

Views on Open Data Business from Software Development Companies

Antti Herala¹, Jussi Kasurinen², and Erno Vanhala³

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

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].

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.

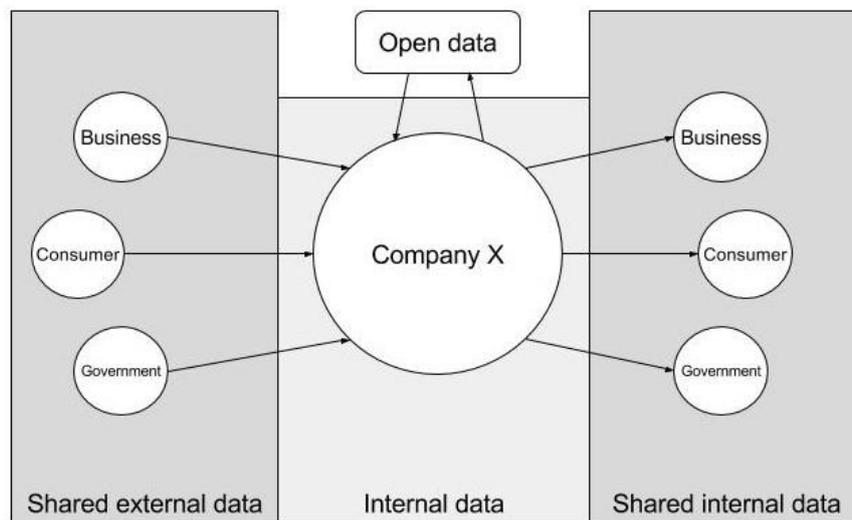


Figure 1: Available external data (L) and internal data (R), based on [15], [25], [28]

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.

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.

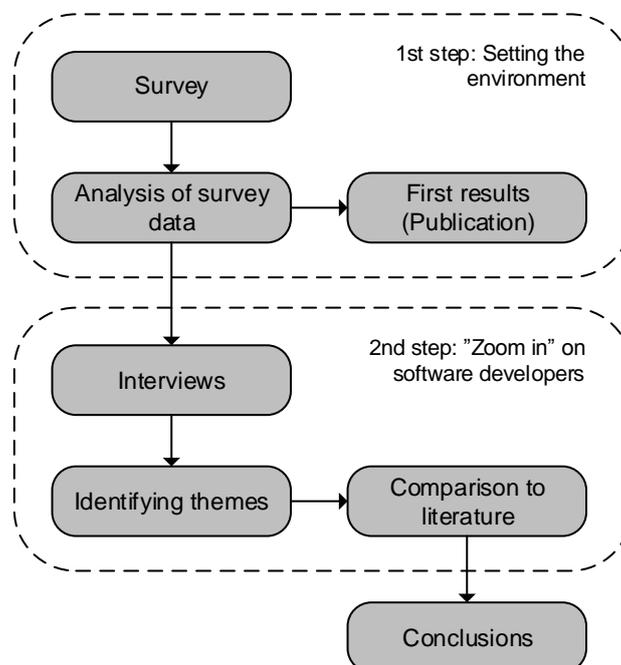


Figure 2: Data collection and research process

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

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

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.

