI Computing Conference (SAI), London, 2016, pp. 715-724.
[10] A. Immonen, M. Palviainen and E. Ovaska, Towards open data based business: Survey on usage of open data in digital services, International Journal of Research in Business and Technology, vol. 4, no. 1, pp. 286-295, 2014.
[11] M. Janssen, Y. Charalabidis and A. Zuiderwijk, Benefits, adoption barriers and myths of open data and open government, Information Systems Management, vol. 29, no. 4, pp. 258-268, 2012.
[12] J. Kasurinen, A. Maglyas and K. Smolander, Is requirements engineering useless in game development?, in Requirements Engineering: Foundation for Software Quality, vol. 8396 (C. Salinesi and I. van de Weerd, Eds.). Cham: Springer International Publishing, 2014, pp. 1-16.
[13] J. Lindman, Similarities of open data and open source: Impacts on business, Journal of Theoretical and Applied Electronic Commerce Research, vol. 9, no. 3, pp. 46-70, 2014.
[14] J. Lindman, T. Kinnari and M. Rossi, Industrial open data: Case studies of early open data entrepreneurs, in Proceedings System Sciences (HICSS), 2014 47th Hawaii International Conference on, Hawaii, 2014, pp. 739-748.
[15] T. McLaren, M. Head and Y. Yuan, Supply chain collaboration alternatives: Understanding the expected costs and benefits, Internet Research, vol. 12, no. 4, pp. 348-364, 2002.
[16] M. B. Miles and A. M. Huberman, Qualitative Data Analysis: An Expanded Sourcebook. Thousand Oaks, CA: Sage, 1994.
[17] M. Mun et al., Personal data vaults: A locus of control for personal data streams, in Proceedings of the 6th International Conference, New York, NY, USA, 2010, p. 17:1-17:12.
[18] M. M. Najafabadi and L. F. Luna-Reyes, Open Government Data Ecosystems: A Closed-Loop Perspective, in Proceedings of the 50th Hawaii International Conference on System Sciences, Hawaii, 2017, pp. 2711-2720.
[19] N. M. O'Boyle et al., Open data, open source and open standards in chemistry: The blue obelisk five years on, Journal of cheminformatics, vol. 3, no. 1, p. 37, 2011.
[20] J. Onwuegbuzie and N. L. Leech, Validity and qualitative research: An oxymoron?, Quality & Quantity, vol. 41, no. 2, pp. 233-249, 2007.
[21] Open Knowledge Foundation. (2011, January) What is open data?, Open Data Handbook. [Online]. Available: http://opendatahandbook.org/guide/en/what-is-open-data/.
[22] B. Otjacques, P. Hitzelberger and F. Feltz, Interoperability of e-government information systems: Issues of identification and data sharing, Journal of Management Information Systems, vol. 23, no. 4, pp. 29-51, 2007.
[23] H. A. Piwowar, R. S. Day and D. B. Fridsma, Sharing detailed research data is associated with increased citation rate, PLOS ONE, vol. 2, no. 3, p. e308, 2007.
[24] K. A. Poikola, K. Kuikkaniemi, and H. Honko. (2015, March) Mydata - A Nordic model for human-centered personal data management and processing. VALTO. [Online]. Available: http://urn.fi/URN:ISBN:978-952-243-455-5.
[25] D. Praditya, M. Janssen and R. Sulastri, Determinants of business-to-government information sharing arrangements, The Electronic Journal of e-Government, vol. 15, no. 1, pp. 44-55, 2017.
[26] C. Robson and K McCartan, Real world research. United Kingdom: John Wiley & Sons, 2016.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 13 / ISSUE 1 / JANUARY 2018 / 91-105

This paper is available online at
www.jtaer.com
DOI: 10.4067/S0718-18762018000100106

[27] A. Rosenthal, P. Mork, M. H. Li, J. Stanford, D. Koester, and P. Reynolds, Cloud computing: A new business paradigm for biomedical information sharing, Journal of Biomedical Informatics, vol. 43, no. 2, pp. 342-353, 2010.
[28] S. Sayogo and T. A. Pardo, Understanding smart data disclosure policy success: The case of green button, in Proceedings of the 14th Annual International Conference of Digital Government Research, QC, Canada, 2013, p. 72.
[29] G. Stefansson, Business-to-business data sharing: A source for integration of supply chains, International Journal of Production Economics, vol. 75, no. 1-2, pp. 135-146, 2002.
[30] A. Strauss and J. Corbin, Basics of Qualitative Research, vol. 15. Newbury Park, CA: Sage, 1990.
[31] S. Verhulst and R. Caplan. (2015, April) Open data: A twenty-first century asset for small and medium-sized enterprise. GOBLAV. [Online]. Available: http://thegovlab.org/open-data-a-21st-century-asset-for-small-and-medium-sized-enterprises/.
[32] J. Xu, F. C. Tong and C. J. Tan, Auto-ID enabled tracking and tracing data sharing over dynamic B2B and B2G relationships, in Proceedings RFID-Technologies and Applications (RFID-TA), 2011 IEEE International Conference on, Spain, 2011, pp. 394-401.
[33] A. Zaballos, A. Vallejo and J. M. Selga, Heterogeneous communication architecture for the smart grid, IEEE Network, vol. 25, no. 5, pp. 30-37, 2011.
[34] F. A. Zeleti, A. Ojo and E. Curry, Emerging business models for the open data industry: Characterization and analysis, in Proceedings of the 15th Annual International Conference on Digital Government Research Aguascalientes, Mexico, 2014, pp. 215-226.
[35] A. Zuiderwijk, N. Helbig, J. R. Gil-García, and M. Janssen, Special issue on innovation through open data: Guest editors' introduction, Journal of Theoretical and Applied Electronic Commerce Research, vol. 9, no. 2, pp. I-XIII, 2014.
[36] A. Zuiderwijk, M. Janssen, K. Poulis, and G. van de Kaa, Open data for competitive advantage: Insights from open data use by companies, in Proceedings of the 16th Annual International Conference on Digital Government Research, Phoenix, AZ, USA, 2015, pp. 79-88.

Publication IV

Herala, A., Kokkola, J., Kasurinen, J., and Vanhala, E.
Strategy for Data: Open It or Hack It?

Journal of Theoretical and Applied Electronic Commerce Research
Forthcoming

Journal of Theoretical and Applied Electronic Commerce Research
Draft

This paper is available online at
www.jtaer.com

Strategy for Data: Open It or Hack It?

Antti Herala[1], Juho Kokkola[2], Jussi Kasurinen[3], and Erno Vanhala[4]

[1] Lappeenranta University of Technology, School of Business and Management, Lappeenranta, Finland, antti.herala@lut.fi
[2] Aalto University, School of Science, Helsinki, Finland, juho.kokkola@aalto.fi
[3] Southern-Eastern Finland University of Applied Sciences, Kotka, Finland, jussi.kasurinen@xamk.fi
[4] University of Tampere, Tampere, Finland, erno.vanhala@staff.uta.fi

Abstract

The open data ventures can be summarized in a way that companies are reluctant to share their data with anyone, whereas governments open their data for citizens, institutions, and businesses as much as they can. However, this principle is changing, since there is added value in the digital information and datasets the companies possess and they are slowly understanding the value of crowdsourcing. In order to engage external experts, companies are reluctant to open their data, but they are interested in hosting hackathons. Hackathons are seen as a valuable direction to engage developers with private data. In this article, we observed and analyzed different industry cases for strategies and opinions on how and why organizations arrange hackathon events to extract information from their data, and how this relates to the popular open data movement. Our results indicate that hackathons offer more control and practical solutions over the fundamental open data approach. It would seem, that hackathons provide better inroads for the companies to monetize their datasets and information assets, while open data could bring more visibility to the brand.

Keywords: Digital information, Hackathons, Open data, Case study, Business strategy, Product development, Information extraction

1 Introduction

The current views in business tend to suggest that nothing should be shared or given away, and this strategy is applied strictly, especially to data. While the practices of data management, and the world in general, are moving towards more open direction with, for example, open innovation and open source software, organizations, and especially companies, have begun to realize the power of the masses and crowdsourcing. Hackathons, short and intensive programming events, have become a popular method to further engage crowdsourcing for product development, while also generating interest towards developing business from the existing closed data [4]. Hackathons allow organizations to share their data with a group of participants and work with them, which allows communities and other interest groups to innovate in a concentrated manner [19].