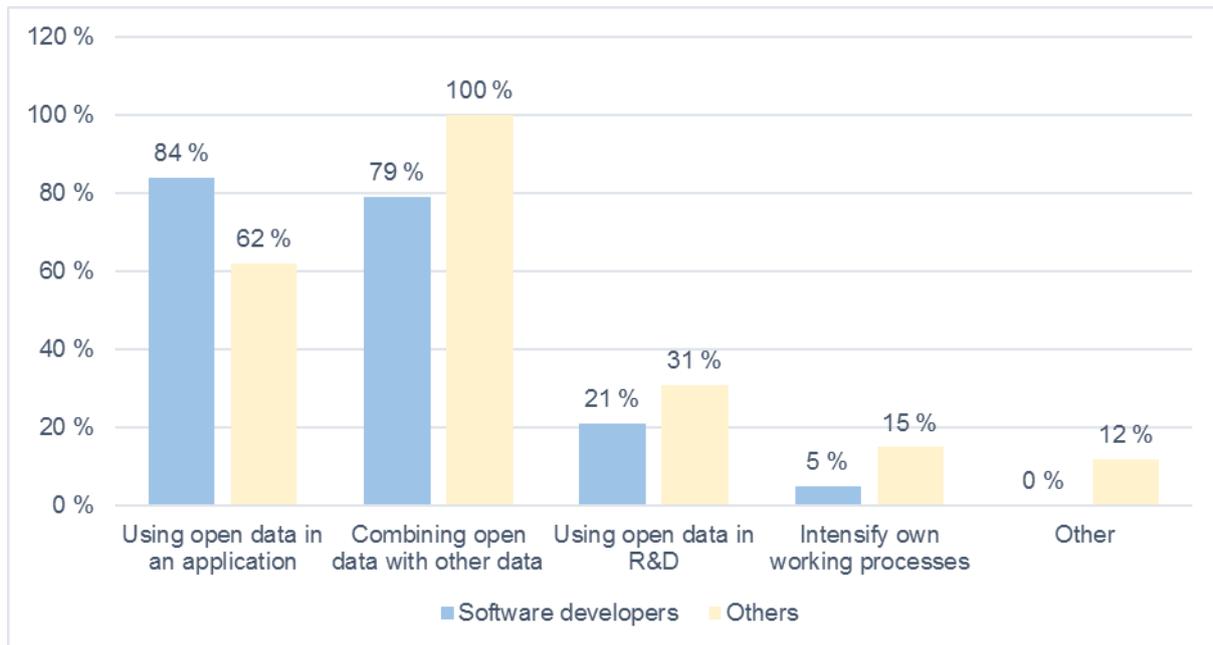


Figure 3: The added value of open data for software developers and other areas of business

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.

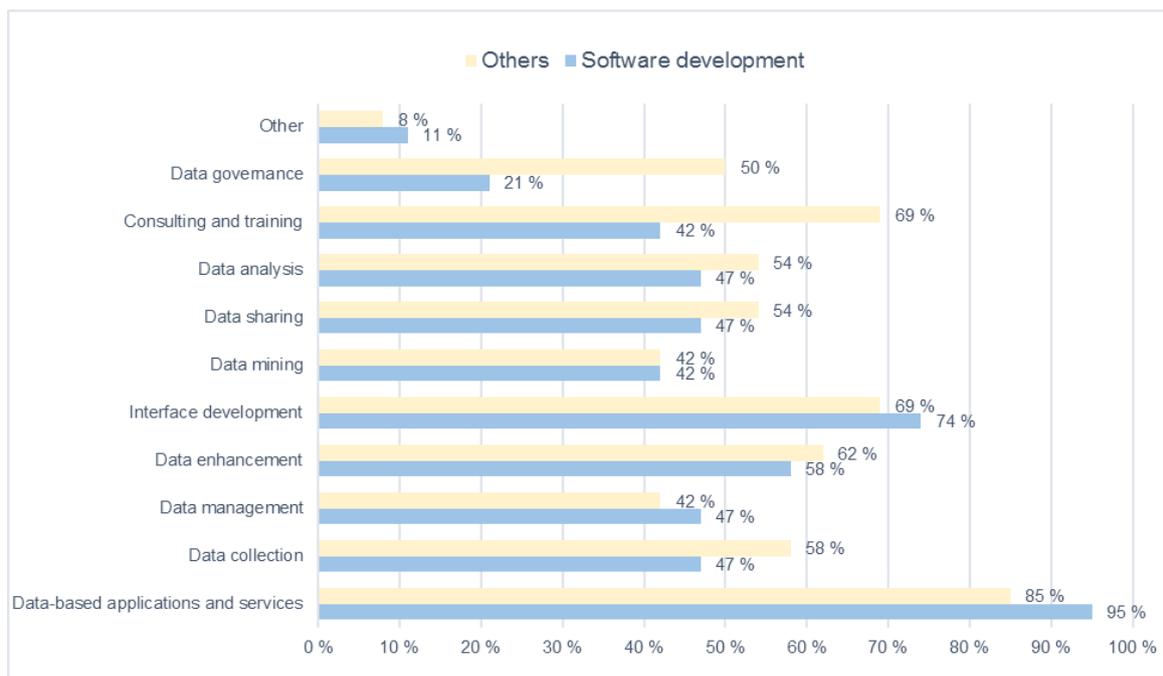


Figure 4: Company interest towards different segments of data business

These results 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.

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.

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-

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.

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

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).

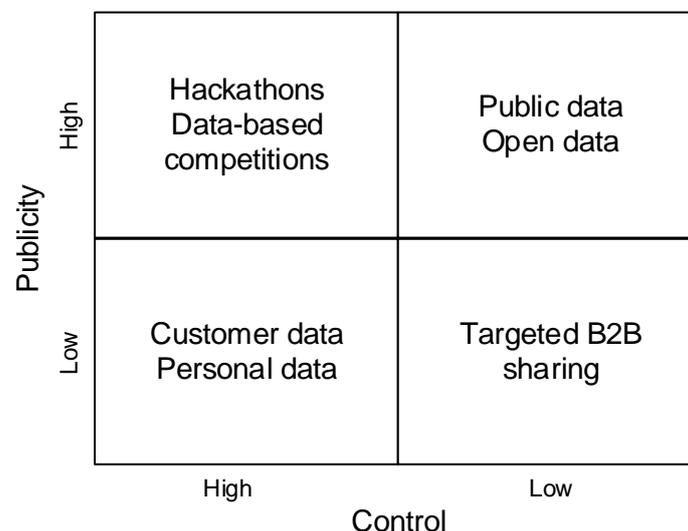


Figure 5: Strategies to share data

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.

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.

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.

- [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.