A manifestation of openness in the field of data management is the open data; data that can be used by anyone for anything. Open data is an important tool for governments in order to increase their transparency and engage citizens, encouraging them to join in the public decision-making and even innovate using the opened data [17]. In the field of open data, open business data is seen as a valuable source of information that could be used in a combination with government data [16], [28]. However, while the companies see the value of crowdsourcing, they are not thrilled with the aspect of opening their own data [12].

In our previous research into the open data in the private sector concentrating on opening their data, it was found that opening data can bring benefits to privately owned organizations [13]. This result was achieved with a systematic literature review, which analyzed the relevant literature in the field and covered different benefits and drawbacks of opening data to private organizations. Continuing into the topic of openness and open data, our continuance study [12] uncovered cases where the data had been opened by a private organization, but the actual applications of the data proved difficult because of regulations and issues in the data usability. Because of this, the open data activities were considered irrelevant in the business scope. However, when it came to data sharing, multiple interviewees mentioned hackathons as a popular method of businesses to engage external actors.

Both methods, hackathons and open data, can be seen as a manifestation of data collaboratives [28] and can be used as a tool for innovation using external resources and development, separately or simultaneously. Because of their nature, these activities can be described either as competitive or complimentary. In some cases, it is possible to use either to achieve goals of the data owner [1]. On the other hand, it is possible to apply hackathons in a relation to open data, in order to promote the data, and further engage the citizens and developers [30], or use the open data as a part of a hackathon [19].

The popularity of hackathons and the unpopularity of open data in a business context motivated the authors to ask a question: "*How do the benefits and drawbacks of organizing a hackathon differentiate from opening data?*". In order to determine this, a sub-question was formed: "*What benefits and challenges are there in organizing a hackathon?*". Since the findings concerning the benefits of opening data have already been published, it is possible to compare the empirical findings from the hackathons to the findings of the systematic literature review, in a form of continuation study.

The rest of the paper is constructed as follows: Chapter 2 discusses prior research related to this study, and Chapter 3 discusses the applied research methods. Chapter 4 summarizes the results of the study, and Chapter 5 discusses the implications, limitations, and validity of the results. Finally, Chapter 6 summarizes the paper with the conclusions.

2 Related Research

This section presents the relevant literature about hackathons and also outlines the findings from the previous literature review about the effects of opening data.

2.1 Hackathons

Hackathons are a relatively new operating method since the term originated from the open source software developers in 1999 [4]. Hackathons lack a strict and uniform definition, making the event format rather vague; while hackathons in general are considered technology-driven events, some non-programming events have also been called hackathons [19] since they refer to hackathons as an event of conducting open innovation. In any case, a hackathon is a method where an organization utilizes external resources in addition to internal innovation [5]. This usage of external resources or a customer-driven innovation is also supported by Desouza et al. [6], as they describe how the industry has moved from the customer-centric innovation to the customer-focused and further, customer-driven, innovation practices. Hackathons can be seen as a straightforward method to implement the customer-driven innovation and acquire external feedback on the business practices.

The format and setting of the hackathon can vary based on the objectives and the requirements of the event [4]. Hackathons have been used in multiple contexts, such as educational hackathons [22], civic hackathons using open data and public resources [18], culture hacks [3], and industrial hackathons, where the hosting organization is a company [24]. In addition, large hackathon events can be organized by a specialized entity, where the challenges are introduced by the sponsors and other participating organizations. Such organizers are for example Ultra Hack (Site 1) and Junction (Site 2). The attendance can either be open to anyone who wants to participate, or the organizers select the participants based on their applications. These events, called public hackathons in this research, can combine other forms of hackathons into one event, since the participating organizations have their own objectives.

Briscoe and Mulligan [4] have categorized the different hackathons into two main subgroups: tech-centric and focus-centric. The tech-centric hackathons are events, which concentrate on the software development by applying specific technologies and tools. They are further divided into three groups: Single-Application (focus to improve one application), Application-Type (focus on specific platform or genre), and Technology-Specific (focus on one technology). The focus-centric events are solving complex problems, such as social issues, or business objectives, and are further divided into three groups: Socially-Oriented (social concerns), Demographic-Specific (for a specific group of programmers), and Company-Internal (internal staff).

As for the benefits and challenges of hackathons, a previous study by Komssi et al. [19] focused on hackathon events in F-Secure, a software company specializing in cybersecurity. Their study collected data from five hackathons and presented findings to describe the benefits and the challenges of hackathons. They outline the benefits, such as new products and new features to the existing products, radical ideas that would not be normally feasible in the scope of the company, and internal and external communication between the developers and other fields of expertise, as well as external collaboration. The detected challenges were the challenge of communication between fields, the continuance process with the prototype development post-event, and the intellectual property rights. They finish their research with the hackathon paradox: during and immediately after the event the audiences, the developers, and the other stakeholders have been satisfied with the event and the outcomes, but the prototypes are rarely being taken forward and commercialized, as if the hackathons are lacking as an innovation method.

As the very definition of a hackathon varies, the benefits and challenges of a hackathon are not that well understood. While the research done by Komssi et al. [19] outlines benefits and challenges, their case hackathons concentrate mostly on internal development in a company, which are focus-centric events by the classification of Briscoe and Mulligan [4]. Even the categorization of hackathons is difficult since while a hackathon event may be focus-centric, it is possible to have tech-centric tasks within the hackathon. In this research, the given tasks in each hackathon are classified based on the work of Briscoe and Mulligan [4], but the hackathons themselves are categorized based on their context.

2.2 Open data

Open data is data, which is published for everyone to use, for any reason. Because of this, it is a powerful resource for innovation, since the availability of data invites innovation and combination of datasets [20] and freely accessible data, making it a powerful business resource [9]. However, even though there is interest towards open data [11], the value is somewhat complex to attain.

While the value generation from the open data can be seen as a complex problem [32], research shows that the value is often indirect [30] and a definite return of investment is difficult to calculate [17]. In their research Janssen et al. [17] found multiple economic benefits of open data, such as economic growth, stimulation to innovation, development of new products and services, crowdsourcing, to name a few. Similarly, Lindman et al. [20] outline benefits, such as cost savings from crowdsourcing, increasing transparency, and new service innovations.

Because the return of investment is difficult to estimate, companies are not interested in opening their data. In the research by Immonen et al. [15], companies were interviewed and they were asked to publish their data as an open data source. The results were that the companies wanted to get financial benefits from their data, because of the costs of data collection. While there are business models for open data, companies require assurances that they will benefit from their data [12]. From the economical point of view, the biggest issue of open data is the fact, that it is difficult to make money with [30], which is why it is generally avoided [12].

But when data is opened by privately owned companies, there are benefits but also drawbacks. An extensive literature review into the benefits and drawbacks of opening data from the private sector was published as prior work [13]. Based on this work, a list of six benefits and four drawbacks were found in the open data activities by private organizations (Table 1).

Table 1: Benefits and drawbacks of opening data to private organizations

Benefits	
Collaborative actions	Through opening their data, companies can collaborate more with other companies and consumers. Especially the long term benefits of opening data enhances industry collaboration and information sharing [7]. The consumers can also be engaged better, allowing them to make informed decisions based on company data [26].
Competitiveness	Transparency and openness allow smaller companies to compete with larger actors, who have more resources and an advantage in a brand [26]. Competitiveness can also be enhanced by the right balance between protecting and sharing assets [10].
Ecosystem-wide collaboration and communication	If one company within an ecosystem is starting to open data, it can possibly affect the whole ecosystem [26]. However, the disclosure can allow informal dialogue, information sharing, and collaboration within the ecosystem [7], [10].
Innovation and development	The innovation based on open data can either create completely new products through the combination of datasets, but also incremental development and new features to existing products [30], [26]. Open data can be used for innovation in multiple levels: by companies, by governments, but also by individuals [9], possibly for the benefit of the data provider.
Internal change	Supplying open data can provide an incentive for a company to improve their processes if the data that is opened reflects poorly on their own requirements [7]. Through opening data, the organization can see possible flaws in their processes and improve them [14].
Public image	Sustainability and ethicality of products and processes are aspects that the consumers and governments are demanding. Through transparency, that open data advances, companies can demonstrate these aspects to the public directly, instead of third party certifications, improving their public image [7], [26].
Drawbacks	
Decrease in efficiency	When data openness is being implemented, the conflicts between data management strategies can cause a division of available resources, making data management processes ineffective [14]. Also, exiting conflicts between organizations [1], [7], or the unclear ownership of data [12] can either make the opening ineffective or prevent the disclosure altogether.
Increased costs	Opening data tends to require investments in order to function properly, investments to both, the technology and the training of the staff [14], [26]. Additionally, the data may require additional material or metadata, documentations, and records available for public data users [2].
Public data increases problems	While publicity of data may be seen as a benefit, it depends on the quality and the content of data. If the data is privacy-sensitive, it can be used to affect the company brand [31], and the release of confidential data can lead to privacy infringements through de-anonymization [2]. Because the lack of control, open data is vulnerable to misuses and misunderstandings [31].
Required changes	Opening data requires a multitude of changes inside and outside the company in the form of policies and focus [31], processes, employees, and their mentality [14], and even business models [30]. Other companies within the ecosystem are also subject to a change, especially if they provide data to the company, where data is being opened [26]. In an unsuitable culture, these changes are difficult to implement and may lead to unwanted aspects of company culture [2].

In this work, the objective is to assess these benefits and drawbacks in the hackathon events to understand and assess, why hackathons are applied more in a current business environment that opening data, while in a general sense the goals are similar. The research is focused on hackathons and the findings are then compared to the benefits and challenges found in the systematic literature review.

3 Research process

In this section, the research methods used in this study are presented. The data was collected via interviews, and the data was qualitatively analyzed in order to observe the benefits and challenges that the hackathons yielded for the organizers. Each hackathon was treated as a singular case and finally synthesized using cross-case synthesis, recommended by Yin [29]. After the results about the hackathons had been reached, they were compared to the benefits and drawbacks of opening data.

3.1 Data collection

The data for this research was qualitative data, collected through interviews of the representatives of the hackathon organizers. The data was collected from organizations, who had hosted a hackathon in 2016 or 2017 and working with a consultation company, we were able to interview these organizations timely, soon after the event. Eight of the interviewed organizations had hosted their own hackathon, using external teams as participants and had used hackathon consultants to set up the event. Each organization had a representative, or a team, organizing the company-end of the event. Depending on a case, either the key representative or personnel from the team was chosen for the interview session. Before the interview, it was made sure that the interviewee had sufficient knowledge about the event. If not, the interview was rescheduled with a more appropriate interviewee.

We did not want to limit this research to only highly controlled industrial cases. Because of this, we added two organizations who participated in a public hackathon by offering challenges to multiple hacking teams through a large hackathon event with multiple organizations offering their challenges. It should be noted, that public hackathon does not require a public organization, privately owned companies can also offer challenges through these events. They were used as control cases in order to determine, if the found benefits and challenges would be only relevant for the industrial hackathons, or if these effects could be reached through other venues. Similarly, the interviewees were selected from the team of organizing personnel within the organization.

As mentioned, out of the ten hackathons, eight of them are categorized as industrial hackathons, where the organization was the direct host and in control of the participating teams, to a certain extent. Two of the events are categorized as public hackathons, where the organization offered challenges to a larger event with less control over the teams. The number of offered challenges, or tasks, ranged from one to four per event (Table 2). In some cases, such as Case C or Case F, the organization had hosted two separate events with similar, but separate themes before the interviews. The agenda for the hackathons varied, some were interested in an abstract concerning the business domain, while others offered a specialized technology and datasets to develop practical constructs. The challenges, that were offered, are categorized by the groupings of Briscoe and Mulligan [4] described in the previous section. In order to protect the anonymity of the participating organizations, the challenges are not described further because of the publicity and findability of the events.

Table 2: Participating organizations

	Type of hackathon	Types of challenges
Case A	Industrial	One Single-Application and two Technology-Specifics
Case B	Industrial	Two Single-Applications and one Application-Type
Case C	Industrial	1st hack: Two Single-Applications and two Application-Types, 2nd hack: One Single-Application and two Technology-Specific
Case D	Industrial	One Single-Application and two Technology-Specifics
Case E	Industrial	One Single-Application, one Application-Type, and two Technology-Specifics
Case F	Industrial	1st hack: One Single-Application and two Application-Types, 2nd hack: Two Single-Applications and two Application-Types
Case G	Industrial	One Application-Type
Case H	Industrial	Two Single-Applications and one Application-Type
Case I	Public	One Socially-Oriented
Case J	Public	One Application-Type and one Socially-Oriented

The interviews were constructed based on three major themes; the motivations and the goals of the event, the realization of the goals throughout the event, and the actions after the event. As a final topic, the interviewees discussed the whole process, their views on the benefits and challenges of this approach, and the future aspects, such as improvement proposals for the future hackathons.

The interviews were held as semi-structured: the interviewer had a list of questions prepared based on the themes of the interview, but it was only used as a checklist to confirm that the topics were discussed thoroughly. The interviewer was mainly listening to the interviewee and interrupted only to ensure, that the discussion remained within the scope of the research. The interviews lasted around 40-60 minutes and were arranged with one or two representatives from the organization and one or two researchers in a face-to-face meeting or over a video call. The interviews were either tape-recorded, or extensive written notes were taken during the interview to ensure a sufficient level of recorded data.

3.2 Data analysis

The applied analysis method in this study is the cross-case synthesis, which is used in multiple-case studies [29]. The key to this analysis is to provide concise findings by analyzing multiple cases separately and combining them into one case. In this study, each of the hackathons of each organization is analyzed as one case. The results are formed by combining all of the cases, which minimizes the number of outliers and gives more strength to the findings.

The first step of the case analysis is done by constructing matrices based on the subsidiary research question [21]. For each case a matrix is constructed, categorizing the motivations, benefits, and challenges in the analyzed hackathons. The data from the interviews were then mapped into these matrices by using the recordings, transcriptions, and notes available. Each case matrix consisted of statements and their contexts, which were carefully analyzed, understood, and codified into a more concise form. For example, lengthy descriptions about how a prototype was taken into production was codified simply as "New product", while more complex descriptions about the networks and partners, or occurrences within the event, were given longer description, such as "Development in ecosystem with participants from different environments" (Case C) and "Views about what can be done and own people starts to see opportunities" (Case F). Codification is used as a tool in order to demonstrate the core messages in the data and the wording of the codified outcomes were taken from each case separately. This way each of the cases were analyzed separately, minimizing involuntary synthesis before all of the cases were analyzed.

After the cases were analyzed, the matrices from each case were then deconstructed into a final, synthesized matrix, where the findings were distilled through combining similarities and anomalies [29]. Each finding was categorized by abstracting them against all of the other findings. The initial set of categories was collected from one case, recognizing distinct patterns from the codified terms. Similar terms were synthesized under a broader category, and the rest were left as is. These patterns, or terms, were then taken to the next case and compared. Again similar terms were synthesized, and the codified, final term was broadened, if necessary, to fully describe all the used terms. This was done until there were no more cases to analyze and all of the terms from case matrices had been synthesized into the final matrix. For example, the codified result "New product" was abstracted as a part of "New products, projects, and research", while the longer codifications, mentioned in the previous paragraph, were re-codified as "Development in ecosystem" (Case C) and "Internal / organization development" (Case F). The original data from the cases related to each term were then used to make descriptions and examples of these abstractions for the findings. By these actions, the effects of the environment and personnel were minimized in the categorization, but the results stay true to the original data. The data analysis and the order of execution are further visualized for clarity (Figure 1).

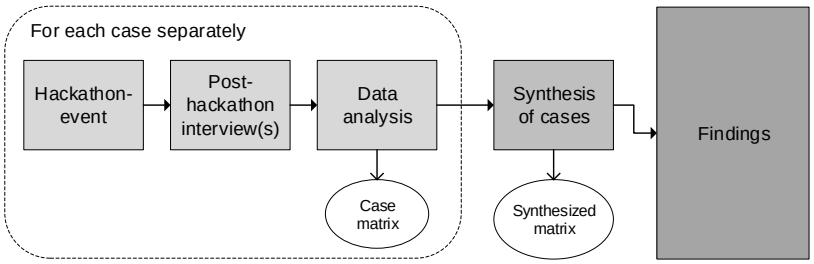

Figure 1: Cross-case synthesis execution

When the complete categories were finished and the findings about the hackathons synthesized, the categories and their descriptions of both, benefits and challenges, were compared to the categories of open data benefits and drawbacks. Utilizing especially the descriptions with similar or identical topics and keywords, the corresponding categories from both fields were taken under further scrutiny. The open data benefits and drawbacks from the literature review were used as the base categories, where the hackathons were mapped into. This allows this research to discuss and bring forward the different aspects of open data and compare the hackathons to them, finding similarities and differences between the practices. While this method of comparing categories is unstructured and superficial, it allows some insights into the fundamental differences of these methods as a tool for innovation in the different business domains. From these differences, conjectures can be drawn to explain, why one method is somewhat popular in the industry and the other one is not. However, it should be noted that these conjectures are not facts, but results from qualitative case data and a crude comparison between two practices. They can be used as the basis for future research in a form of provable hypotheses, but not as definite statements.

4 Results

The main results are divided into benefits and challenges; the resulting categories were created with the cross-case synthesis method as defined in the previous section. In each main result, the occurrences and descriptions are provided to define the observed phenomena.

4.1 Benefits

From the analyzed hackathons, four benefit categories were present in the majority of the cases, with three existing in basically all cases (Table 3). The new ventures in a form of new projects, products, and research were reported in every case, which indicates that each organizer got something practical out of the event. For some organizers, this

was more superficial or incremental within the organization, while to others this meant completely new products that were implemented into business quickly. The second popular benefit was changes or potential changes in the organization. The personnel organizing the event and also the visiting employees describe the hackathon as an eye-opening event, an example that agile processes can be used even in their organization. It also enabled the different departments to work together for a common goal. The third major benefit is the change in the current ecosystem and how the event molded or can mold the existing ecosystem through new partners and collaboration and the use of external professionals to enhance development with an outside point of view.

In Table 3, each case is presented as one column, using an x to demonstrate, that the category was mentioned in the case interviews and was seen as a benefit in the case. After the table, each category is explained with descriptions and examples derived from the case, to give context to the category.

Table 3: Benefits from hosting for case organizations (x = was mentioned in the interview)

Category	A	B	C	D	E	F	G	H	I	J	Sum
New ventures: projects, products, and research	x	x	x	x	x	x	x	x	x	x	10
Internal / organization development	x	x	x	x	x	x	x		x	x	9
Ecosystem development	x	x	x	x		x		x	x	x	8
Feedback and communication		x	x		x	x	x		x		6
Motivated system development	x		x			x			x	x	5
Accelerated R&D	x		x		x	x	x				5
Visibility		x		x	x		x				4
Archive of ideas and concepts	x		x					x			3
Recruitment			x	x							2

New ventures: projects, products, and research combines the new ventures that were achieved. While the main goal of many organizers was to achieve new products, the level of which this was realized varied. In a couple of cases, such as Case G, Case D, and Case J, the ideas and prototypes remained at a concept level and were not yet implemented into business. Some of the organizers started piloting with one (Case H) or more (three in Case F) proof of concepts and in Case B, one prototype had been proceeded into testing in customer site while others remained in discussions. Few cases had already moved even beyond testing: in Case C the prototypes were converted into research projects with the teams, in Case E a product was implemented and was generating revenue at the time of the interview, and in Case A the product had been ready to launch but was withheld because a competitor launched similar product before them. For some organizers, the yield was even more than just isolated products: In Case I one of the major outcomes was that they found new models how to connect with external developers and in Case D it was reported, that new collaboration and projects were being set up after the event, without a clear connection to the prototypes.

The category *Internal / organizational development* was mentioned in nine cases out of ten, perceived on different levels. On an individual level, an energizing and an eye-opening impact was reported in multiple cases (cases C, J, B, and G), the employees were excited and enthusiastic about the new ways a prototype can be developed. The hackathon also brought about development towards a more open culture within the organization (cases J and E) and allowed everyone the chance to innovate and see different opportunities outside their current positions (cases G and F). A change was also reported in the communication between the departments in cases C and D: through the event multiple departments were activated, the information flow strengthened, and the feeling of togetherness was increased. The increased communications within the organization also allowed the organizers to take a new approach to in-house product development, changing the process how innovation has been done within the organization (cases A, I, D, and C).

Ecosystem development covers a multitude of actors from external developers to other partners and collaborators within the industry. In multiple cases (A, B, D, I), it was mentioned how hackathon can be used to meeting new partners and suppliers through the common interests. Instead of a random sampling from the field, the event can be used to focus on a specific group of external developers (cases A, C, and J), who would be difficult to meet otherwise, and also the organizer can see, what is currently going on in the field (Case F). The event was also used to market the organization: in Case B the hackathon organizer wanted to work with start-ups and the event was used as a marketing tool to show how. Lastly, in some cases (C, F, J, B, and H) the hackathon was seen as a method to create a new form of partnerships, where a larger portion of development and conceptualization would be shared, gaining an edge through a large network of partners and individuals.

The category *Feedback and communication* is used to describe the effects of face-to-face meetings in the event and also the new knowledge that was brought inside the company by the external developers. In cases E and I, the organizers reported usable feedback about their data and how could the data be improved in the future. On the other hand, through discussion the organizer was able to offer a view of the industry, what are the current issues and challenges (cases B and C) and the participating teams were able to view the issues from an external point of view and apply new technologies and solutions to the problems (cases E and F). Finally, in Case G, the organizer reported that the hackathon was used as a marketing material, giving empirical talks based on the event and company strategy.

The category *Motivated system development* refers to three stages: before, during, and after the event. Before the event, the organizer had to make sure, that the technologies, systems, and data were sufficient for the external developers, and the event offered a suitable deadline (cases F and I). During the event, in cases A and C, the teams were helping the organizer to develop and enhance their systems, while there was no direct business benefit in these incremental developments. Additionally, the organizers got feedback about their systems from an external point of view, which allowed them to realize faults and points of improvement in their systems (cases F and I). After the evet was over, the organizer was left with new motivation to improve the systems even further (cases I and J).

Accelerated R&D refers to the possibility of increasing the velocity of the innovation process. In multiple cases (A, C, E, G) the organizer was impressed by the speed in which the prototypes were built and their overall quality. For example in Case C, a concept was thought of for three years, it was developed for half a year, and implemented by outsiders within two days. In Case F it was also highlighted, that the engineers had to crystallize their results within a couple of minutes and this was seen as a significant improvement.

Visibility was mentioned as a positive benefit since the events had been noticed by a multitude of possible partners, which had then contacted the organizer and referenced the hackathon as a point of interest (cases B, D, E, and G). In Case D, the organizer highlighted the fact, that the hackathon was used to show partners and other organizations, that they are open to working in different ways.

Some organizers (cases A, C, and H) reported that they had built an *Archive of ideas and concepts* from the event. They were not capable of implementing all of the concepts immediately, and the unused concepts were stored with the intention of implementation whenever possible, depending on markets and available resources. However, it was mentioned in case C, all of these ideas did not come from the hackathons, but from within the organization when preparing for the event or after it.

The least mentioned category was *Recruitment*, where the organizers used the event in order to recruit new talents. For instance, in Case D the organizer was recruiting software developers for their subsidiary and was also building their brand for the possible experts looking for new positions.

4.2 Challenges

The challenges of hackathons did not offer as uniform results as the benefits (Table 4). Eight out of ten organizations reported that they had issues with their own internal processes or in general in the organization of the activities when hosting the event. This is a major issue, while not an unexpected one since the large organizations tend to have more detailed and defined management processes. Introducing agile development into these processes was sometimes considered problematic, although in some cases it was one of the major benefits as well.

Some challenges correspond to the benefits, and imply a very polarized view of the hackathon experiences, but also indicate the motivation to improve based on the challenges. Challenges such as the lack of communication and technological issues, or resource and cost management are likely to be not caused by the hackathons but become issues because of the organizer's own internal need for development. However, some challenges emerged because of the event and did not have a direct link to the organizations, such as preparatory actions and scheduling.

Table 4: Challenges in hosting an event (x = was mentioned in the interview)

Category	A	B	C	D	E	F	G	H	I	J	Sum
Process and organizational issues	x	x	x	x		x	x	x		x	8
Resource and cost management		x	x	x		x	x			x	6
Difficulty of communication			x	x				x	x	x	5
Technological issues	x				x	x		x	x		5
Challenges in preparation	x		x	x		x		x			5
Team commitment		x	x					x	x		4
Schedule			x	x				x			3
Ownership			x		x		x				3
Lack of follow-up							x			x	2

The category *Process and organizational issues* was mentioned the most. The category refers to the issues noticed in the organization, how the culture of the organization was not suitable for hackathon as an innovation method, or the processes within the organization did not support the agile development principles. In cases C, D, G, and J, the culture was mentioned as a major hindrance, it was difficult to engage individuals in the development and management to take the projects forward. Another issue was the innovation process within the organization, mentioned in cases A, B, D, and F. After the hackathon, the projects are supposed to be taken under further development and commercialization, but in these cases, getting the project through the project office and securing funding was a major hindrance and a cause behind losing the momentum. Additionally, in Case H it was mentioned,

that the number of business secrets, related to the functions of the organization, were seen as a limiting factor, losing the competitive edge was seen as a major challenge.

The category *Resource and cost management* represents the issues with funding and the strain on the employees, who are running the day-to-day business in the organization, even though in Case B the interviewee mentioned that in hindsight, focusing resources to the event was somewhat irrelevant, since the event only served as a starting point. In cases B, C, D, and F, the organizer reported a shortage of manpower to make the preparation for the event, especially because of the scale of the event: larger event requires more people to prepare and manage it, in addition to the day-to-day responsibilities. On the other hand, in cases C, F, G, and J the funding was seen as an issue and an extra cost of the event. In Case F, one of the prototypes was discarded because of the scale it needed, there was not enough budget or other resources to create a project. In addition, in Case G the funding was even more difficult, since the nature of their business leans towards design and physical products, the benefits from hackathons could not be measured in direct sales, which made it difficult to justify funds for the winner and other costs.

The category *Difficulty of communication* refers to the issues with the language barriers (e.g. lack of English skills from presenters in cases E and H), whereas some teams were just challenging to contact in general. In cases I and J, the lack of communication and mentoring during the event was noticed, which caused a mismatch between the initial goals and the solutions. This could be an issue in public hackathons because the teams are not as controlled as they are in industrial events, but similar problems were mentioned in Case E, that there were not enough updates on the progress of the prototype, which also led into a mismatch of interests. In Case D, the interviewee highlighted the fact, that the communications have to be planned beforehand, a sentiment also mentioned in Case J.

The category *Technological issues* describes the practical issues during the events. In some organizations the access to data was limited and also there was a lack of suitable data, which caused the solutions to drift into a certain direction, while the organizer would have wanted more variance (cases A, I, and E). Another key issue was noticed with the current systems, the organizers did not have sufficient IT systems to support the hackathon, which limited the information flow to the teams and the prototypes were limited in this sense (cases E and H). In Case F it was mentioned, that because of interfacing problems with some teams, it was necessary to postpone the development altogether.

Challenges in preparation were noticed by the organizations: they recognized issues in preparation and were able to cancel out the possible risks and pitfalls. When preparing for the event, the challenges were formulated and resources were prepared for future development. Many organizers reported issues that were natural for a first-time organizer: adjustment of time, technology, resources, and winning criteria, as well as the selection of teams and participants (cases A, C, and F). Another issue was recognized in the environment and the scope of business of the organizer: the hackathon is not an optimal tool for every organization because the themes can be intangible, and the environments constrained, which can limit the level of innovation that the external professionals can bring into the event (cases C and H). Finally, in Case D the publicity was mentioned as a high risk. If something went wrong with the preparations, the combination of the visibility of the event and social media could cause irreparable damage to the company brand.

Some organizers experience issues with the *Team commitment*, which refers to the readiness for follow-ups after the hackathon. The teams were reported to be slightly disappointing to the organizer, because of few reasons: either the teams were not ready to commit to proper collaboration and only wanted to sell their concept (cases B and C), the developers were not that interested in the data of the organizer (Case I), the size and scale of the team and the backing organization was not sufficient for the scope of their prototype (Case H), and the teams did not change their visions based on expert comments or were unable to present their concept sufficiently (Case H).

The category *Schedule* did not appear as much as was expected: Since the hackathon events tend to last for a couple of days, it does not offer enough time to develop a highly refined solution for the complex problems and objectives. In Case H, the scheduling was especially difficult since the environment was changing during the event, and finding a date that could work for everyone involved was challenging. In cases C and D, the organizers were even contemplating that in hindsight there could have been multiple things that could have failed because of the tight time frame.

The *Ownership* of the product inside the organization is an issue since someone has to be responsible for the further development and the rights for it. Because of the common deals beforehand, where the property rights are agreed upon, the rights were an issue only in Case C, where two companies did not agree with the organizer. The results in an industry, which relies heavily on the visual design (Case G), allowing anything to leave the organization was not an option which was difficult situation for the teams. In Case E, it was highlighted that the ownership within the organization is extremely essential, and even the profitable concepts would be buried if the ownership is unclear in any way.

The events that were held as a curiosity show that the *Lack of follow-up* is a challenge. Every event, where something was created but nothing was taken forward, fundamentally only wasted resources. For instance, in Case G, the results were shared internally, they did receive positive comments, but nothing else happened. This happened only in two cases and in situations, where there were severe environmental or organizational issues preventing the

development. In Case J, the interviewee mentioned that the advertisements of the prototypes developed in their hackathon were seen in other events and seminars; the teams were developing the prototypes themselves.

5 Discussion

In this article, the main research question was "*How do the benefits and drawbacks of organizing a hackathon differentiate from opening data?*". In the previous section, the benefits and challenges of organizing hackathons are presented as based on the observations from the data and supported by findings from [19]. In this section, the goal is to further elaborate on the research question by comparing these results to the previous results [13] concerning the benefits and drawbacks of opening data.

5.1 Benefits from hackathons and opening data

The comparison of benefits is done by comparing the results of the systematic literature review about open data benefits [13], described in section 2, to the results of this research. In this comparison, the six business benefits found from the literature are used as the base categories, and the hackathon benefits are then placed into the categories based on their similarity. This means, that one benefit of opening data may have more than one corresponding benefit from hackathons. The same method is then used when comparing the recognized issues in opening data and hackathons.

Collaborative actions: There were similar elements in the collaborative actions emerged from opening data and hackathons. With open data, the collaboration between the consumers and the companies was triggered and enhanced through the data, allowing communication, feedback, and information sharing [7]. Hackathons, on the other hand, offer significantly more relevant communication and feedback, since the participants are familiarized to the environment, they are experts from relevant fields, and the solutions and development can be discussed face-to-face during the event (*Feedback and communication*). Because of this, the insights towards products and services are more useful and significant [18]. Opening the data can bring incremental development through two-way dialogue, while hackathons can introduce new technologies and techniques to the organization in a controlled environment and promote collaboration with teams, current partners, and also with new, or potential, partners (*Ecosystem development*).

Competitiveness: Easily measured effects towards competitiveness are not found from opening data since the nature of the initiative demands that the data is given away. Unlike hackathons, open data does not necessarily offer any new ventures to the organization, but enables outsiders to use the data for their own purposes, while the hackathon organizer was able to gain new ventures in every case in this research (*New ventures: projects, products, and research*). In some cases, the transparency and information sharing is a tool that the smaller companies use to enhance their competitiveness against the larger organizations [26]. However, sharing data without limits requires a suitable balance, since even one mistake can have negative impact or even prove fatal to the data provider [2]. To counteract this, the control in hackathons is necessary for the organizations to withhold their data and direct the innovations with the freedom to choose their participants and data users, which does not expose them to the competitors. There is also the aspect of recruitment, (e.g. [22]) since opening data does not bring new people to the organization, except for some rare, special cases, while hackathons allow the creation of personal contacts (*Recruitment*).

Ecosystem-wide engagement and communication: While opening data may increase the communication inside an ecosystem, the hackathons take this further (*Ecosystem development*). When data is opened, the data provider has limited control over the data usage and discoverability. Through hackathon, the participants are screened and motivated to use the given data in the event to nurture their concepts and ideas further. The organizer can monitor the participating teams, make suggestions, and even develop projects with multiple teams. By meeting with the developers, the communication is more coherent and through control, both sides can create something practical, while with open data the communication is limited to data [11]. Through hackathons, the roles in the ecosystems do not change, but opening data may require additional changes to the data provider, and also for other organizations in their ecosystem [26]. It was also noticed that the hackathons offered new partners and suppliers for the organizer, widening their ecosystem, which was not mentioned as a benefit of open data.

Innovation and development: The innovations based on open data, incremental or radical, can be made by a variety of developers because of the availability of open data. However, in the scope of one organization, the innovations based on open data may or may not aid the data publisher, since the data has already been opened and the innovation is done by someone else. The control of the data, in this case, is crucial in order to oblige the innovator to collaborate with the publisher. Without control, the innovation, radical or incremental, is difficult to commercialize by the original manufacturer and the data provider. The hackathons, on the other hand, offer more practical cases, because the participants are familiarized with the industry and the systems, so the innovations and development can be focused further and it serves as a base (*Archive of ideas and concepts*) to relevant innovations that the organization can use [24]. With hackathons, the rights and ownership of the created solutions are agreed upon beforehand, but with innovations based on the open data can be difficult to manage, as was mentioned earlier.

Hackathons were also linked to the developments in the innovation process (*Accelerated R&D*) within the organization, an aspect that opening data does not promote.

Internal change: There is quite the contrast between opening data and organizing a hackathon in a sense of internal development, since opening data allows the organization to see flaws and issues in their processes, while through hackathons the company can develop their systems (*Motivated system development*) and processes in a controlled manner towards more agile development (*Internal / organizational development*). Another contrast is in the scope of cultural change inside the organization. When data is opened, the change is reported as being a necessity [17], while with the hackathon it was found as an additional benefit born from enthusiasm, although not in every case. In a hackathon, the departments are creating new collaboration and increasing communication between themselves voluntarily, which can increase efficiency and change the ways things are done within the organization.

Public image: The visibility from hackathons is limited to knowing about the event and the event description, but the outcomes are not always disclosed (*Visibility*). Opening data, on the other hand, can be utilized by anyone and through use, the data gains more visibility. Visualizing opened data for consumers is a powerful method to deliver useful information to environments and adds value to the transaction. Additionally, depending on the industry, companies with a large consumer market can use the publicity and transparency to stand out from competitors [26].

5.2 Challenges in hackathons and opening data

The comparison of drawbacks and challenges is not as straightforward as the comparison between the benefits. Some of the challenges in hackathons are about the issues from third parties, such as team commitment, or issues born from the very nature of a hackathon, such as a schedule for the event, preparations, and follow-up, issues also recognized in [24]. Because of this, they do not correspond well with opening data.

Decrease in efficiency: One major issue with opening data is the duplication of processes and ownership of the data. As it was mentioned, the efficiency of opening data can hinder, because the processes are not updated and the organization is using duplicate processes for the same tasks [14]. With hackathons the nature of this issue is different since the organization does not commit assets; much of the development is done by external professionals and ownerships are negotiated beforehand. After the event, the follow-up is done by following the existing processes. If the follow-ups require creating some changes in the organization, the changes are incremental and manageable and they do not require large changes in the current state, even though it was seen as a challenge (*Process and organizational issues*). Another issue is the ownership of data: if the ownership is not clear, for example, the data is owned by a third party, this can make opening data ineffective or even prevent opening of data completely [12]. In hackathons, the ownership of the prototypes and further, the product ownership was seen as a major issue (*Ownership*), since clear ownership would prevent the prototype from further development, and would not be done.

Increased costs: The costs of opening data originate from collecting the data, data visualization, and training, but also from managing the systems. Additional costs may also be caused by the documentations, metadata, and other additional materials that could be necessary for the data usage [2]. In hackathons, a major issue was the availability of resources and funds for the event (*Resource and cost management*), in order to host the event, but also to move the prototypes forward into products. However, in hackathons, the data does not need to move outside the organization and the participants are on the premises, which allows the organizer to use the pre-existing systems, so the technologies do not have to be developed to very advanced degree. While the event does cause a strain on the resources, in most cases it is only a short and temporary need, while opening data would require more financial, technological, and ideological long-term commitment.

Public data increases problems: While there is not a category from the hackathons that matches directly with this category, there was a mention in Case D in the category *Challenges in preparation*. If the hackathon is not sufficiently prepared for and something goes wrong, the combination of publicity and social media may cause issues to the organizer and their brand. This issue seems as a minor hindrance since open data can be used to purposefully misunderstand facts, used against the company, or used to invade privacy, which can affect the brand even more [31]. Hackathons also had an issue with the communication during the event with the teams (*Difficulty of communications*), causing the organizer to receive solutions, which did not match the initial goals. However, during the hackathon, the lack of communication can be an issue, but crucial misunderstanding and misuses can be prevented through control, which is not an option with the open data.

Required change: In the literature review and in this research, the processes and culture of the organization were deemed an issue (also [14], [17]), but also one of the most potential subjects of development. With opening data, the amount of change for the organization and their ecosystem was noted to be substantial, requiring changes in the policies, focus, business models, and ideologies and not only for the organization opening data, but possibly for their partners as well. While opening data requires a substantial commitment to the open data initiative and changes in the organization, the hackathon is contained within the organization and rarely requires changes to the organization or their ecosystem, while they could be heavy on resources (*Resource and cost management*). This decreases the opposition and makes the changes more relevant and fits better into the culture of the organization. Applying a process for opening data in an unsuitable culture can lead to harmful actions within the organization [2], especially if

opening data is applied forcefully, while in hackathons the culture change emerges from the individuals, not from the strategy. In addition, the hackathon may require some technological upgrades, as was noticed in half of the cases (*Technological issues*), but opening data requires a whole new set of systems.

5.3 Conjectures from the study

In conclusion, the benefits and drawbacks of applying hackathons as an approach to open and share company data are summarized. When comparing opening data to hackathons from the viewpoint of the organization, there seem to be multiple benefits, but scarce evidence about why opening data could be more profitable. Hackathons offer:

1. More control over the whole innovation process than opening data

The organization that organizes the hackathon has control over who participates in the event. This allows control over the teams and ultimately the solutions they are going to develop during and after the event. The innovation can also be influenced by the event because teams can communicate with the organizer and their experts in order to gain insights into the industry. This decreases the risk of misunderstandings and the misuses of data.

2. Practical business cases and enthusiasm towards improving the culture within the organization

Open data as an initiative has its largest drawbacks on when the organization is publishing data in order to generate new business. The hackathons in this research have shown, that through the events the host gets something new in the form of projects, products, and/or research, while through opening data this is not guaranteed. Some of the organizers are already reporting profit from these innovations, while some organizers utilize the results as incremental development. It was also noticed, how opening data requires systematic change, while hackathons inspire and motivate the employees to change the culture through actions, not because of a strategy.

3. Engagement and communication with educated experts better than through open data

Data is a lacking medium to hold a discussion or other form of communications, but it can be used as the base for discussion. In order to engage the experts and developers, it seems to be more beneficial to meet face-to-face and discuss the topic in a suitable environment. This way the communications are clear and misunderstanding can be resolved quickly, even though this was not a complete success with hackathons either. In the event, it is also possible to motivate the experts and assess their skills, which may even lead to new recruitments.

4. Fewer changes in day-to-day processes without external changes

Unlike opening data, organizing a hackathon does not require a sustained strain on the internal resources, since the event takes only a few days and the collaboration with the teams does not change anything within the organization, since the teams are at most treated as another partner. Working with external teams also allows flexibility to the organizer, since they do not need to commit resources and systems as much as they would need to with opening data. The collaboration with the teams may require some changes to the software systems, but they are usually incremental changes and do not require new systems. What hackathons also offer when compared to opening data is that the necessary changes are confined to the organization and it does not affect the existing partnerships in a way that opening data might change the environment. Opening data also requires commitment and if the company culture does not adopt or adapt to opening data, the changes can be difficult and possibly costly to either roll back or enforce.

As mentioned, when comparing the hackathons and opening data, in this research opening data is seen to have only one tangible reason, why it seems to be the more efficient solution. That is:

5. Open data attracts more public visibility than hackathons

The hackathon is, by nature, a closed event, usually accessible only through a screening process. Open data, on the other hand, is by definition accessible to anyone and it can be used for any legal objective imaginable. In hackathons, the organizer can limit the scope of the solutions by determining goals and limitations because the event is a competition between teams. These aspects show, how open data has the potential to reach any number of individuals and serve as a free advertisement to the data publisher; the data can be also used more freely in order to develop new products and services by consumers. While hackathons offer strict visibility to partners, open data allows widespread publicity through data. Although, depending on the nature of an organization, the type of visibility that a hackathon offers may be more suitable for their goals. It was also mentioned, that the visibility of open data can be a benefit, but it can also be a drawback, depending on how the data is used and by whom. Because of these reasons, even the extended visibility through open data is something that organizations may not want and they could prefer a more limited scope of advertisement.

While this research cannot be generalized to public institutions and the scope of this research is private organizations, some lessons learned can be drawn from here. Opening data is another strategy to allow access outside the organization, but the critical control and harvesting of benefits are more efficiently achieved through the hackathons than the blind opening of data, possibly because of the planning and organizing that hackathons require. It would seem that in comparison to open data, hackathons are more beneficial to companies because the return of investment can be calculated more accurately and there seems to be trust that hackathons will deliver benefits.

5.4 Limitations of this study

In qualitative research, there exists a multitude of risks towards the credibility of the research [23]. For instance, the researcher bias is troublesome in this type of research, where the researcher can affect the interviewee through active or passive actions. Such instances can occur when constructing the questionnaire while collecting the data and during the data analysis activities. In this study, this issue was taken into account by using two researchers to construct the questionnaire and consult the organizations, which were developing and arranging hackathon events, and by using four researchers representing four different research groups in the analysis and publication of the data. In addition, the interviews were held in the native language of the interviewee in order to allow informal discussion, to ensure that the interviewee was able to speak their mind, and to ensure that the interviewees understood the questions correctly. Finally, the data were analyzed by using cross-case synthesis method, which allows more robustness because of the abstraction of multiple cases [29]. In general, the principles defined by, for example [8], [25], were followed to enhance the trustworthiness, rigor and the overall quality of this study. In any case, these types of qualitative studies cannot offer any generalization in a mathematical sense, but rather offer a generalization of observations, which can be applied as guidelines or considerations for best practices when outside of the study scope [27].

6 Conclusions

In this article, we present the results from a qualitative case study observing the benefits and different opinions related to the hackathons, comparing them to opening of the corporate data to the third parties in general. The study observed and interviewed organizers from ten hackathon events, analyzing the beneficial and counterproductive outcomes from the viewpoint of the hosting organization.

Based on the observations, the hackathon event benefits are similar to the general benefits of opening data for third parties, engaging collaboration and innovation from external professionals. It seems that the hackathon events are more beneficial to the organization in question, while opening data could benefit the larger number of actors, markets, and industries as a whole. Some of the benefits from hackathons in comparison of opening data are practical business cases, a higher level of control, more natural communications, and incremental improvements of systems. In the scope of the organization, opening data offers more visibility to the data and the organization.

In future work, the aim is to observe the hackathons and study the open data aspects in order to better understand these two initiatives. For example, what happens in practice with the hackathon results, how successful are the actual product implementations, and how could opening data match these results.

Acknowledgments

This research work was partially funded by DIMECC and S4Fleet research grants. We would like to thank all participating organizations.

Websites List

Site 1: Ultra Hack
https://ultrahack.org/

Site 2: Junction | Europe's largest hackathon
http://www.hackjunction.com/

References

[1] E. Almirall, M. Lee, and A. Majchrzak, Open innovation requires integrated competition-community ecosystems: Lessons learned from civic open innovation, Business Horizons, vol. 57, no. 3, pp. 391-400, 2014.
[2] F. Bannister, and R. Connolly, The trouble with transparency: a critical review of openness in e-government, Policy & Internet, vol. 3, no. 1, pp. 1-30, 2011.
[3] G. Briscoe and J. X. Hon. (2014) Choreographons: hackathons for dance composition. [Online]. Available: http://webspace.qmul.ac.uk/gbriscoe-disabled/site/Publications_files/Choreographons-%20Hackathons%20for%20Dance.pdf.
[4] G. Briscoe and C. Mulligan, Digital innovation: The hackathon phenomenon, London: Creativeworks London Work Paper, vol. 6, pp.1-13, 2014.
[5] H. W. Chesbrough, Open innovation: The new imperative for creating and profiting from technology. Boston: Harvard Business Press, 2006.
[6] K. C. Desouza, Y. Awazu, S. Jha, C. Dombrowski, S. Papagari, P. Baloh, and J. Kim, Customer-driven innovation, Research-Technology Management, vol. 51, no. 3, pp. 35–44, 2008.

[7] D. J. Doorey, "The transparent supply chain: From resistance to implementation at Nike and Levi-Strauss," Journal of Business Ethics, vol. 103, no. 4, pp. 587–603, 2011.
[8] N. Golafshani, Understanding reliability and validity in qualitative research, The qualitative report, vol. 8, no. 4, pp. 597–606, 2003.
[9] J. Gurin, Open Governments, Open Data: A New Lever for Transparency, Citizen Engagement, and Economic Growth, The SAIS Review of International Affairs, vol. 34, no. 1, pp. 71–82, 2014.
[10] J. Henkel, Selective revealing in open innovation processes: The case of embedded Linux, Research policy, vol. 35, no. 7, pp. 953-969, 2006.
[11] A. Herala, J. Kasurinen, and E. Vanhala, Current status and the future directions of open data: perceptions from the finnish industry, in Proceedings of the 20th International Academic Mindtrek Conference, Finland, 2016, pp. 68–77.
[12] A. Herala, J. Kasurinen, and E. Vanhala, Views on Open Data Business from Software Development Companies, Journal of Theoretical and Applied Electronic Commerce Research, vol. 13, no. 1, pp. 91-105, 2018
[13] A. Herala, E. Vanhala, J. Porras, and T. Kärri, Experiences about opening data in private sector: A systematic literature review, in 2016 SAI Computing Conference (SAI), London, 2016, pp. 715–724.
[14] N. Huijboom and T. Van den Broek, Open data: an international comparison of strategies, European journal of ePractice, vol. 12, no. 1, pp. 4–16, 2011.
[15] A. Immonen, M. Palviainen, and E. Ovaska, Towards Open Data Based Business: Survey on Usage of Open Data in Digital Services. International Journal of Research in Business and Technology, vol. 4, no. 1, pp. 286-295, 2014.
[16] H. Jaakkola, T. Mäkinen, and A. Eteläaho, Open data: opportunities and challenges, in Proceedings of the 15th International Conference on Computer Systems and Technologies, Bulgaria, 2014, pp. 25-39.
[17] M. Janssen, Y. Charalabidis, and A. Zuiderwijk, Benefits, Adoption Barriers and Myths of Open Data and Open Government, Information Systems Management, vol. 29, no. 4, pp. 258–268, 2012.
[18] P. Johnson and P. Robinson, Civic hackathons: Innovation, procurement, or civic engagement?, Review of Policy Research, vol. 31, no. 4, pp. 349–357, 2014.
[19] M. Komssi, D. Pichlis, M. Raatikainen, K. Kindström, and J. Järvinen, What are Hackathons for?, IEEE Software, vol. 32, no. 5, pp. 60–67, 2015.
[20] J. Lindman, T. Kinnari, and M. Rossi, Industrial Open Data: Case Studies of Early Open Data Entrepreneurs, in 2014 47th Hawaii International Conference on System Sciences (HICSS), Hawaii, 2014, pp. 739-748.
[21] M. B. Miles and A. M. Huberman, Qualitative data analysis: An expanded sourcebook. Thousand Oaks, CA: Sage, 1994.
[22] A. Nandi and M. Mandernach, Hackathons As an Informal Learning Platform, in Proceedings of the 47th ACM Technical Symposium on Computing Science Education, New York, NY, USA, 2016, pp. 346–351.
[23] A. J. Onwuegbuzie and N. L. Leech, Validity and Qualitative Research: An Oxymoron?, Quality & Quantity, vol. 41, no. 2, pp. 233–249, 2007.
[24] M. Raatikainen, M. Komssi, V. d Bianco, K. Kindstöm, and J. Järvinen, Industrial Experiences of Organizing a Hackathon to Assess a Device-centric Cloud Ecosystem, in 2013 IEEE 37th Annual Computer Software and Applications Conference, Japan, 2013, pp. 790–799.
[25] C. Robson, Real world research. United Kingdom: John Wiley & Sons, 2002.
[26] D. S. Sayogo, J. Zhang, T. A. Pardo, G. K. Tayi, J. Hrdinova, D. F. Andersen, and L. F. Luna-Reyes, Going Beyond Open Data: Challenges and Motivations for Smart Disclosure in Ethical Consumption, Journal of Theoretical and Applied Electronic Commerce Research, vol. 9, no. 2, pp. 1–16, 2014.
[27] A. Strauss and J. Corbin, Basics of qualitative research. Newbury Park, CA: Sage, 1990.
[28] I. Susha, M. Janssen, and S. Verhulst, Data collaboratives as "bazaars"? A review of coordination problems and mechanisms to match demand for data with supply, Transforming Government: People, Process and Policy, vol. 11, no. 1, pp. 157-172, 2017.
[29] R. K. Yin, Case study research: Design and methods. Thousand Oaks, CA: Sage, 2013.
[30] A. Zuiderwijk, N. Helbig, J. R. Gil-García, and M. Janssen, Special Issue on Innovation through Open Data: Guest Editors' Introduction. Journal of Theoretical and Applied Electronic Commerce Research, vol. 9, no.2, pp. i-xiii, 2014
[31] A. Zuiderwijk, M. Janssen, R. Meijer, S. Choenni, Y. Charalabidis, and K. Jeffery, Issues and guiding principles for opening governmental judicial research data, in International Conference on Electronic Government, Norway, 2012, pp. 90-101.
[32] A. Zuiderwijk, M. Janssen, G. van de Kaa, and K. Poulis, The Wicked Problem of Commercial Value Creation in Open Data Ecosystems: Policy Guidelines for Governments. Information Polity, vol. 21, no. 3, pp. 223-236, 2016.

ACTA UNIVERSITATIS LAPPEENRANTAENSIS

771. STADE, SAM. Examination of the compaction of ultrafiltration membranes with ultrasonic time-domain reflectometry. 2017. Diss.

772. KOZLOVA, MARIIA. Analyzing the effects of a renewable energy support mechanism on investments under uncertainty: case of Russia. 2017. Diss.

773. KURAMA, ONESFOLE. Similarity based classification methods with different aggregation operators. 2017. Diss.

774. LYYTIKÄINEN, KATJA. Removal of xylan from birch kraft pulps and the effect of its removal on fiber properties, colloidal interactions and retention in papermaking. 2017. Diss.

775. GAFUROV, SALIMZHAN. Theoretical and experimental analysis of dynamic loading of a two-stage aircraft engine fuel pump and methods for its decreasing. 2017. Diss.

776. KULESHOV, DMITRII. Modelling the operation of short-term electricity market in Russia. 2017. Diss.

777. SAARI, JUSSI. Improving the effectiveness and profitability of thermal conversion of biomass. 2017. Diss.

778. ZHAO, FEIPING. Cross-linked chitosan and β-cyclodextrin as functional adsorbents in water treatment. 2017. Diss.

779. KORHONEN, ILKKA. Mobile sensor for measurements inside combustion chamber – preliminary study. 2017. Diss.

780. SIKIÖ, PÄIVI. Dynamical tree models for high Reynolds number turbulence applied in fluid-solid systems of 1D-space and time. 2017. Diss.

781. ROMANENKO, ALEKSEI. Study of inverter-induced bearing damage monitoring in variable-speed-driven motor systems. 2017. Diss.

782. SIPILÄ, JENNI. The many faces of ambivalence in the decision-making process. 2017. Diss.

783. HAN, MEI. Hydrodynamics and mass transfer in airlift bioreactors; experimental and numerical simulation analysis. 2017. Diss.

784. ESCALANTE, JOHN BRUZZO. Dynamic simulation of cross-country skiing. 2017. Diss.

785. NOKKA, JARKKO. Energy efficiency analyses of hybrid non-road mobile machinery by real-time virtual prototyping. 2018. Diss.

786. VUORIO, ANNA. Opportunity-specific entrepreneurial intentions in sustainable entrepreneurship. 2018. Diss.

787. PULKKINEN, AKI. Towards a better understanding of activity and selectivity trends involving K and O adsorption on selected metal surfaces. 2017. Diss.

788. ZHAO, WENLONG. Reliability based research on design, analysis and control of the remote handling maintenance system for fusion reactor. 2018. Diss.

789. IAKOVLEVA, EVGENIA. Novel sorbents from low-cost materials for water treatment. 2018. Diss.

790. KEDZIORA, DAMIAN. Service offshoring industry: systems engineering approach to its transitional challenges. 2018. Diss.

791. WU, JING. Soft computing methods for performance improvement of EAMA robot in fusion reactor application. 2018. Diss.

792. VOSTATEK, PAVEL. Blood vessel segmentation in the analysis of retinal and diaphragm images. 2018. Diss.

793. AJO, PETRI. Hydroxyl radical behavior in water treatment with gas-phase pulsed corona discharge. 2018. Diss.

794. BANAEIANJAHROMI, NEGIN. On the role of enterprise architecture in enterprise integration. 2018. Diss.

795. HASHEELA-MUFETI, VICTORIA TULIVAYE. Empirical studies on the adoption and implementation of ERP in SMEs in developing countries. 2018. Diss.

796. JANHUNEN, SARI. Determinants of the local acceptability of wind power in Finland. 2018. Diss.

797. TEPLOV, ROMAN. A holistic approach to measuring open innovation: contribution to theory development. 2018. Diss.

798. ALBATS, EKATERINA. Facilitating university-industry collaboration with a multi-level stakeholder perspective. 2018. Diss.

799. TURA, NINA. Value creation for sustainability-oriented innovations: challenges and supporting methods. 2018. Diss.

800. TALIKKA, MARJA. Recognizing required changes to higher education engineering programs' information literacy education as a consequence of research problems becoming more complex. 2018. Diss.

801. MATTSSON, ALEKSI. Design of customer-end converter systems for low voltage DC distribution from a life cycle cost perspective. 2018. Diss.

802. JÄRVI, HENNA. Customer engagement, a friend or a foe? Investigating the relationship between customer engagement and value co-destruction. 2018. Diss.

803. DABROWSKA, JUSTYNA. Organizing for open innovation: adding the human element. 2018. Diss.

804. TIAINEN, JONNA. Losses in low-Reynolds-number centrifugal compressors. 2018. Diss.

805. GYASI, EMMANUEL AFRANE. On adaptive intelligent welding: Technique feasibility in weld quality assurance for advanced steels. 2018. Diss.

806. PROSKURINA, SVETLANA. International trade in biomass for energy production: The local and global context. 2018. Diss.

807. DABIRI, MUHAMMAD. The low-cycle fatigue of S960 MC direct-quenched high-strenght steel. 2018. Diss.

808. KOSKELA, VIRPI. Tapping experiences of presence to connect people and organizational creativity. 2018. Diss.

Acta Universitatis
Lappeenrantaensis
809

LUT University

ISBN 978-952-335-261-2
ISBN 978-952-335-262-9 (PDF)
ISSN-L 1456-4491
ISSN 1456-4491
Lappeenranta 2018